

RAPPORT D'ETAPE (2015-2016)

Septembre 2016

Iwona DUDEK-BLAISE
UMR CNRS/MCC 3495 MAP



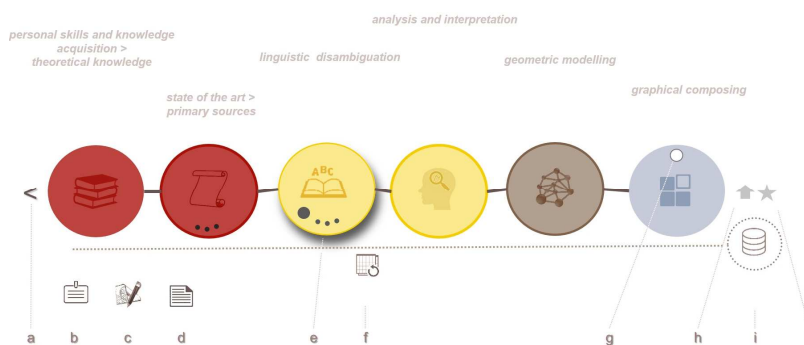
A.1 - OBJECTIFS	1
A.2 - MOYENS HUMAINS ET MATÉRIELS	2
A.3 - DÉVELOPPEMENTS (2015-2016) DU PROJET	4
A3.1 RÉSULTATS OBTENUS.....	6
<i>Model conceptuel, logique et physique de la structure de la base de données</i>	<i>7</i>
<i>Vocabulaire normalisé</i>	<i>8</i>
<i>Identification et structuration des activités</i>	<i>11</i>
<i>Identification, classification et validation des champs d'expertise</i>	<i>16</i>
<i>Système d'information MEMORIA: l'interface utilisateur</i>	<i>18</i>
<i>Un langage visuel pour l'analyse de motifs et d'exceptions.....</i>	<i>22</i>
<i>Site Web de présentation du projet MEMORIA.....</i>	<i>26</i>
A.2 - PUBLICATIONS	28
A.3 - RESSOURCES EN LIGNE	29
A.4 - COLLABORATIONS ET LIENS SCIENTIFIQUES	30
ANNEXES	

A.1 - Objectifs

Le projet MEMORIA est un projet à long terme, visant à développer un système d'information permettant la description, la structuration et l'archivage des ressources produites dans notre laboratoire.

L'objectif sous-jacent du projet est de mémoriser non seulement un résultat (cf. Fig.1) – *i.e.* une ressource numérique ou non - mais aussi et surtout son historique de production, autrement dit la façon dont ce résultat a été obtenu.

À cette fin, les ressources produites seront d'une part décrites par une liste de descripteurs classiques (du type format, auteurs, date de production, *etc.*) et d'autre part associées à des *processus* (notion à comprendre comme une chaîne d'activités). Chaque *processus* gardera trace du cadre institutionnel dans lequel les travaux se sont déroulés (organisations, projets, personnel employé, ...), des sources externes employées (dans une phase d'analyse par exemple), ou encore des techniques et outils utilisés pour produire les ressources numériques. (cf. Fig.2)



Un point central de la démarche adoptée est la volonté de mettre au point des interfaces visuelles. Ces interfaces doivent par exemple donner accès aux résultats de requêtes sur des ressources numériques (triées par objet d'étude, projet, processus de production, *etc.*) (cf. Fig. 3), mais aussi montrer l'évolution des méthodes de travail ou des techniques et outils utilisées au cours du temps, ainsi que les types d'activités mobilisées pour produire une ressource.

À terme, le projet vise à associer à une ressource numérique quelle qu'elle soit (ex. maquette 3D, cartographie 2D, répertoire de références) une série de descripteurs permettant de retracer l'ensemble des actions mobilisées pour la produire, et donc de mémoriser au-delà d'un document numérique « final » une démarche cognitive, ou tout au moins une démarche de production.

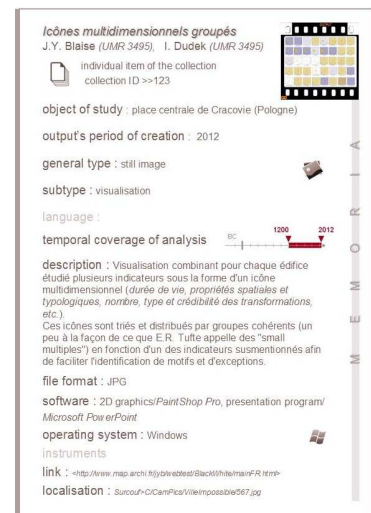


Fig. 1 Exemple d'une fiche de description d'un extrait

Fig. 2 Exemple d'un processus ordonné s'appuyant sur les résultats d'un processus précédent et sur une infrastructure externe.

Trois types de résultats issus de ce processus ont été enregistrés : g - output(s), h - publication(s) et j - composition(s). a - processus précédent, b - nom du processus, c - projet dans lequel le processus actuel a été réalisé, d - remarques, e - une activité récurrente, f - le processus a entraîné une modification de la structure de l'infrastructure externe, g - lien vers les extraits issus du processus, h - lien vers les publications issues du processus, i - informations sur l'infrastructure qui a été engagée dans le processus, j - lien vers les compositions issues du processus

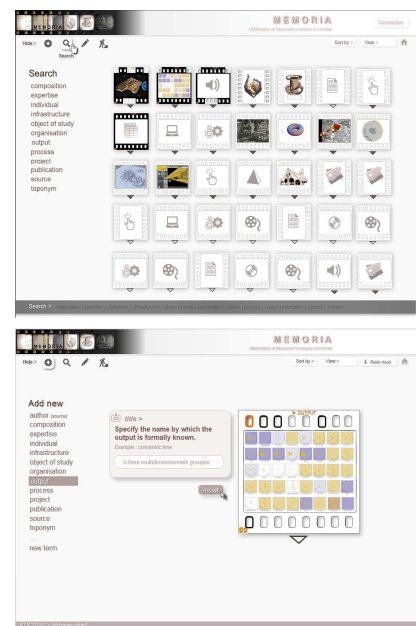


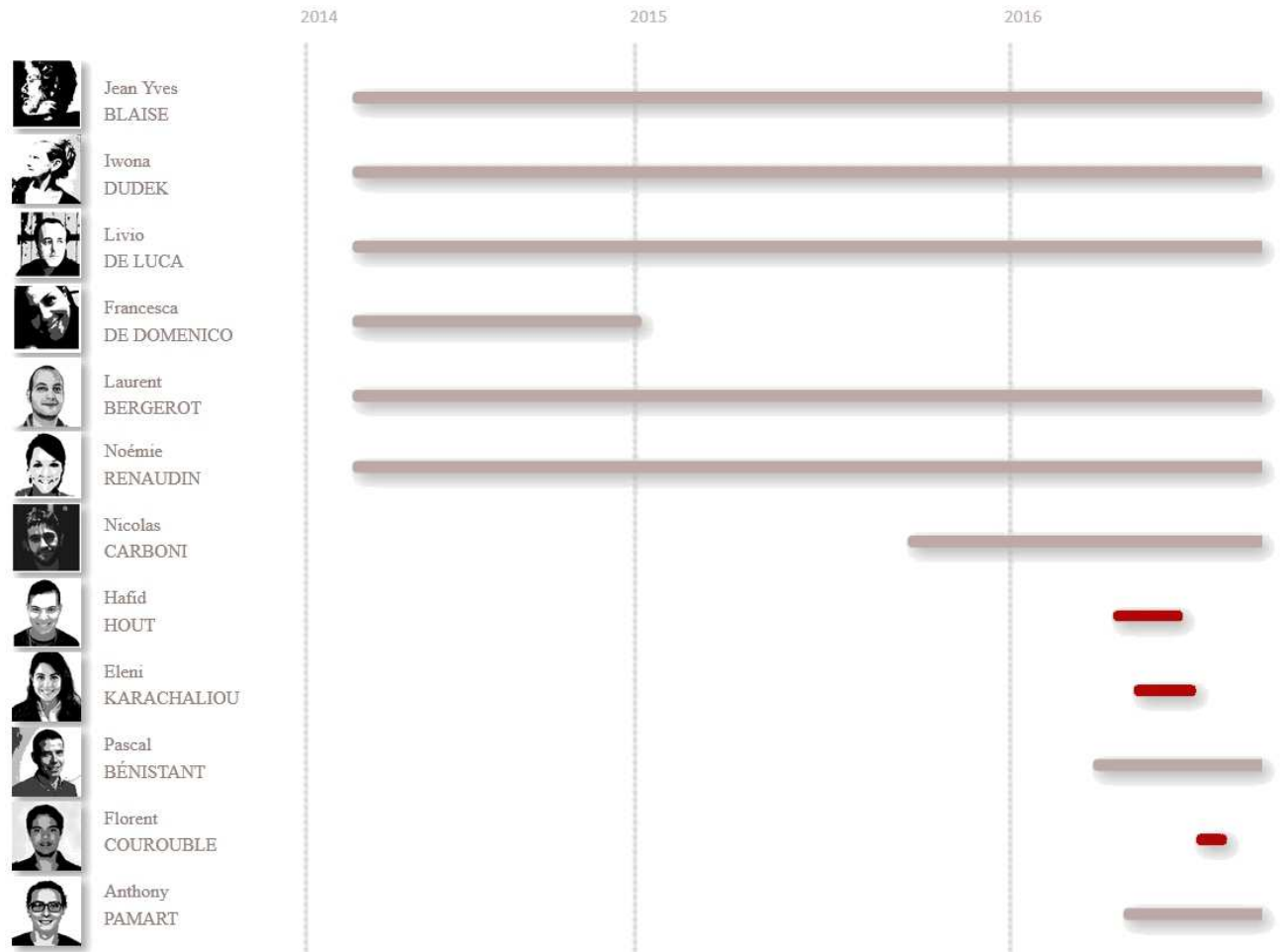
Fig. 3 Principe graphique de l'interface

A.2 – Moyens humains et matériels

Le projet MEMORIA est un projet de long terme initié et porté par l'UMR 3495 CNRS/MCC MAP.

Il se déploie en s'appuyant sur les ressources propres du laboratoire (le matériel et le personnel) avec le soutien du Département de la Recherche, de l'Enseignement Supérieur et de la Technologie du Ministère de la Culture et de la Communication.

Fig. 4 Moyens humains engagés dans le projet. En gris le personnel de l'UMR MAP, en rouge les stagiaires employés pour remplir des tâches particulières (*i.e.* implémentation du Site Web de présentation, expérimentation d'une plateforme cartographique). Septembre 2016



A2.1 Responsabilités et rôles attribués aux membres de l'équipe en 2015-2016

- chef de projet - I. Dudek-Blaise
- comité de pilotage - J.Y. Blaise, L. De Luca, I. Dudek-Blaise
- acquisition, filtrage et traitement des données - I. Dudek-Blaise
- modélisation des connaissances - J.Y. Blaise, I. Dudek-Blaise
- explicitation des connaissances - J.Y. Blaise, I. Dudek-Blaise
- explicitation des connaissances (participation) – L. Bergerot, L. De Luca, N. Renaudin, N. Carboni, A. Pamart
- encadrement des stagiaires - P. Bénistant, J.Y. Blaise, I. Dudek-Blaise
- conception des interfaces - I. Dudek-Blaise
- administration système - P. Bénistant

test d'implémentation de l'interface SI - L. Bergerot
implémentation Site Web de présentation - H. Hout, F. Courouble
test d'implémentation d'une plateforme cartographique – E. Karachaliou

A2.2 Moyens techniques

Du point de vue des choix technologiques, nous nous appuyons sur des solutions, des standards et des formalismes issus du monde du logiciel libre. Ces technologies sont par ailleurs conçues pour le développement d'applications sur Internet (Base de données MySQL, Serveur Apache, Scripts PHP, SVG, Librairie Leaflet, cartographie OSM, ...).

Les composants du projet sont stockés au fil de l'eau dans une zone d'archivage et de travail propre au projet MEMORIA. (cf. Fig. 5)

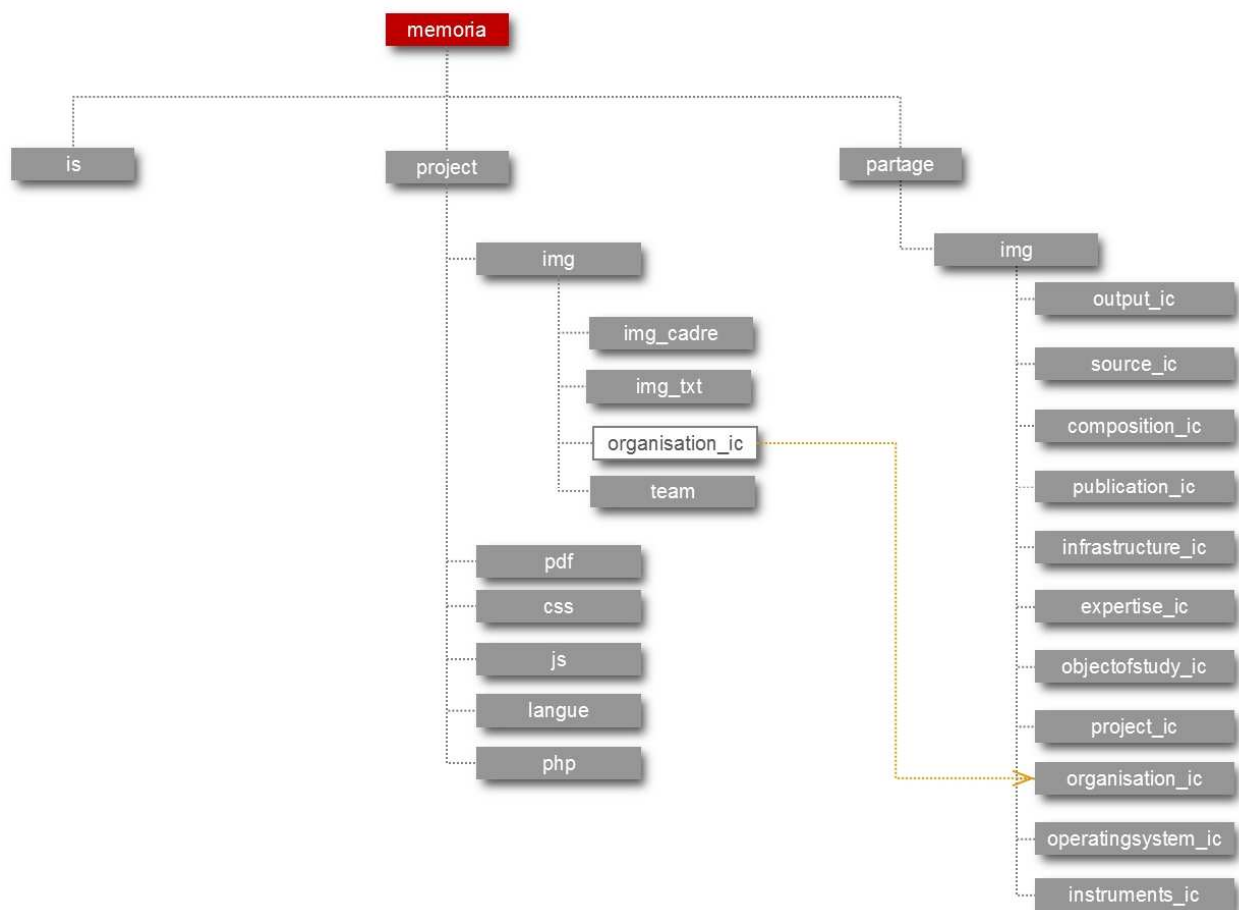


Fig. 5 Structure des fichiers de la zone de stockage

A.3 – Développements (2015-2016) du projet

Les différentes actions menées dans le développement du projet MEMORIA sur la période sont ici répertoriées en suivant la grille d'analyse des activités telle que formalisée à travers le projet. Une description plus classique des résultats obtenus est proposée dans la sous-section suivante. (cf. Annexe 8)

1. Acquisition de données (A)

- acquisition indirecte (A2)
 - état de l'art (A2.1)
 - connaissances explicites et déclaratives (A2.1.3)
 - acquisition de connaissances (A2.2.2)

2. Filtrage et traitement de données (B)

- traitements préliminaires (B1)
 - sélection de données (B1.1)
 - conversion de données (B1.2)
 - traduction (B1.2.4)
 - extraction de données (B1.3) (*taxonomies, métadonnées*)
 - extraction manuelle (B1.3.1)
 - prétraitement de données (B1.4)
 - *data cleaning* (B1.4.1)
 - *data integration* (B1.4.2)
 - *data transformation* (B1.4.3) (*normalisation, user specific clustering*)

3. Méthodes d'analyse (C)

- évaluation des données (C1)
 - évaluation de crédibilité des informations (C1.1.1)
 - résolution des ambiguïtés linguistiques (C1.2)
 - désambiguïtation sémantique (C1.2.3)
 - désambiguïtation contextuelle (C1.2.4)
- modélisation conceptuelle (C2)
 - modélisation des données/information (C2.2)
 - *modèle conceptuel de données* (C2.2.1), *modèle logique de données* (C2.2.2), *modèle physique de données* (C2.2.3)
 - modélisation du système (C2.3)
 - modélisation scientifique (C2.1)
 - *domain knowledge modelling* (C2.1.5)

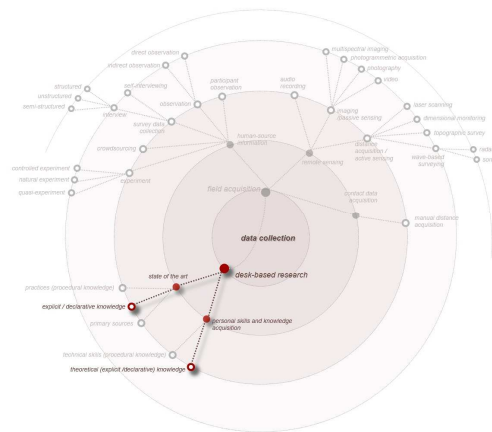


Fig. 6 Groupe d'activités (A) liées à l'acquisition des données - le schéma souligne les activités menées dans le projet MEMORIA (2015-2016).

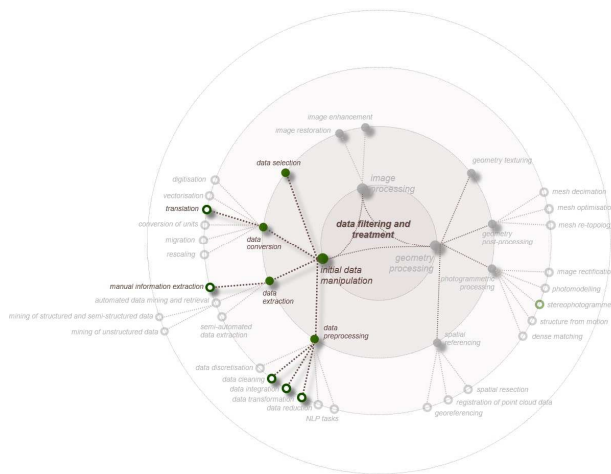


Fig. 7 Groupe d'activités (B) liées au filtrage et au traitement des données - le schéma souligne les activités menées dans le projet MEMORIA (2015-2016).

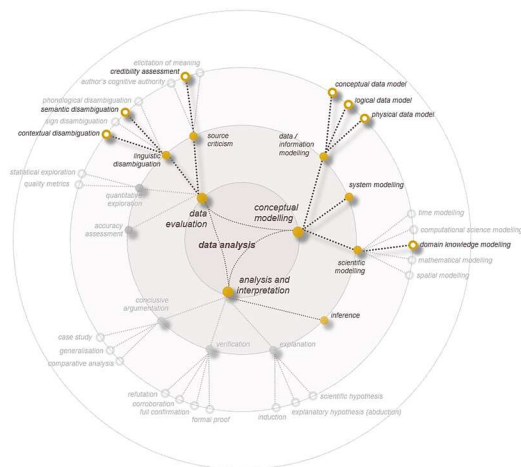


Fig. 8 Groupe d'activités (C) liées à l'analyse des données - le schéma souligne les activités menées dans le projet MEMORIA (2015-2016).

A3.1 Résultats obtenus

Tout d'abord nous avons défini les attentes, les besoins et les priorités concernant le système d'informations à construire. En quelques mots nous pouvons les résumer comme suit :

Le système doit permettre de décrire, d'enregistrer et d'archiver de manière structurée les résultats de notre travail. Cependant il ne s'agit pas ici de sauvegarder ou de décrire des fichiers contenant ces résultats - qui est un sujet différent - mais d'identifier, de structurer et de préserver des informations sur les objectifs, sur la manière (méthodes, techniques, outils, ...), et sur le contexte de création de ces résultats. De ce point de vue, un résultat que nous appelons un **output**, est l'objet central de notre système.

La façon dont un *output* a été produit doit être exprimée sous la forme d'un **processus**, lui-même composé d'une chaîne ou d'un ensemble d'**activités**. Ces dernières doivent fédérer des informations sur les individus et les organisations engagés dans la production d'un output, comme sur leur cadre formel (ex. un *projet*).

La production d'un output s'appuie inéluctablement sur des sources, des techniques, des connaissances et du savoir-faire. Nous devons être en mesure de les documenter.

Par ailleurs, chaque output a son objet d'étude (ou ses objets d'études) c'est à dire son sujet, qui mérite d'être clairement identifié.

Et enfin, une fois produit, un output – et particulièrement un output numérique - ne devient pas immuable. Pour une multitude de raisons il peut être sujet à modifications ou réarrangements. Il nous semble utile de pouvoir garder les traces des résultats de ces variations.

Naturellement tous les éléments cités ci-dessus n'ont pas un caractère obligatoire. Il y a des cas où nous ne disposons pas de toutes ces informations, comme il y a des cas où certaines d'entre elles ne sont pas pertinentes. Une des caractéristiques essentielles du système que nous voulons construire est d'assurer sa souplesse (ex. minimiser la quantité d'informations obligatoires, permettre de varier l'ordre dans lequel les informations sont renseignées).

Un autre point fondamental est la volonté de visualiser [1] le niveau d'incomplétude de renseignements rassemblés dans le système pour pouvoir analyser et mieux comprendre la nature de ces manques.

Les résultats concrets obtenus à ce jour sont discutés ci-dessous.

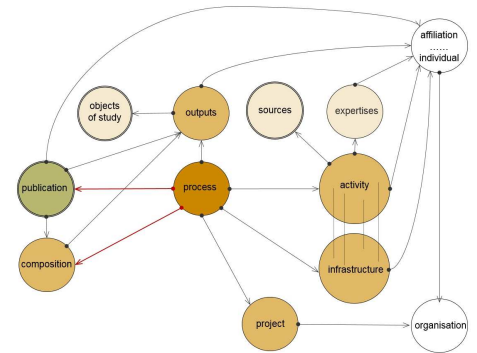


Fig. 11 Représentation schématique de la structure du système MEMORIA. (cf. Annexe 1)

Note 1. 'Visualiser' au sens qu'a ce terme en Visualisation d'Informations (InfoViz) - ...The use of abstract visual representations to aid in the analysis of quantitative or qualitative information and support a cognitive activity. ...

cf. W. Kienreich, Information and Knowledge Visualisation: an oblique view, MiaJournal vol0, 2006

Modèle conceptuel, logique et physique de la structure de la base de données

La structure de la base de données est aujourd'hui pleinement implémentée.

La phase la plus longue de développement a été le processus de construction d'un modèle conceptuel stable. (cf. Fig. 12) Comme illustré ci-dessous ce processus a été marqué par beaucoup d'activités récurrentes (ombrées sur la figure) et de chaînes d'activités récurrentes (attachées par une ligne pointillée). Les activités conduites au cours de ce processus sont en plus difficiles à organiser chronologiquement car elles se sont souvent superposées dans le temps.



Fig. 12 Construction de la base de données MEMORIA : processus et activités.

Des documents relatifs à ce travail sont disponibles dans la zone d'archivage du projet MEMORIA:

1. Usage scenario and database structure (EN), 65 pages
Memoria/RAPPORT_sept2016/02_Memoria_schemBD.pptx
2. Database structure (EN), 72 pages (cf. Annexe 1)
Memoria/RAPPORT_sept2016/01_DB_structure_2016.docx

Vocabulaire normalisé

La base de données et le système d'informations MEMORIA nécessitent un ensemble de listes fixant un vocabulaire normalisé. L'effort de normalisation qui a permis de constituer ces listes a été considérable.

Pour assurer la stabilité et la cohérence du système d'informations que nous mettons en place, nous avons identifié 20 listes de vocabulaires normalisés. (cf. Fig. 13, Tab. 1) À l'intérieur de chaque liste chaque terme est défini, exemplifié et documenté.

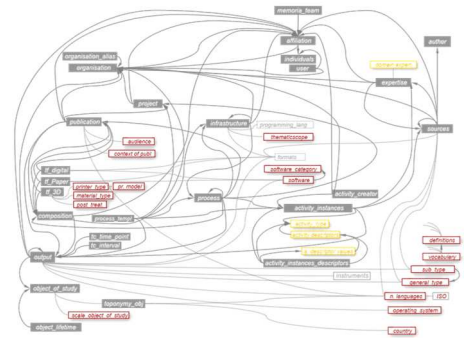


Fig. 13 Schéma de la base de données. Les listes de vocabulaire normalisé sont 'surlignées' – les listes pleinement implémentées en rouge, les listes en attente d'implémentation ou de validation en jaune.

Nom de la liste	No terms	Nom de la liste	No terms
thematic scope	10	formats	85
audience	15	instruments	49
context of publication	25	programming languages	26
software category	20	ISO 639-2	
software	140		
printer model	2		
printer type	7		
material type	11		
post treatments	8		
definitions	128		
subtype	55		
general type	10		
natural languages	19		
operating system	37		
scale of object of study	11		
country	11		

Tab. 1 Nombre de vocables pour chaque liste à ce stade.

La majorité de ces listes a d'ores et déjà été formalisée (16 listes) et validée lors de sessions de travail collectives, puis implémentée (*i.e.* les vocables et leurs définitions ont été incorporés dans la base de données). Les quatre groupes de vocables restants demandent encore un peu de travail.

Au tout début du projet nous avons fait le choix de la langue anglaise pour notre interface, nous avons donc continué nos travaux dans cette direction. Tous les termes, les définitions et les exemples sont en anglais. (cf. Annexe 3) En conséquence il nous semble qu'un effort additionnel, un effort de coordination et de vérification linguistique, sera nécessaire pour éliminer d'éventuelles erreurs de traduction résiduelles et pour éviter des ambiguïtés ou des confusions, des effets de polysémie contreproductifs.

La totalité du vocabulaire normalisé utilisé dans le projet MEMORIA sera rapidement consultable sur le site de présentation du projet. (cf. Fig. 14) L'implémentation de ces fonctionnalités a été commencée, mais l'interface n'est pas encore pleinement fonctionnelle.

<<http://memoria.gamsau.archi.fr/projet/vocabulary.php?lang=en>>

Les travaux menés sur le vocabulaire peuvent être décrits sous la forme des trois processus (cf. Fig. 15) – ces trois mêmes processus ont été nécessaires pour chaque groupe de vocables, à l'exception de la liste 'country' basée sur ISO 639-2.

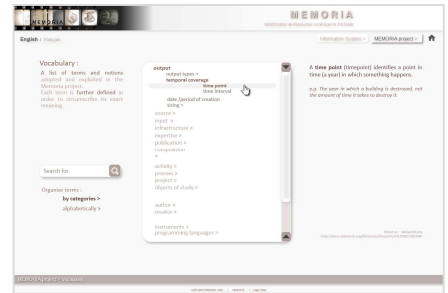


Fig. 14 Interface de consultation des vocabulaires normalisés sur le site de présentation du projet MEMORIA – telle que projetée.

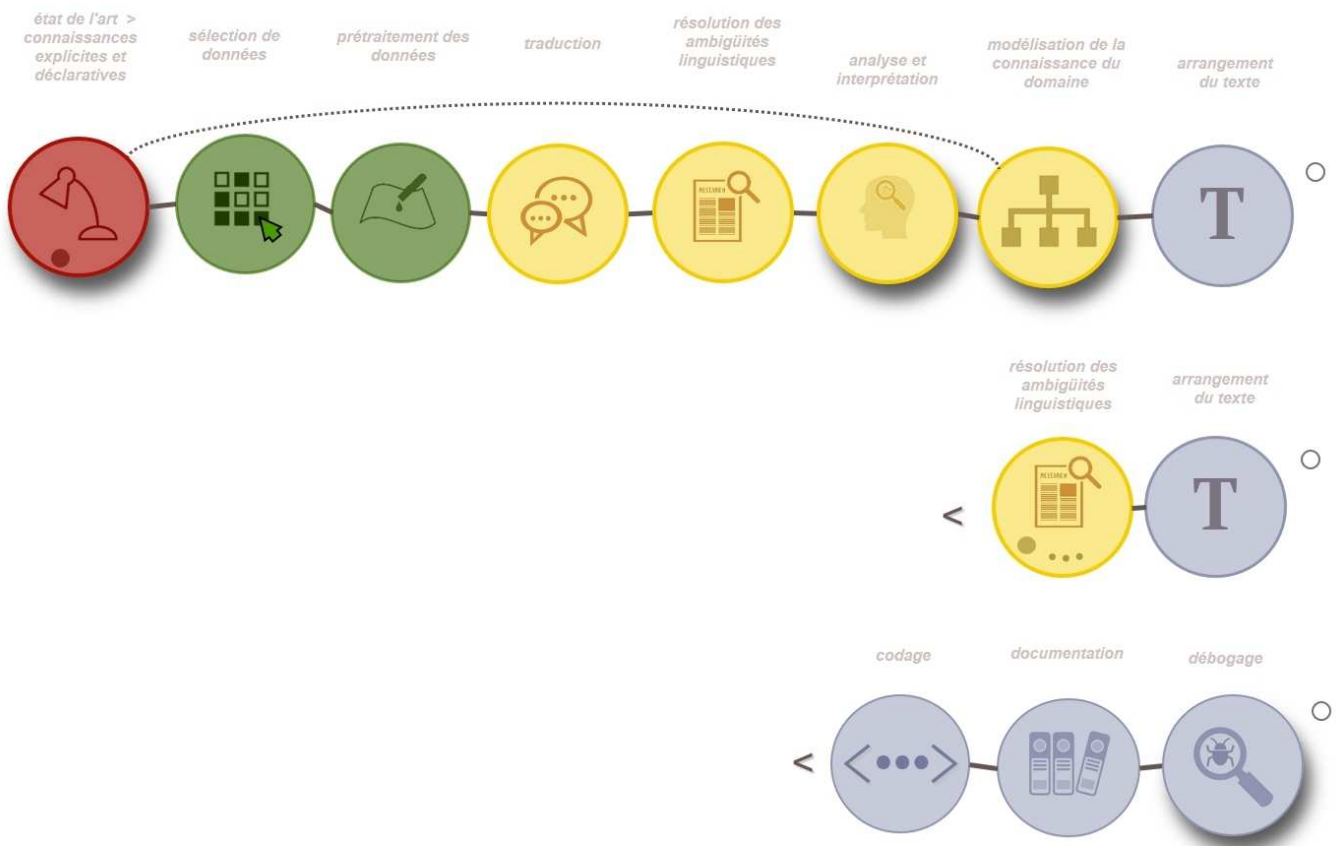


Fig. 15 Représentation schématique de la phase de normalisation du vocabulaire sous la forme de processus.

Des documents relatifs à ce travail sont disponibles dans la zone d'archivage du projet MEMORIA:

1. Lists of controlled vocabulary (EN), 48 pages
Memoria/ RAPPORT_sept2016/controlled_Voc/ControlledVocab_MEMORIA_now.docx
2. Output and Source -types and subtypes (EN), 13 pages
Memoria/RAPPORT_sept2016/Controlled_Voc/GeneralandSubTypes_MEMORIA.docx
3. Printer types and materials (EN), 3 pages
Memoria/ RAPPORT_sept2016/controlled_Voc/ok_printerT_material_paper_posttreatment.docx

4. List of instruments of MAP (EN), 7 pages
Memoria/RAPPORT_sept2016/controlled_Voc/instruments_MEMORIA.docx
5. List of scales of objects of study relevant for MAP (EN), 2 pages
Memoria/RAPPORT_sept2016/controlled_Voc/ok_scales_MEMORIA_now.docx
6. List of file type extensions (EN), 5 pages
Memoria/ RAPPORT_sept2016/controlled_Voc/fileFormats_MEMORIA.docx
7. List of event types (EN), 3 pages
Memoria/RAPPORT_sept2016/Controlled_Voc/ok_publicationContext_MEMORIA.docx
8. List of thematic scope (EN), 1 page
Memoria/RAPPORT_sept2016/controlled_Voc/ok_thematicScope_MEMORIA.docx
9. Liste of audience type (EN), 1 page
Memoria/ RAPPORT_sept2016/controlled_Voc/ok_audience_MEMORIA.doc
10. List of operating systems (EN), 1 page
Memoria/RAPPORT_sept2016/controlled_Voc/ok_operatingSystems_MEMORIA.docx
11. List of software (EN), 6 pages
Memoria/RAPPORT_sept2016/Controlled_Voc/ok_software_MEMORIA.docx
12. List of languages of the intellectual content (EN), 18 pages
Memoria/RAPPORT_sept2016/controlled_Voc/ok_language ISO.docx
13. List of programming languages (EN), 3 pages
Memoria/RAPPORT_sept2016/Controlled_Voc/programingLang_MEMORIA_now.docx
14. List of natural languages (EN), 1 page
Memoria/RAPPORT_sept2016/controlled_Voc/ok_natural_languages_MEMORIA_.docx
15. Database structure and lists controlled vocabulary (EN), 6 page
Memoria/RAPPORT_sept2016/controlled_Voc/ 03_DB_vocabulary.pptx

Identification et structuration des activités

La phase d'identification et de structuration de connaissances non encore formalisées, ou formalisées de façon imparfaite voire ambiguë, est souvent mentionné comme phase *d'élicitation des connaissances* – *i.e.* un processus au travers duquel les activités humaines sont différenciées, nommées, décrites, et finalement organisées ou modélisées (souvent sous la forme d'ontologies).

Une **activité** – telle que définie dans le projet MEMORIA - recense un ensemble d'actions mobilisées pour produire une ressource, actions dont on souhaite garder trace comme « *historique de production* ». Toutes les actions mobilisées, ou tous les détails de ces actions, n'ont pas vocation à être enregistrés.

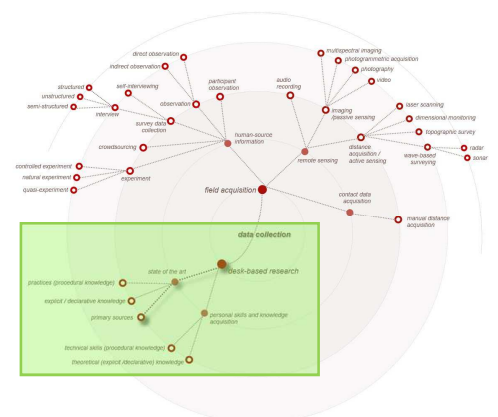
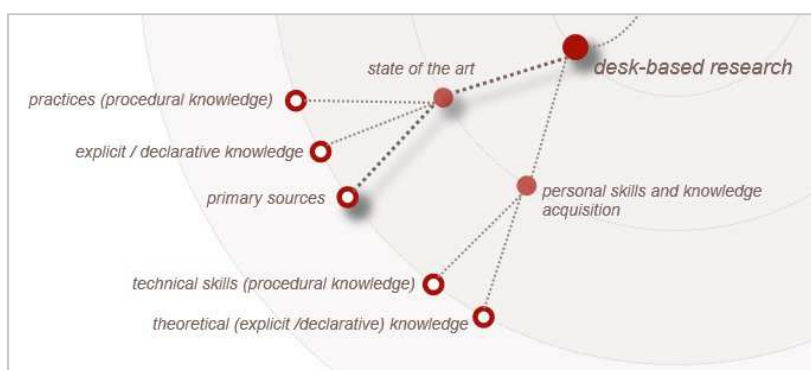
Une des étapes préliminaires et essentielles du projet était donc une phase d'élicitation des connaissances liées aux activités mobilisées pour produire une ressource, dans le cadre des travaux de notre laboratoire.

L'identification des activités et la validation de « classes » d'activités (au sens structure ontologique en informatique) se sont étalées tout au long de la durée du projet, et au moment de l'écriture du présent rapport le travail sur trois groupes d'activités a été mené à bien. Les deux derniers groupes d'activités identifiées en sont à l'étape de validation.

Les 5 groupes d'activités identifiés correspondent à des activités qui sont de natures différentes – nous en donnons ici les titres et un court descriptif en anglais (définitions telles que présentées dans le système) :

- acquisition de données (*data collection*) (cf. Fig. 16)

Gathering data which is to become the subject of further analysis, filtering and processing.



- filtrage et traitement de données (*data filtering and treatment*)

Transforming the raw data into a suitable form with regards to analysis, output production or finalisation needs, either when accessing the data for the first time, or in subsequent data steps. Editing, cleaning or modifying the raw data results in processed data.

Fig. 16 Schéma de la structure ontologique des activités du groupe acquisition de données – (à droite) vue globale de la hiérarchie, (à gauche) une vue partielle.

- méthodes d'analyse (*data analysis*) (cf. Fig. 17)
Methods of acquisition or gaining of scientific - theoretical, explicit - knowledge, as well as manners of its articulation and transmission in a formal language.
- protocoles d'exploitation de la donnée (*added value procedural activities*)
A phase of research centred on the use of procedural knowledge, such as scientific procedures and technological protocols, and implicating the use of technical skills and abilities acquired and developed by training or practice.
- finalisation (*finalisation*)
Finalisation activities correspond to those stages in a research process that are specifically undertaken in such contexts as publication, communication, dissemination, etc. These activities focus on presenting, disseminating, transmitting research results to various audiences.
Finalisation encompasses activities that lead to the reprocessing of existing outputs (modification, adjustment, reformation, optimisation, adaptation ...) or activities that lead to the creation of new ones (video capture, voice-over narration ...).

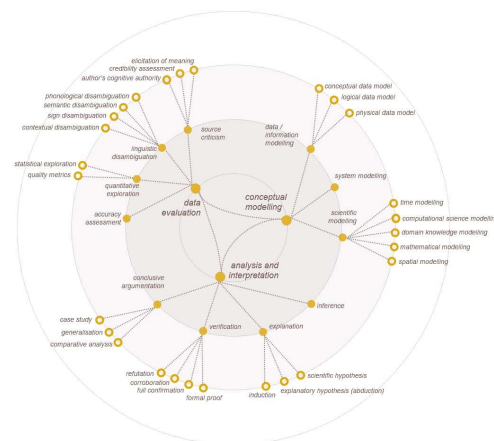


Fig. 17 Schéma de la structure ontologique des activités du groupe méthodes d'analyse.

Étape d'élicitation des connaissances – extraction et analyse des activités

La première étape, celle de l'*extraction des connaissances* consiste à rechercher et organiser les termes et définitions présents dans la littérature scientifique ou dans les pratiques de travail existantes pour fournir des *représentations verbales* correspondants aux activités et protocoles de recherche qui nous intéressent.

Dans la deuxième étape, étape *d'analyse de concepts*, un effort important a dû être entrepris afin de filtrer la liste des termes et des notions, d'interpréter et d'évaluer les relations entre les activités au niveau conceptuel et pratique. Les relations potentielles entre les activités ont été mises en évidence, et les activités appartenant à un même « sujet de préoccupation » ont été regroupées et organisées dans une structure ontologique identifiant différents niveaux de spécialisation. Les activités les plus génériques (ex. *visual outputs production*) se déclinent en activités plus spécifiques (ex. *visualisation*, puis *visual encoding*, etc.). (cf. Fig. 17, 18)

Cette étape explicite chaque activité en « *intension* » en l'associant avec diverses définitions récoltées à partir de la littérature scientifique, mais pointant vers une « *activité concept* ».

La structure ontologique finale est visualisée sous la forme d'une « *roue d'activités* » (cf. Fig. 16, 17 et Annexe 4). Une liste d'attributs est ensuite spécifiée.

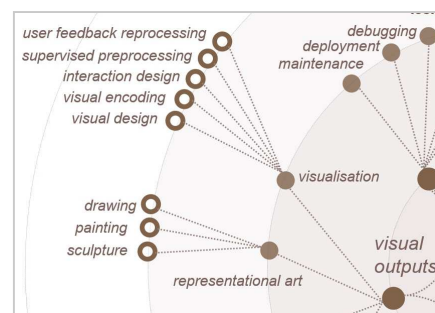


Fig. 18 Visualisation partielle de la structure ontologique des activités du groupe protocoles d'exploitation de la donnée.

La structure hiérarchique des activités doit permettre d'identifier chaque activité sans laisser la place à des ambiguïtés qui se traduiraient par la nécessité d'opérer des choix multiples au moment de la sélection d'une activité. (cf. Fig. 19) Si une telle nécessité de choix multiple apparaît, cela traduit une faute dans l'identification de l'activité – *i.e.* ce que l'on vient d'identifier n'est pas une *activité*, mais un *attribut* d'une activité.

Une activité peut être caractérisée par un certain nombre d'attributs qui décrivent différents aspects de cette activité (par exemple instruments utilisés, réglage, position) et rendent la définition d'une activité plus étroite ou plus spécifique. Nous avons appelé ces attributs les *descripteurs d'activité*.

Pour chaque descripteur il peut y avoir plusieurs valeurs (ex. activité : *acquisition photogrammétrique* > descripteur: *orientation de la caméra* > valeurs de ce descripteur: *horizontale, verticale, oblique*). (cf. Fig. 20)

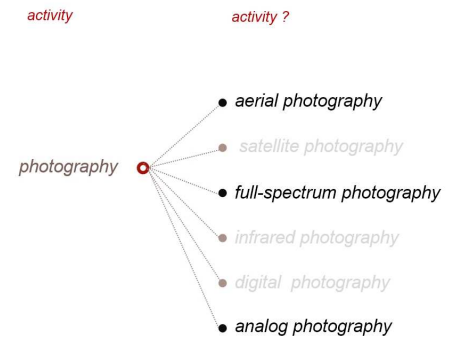
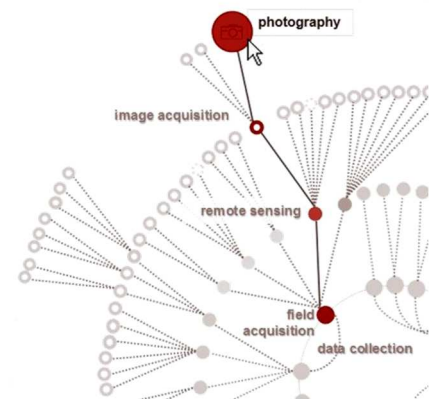
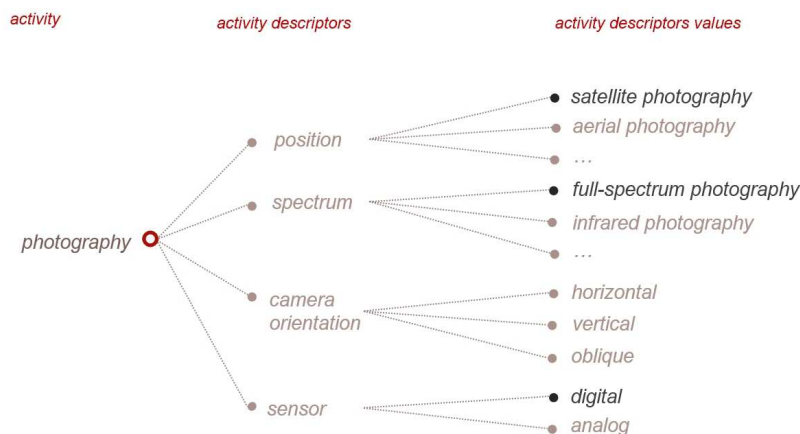


Fig. 19 Exemple d'une identification erronée d'activités - les termes identifiés à droite, ne s'excluent pas les uns les autres. Par conséquent, ils doivent être considérés comme des attributs de l'activité (à gauche) et non comme des activités plus spécifiques.



Pour les 236 activités nous avons identifié 91 descripteurs et 402 valeurs.

En bref, l'étape d'analyse de concept se déroule en trois sous-étapes récurrentes:

- le filtrage des définitions verbales de chaque activité ;
- l'organisation des relations entre activités ;
- l'identification des attributs relatifs aux activités.

Il faut préciser que ce processus s'est avéré profondément itératif, avec un certain nombre de boucles de rétroaction tout au long de la construction de la structure ontologique à travers lesquelles l'étape d'extraction supposée initiale peut être remise en question et réexaminée à tout moment.

Fig. 20 Les descripteurs rendent la définition d'une activité plus étroite ou plus spécifique. Plusieurs descripteurs peuvent être employés pour caractériser une activité. Les valeurs d'un descripteur peuvent être exclusives (ex. digital / analog) ou non.

(gauche) exemple de descripteurs et leurs valeurs pour l'activité « *photography* »

(droite) structure ontologique correspondant au niveau de spécialisation de l'activité « *photography* »

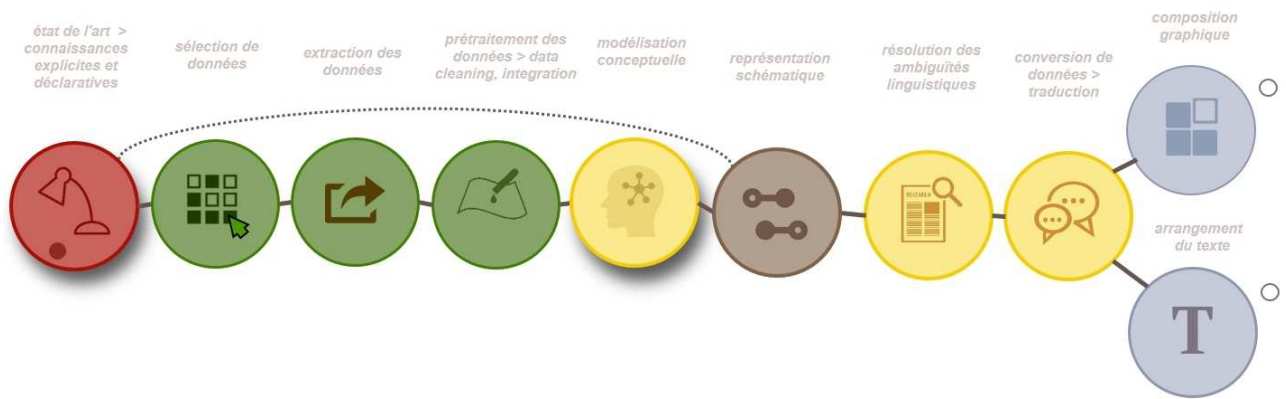


Fig. 21 Représentation schématique sous la forme d'un processus de la phase d'extraction des connaissances et d'analyse de concepts.

La troisième étape a consisté à évaluer collectivement la structure ontologique définie dans la phase précédente. Dans cette phase un *éliciteur* de connaissances et de petits groupes de participants se réunissent pour des séances d'une demi-journée de travail centrées sur des groupes d'activités ou de compétences. Ensemble, les acteurs parcourent une hiérarchie d'activités, discutent des définitions données à une activité, reformulent ou combinent ces définitions jusqu'à ce qu'un consensus soit atteint.

Des exemples sont commentés aussi, et dans de nombreux cas, les participants sont invités à contribuer en ajoutant un nouvel exemple. Cette exemplification d'une activité par les participants, à rapprocher de la notion de *définition ostensive*, est souvent absolument cruciale dans la façon dont les participants acquièrent une compréhension personnelle de ce que l'activité englobe. En cas de besoin des termes qui sont jugés comme équivalents sont cités ou désambiguïsés dans les définitions.

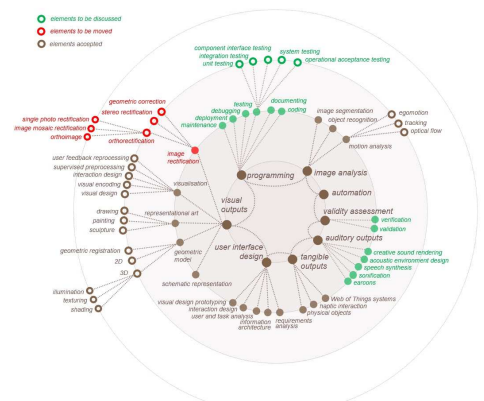


Fig. 22 Des représentations schématiques appelées « roues d'activités » ont été utilisées dans la phase d'organisation des relations entre les activités pour clarifier la position de chaque activité dans la hiérarchie et faciliter les discussions du groupe de travail.

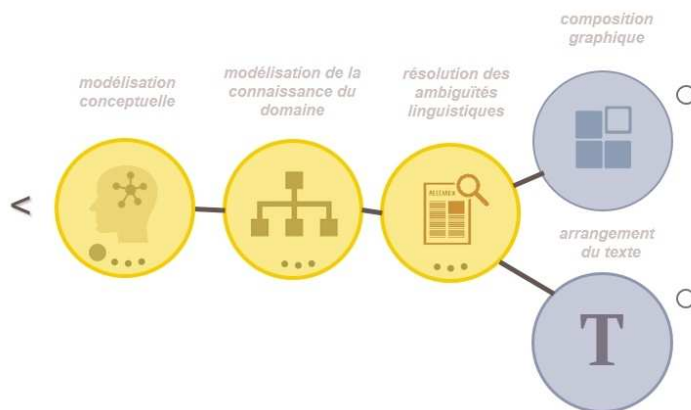


Fig. 23 Représentation schématique de la phase d'évaluation collective de la structure ontologique des activités et des descripteurs sous la forme d'un processus.

Au total, nous avons identifié, défini et structuré 236 activités.

Comme dans le cas du *vocabulaire normalisé* les *activités* sont exprimées en anglais, en conséquence un effort de vérification linguistique sera nécessaire avant la phase d'implémentation finale. Mais avant cela, un travail de relecture et de validation sera nécessaire pour tous les termes qui dans la langue courante, peuvent désigner à la fois une activité et le résultat d'une activité (ex. *'state of the art'* : confusion possible entre l'activité et son résultat).

Des documents relatifs à ce travail sont disponibles dans la zone d'archivage du projet MEMORIA:

1. activities - summary (EN), 1 page
Memoria/RAPPORT_sept2016/05_Activities.pptx
2. data collection group – hierarchy of activities, definitions, examples and descriptors (EN), 20 pages
Memoria/RAPPORT_sept2016/11_Activities_dataCollection.docx
3. data collection group – hierarchy of activities and descriptors - (EN), 9 pages
Memoria/RAPPORT_sept2016/11_Activities_dataCollection.pptx
4. data filtering and treatment – hierarchy of activities, definitions, examples and descriptors (EN), 20 pages
Memoria/RAPPORT_sept2016/12_Activities_dataTraitement.docx
5. data filtering and treatment – hierarchy of activities and descriptors (EN), 8 pages
Memoria/RAPPORT_sept2016/12_Activities_dataTraitement.pptx
6. data analysis – hierarchy of activities, definitions, examples and descriptors (EN), 27 pages
Memoria/RAPPORT_sept2016/13_Activities_dataAnalysis.docx
7. activity descriptors - (EN), 11 pages
Memoria/RAPPORT_sept2016/13_Activities_dataAnalysis.pptx
8. added value procedural activities - hierarchy of activities, definitions, examples and descriptors (EN), 40 pages
Memoria/RAPPORT_sept2016/14_Activities_AddedValueProceduralActivities.docx
9. added value procedural activities - hierarchy of activities and descriptors (EN), 13 pages
Memoria/RAPPORT_sept2016/14_Activities_AddedValueProceduralActivities.pptx
10. finalisation - hierarchy of activities, definitions, examples and descriptors (EN), 34 pages
Memoria/RAPPORT_sept2016/15_Activities_Finalisation.docx
11. finalisation - hierarchy of activities and descriptors (EN), 14 pages
Memoria/RAPPORT_sept2016/15_Activities_Finalisation.pptx
12. finalisation - hierarchy of activities and descriptors (EN), 11 pages
Memoria/RAPPORT_sept2016/08_activitiesETdescripteurs.pptx

Identification, classification et validation des champs d'expertise

Une *activité* peut être accompagnée par toutes sortes de conseils d'experts externes. Dans le projet MEMORIA, une *expertise* est considérée comme ... *une intervention [...] d'un expert externe, qui influence le contenu de l'output. ...*

Un des paramètres permettant d'identifier et de décrire cette intervention nommée *expertise* est le '*champ d'expertise*' correspondant, défini à son tour comme ... *domaine de compétences des experts ou des organisations ...* Mais comme beaucoup de notions qui semblent nettes et évidentes au premier abord, '*champs d'expertise*' pose de sérieux problèmes d'identification et de classification. Il existe plusieurs manières de classifier différentes disciplines scientifiques, artistiques ou tout simplement différentes compétences techniques.

Après une étape d'acquisition et d'analyse de l'état des connaissances dans le domaine, nous avons construit deux structures hiérarchiques correspondant à deux visions distinctes: une plutôt classique identifiant quatre grands familles - *science, ingénierie, philosophie, arts* (cf. Fig. 24), et une deuxième qui distingue les champs d'expertise en fonction de la méthode de raisonnement mobilisée (proposée par le philosophe T. Kotarbiński) (cf. Fig. 25).

Dans chacune de deux classifications il y a le même nombre de domaines de compétences (nous avons identifié, défini et structuré 199 champs d'expertises). Les différences apparaissent dans leurs définitions et dans leurs assignations dans des groupes plus ou moins génériques. L'utilisateur du système aura le choix de sélectionner une de ces deux classifications pour répertorier et visualiser les compétences des organisations et des individus. (cf. Annexe 5) Il est important de noter qu'une discipline (ex. architecture) peut en réalité renvoyer vers des champs d'expertises distribués dans des sous-hiérarchies très différentes (*architecture* > *sciences sociales* / *génie civil* / *arts visuels*)

Les étapes que nous avons mises en place pour aboutir à ce résultat ont été similaires à celles nécessitées par le travail d'extraction et d'analyse des activités – *i.e.* recueil des termes et des définitions préliminaires de la littérature scientifique et des pratiques, sélection des termes, extraction et prétraitement des données acquises, puis analyse et interprétation de ces données, et phase de modélisation de la connaissance du domaine. Ensuite la structure ontologique résultante a été représentée schématiquement, pour être réexaminée et convertie dans des formats de publication finaux. (cf. Fig. 26).

Avant la phase d'implémentation, cette double structure ontologique nécessitera encore une validation collaborative pour

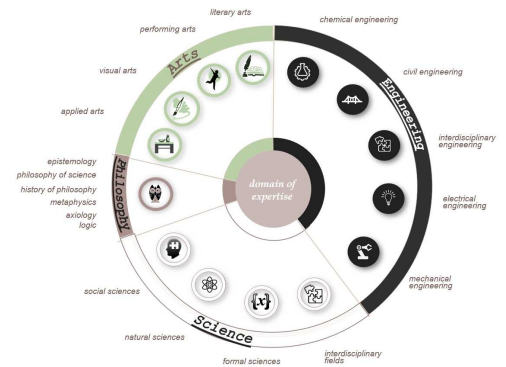


Fig. 24 Classification par quatre grands domaines de compétences.

Les domaines de compétences sont divisés en quatre catégories principales (arts, science, ingénierie, philosophie). Chaque catégorie est ensuite organisée dans une structure hiérarchique qui définit formellement des relations entre différents domaines d'expertise. Le niveau le plus élevé dans l'organisation hiérarchique correspond aux domaines avec le champ sémantique le plus large.

ex. *ingénierie* > *génie civil* > *géomatique*

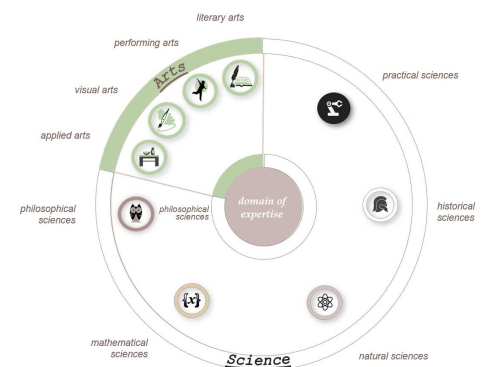


Fig. 25 Classification par la méthode de raisonnement utilisée (T. Kotarbiński)

Les domaines de compétences sont divisés en deux catégories principales (les arts et la science). Chaque catégorie est ensuite organisée dans une structure hiérarchique qui définit formellement des relations entre différents domaines d'expertise.

ex : *sciences* > *sciences philosophiques* > *épistémologie* > *philosophie de la langue*

s'assurer de l'intersubjectivité des classifications proposées et pour réduire des ambiguïtés potentielles. (cf. Fig. 26).

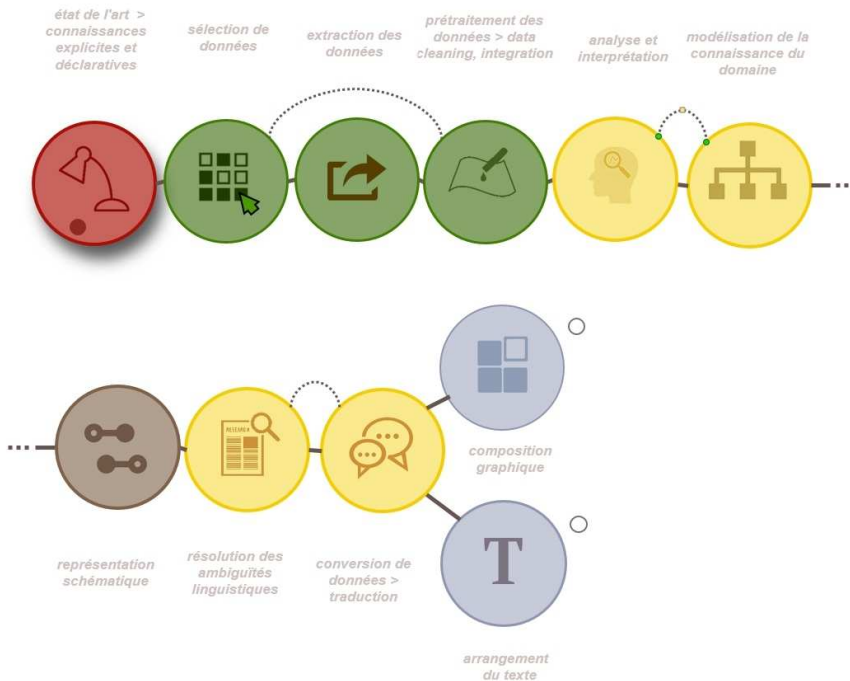


Fig. 26 Représentation schématique sous la forme d'un processus de la phase d'identification et de validation des champs d'expertise.

Des documents relatifs à ce travail sont disponibles dans la zone d'archivage du projet MEMORIA:

1. domains of expertise (EN), 48 pages
Memoria/RAPPORT_sept2016/controlled_Voc/domains_ofExpertise_MEMORIA_tab.docx
2. hierarchy of domains of Expertise (EN), 13 pages
Memoria/RAPPORT_sept2016/controlled_Voc/DomainesOfExsp_Memoria.pptx

Système d'information MEMORIA: l'interface utilisateur

Conception de l'interface – les points clés

Nous avons défini un *scénario d'utilisation* spécifique au système d'informations MEMORIA. Il se différencie des scénarios de « remplissage de formulaires chaînés » classiques par le fait qu'il n'est pas demandé à l'utilisateur de suivre un ordre systématique lors de l'enregistrement des informations. Il y a quand même certaines étapes qui nécessitent un remplissage prioritaire, mais globalement le système est prévu pour éviter d'imposer une suite de formulaires rigide.

L'interface est conçue pour permettre un remplissage du système « au fil de l'eau » : à tout moment pendant son flux de travail un acteur peut insérer les informations dont il dispose.

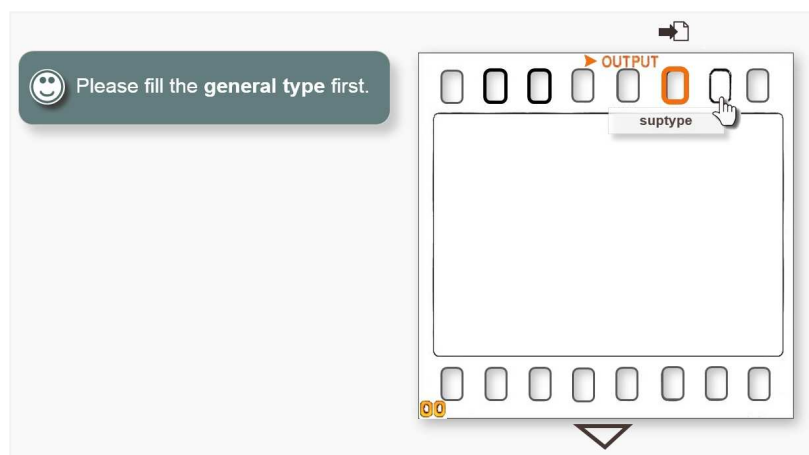


Fig. 27 Design graphique de l'interface du Système d'information MEMORIA - mode navigation.

Fig. 28 Design graphique de l'interface d'indexation d'un élément de type *output*. Les perforations dans le cadre signalent des entrées à remplir. Les champs obligatoires sont signalés par des bordures épaisses. L'ordre de remplissage des informations est libre, mais l'utilisateur sera amené à respecter des champs prioritaires. Dans cet exemple, le champ *general type* a priorité devant le champ *subtype* : le premier devra être rempli avant le second.

Dans la base de données MEMORIA il y a trois types de champs : ceux qui se remplissent *automatiquement*, ceux qui sont *obligatoires* et ceux qui sont *facultatifs*. Comme nous l'avons déjà mentionné plus haut, l'ordre dans lequel des informations sont enregistrées reste relativement libre, mais dans la limite du raisonnable (ex. des formulaires qui utilisent des données provenant d'autres formulaires vont déclencher une demande de remplissage de ces derniers). (cf. Fig. 28)

Une entrée ne peut pas être enregistrée si tous les éléments obligatoires ne sont pas renseignés. (cf. Fig. 29a)

Un remplissage de tous les champs obligatoires suffit pour valider un formulaire : le formulaire est alors sauvegardé dans l'espace personnel de l'utilisateur (*personal work zone*). (cf. Fig. 29b). À tout moment l'utilisateur pourra renseigner les champs manquants, ou choisir d'enregistrer définitivement le formulaire tel quel. (cf. Fig. 29c)

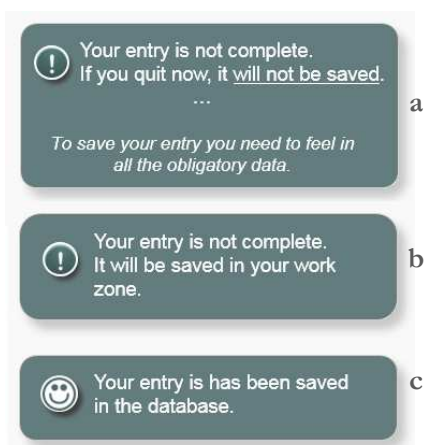


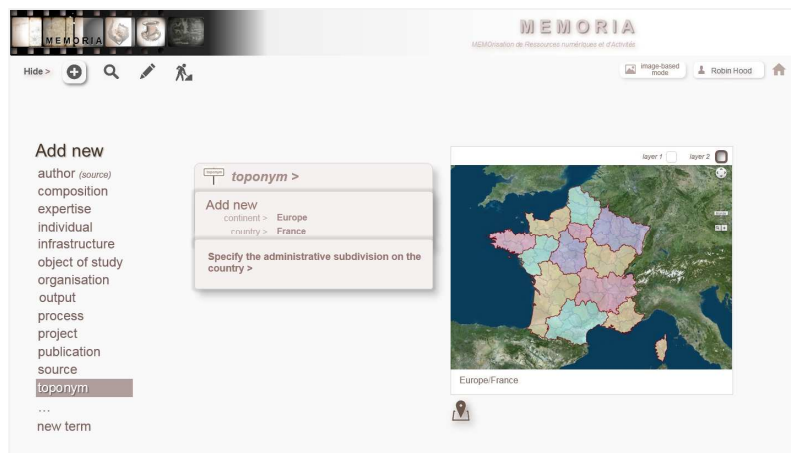
Fig. 29 Les trois types de messages du système informant sur le statut d'un enregistrement. (cf. Annexe 6e)

Modalités d'interrogation et de mise à jour du système

Le SI MEMORLA peut être interrogé sans identification d'un utilisateur en mode navigation (cf. Fig. 27, Annexe 6a), par contre l'identification est obligatoire dans le mode indexation et mise à jour du système - *add new*, *modify*, et *work zone*. (cf. Annexes 6b, 6c, 6d)

Selon les droits qui lui sont accordés, un utilisateur peut avoir un control total ou limité du système. [3]

La charte graphique de l'interface est la même pour tous les modes d'accès, mais pour des raisons de lisibilité la coloristique des formulaires est légèrement modifiée. (cf. Fig. 30)



Le processus de renseignement ou d'interrogation commence par l'identification de l'élément que l'on souhaite décrire voire modifier (ex. expertise, output, source externe, auteur, institution, projet, publication, etc.). (cf. Fig. 31)

Prenons l'exemple de l'indexation d'un nouvel output.

L'output peut être un document unique (un modèle 3D par exemple) ou un ensemble de documents (par exemple un site Web entier). Par la suite plusieurs étapes peuvent être combinées:

- identification des activités qui ont été menées ;
- enchaînement des activités pour former un processus ordonné ou non;
- repérage des activités ou des chaînes d'activités itératives à l'intérieur du processus ;
- référencement des infrastructures sous-jacentes ;
- référencement d'un processus précédent ;
- renseignement de la liste des sources externes nécessaires comme condition préalable à la réalisation des activités ;

Note 3. Nous avons défini quatre types d'utilisateurs : *administrator*, *editor*, *contributor* et *viewer*.

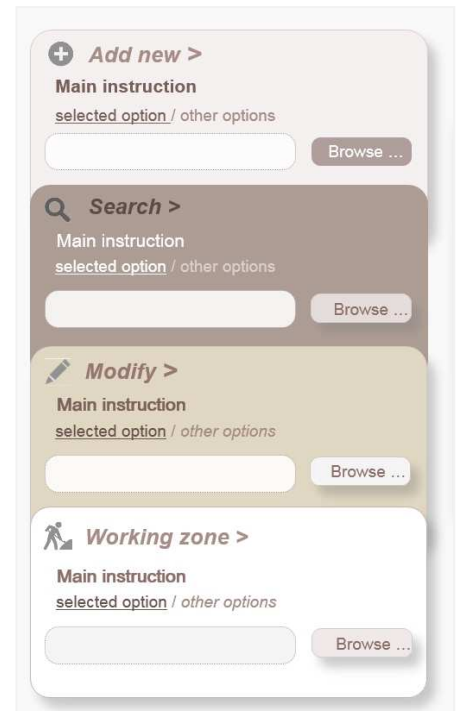


Fig. 30 Design graphique de l'interface. À gauche, mode d'indexation > *add new toponym* ; À droite, quatre gammes de couleurs choisies pour les différents modes d'accès.

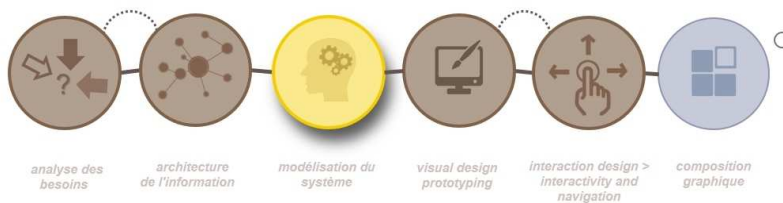


Fig. 31 Design graphique de l'interface. mode d'indexation > *add new source* ; L'interface visuelle et textuelle permettant la sélection de l'élément que l'on souhaite décrire.

- référencement des interventions extérieures (expertises) ;
- description des institutions, des projets, des acteurs et des objets d'étude (cf. Fig. 32) ;
- description des publications et des compositions dans lesquelles l'output est utilisé ou réutilisé ;
- extension des 20 listes de vocabulaire contrôlé en cas de besoin (cf. Fig. 33).

Dans un contexte applicatif où les flux de travail comprennent des décisions et des savoir-faire personnels, subjectifs ; où les processus de travail peuvent inclure des protocoles d'essai-erreur nous considérons que l'option de flexibilité choisie dans le projet MEMORIA est plus adéquate qu'une option qui serait de compter sur des processus reproductibles à l'identique pour tous les utilisateurs.

Il est important de préciser que le projet MEMORIA n'a pas vocation à assurer le suivi des activités quotidiennes du personnel : il se veut un moyen offert aux générations futures de comprendre et de réutiliser la production scientifique d'aujourd'hui - et cela fait une différence notable.



Implémentation

L'implémentation de l'interface est depuis décembre 2015 dans une phase expérimentale. Pour pouvoir progresser plus rapidement nous souhaitons associer à l'équipe un développeur Web expérimenté et efficace.

Sans ces compétences dans le projet, seule la partie conceptuelle du projet peut encore avancer. (cf. Annexe 6f)

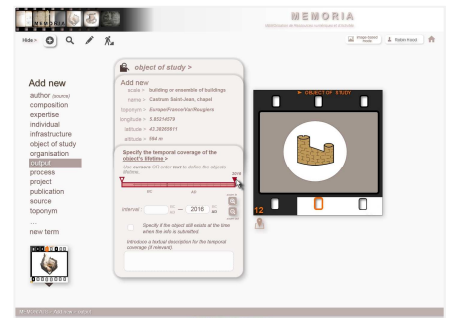


Fig. 32 Design graphique de l'interface. mode d'indexation > add new object of study.

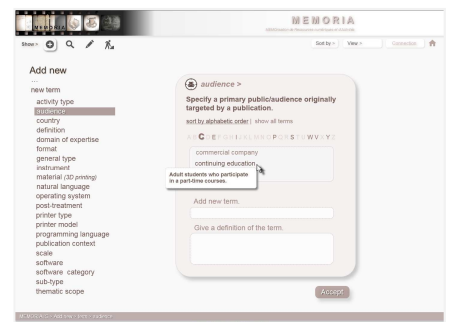


Fig. 33 Design graphique de l'interface. mode d'indexation > add new term > audience.

Fig. 34 Représentation schématique sous la forme d'un processus de la phase de conception de l'interface.

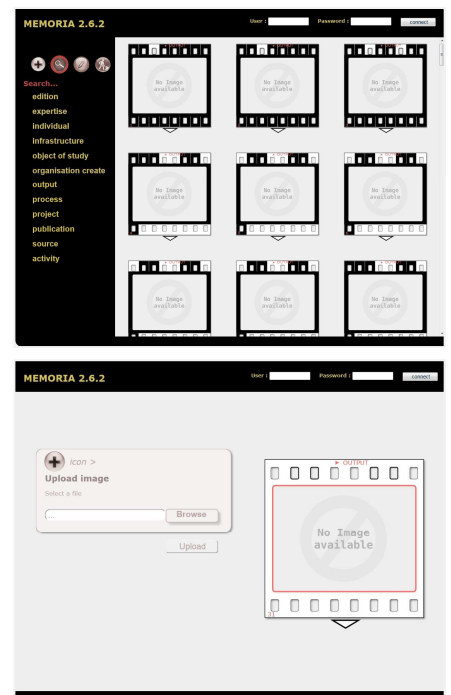


Fig. 35 Captures d'écran de la mise en œuvre de test.

Des documents relatifs à ce travail sont disponibles dans la zone d'archivage du projet MEMORIA:

1. Usage scenario (EN), 54 pages
Memoria/RAPPORT_sept2016/03_Memoria_Scenario1.pptx
2. MEMORIA IS and usage scenario (EN), 120 pages
Memoria/RAPPORT_sept2016/03_Memoria_system_DEM2.pptx
3. Addition mode - add new term, add new output (EN), 125 pages
Memoria/RAPPORT_sept2016/03_Memoria_diapo_interface2b_AddNewTerm_Output.pptx
4. Addition mode - add new object of study, add new toponym (EN), 198 pages
Memoria/RAPPORT_sept2016/03_Memoria_interfaceCART_addNewObjectofStudy_Toponym.pptx
5. Users workzone (EN), 37 pages
Memoria/RAPPORT_sept2016/03_Memoria_diapo_interface2c_workzone.pptx
6. System modes and messages (EN), 19 pages
Memoria/RAPPORT_sept2016/03_Memoria_SYSTMessages.pptx
7. Implémentation: état d'avancement (FR), 5 pages
Memoria/RAPPORT_sept2016/00_EtatdAvancement_implementation.pptx

Un langage visuel pour l'analyse de motifs et d'exceptions

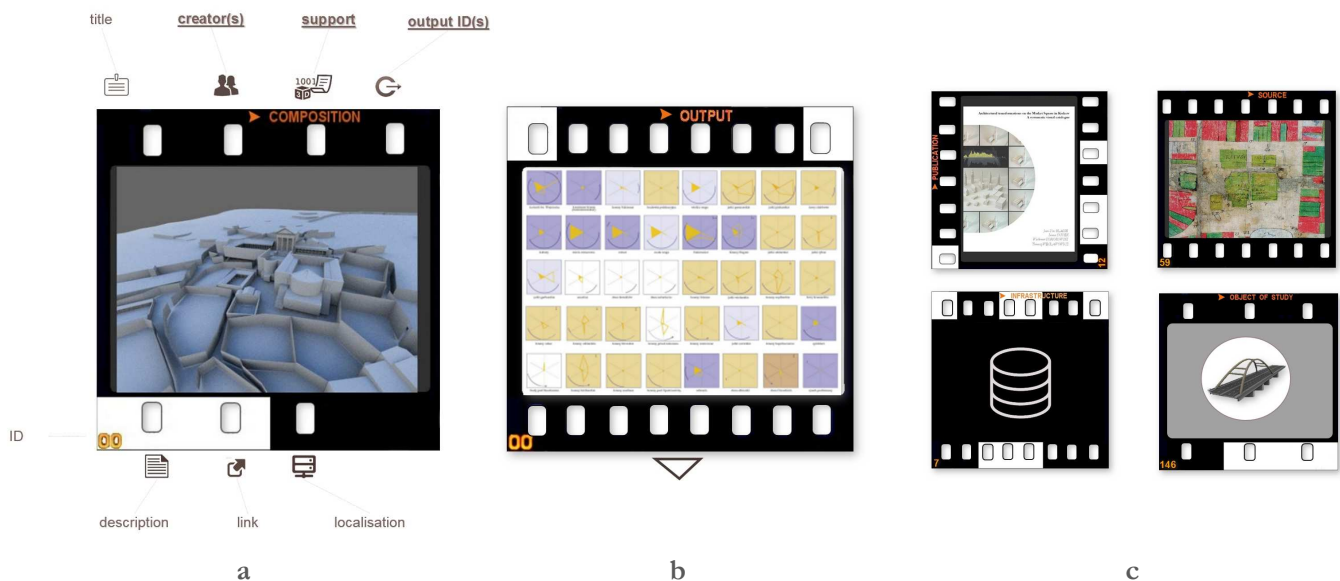
Un des aspects cruciaux dans le projet MEMORIA est le développement d'une interface visuelle qui puisse améliorer la lisibilité des requêtes et faciliter la lecture et l'analyse des résultats. (cf. Fig. 36)

Ce que l'on entend ici par *langage visuel* est le résultat d'un effort spécifique que les concepteurs de SI peuvent faire afin d'assurer la cohérence entre d'une part la sémantique derrière le SI, et d'autre part, les modalités offertes aux utilisateurs en terme d'interaction.

Notre première priorité a été la conception d'une métaphore visuelle, qui reprend l'image d'une pellicule de film analogique. Cette métaphore est utilisée pour transmettre un ensemble d'informations de façon synthétique. (cf. Fig. 37b, Annexe 7a) Elle résume et communique visuellement un état du système (*i.e.* niveau d'avancement dans l'effort d'indexation, manques en terme de documentation – que ces manques soient transitoires ou définitifs). La métaphore du film marque l'identité visuelle du projet, elle est réutilisée dans la plupart des composants du système – un output, une source, une composition, une infrastructure, *etc.* (cf. Fig. 37)



Fig. 36 Interface - mode d'indexation.



Les perforations du film correspondent aux attributs d'un composant - la « perforation » apparaît sur un fond blanc jusqu'à ce que l'attribut soit renseigné par le contributeur. Un événement *onmouseover* sur chaque perforation déclenche l'ouverture d'un formulaire correspondant à l'attribut.

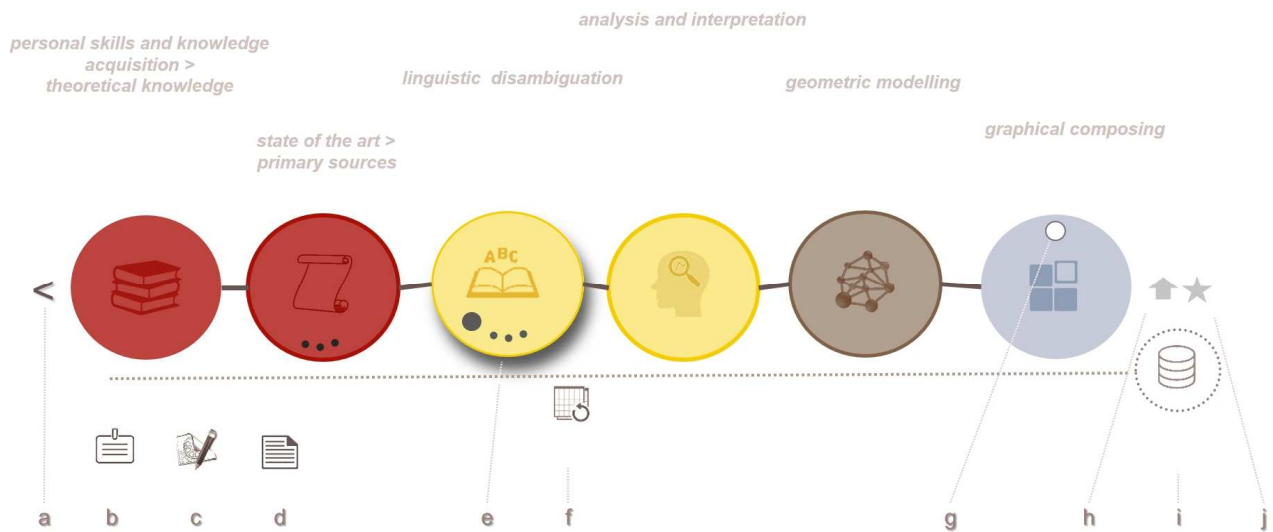
La métaphore du film perforé transmet des informations sur le niveau d'avancement atteint dans la description des outputs, des projets, des expertises, des compositions, des publications, des objets d'études, des infrastructures, des sources et des toponymes. Elle ne concerne pas les activités en tant que telles : les activités

Fig. 37 Plusieurs utilisations de la métaphore du « film perforé » pour résumer le niveau d'avancement dans l'effort d'indexation.

- (a) une composition,
- (b) un output - la flèche vers le bas est laissée vacante, soulignant le fait que l'output n'a pas encore été associé à un processus,
- (c) d'autres composants - une publication, une source externe, une infrastructure, un objet d'étude.

et les processus sont visualisés par des glyphes qui réutilisent les codes de couleur mentionnés dans la section A.3 (correspondant aux catégories d'activités).

Fig. 38 Un processus est défini comme une chaîne ou simplement comme un ensemble d'activités.



Chaque activité contribuant à un processus est représentée par une icône multidimensionnelle qui montre la catégorie de l'activité (couleur) et sa position dans la sous-hiérarchie de la catégorie (glyphe). Les icônes représentant un ensemble d'activités sont groupées et ordonnées si un ordre peut être identifié (y compris les liens potentiels aux processus précédent). (cf. Fig. 38, Annexe 7b)

Exemple de processus ordonné, s'appuyant sur les résultats d'un processus précédent et sur une infrastructure externe. Trois types de résultats issus de ce processus ont été enregistrés : extrait(s), publication(s) et composition(s). a – processus précédent, b – nom du processus, c – projet dans lequel le processus actuel a été réalisé, d – remarques, e – une activité récurrente, f – le processus a entraîné une modification de la structure de l'infrastructure externe, g – lien vers les extraits issus du processus, h - lien vers les publications issues du processus, i – informations sur l'infrastructure qui a été engagée dans le processus, j - lien vers les compositions issues du processus ;

Un ensemble de symboles complète le vocabulaire graphique associé à la représentation des chaînes d'activités permettant de repérer visuellement des particularités d'un processus (ex. chaînes ou activités itératives). (cf. Fig. 39).

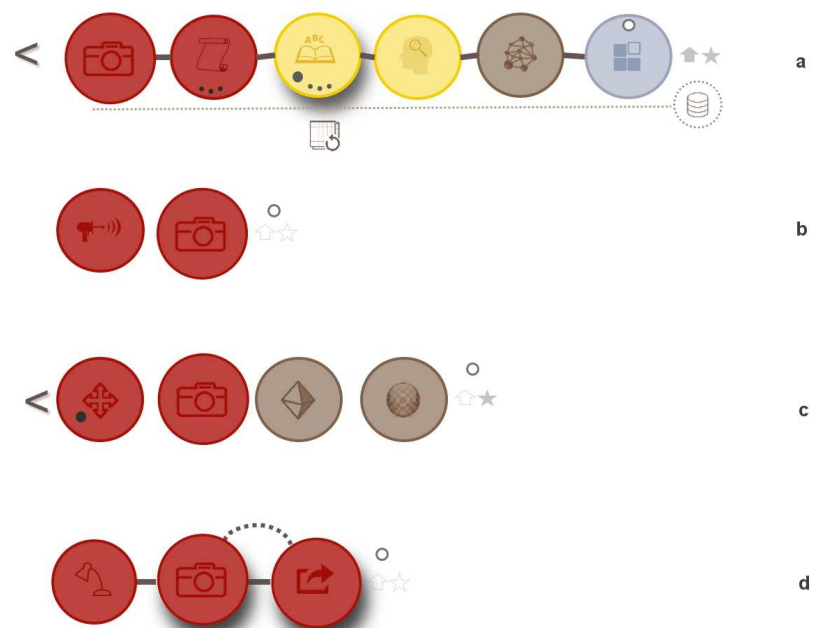


Fig. 39 Quatre processus différents - des processus ordonnés (a, d) et non ordonnés (b, c), utilisant l'infrastructure externe (a), s'appuyant sur les résultats d'un processus précédent (a, c), contenant une activité récurrente (a) ou une chaîne récurrente d'activités (d), avec des extraits attachés à une activité précise (a) ou au processus entier (b, c, d), avec publications comme résultats (a), avec compositions comme résultats (a, b);

Différents processus peuvent donc être comparés visuellement, récapitulant d'une manière visuelle l'ensemble des données sur la façon dont les résultats ont été produits – par exemple le nombre d'étapes, les activités itératives, *etc.*

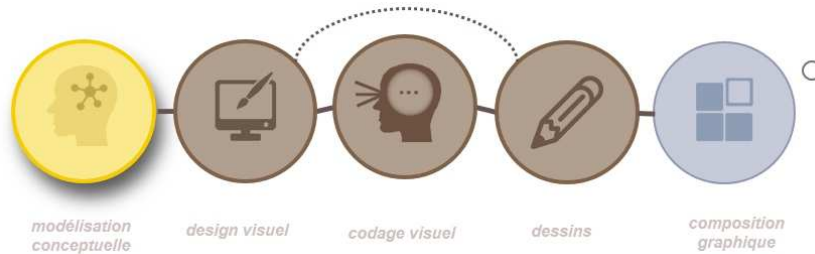


Fig. 40 Représentation schématique de la phase de la conception des outils visuels sous la forme d'un processus.

D'autres formalismes visuels sont introduits afin de représenter l'ensemble des informations concernant les organisations et les individus. (*cf.* Fig. 41 et 42)



Fig. 41 Le formalisme visuel utilisé pour résumer des informations concernant des organisations.

L'épaisseur du livre indique le nombre de projets auxquels l'organisation a participé (décrits dans le système) et permet de les consulter.

Les codes de couleur utilisés pour distinguer les catégories d'activités sont réemployés comme couleurs des marque-pages.

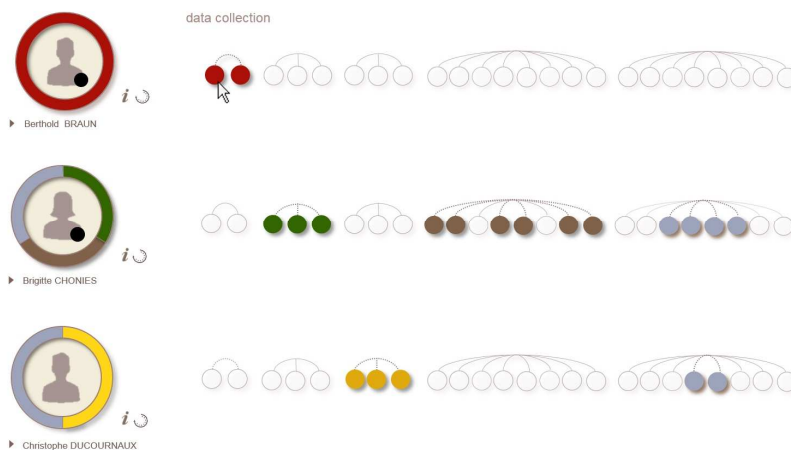


Fig. 42 Le formalisme visuel visant à faciliter la lecture de l'évolution des activités développées par des acteurs – comparaison simplifiée de plusieurs individus.

Noter la réutilisation des codes de couleur utilisés pour distinguer les catégories d'activités.

Un des objectifs principaux de l'approche MEMORIA est d'aider à résumer visuellement l'évolution des méthodes, techniques et outils utilisés au fil du temps, ainsi que les types d'activités mobilisées pour produire un output, la migration thématique des équipes ou des individus, l'histoire d'une coopération, *etc.*

En ce sens, l'approche MEMORIA, et en particulier ses étapes de visualisation, peut contribuer à renouveler notre compréhension des contraintes temporelles liées à chaque activité, des impasses techniques, des choix et évolutions méthodologiques, *etc.*

Icônes

L'interface visuelle du système d'informations MEMORIA nécessite un grand nombre d'icônes (des icônes classiques «système» et des icônes dessinées spécifiquement pour tous les outils visuels).

Les icônes doivent respecter la charte graphique et maximiser le '*data-ink*'. Pour le moment nous avons identifié et dessiné plus de trois cents pictogrammes, mais la majorité des activités et des champs d'expertise reste à iconifier.

La totalité de ces icônes doit être reformatée pour les besoins de l'interface finale.

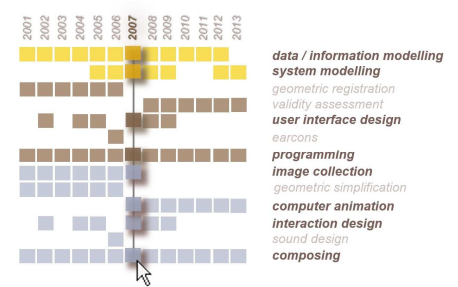


Fig. 43 Un outil visuel visant à faciliter la lecture de l'évolution des activités développées par une organisation, un acteur, dans un projet, *etc.* – comparaison simplifiée sur plusieurs années.

Des documents relatifs à ce travail sont disponibles dans la zone d'archivage du projet MEMORIA:

1. MEMORIA IS and usage scenario (EN), 120 pages
Memoria/RAPPORT_sept2016/03_Memoria_system_DEM2.pptx
2. Addition mode - add new term, add new output (EN), 125 pages
Memoria/RAPPORT_sept2016/03_Memoria_diapo_interface2b_AddNewTerm_Output.pptx
3. Addition mode - add new object of study, add new toponym (EN), 198 pages
Memoria/RAPPORT_sept2016/03_Memoria_interfaceCART_addNewObjectofStydy_Toponym.pptx
4. System modes et messages (EN), 19 pages
Memoria/RAPPORT_sept2016/03_Memoria_SYSTmessages.pptx
5. Pictogrammes
Memoria/RAPPORT_sept2016/ic_ons

Site Web de présentation du projet MEMORIA

La conception et la réalisation

Pour présenter le projet MEMORIA j'ai conçu un Site Web dont la réalisation a été confiée dans un premier temps à un stagiaire de l'Institut Universitaire de Technologie d'Aix-Marseille (H. Hout) puis à un assistant ingénieur en CDD (F. Courouble). (cf. Annexe 2)

Ce Site contient des informations (en anglais et en français) sur :

- les objectifs du projet MEMORIA ;
- l'équipe engagée dans le projet ;
- les collaborations institutionnelles ;
- le soutien financier au projet ;
- les publications liées au projet;
- le vocabulaire - liste de termes et de notions adoptée et exploitée dans le projet ;
- les personnes à contacter.

Il est actuellement accessible et consultable en ligne.
<<http://memoria.gamsau.archi.fr/projet/>>



Fig. 44 La page d'accueil du projet MEMORIA permet à l'utilisateur de sélectionner soit le Site de présentation du projet, soit le système d'informations.
<<http://memoria.gamsau.archi.fr/projet/>>



a



b



c

La dernière partie (*i.e. le vocabulaire*) est fournie uniquement en anglais puisqu'elle présente les termes et leurs définitions tels qu'ils vont apparaître dans le Système d'Informations – *i.e.* en anglais. Cette partie du Site est encore en phase de développement.

Fig. 45 Sélection de pages du site.

- a) présentation des objectifs du projet
b) présentation de l'équipe engagée dans le projet
c) adresse des responsables du projet et plan d'accès

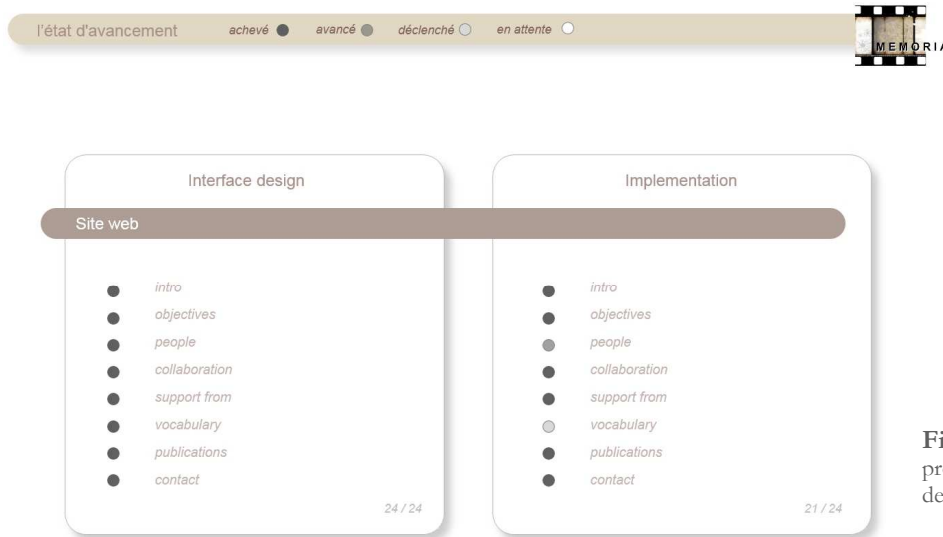


Fig. 46 Site Web de présentation du projet MEMORIA – état d'avancement des travaux.

Des documents relatifs à ce travail sont disponibles dans la zone d'archivage du projet MEMORIA:

1. Site Web de présentation (EN/FR), 29 pages
Memoria/RAPPORT_sept2016/03_Memoria_diapo_interface2a_siteWeb.pptx
2. Texte de présentation (EN/FR), 4 pages
Memoria/RAPPORT_sept2016/SiteWeb/objectivesText_18_04.docx
3. Site Web de présentation – cahier des charges détaillé (EN), 14 pages
Memoria/RAPPORT_sept2016/SiteWeb/Memoria_structure02V.pptx
4. Plan d'accès (FR/EN), 1 pages
Memoria/RAPPORT_sept2016/SiteWeb/planFR_EN.ppt
5. Images
Memoria/RAPPORT_sept2016/SiteWeb/img

A.2 - Publications

publications dans des revues avec comité de lecture

I. Dudek, J.Y. Blaise, L. De Luca, L. Bergerot, N. Renaudin
How Was This Done? An Attempt at Formalising and Memorising a Digital Asset's Making-of.

Proceedings of the 2nd International Congress on Digital Heritage 2015,
Vol. 2, Assessment of Methodologies and Tools in DH,
IEEE Computer Society, 2015, ISBN: 978-1-5090-0254-2, pp. 343 - 346
(cf. Annexe 9)

<<http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=7419519>>

communications à des colloques (publiés), congrès ou symposiums

MEMORIA

Présentation du projet au séminaire *L'imagerie 3D au service de la connaissance et de la gestion du patrimoine*,
Marseille, 30 Octobre 2015 (L. De Luca)

séminaires, workshops

MEMORIA projet

Présentation du projet aux responsables du CNP (Centre National de Préhistoire)
Marseille, 08 Septembre 2015 (I. Dudek)

A.3 - Ressources en ligne

- Site Web de présentation du projet MEMORIA (anglais/français) (Fig. 47)
<<http://memoria.gamsau.archi.fr/projet/>>
- MEMORIA portail de développement (cf. Fig. 48)
- Base de données en accès local (cf. Fig. 49)
<<http://boh30.gamsau.archi.fr/phpmyadmin/index.php>>



Fig. 47 Site Web de présentation du projet MEMORIA - page présentant les objectifs du projet.



Fig. 48 Capture d'écran de la mise en œuvre de test (portail de développement).

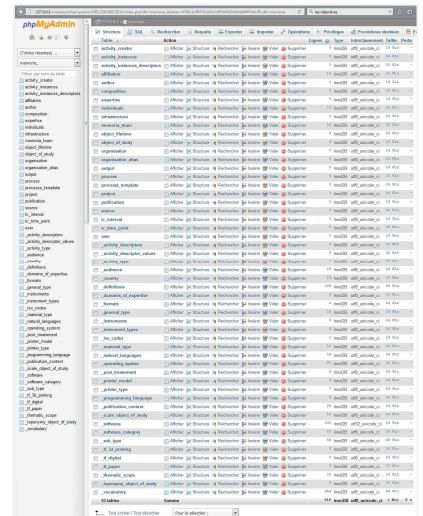


Fig. 49 Capture d'écran de la base de données MEMORIA.

A.4 - Collaborations et liens scientifiques

AUTh - School of Spatial Planning and Development, Faculty of Engineering, Aristotle University of Thessaloniki, EL

domaine : gestion de l'information géographique

nature de l'activité : accueil de stagiaires



IUT AM - l'Institut Universitaire de Technologie d'Aix-Marseille, FR

domaine : informatique

nature de l'activité : accueil de stagiaires



CNP - Centre National de Préhistoire, FR

domaine : gestion de ressources numériques

nature de l'activité : expertise



RAPPORT D'ETAPE (2015-2016)

ANNEXES

Septembre 2016

Iwona DUDEK-BLAISE
UMR CNRS/MCC 3495 MAP



ANNEXE 1A	STRUCTURE DE LA BASE DE DONNÉES – SCHÉMA.....	3
ANNEXE 1B	STRUCTURE DE LA BASE DE DONNÉES – TABLES ET ATTRIBUTS.....	4
ANNEXE 2	SUJET DU STAGE « IMPLÉMENTATION D'UNE PLATEFORME WEB »	5
ANNEXE 3	MEMORIA – DÉFINITION DES NOTIONS PRINCIPALES	6
ANNEXE 4A	IDENTIFICATION ET STRUCTURATION DES ACTIVITÉS.....	7
ANNEXE 4B	IDENTIFICATION ET STRUCTURATION DES DESCRIPTEURS DES ACTIVITÉS.....	8
ANNEXE 5	IDENTIFICATION ET STRUCTURATION DES CHAMPS D'EXPERTISE.....	9
ANNEXE 6A	INTERFACE UTILISATEUR – MODE D'INDEXATION.....	10
ANNEXE 6B	INTERFACE UTILISATEUR – MODE D'INDEXATION	11
ANNEXE 6C	INTERFACE UTILISATEUR - MODE DE MODIFICATION	12
ANNEXE 6D	INTERFACE UTILISATEUR – WORK ZONE	13
ANNEXE 6E	INTERFACE UTILISATEUR – MESSAGES DU SYSTÈME	14
ANNEXE 6F	INTERFACE UTILISATEUR – IMPLÉMENTATION.....	15
ANNEXE 7A	LANGAGE VISUEL – MÉTAPHORE DU FILM PERFORÉ	16
ANNEXE 7B	LANGAGE VISUEL – PROCESSUS	17
ANNEXE 8	ÉTAT D'AVANCEMENT DU PROJET MEMORIA - SEPTEMBRE 2016.....	18
ANNEXE 9	PUBLICATION.....	19

Annexe 1b Structure de la base de données – tables et attributs

MEMORIA/RAPPORT_sept2016/02_Memoria_schemBD.pptx



Figure 2 : Schéma de la base de données – tables et attributs

Annexe 3 MEMORIA – définition des notions principales

MEMORIA/RAPPORT_sept2016/controlled_Voc/controlledVocab_MEMORIA_now.docx

<i>term</i>	<i>definition</i>	<i>example</i>
activity	An activity identifies a series of actions undertaken in order to produce resources all along a project's workflow. When the resource is considered for storage in the MEMORIA infrastructure (this is an analyst's decision), this resource is called an <i>output</i> .	<i>e.g. imaging, phonological disambiguation, data conversion, graphical composing</i>
output	A resource, usually digital or digital-born, resulting from one or more <i>activities</i> stored in form of a <i>process</i> . An output can be a simple document (a screenshot, a 3D model, a video, <i>etc.</i>) or a set of documents (a collection of models, charts, <i>etc.</i>). Categories of outputs are differentiated through a set of classes defining types and subtypes. It can be associated with one or more objects of study. Outputs reused in subsequent activities are named ' <i>inputs</i> '.	<i>e.g. a screenshot, a 3D model, a video, a collection of models, a collection of charts</i>
process	The notion of process is used to represent the <i>chain of activities</i> mobilised to produce an <i>output</i> , a <i>publication</i> or a <i>composition</i> . A process may include only one activity. The order of execution of the various activities within a process can be specified or left unspecified. The same activity can contribute to several independent processes.	<i>e.g. successive steps within a survey campaign, that lead to the production of a 3D point cloud.</i>
input	An activity can be based on one or more <i>sources</i> (external resources), as well as it can be based on one or more <i>outputs</i> produced previously and described by a separate <i>process</i> . To avoid confusion the outputs used in an activity as initial root elements will be called inputs.	<i>e.g. a 3D reconstruction of a Greek temple produced previously, reused in a 3D printing process</i>
publication	One or several outputs delivered to a given public in the context of a formal event. A publication is always public and context driven. It combines outputs according to an objective defined by: a context of publication, an audience, final users, a project, or a technological framework.	<i>e.g. a journal article, a book, a commercial CD-ROM, an on-line website, a conference poster, etc.</i>
composition	A composition corresponds to the results of the steps in a workflow where one or several previously produced outputs can be combined or revised. It differs from a <i>publication</i> by the fact that it does not necessarily target a given public, a given context. It should not be considered an official or formal end product. Such results can be digital documents (e.g. a digital image) or physical objects (e.g. a physical model, a paper poster).	<i>e.g. printing of proof-of-concept 3D physical models, presentations supporting informal workgroup discussions, etc.</i>
object of study	Any natural or manufactured object of heritage significance that has been studied and represented in an <i>output</i> . Each object of study is designated by one of eleven categories corresponding to the concept of scale (<i>movable objects, earthworks, grottos and mines, architectural components, urban fabric and its components, ...</i>). Objects of study are characterised by some basic taxonomical and geographical information, as well as temporal indicators, allowing to query the system on space+time search criteria.	<i>e.g. ensemble of edifices of the Main Market Square in Krakow, Pont Saint-Bénézet – Avignon, pieces of furniture of the Trianon, Versailles, etc.</i>

Annexe 4a Identification et structuration des activités

MEMORIA/RAPPORT_sept2016/05_Activities.pptx

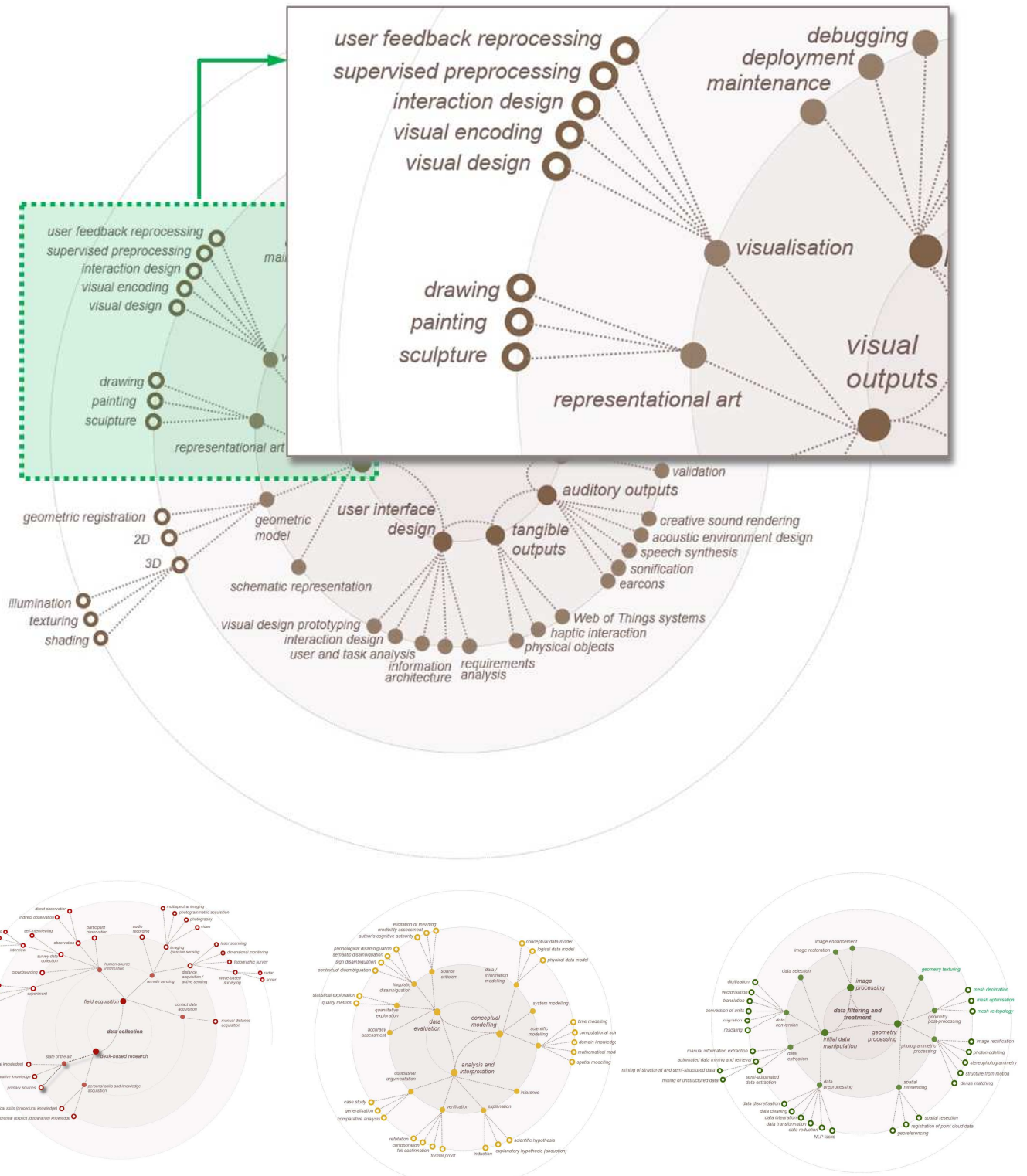


Figure 4 : Structure ontologique représentant les activités – visualisation sous la forme d’une « roue d’activités »

Annexe 5 Identification et structuration des champs d'expertise

MEMORIA/RAPPORT_sept2016/05_Activities.pptx

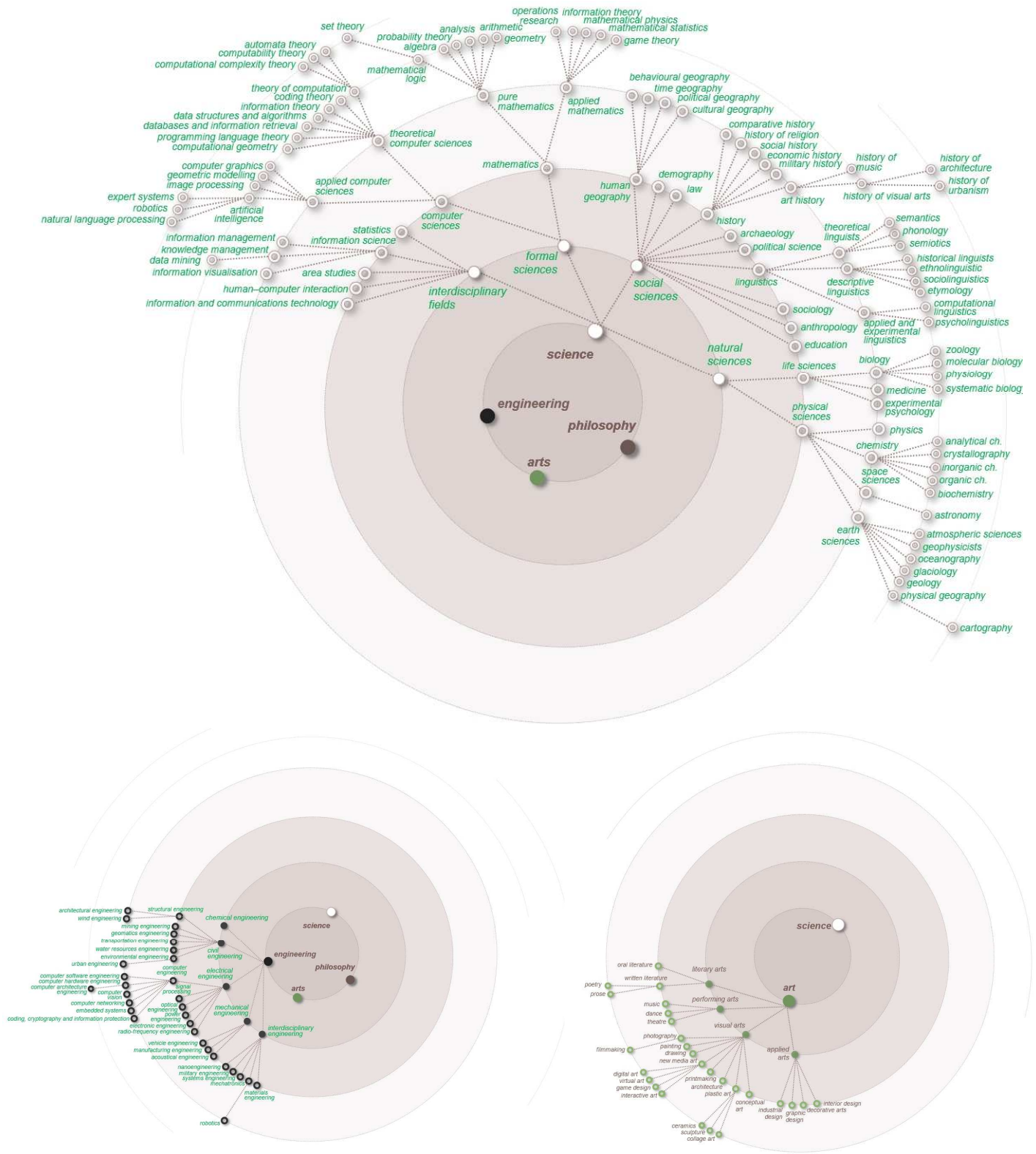


Figure 6 : Classification par quatre grands domaines de compétences (haut et gauche), classification par la méthode de raisonnement utilisée (bas droite).

Annexe 6a Interface utilisateur – mode d'indexation

MEMORIA/RAPPORT_sept2016/03_Memoria_diapo_interface2c_workzone.pptx



Figure 7 : Interface MEMORIA dans le mode d'indexation. (ici une modalité accessible sans nécessité d'identification).

En haut, recherche simple – toutes les éléments sont montrés (i.e. *output*, *composition*, *source*, *infrastructure*, ...).

En bas, le volet latéral est caché, le menu permettant un filtrage des données est activé.

Annexe 6b Interface utilisateur – mode d'indexation

MEMORIA/RAPPORT_sept2016/03_Memoria_diapo_interface2b_AddNewTerm_Output.pptx

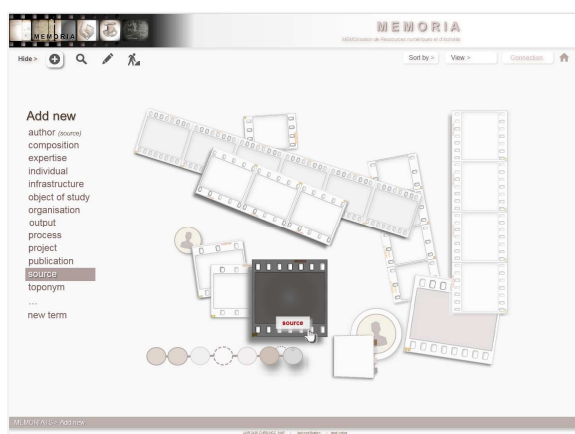
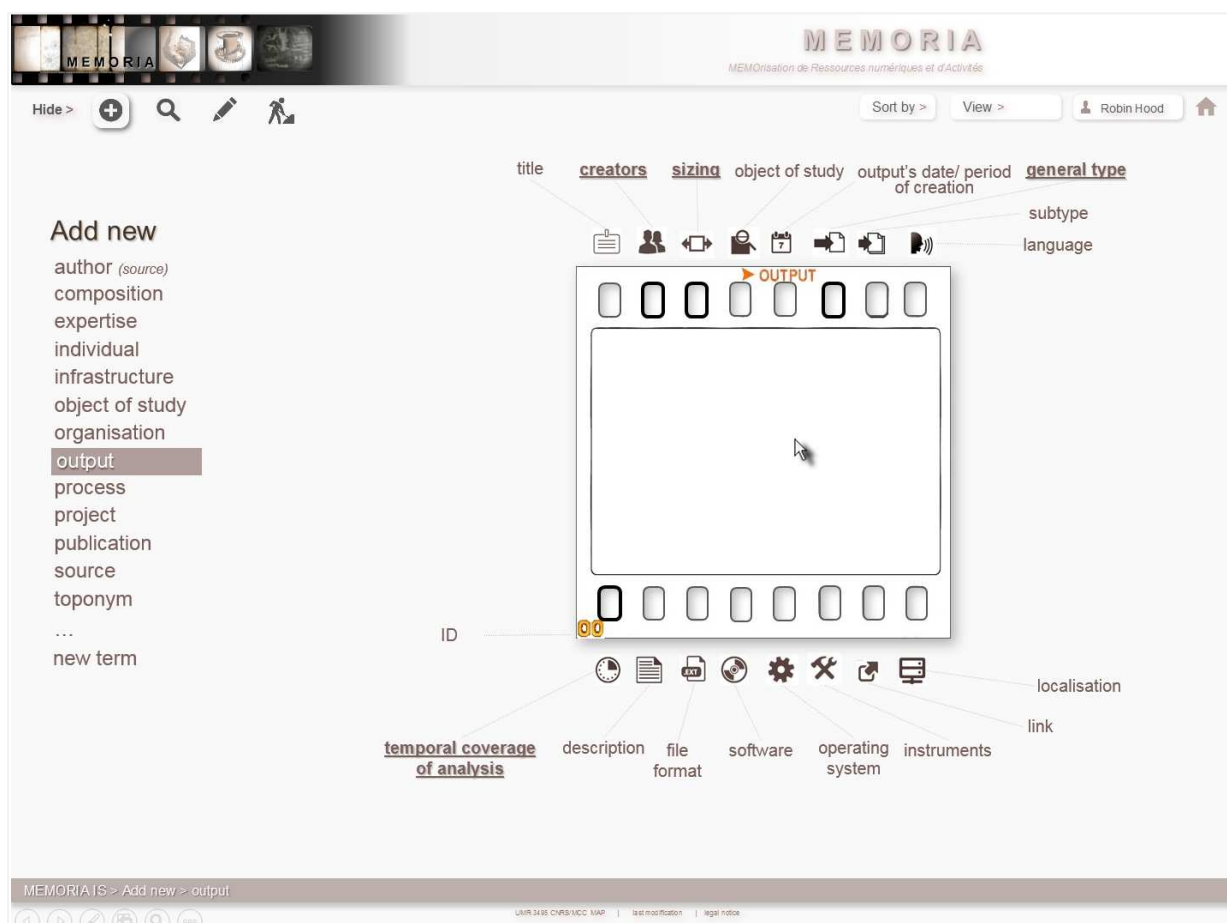


Figure 8 : Interface MEMORIA dans le mode d'indexation (ici une modalité qui nécessite une identification préalable de l'utilisateur).

En haut, insertion d'un nouvel output – tous les descripteurs disponibles sont affichés, les 'perforations' entourées en gras montrent les descripteurs obligatoires.

Bas gauche, un élément à renseigner peut être sélectionné soit dans le volet textuel situé à gauche, soit dans la zone visuelle centrale.

Bas-droite, insertion d'un nouvel objet d'étude et formulaire en attente (coin bas gauche).

Annexe 6c Interface utilisateur - mode de modification

MEMORIA/RAPPORT_sept2016/03_Memoria_diapo_interface2c_workzone.pptx

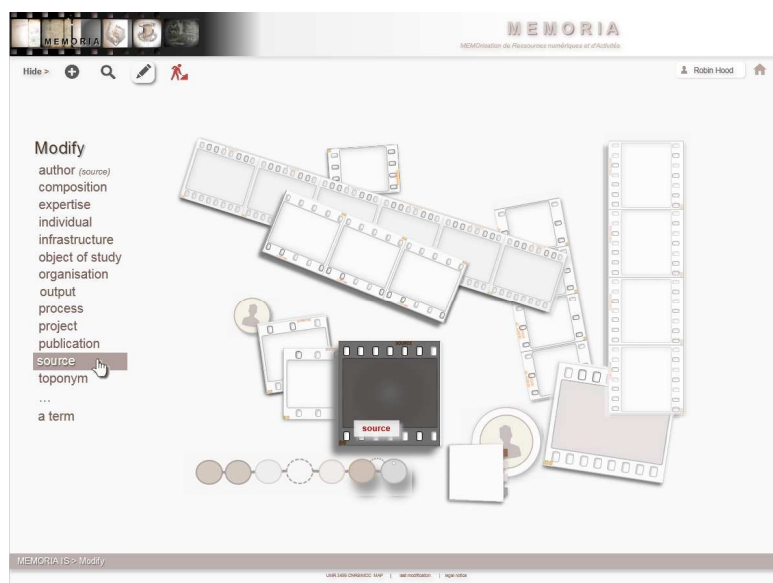
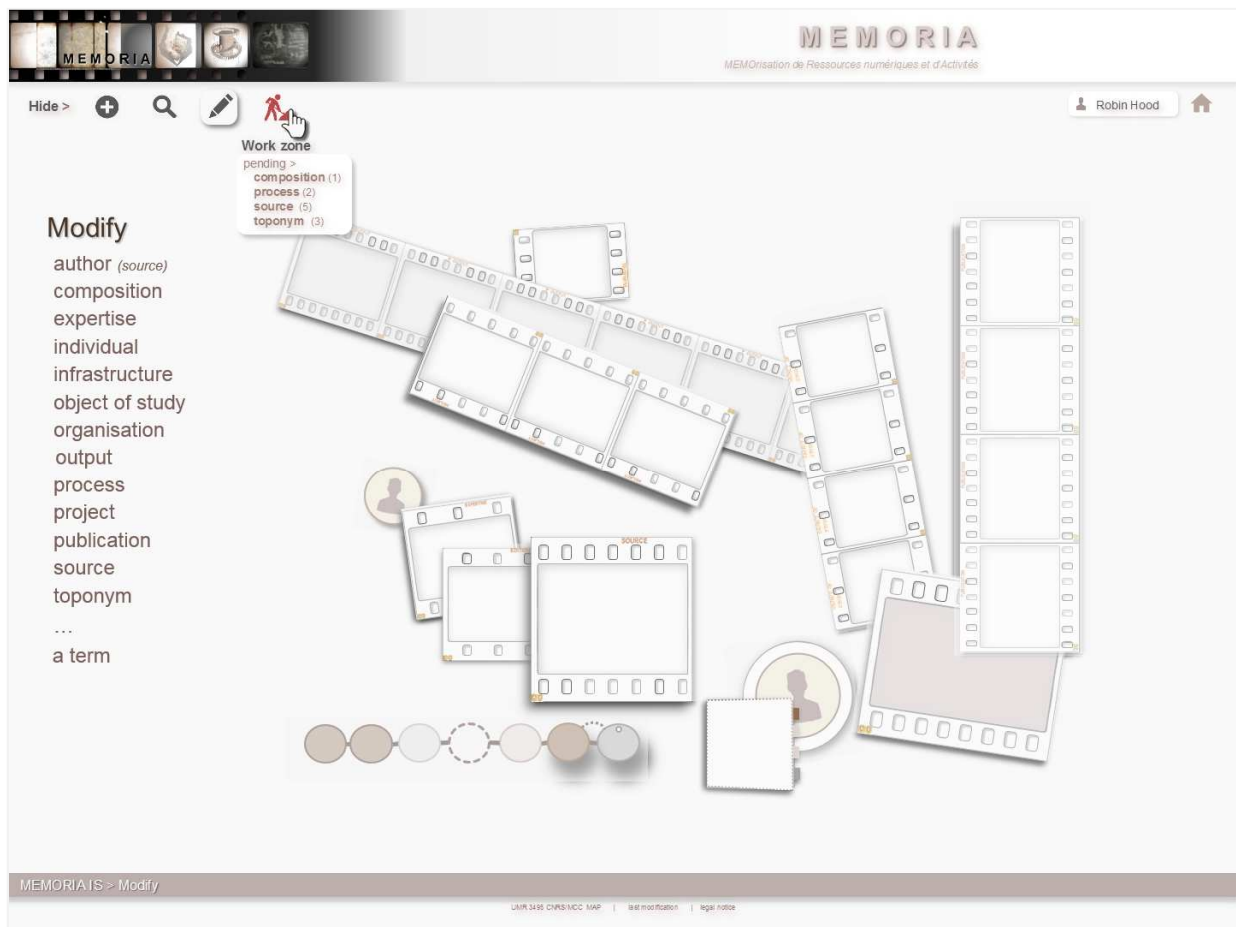


Figure 9: Interface MEMORIA dans le mode de modification.

Ce mode permet soit la modification de n'importe quel objet (droits *éditeur* ou *administrateur*) soit la modification des objets que l'on a créés soi-même (droits *utilisateur*)

Un élément à modifier peut être sélectionné soit dans le volet textuel situé à gauche, soit dans la zone visuelle centrale.

Annexe 6d Interface utilisateur – *work zone*

MEMORIA/RAPPORT_sept2016/03_Memoria_diapo_interface2c_workzone.pptx



Figure 10 : Interface MEMORIA – la zone personnelle de travail.

Le système garde dans la *working zone*, tous les formulaires qui n'ont pas été finalisés par l'utilisateur. Ces formulaires peuvent être repris puis finalisés ultérieurement. Ils sont ensuite enlevés de la zone personnelle du travail et enregistrés définitivement.

Annexe 6e Interface utilisateur – messages du système

MEMORIA/RAPPORT_sept2016/03_Memoria_SYSTMessages.pptx

La phase de conception de l'interface utilisateur s'est traduite par plusieurs étapes: définition de la structure de navigation, des interactions, structure des écrans (ergonomie « de surface », mise en page, menus), style visuel (charte graphique, couleurs, polices de caractères), éléments iconographiques (icônes, pictogrammes, images, illustrations) et systèmes d'aide.

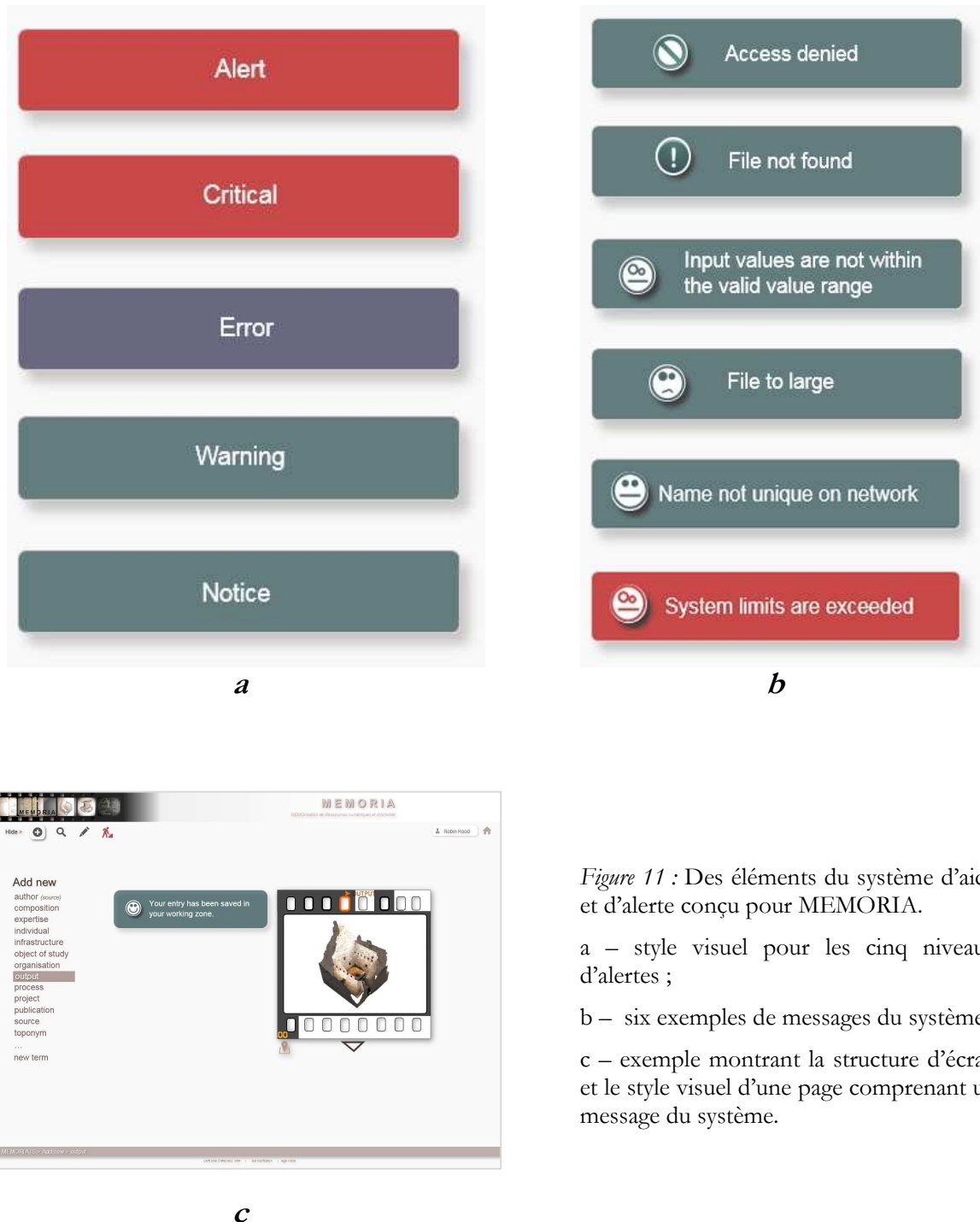


Figure 11 : Des éléments du système d'aide et d'alerte conçu pour MEMORIA.

a – style visuel pour les cinq niveaux d'alertes ;

b – six exemples de messages du système ;

c – exemple montrant la structure d'écran et le style visuel d'une page comprenant un message du système.

Annexe 6f Interface utilisateur – implémentation

MEMORIA/RAPPORT_sept2016/00_EtatdAvancement_implementation.pptx

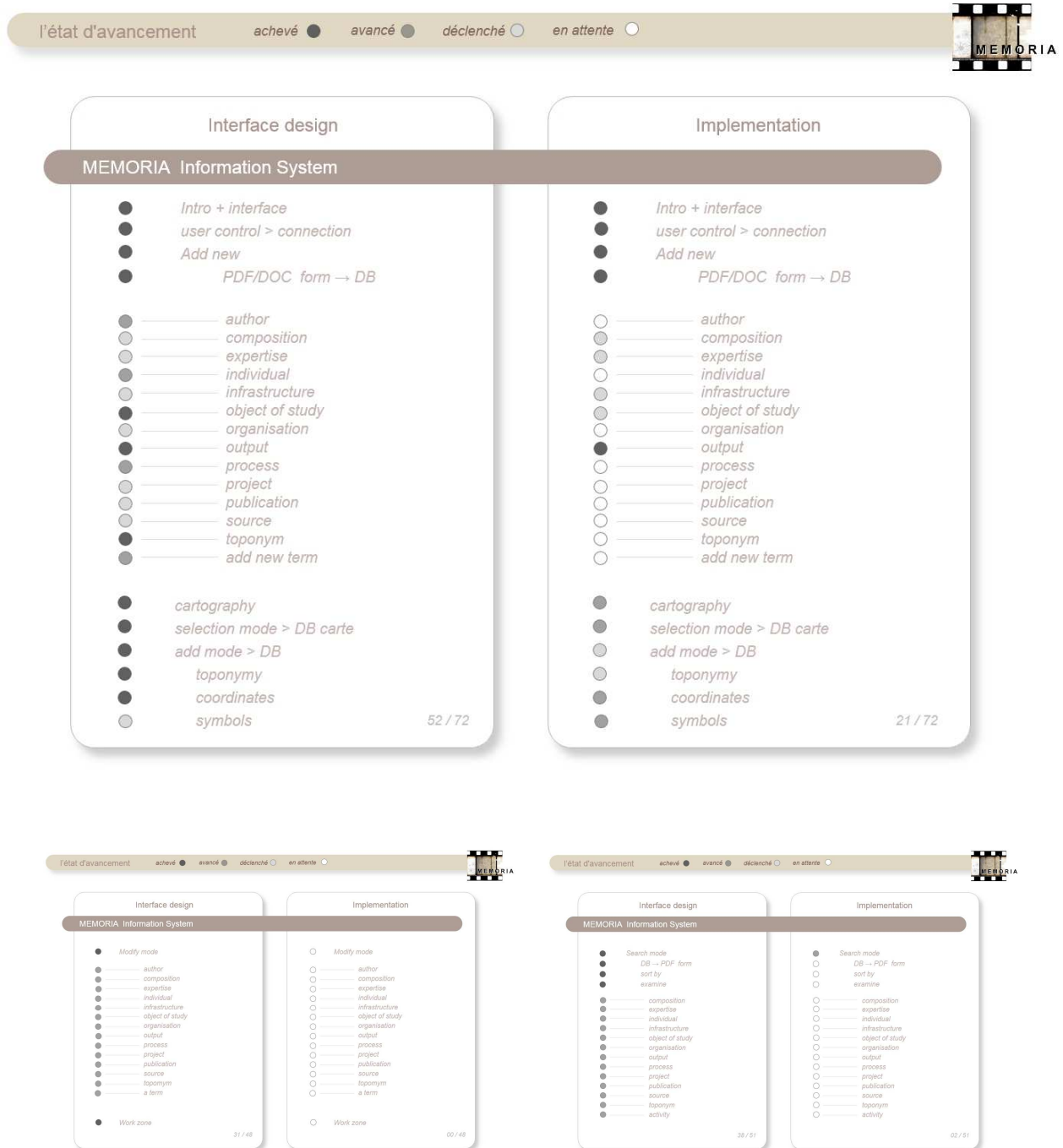


Figure 12 : État d'avancement de l'implémentation - trois tableaux récapitulatifs

Annexe 7a Langage visuel – métaphore du film perforé

MEMORIA/RAPPORT_sept2016/ 03_Memoria_system_DEM2.pptx

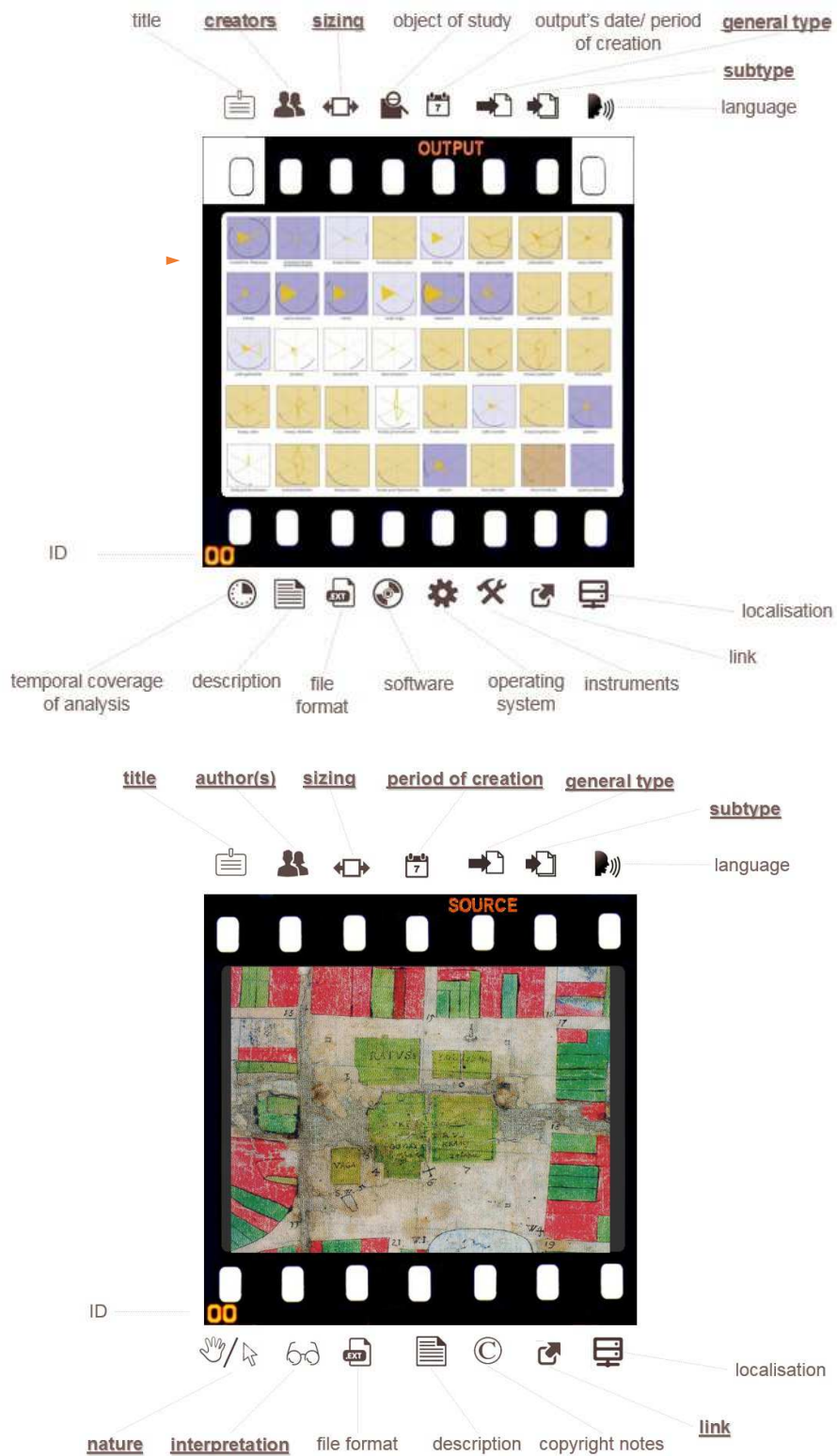


Figure 13 : Métaphore visuelle exploitée pour l'interface utilisateur.
 Les « perforations » sur fond blanc correspondent à des formulaires contenant des informations manquantes.

Annexe 7b Langage visuel – processus

MEMORIA/RAPPORT_sept2016/ 03_Memoria_system_DEM2.pptx

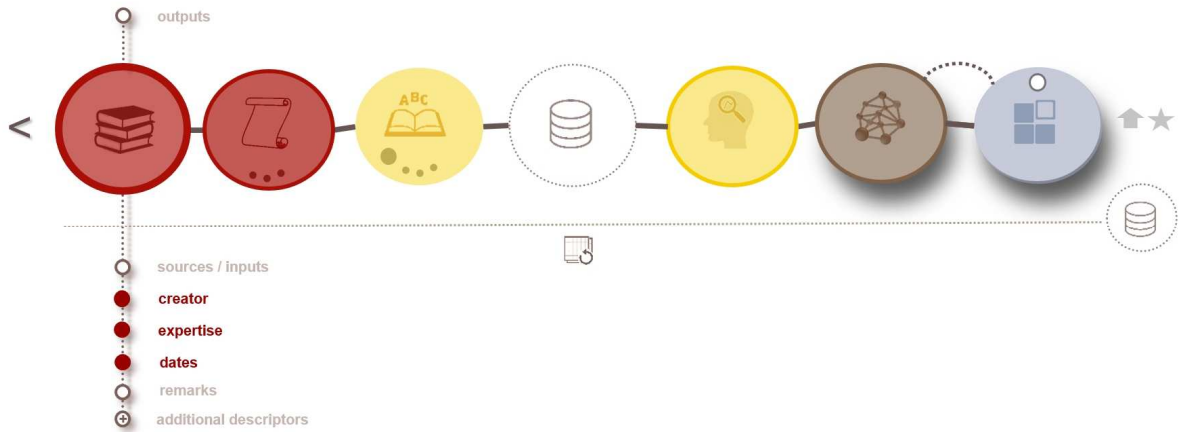


Figure 14 : Exemple d'un processus ordonné – types d'informations pouvant caractériser une activité.

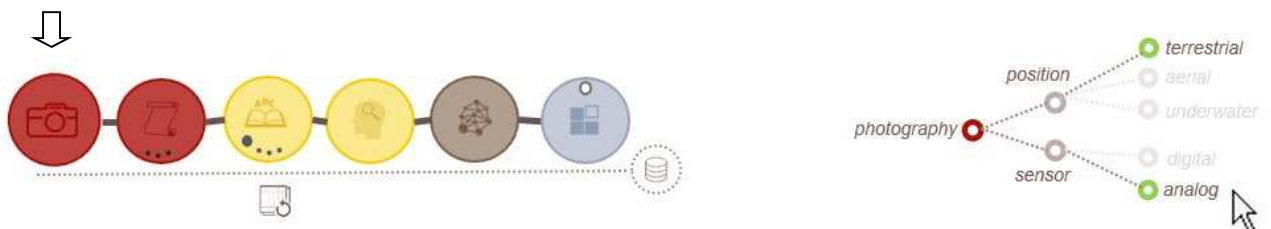


Figure 15 : Groupes des descripteurs associés à une activité – ici l'exemple de prise de photo.

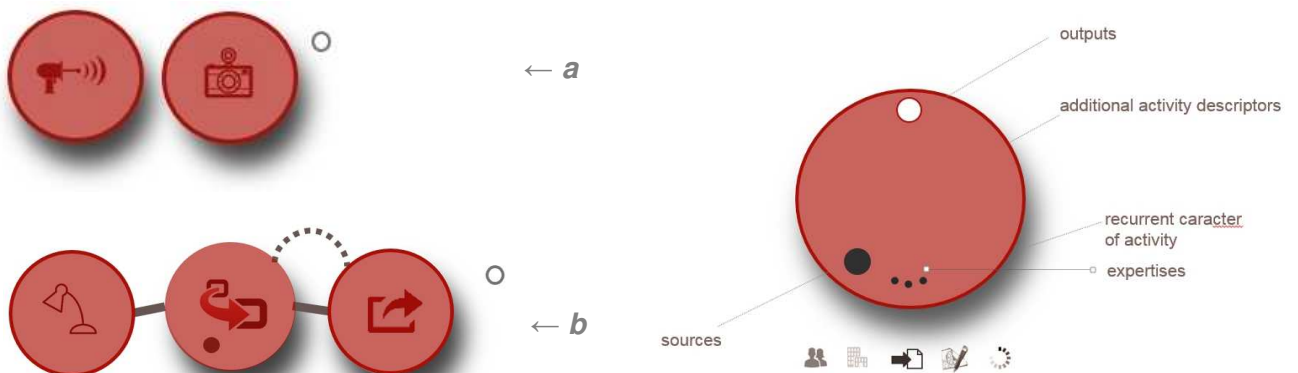
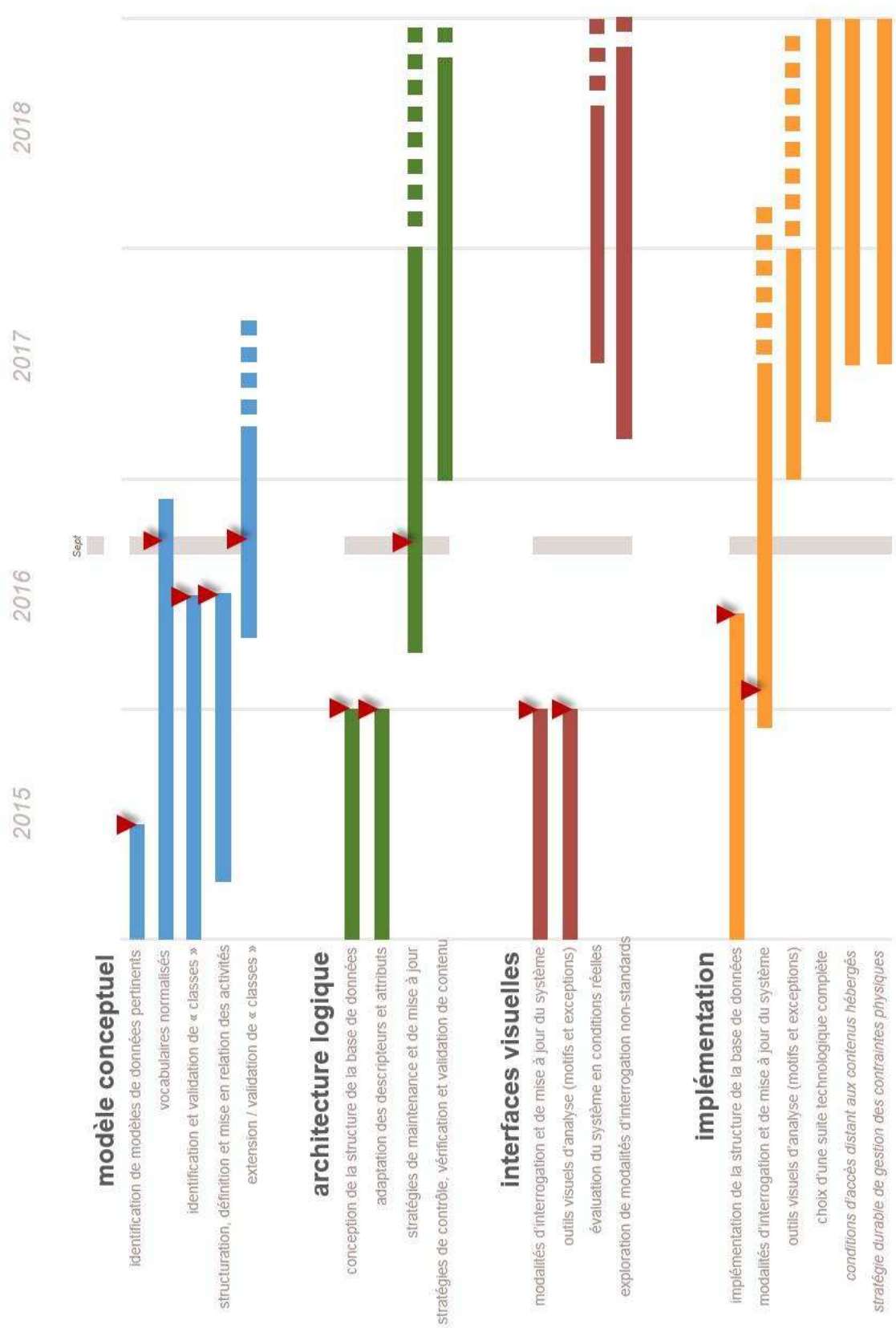


Figure 16 : Les utilisateurs ont la possibilité d'enregistrer les processus comme des chaînes ordonnées (les activités sont censées avoir eu lieu dans un ordre précis), ou comme des ensembles libres (« sac d'activités »).

Annexe 8 État d'avancement du projet MEMORIA - Septembre 2016



Annexe 9 Publication

*Proceedings of the 2nd International Congress on Digital Heritage 2015,
Vol. 2, Assessment of Methodologies and Tools in DH,
IEEE Computer Society, 2015, ISBN: 978-1-5090-0254-2, pp. 343 - 346*

How Was This Done? An attempt at Formalising and Memorising a Digital Asset's Making-Of

Iwona Dudek, Jean-Yves Blaise, Livio De Luca, Laurent Bergerot, Noémie Renaudin

UMR CNRS/MCC 3495 MAP

Marseilles, France

iwona.dudek, jean-yves.blaise, livio.deluca, laurent.bergerot, noemie.renaudin@map.cnrs.fr

Abstract— This paper introduces the early results of a research programme called MEMORIA that aims at developing an information system enabling the description, structuring and storage of digital outputs produced in the course of Heritage Architecture studies. Our objective is to memorize not only a given result – *i.e.* a digital asset - but its *making-of* - in other words to record and share with future generations a *work process* rather than solely its outcomes. Digital assets are on the one hand described by a set of “classic” parameters (e.g. format, authors, creation date, *etc.*) and on the other hand associated with a process (concept that should be understood as a chain of activities). Ultimately, the project investigates how a digital resource resulting from a human-birthing cognitive process can be associated with descriptors ensuring that all actions mobilised to produce the resource are recorded, and therefore ensuring a sort-of scientific traceability of the “final” digital document.

Index Terms—Heritage architecture, Methodology, Work processes, Documentation.

I. MOTIVATION

Massive digitisation campaigns carried out over the past decades in the Digital Humanities community, and notably in the cultural and heritage sectors, do result in a positive move towards more content available for analysts, more cross-analyses made possible, more knowledge collected, structured and shared [1] and generally speaking more focus potentially put on these sectors thanks to end-line products or valuation initiatives. In parallel, the notion of open science [2] has paved its way into the work practices of scholars and collection holders, with today a real concern in the above mentioned sectors for interoperability, open access, standards, metadata (*etc.*). As a result, a number of initiatives such as [3] or [4] do offer an unprecedented opportunity for web users to access without charge to digital content. In addition, the notion of *digital asset* – *i.e.* a digital content *along with* the right to use it – has been pinpointed as a mean to create value by [5], thereby adding economy to science in terms of perspectives.

However, with more and more content digitised, more and more content e-published, digital humanities actors are now clearly confronted to a challenge initially picked up in information sciences: *Big Data* (big volumes of data, dynamically changing data, as well as high variety, highly heterogeneous data). This is visible in many research programmes where tools and formalisms stemming from the computer science side are confronted to analysis needs

stemming from the Humanities side – literary analysis is a classic example [6], [7], [8].

In the digital heritage, more data sets, more data acquisition procedures, more heterogeneity appears at first glance as a positive move – yet actors very well know that it also means more choices to make in terms of tools and methods, and more interpretation steps to memorise if wanting to ensure scientific validity. A quick overview of research programmes focusing on digital heritage assets shows that, beyond the digitisation effort itself, three challenges are at this stage on top of the research agenda:

- developing standards / data models in order to facilitate information sharing – a typical example is the CIDOC Conceptual Reference Model [9],
- sensemaking in massive data sets – a challenge picked up prominently by the fields of Visual Analytics and by the data mining community [10],
- allowing for an open access to the data, and ultimately considering it as a repurposable material ([11]).

But in some application fields - and in particular in heritage architecture – the digitisation effort **most often leads to the creation of a new, highly interpretative, digital asset** (typically a virtual reconstruction). This new *output* of the analysis process is considered by many as an *extension* and an *enrichment* of the initial collection. Hence a fourth, emerging challenge: describing not only the asset as such (“this is a virtual model of building B created by actors A1 and A2, showing B at time T5”), but memorising the various steps undertaken by the analyst(s) in his/their move from raw data (archival material, 3D point clouds, *etc.*) to interpretation and knowledge. This is obviously key if the asset is to be of any help for future generations, and if we consider a digital asset as potentially of scientific value – what is a conclusion worth if one cannot explain how one reached that conclusion?

This paper introduces an ongoing research programme called MEMORIA, through which we try to preserve not only the *result* of a cognitive process (e.g. “a Collada online 3D model showing a virtual reconstruction of Emperor Augustus’ *Tropaeum Alpium*”) but the *making-of* this result (e.g. *How was the 3D data acquired? What pieces of generic knowledge were used?, What hypotheses were laid down?, What tools were used in the 3D modelling steps?, ...*).

The programme aims at investigating to which extent we can today preserve a *cognitive process* as such, since if and

Annexe 9 Publication

*Proceedings of the 2nd International Congress on Digital Heritage 2015,
Vol. 2, Assessment of Methodologies and Tools in DH,
IEEE Computer Society, 2015, ISBN: 978-1-5090-0254-2, pp. 343 - 346*

only if we manage to do that can digital assets produced daily become a readable, sharable, scientifically relevant *digital heritage*. The approach builds on the idea that the creation of a digital asset results from a series of *activities* (ranging from in situ 3D data acquisition to analogical reasoning basing on archival material) that need to be recorded so as to allow for intersubjective verification. Activities can then be chained to form *processes* (successive moves from a raw cloud of points to a consistent meshed and textured 3D model for instance) that lead to the creation of an *output* (a new digital asset).

This research is conducted as multidisciplinary teamwork in a research unit focusing on heritage architecture analyses from various points of view (computer vision, geometric modelling, information systems, knowledge modelling, time-oriented data visualisation, *etc.*). Accordingly the outputs produced by the team members strongly vary in terms of format, of aim and target, and ultimately of role in the global understanding of how an artefact births and changes over time. We therefore will not claim that the MEMORIA approach is fully representative of our field of concern, but that it is tested on digital content that is heterogeneous enough to provide for a sound test field. Section 2 briefly comments on the notions of activities and processes while section 3 presents our main choices, and introduces the experimental platform itself.

II. RELATED WORKS

The notions of activity and process at the heart of this research are nothing but new or original: readers who once had to produce PERT or Gantt charts do have a clear idea of what is meant here, others will find in [12] a number of convincing examples of how these notions can be applied to real-life situations (SOPO diagrams, PlanningLines, Decision charts, *etc.*). An activity is basically a triplet of values: *who* does *what*, and *when*? Processes act as containers for an ordered chain of individual activities leading from *input* α to *output* Ω .

In short, the notions of activity and process are very present in the knowledge management community (see for instance [13] or [14]), but their application to heritage sciences remains challenging [15]. The MEMORIA initiative can be seen as an opportunity to question the applicability to historical sciences of formalisms that often birthed in the context of business and industry applied research.

III. THE MEMORIA APPROACH

The project builds on four main notions: *outputs*, *activities*, *processes* and *sources/inputs*.

(a) **Outputs** are digital material/resources produced in the course of a study, whatever the media (still image, moving image, sound, *etc.*). Outputs can be an end result (e.g. a 3D model ready for publication on the web) but they can also be an intermediate outcome, ready for future use in alternative contexts (a 3D model needing texturing before web publication, and in the meantime ready for 3D printing). Ultimately, the project aims at associating a digital resource with a set of descriptors in order to trace all actions mobilised to produce the output, and therefore memorize not only the "final" digital document but also the main steps of the

cognitive process that lead to its production. These descriptors are summarised in a visual metaphor illustrated on Figure 1, a visualisation that combines a thumbnail view (or icon) of the output and pseudo film perforations (*perfs*) thanks to which users have access to each standard parameter associated with the output (creator, date of creation, *etc.*).

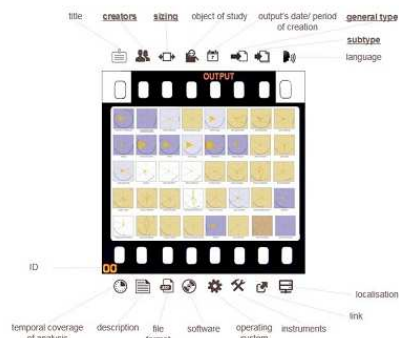


Fig. 1. The visual metaphor used as the user interface for *outputs*. White "perfs" correspond to blank fields (missing information).

A critical feature of the approach should therefore be a visual interface providing user-filtered accesses to the results of queries (outputs sorted by object of study, project, production process, creator, *etc.*). Progressive zooming is one of the solutions we are testing (Figure 2), but there is here clearly here a broader and still open challenge: bridging the gap between "classic" collection browsing modalities, with each individual item present in the display, and Visual Analytics solutions who introduce progressive filtering or collections reading mechanisms that will need to be tested.



Fig. 2. Screenshot of the online test implementation.

(b) **Activities** are actions undertaken by one or several actors of the study in order to move from one point of the overall cognitive process to a new one. An activity is single out as such, and differentiated from a process, basing on the consistency of the method, tools and aim of a set of actions (e.g. *Translating three inventories in a row* is considered as one unique activity, but using these translations in order to *draw plans of buildings* is considered as another activity).

An activity is described by a set of standard parameters (creator, date and/or duration, inputs list, project id if relevant, external expertise called in, *etc.*). Activities are then grouped

Annexe 9 Publication

Proceedings of the 2nd International Congress on Digital Heritage 2015,
 Vol. 2, Assessment of Methodologies and Tools in DH,
 IEEE Computer Society, 2015, ISBN: 978-1-5090-0254-2, pp. 343 - 346

by category (for instance steps dedicated to collecting data vs. steps dedicated to analysing data), with inside each category a specific hierarchy helping users to find their way in the reporting of “what they did” (e.g. activity desk-based research, sub-category handling of primary sources, Figure 3). Finally descriptors that are specific to one particular activity can be added (e.g. activity *laser scanning* > descriptor *measurement method: optical triangulation*, Figure 4).

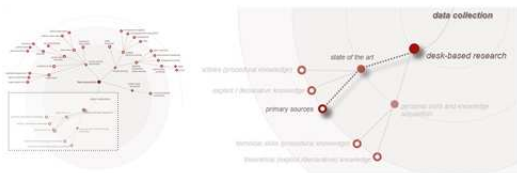


Fig. 3. The data collecting activities hierarchy (an activity identifies a series of actions mobilised to produce a digital resource).

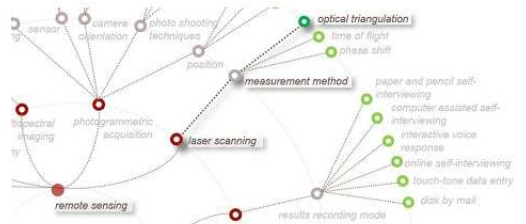


Fig. 4. Activity descriptors for the data collecting group of activities

(c) Processes act as containers for one or several activities. They are singled out basing on the fact that a new output has been created. In other words, each output is associated with a given process, containing one or more activities, and if this output is reused in the context of a new analysis task then the whole set of activities that lead to the creation of the output can be duplicated and memorised in the description of that new analysis task. A process keeps track of the institutional framework in which the work was carried out (organizations, projects, staff employed, etc.), as well as of primary sources used, or of the techniques and tools used. Each activity contributing to the process is represented by a multidimensional icon (Figure 5) that shows the activity’s category (colour), and its position in the category’s sub-hierarchy (glyph). Various processes can accordingly be compared visually, helping analysts and decision makers to get an overall and synthetic view of how outputs were produced (how many steps, what kind of steps, etc.).

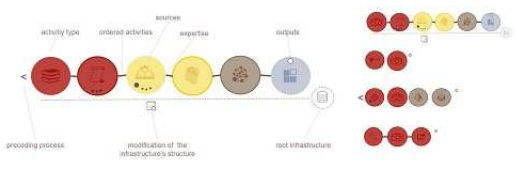


Fig. 5. The notion of process: a chain of activities mobilised to produce one or more outputs (including potential links to preceding processes).

A fruitful side-effect of the introduction of processes is the possibility this notion offers to spot methodological patterns (recurrent chains of activities) developed by a team, an individual, or in relation with a given type of data. As a consequence, with time, the notion of process can contribute to a better depiction of work processes in our field of concern, and ultimately to the creation of visual dashboards [15] helping actors to better explain, compare, quantify work processes.

(d) Sources and inputs

An activity can be based on one or more sources (external resources), as well as on outputs produced previously and described by a separate process. To avoid confusion the outputs used in an activity as initial root elements are called inputs.

Sources and inputs are pieces of data and information (digital or not) that are needed in order to start an activity or a process. They can be raw data (typically a 3D point cloud acquired in situ, or historical evidence retrieved from archives) or an intermediate outcome (e.g. a mesh built from the initial point cloud). Accordingly, they are described in a way that is quite similar to outputs, with a set of standard parameters (creator, date of creation, format, etc.) and will be searchable through an interface (Figure 6) that closely resembles to this that has been implemented for outputs (Figures 1- 2).



Fig. 6. A source is basically a piece of data or information, digital or not, needed to carry out one or more activities. A source is an “external” element on which an activity or process will base on. By contrast *inputs*, are intermediate results, reused in later MEMORIA activities.

Besides these four main elements, the MEMORIA platform allows for the storage of other pieces of information, about *infrastructures*, *actors*, *organisations*, *objects of study*, *expertise*, *context*, etc. as exemplified in Figure 7.

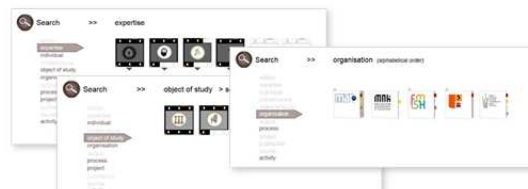


Fig. 7. Visual dashboards for supplementary pieces of information attached to outputs or activities.

Annexe 9 Publication

*Proceedings of the 2nd International Congress on Digital Heritage 2015,
Vol. 2, Assessment of Methodologies and Tools in DH,
IEEE Computer Society, 2015, ISBN: 978-1-5090-0254-2, pp. 343 - 346*

The implementation we are currently developing (both for updating and browsing/searching the collection of outputs) combines an RDBMS and WebGL/Js modules built as pseudo-objects (in the sense of OOP programming). It is likely, depending on potential end-users feedbacks, to include also PDF outputs summarising in a one page "ready to print" document a visual overview of processes that are associated with a given digital output.

One of the acute issues we are facing in the MEMORIA programme is dealing with *existing* outputs: one thing is promoting a "new" preservation policy, another is coping with the digital assets that have *already* been produced. If we ignore, to a large extent, how they were produced, how can we integrate them into the MEMORIA approach? As an answer, the description of activities (as a hierarchy, with growing levels of details) and the corresponding visual interfaces are designed in order to allow for the recording (and searching) of "poorly" known work processes (for instance, recording the fact that a study bases on a 3D photogrammetric survey but not detailing the actual camera used, the date, *etc.*). Users are given a possibility to record processes as ordered chains (activities are known to have occurred in a precise order - Figure 8, left), or as disordered lists (*i.e.* a "bag of activities" (Figure 8, right).

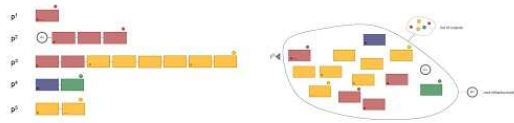


Fig. 8. An illustration of ordered and non-ordered processes.

Finally, it is important to state that the MEMORIA approach can also help showing the evolution of methods, techniques and tools used over time, as well as types of activities mobilised to produce a given type of output, thematic migration of teams or individuals, history of cooperation, *etc.*. In that sense, the MEMORIA approach, and in particular its visualisation steps, can contribute to a better understanding by decision makers of today's work processes, of temporal constraints connected to each and every activity, of technical bottlenecks, and more generally of planning difficulties.

IV. CONCLUSION

This paper's claim is that in some application fields, and in particular in heritage architecture studies, we have come to a point where we need to preserve not only a digital asset as such (format, creator, owner, *etc.*) but also its *making-of* (analogies, tools, decimation or simplification, sources? *etc.*). With (at least) 20 years of digital assets production behind us most actors in the field are now facing hundreds when not thousands of "*semantically orphan*" digital content.

We introduce an experimental platform called MEMORIA thanks to wish we investigate a methodological framework aimed at memorising the various steps (represented as activities and processes) undertaken by human analysts in order to move from an initial state of knowledge to the production of one or several new digital assets. Although this research is still at an

early stage, it underlines the potential benefits of enabling us to answer to the "*How was this done?*" question, and unveils a number of open challenges, for instance in terms of knowledge modelling (e.g. *What is the optimal level of detail in the description of activities/outputs?*) and of visualisation (e.g. *How do we handle a focus+context view over heterogeneous and massive collections?*).

To conclude, our vision is that the approach, primarily intended at ameliorating the readability and usability of digital assets over time for scientists or specialists at large, can also renew the way historical sciences can be presented, or "storytold", to the wide public, at a time when social networking and crowdsourcing significantly infiltrate practices of scientific communities.

REFERENCES

- [1] G. Schuller, "Designing universal knowledge", Lars Müller publishers, Baden 2009.
- [2] Open science, blog-like resources curated by F. Piron
- [3] Europeana project web portal
- [4] 3D-ICONS project web portal
- [5] A. J. Van Niekerk, "The Strategic Management of Media Assets; A Methodological Approach", Allied Academies, New Orleans Congress, 2006.
- [6] R. Vuillemot, T. Clement, C. Plaisant, A. Kumar, "What's being said near 'Martha'? Exploring name entities in literary text collections", IEEE Symposium on Visual Analytics Science and Technology (VAST), 2009, pp.107-114.
- [7] F. Wanner, J. Fuchs, D. Oelke, D.A. Keim, "Are my children old enough to read these books? Age suitability analysis", POLIBITS: Research journal on computer science and computer engineering with applications, 2011.
- [8] S. Koch, J. Markus, M. Wörner, A. Müller, T. Ertl, "VarifocalReader – In-Depth Visual Analysis of Large Text Documents", IEEE Transactions on Visualization and Computer Graphics, Vol. 20, No. 12, 2014.
- [9] <<http://www.cidoc-crm.org>>
- [10] D. Keim, J. Kohlhammer, G. Ellis, F. Mansmann, "Mastering the Information Age – Solving Problems with Visual Analytics", 2010, online at <www.vismaster.eu>.
- [11] J.Y. Blaise, I. Dudek, P. Bénistant, A. Durand, "Online 2D/3D graphic interfaces using XML 'reusable' heritage contents", 12th International Conference on Computer Graphics, Visualisation and Computer Vision, WSCG 2004, Plzen 2004, pp. 39-46.
- [12] W. Aigner, S. Miksch, H. Schumann, C. Tominski, "Visualization of Time-Oriented Data", Human-Computer Interaction Series: Springer, 2011.
- [13] H. Ping Tserng, Yu-Cheng Lin, "Developing an activity-based knowledge management system for contractors", Automation in Construction 13, 2004, pp.781– 802.
- [14] M.J. Eppler, "A Generic Framework for Information Quality in Knowledge-intensive Processes" Proceedings of the Sixth International Conference on Information Quality, 2001.
- [15] J.Y. Blaise, I. Dudek, "Picturing what others know: towards a dashboard for interdisciplinarity", Proc. of the 14th International i-KNOW conference, 2014, Article no 15, ACM 2014.