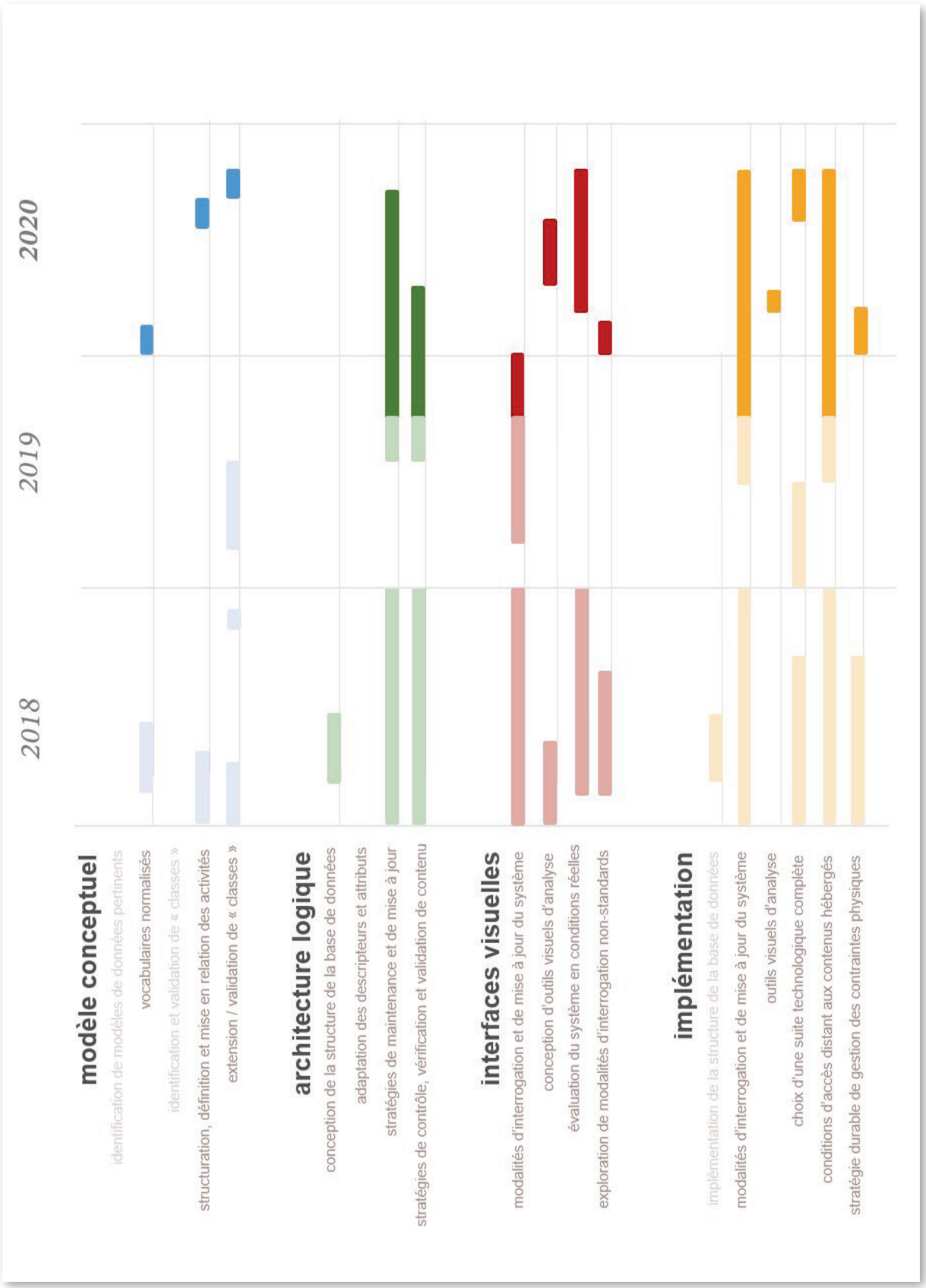
UMR 3495 CNRS/MC MAP

Projet MEMORIA – mémorisation de ressources numériques et d'activités

Bilan 2019/2020

Iwona DUDEK-BLAISE
Chargé de recherche CNRS
30/10/2020

MEMORIA : actions menées entre septembre 2019 et novembre 2020



1 Rappel des objectifs et axes scientifiques et artistiques (le cas échéant)

Le projet MEMORIA cherche à faciliter l'inscription d'activités de recherche dans une logique d'intégrité scientifique et de bonnes pratiques. Il vise à développer un système d'information permettant la description, la structuration et la pérennisation des modalités d'obtention de résultats de recherche (couvrant à la fois données et inférences). À terme, le projet vise à mémoriser au-delà d'un document « final » une démarche cognitive, ou tout au moins une démarche de production.

L'objectif général du projet est d'associer aux résultats scientifiques, en particulier en sciences historiques et patrimoniales, des indicateurs permettant de mieux retracer la façon dont ces résultats sont obtenus. L'enjeu à relever est donc de contribuer à éliciter, structurer, pérenniser et étudier *les processus de production* de ces résultats, en développant méthodes et outils pour :

- assurer la ***transmissibilité intersubjective*** de ces processus;
- permettre ***l'interprétabilité***, la ***vérifiabilité*** et la ***reproductibilité*** des résultats;
- favoriser une lecture ***comparative*** et *cumulative***.**

La démarche adoptée promeut le développement ***d'interfaces graphiques interactives*** facilitant le raisonnement visuel sur :

- nos méthodes de travail, leur évolution, les biais disciplinaires;
- les techniques et les outils utilisés au cours du temps.

verrous scientifiques

- identification des activités : une phase d'élicitation de connaissances (ou de savoir-faire) indispensable pour identifier et structurer les activités;
- calibrage de la granularité d'analyse: un choix de « grain » difficile à opérer pour décrire l'indispensable seulement, pour éviter une surcharge d'informations;
- vocabulaires de spécialité divergents
- travail d'explicitation visuelle privilégiant des modalités de raisonnement non-verbal, afin de maximiser l'utilisabilité des informations recueillies.

verrous technologiques

- élaboration des stratégies de maintenance et des conditions d'accès distant aux contenus hébergés;
- choix d'une stratégie de contrôle, de vérification et de validation/rejet de nouveaux contenus;
- interfaces visuelles : exploration de modalités d'interrogation et de mise à jour du système originales;
- développement de dispositifs de visualisation d'informations et d'analyse innovants;
- stratégie de moissonnage et de profilage de références (bibliographiques et e-sources).

2 Réalisation scientifique et artistique et modes de valorisation

L'année 2019/2020 a été une période de continuation du développement et de l'implémentation du système d'informations MEMORIA. Grâce au recrutement d'un développeur accompli (contrat CDD) le travail a pu faire un grand pas. Ce recrutement a été possible en grande partie grâce à notre engagement dans un projet ANR, Sesames (ANR-18-CE38-0009-01) dans lequel le projet MEMORIA est évalué et mis en confrontation avec des champs de recherche et des pratiques allant au-delà de ce qui est fait dans notre propre laboratoire (archéologie et acoustique).

Actuellement le SI est fonctionnel, et il est évalué dans un cadre interdisciplinaire au travers de sa mise en pratique au sein du projet Sesames (ANR) - mais il est loin d'être définitivement fini. (*cf. 2.2 État d'avancement, p.7*)

2.1 Les grandes lignes de réalisation du projet MEMORIA dans cette année :

Nouvelles étapes de la conceptualisation liées à l'élargissement thématique

- Identification, structuration, désambiguïsation et mise en relation des activités (*i.e.* actions mobilisées pour produire une ressource) et de leurs attributs - filtrage des définitions verbales des activités et identification des attributs pertinents.

  Activités relatives au profil et ou pratiques de travail et ou protocoles de recherche de nos partenaires : UMR 7324 - CItés, TERritoires, Environnement et Sociétés, Laboratoire Archéologie et Territoires, UMR 7061 - P.R.I.S.M. Perception Représentation Image Son Musique

- Analyse du vocabulaire de spécialité pour aboutir à des listes de terminologies normalisées. À l'intérieur de chaque liste chaque terme est défini, exemplifié (*i.e.* définition ostensive) et documenté - continuation.

- Formalisation d'un logigramme du processus d'analyse des sources électroniques et de l'alimentation de la base de données MEMORIA-ref.

  <http://127.0.0.1:8080/projects/BlackWhite/Recher/flowchart_inter_idu2.pdf>
  <http://127.0.0.1:8080/projects/BlackWhite/Recher/flowchart_processEX2019.pdf>

Création des interfaces visuelles

- Création des icônes pour interfaces visuelles

  Icônes représentant 266 types activités (cinq groupes d'activités). Pour chaque activité nous avons conçu et produit 4 types d'icônes représentant des activités avec variation en couleurs et en composition selon la précision atteinte dans l'effort de documentation de l'activité.
  (cf. Annexes pp. 17-18, 28)

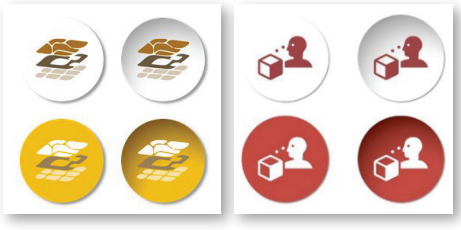


Fig.1 Exemple d'icônes déclaratifs pour énoncer *l'état de description* d'une activité – quatre niveaux de renseignement possibles pour une même activité (gauche : modélisation spatiale, droite: direct observation).

- Conception d'interfaces visuelles en ligne pour alimenter la base de données MEMORIA combinant éléments textuels et graphiques sémantisés qui identifient des éléments *assertifs* (pour donner une information), *directifs* (pour indiquer des actions à

mener et guider l'utilisateur), *déclaratifs* (pour énoncer un fait), et *interrogatifs* (pour demander une information). *(cf. Fig 1, Annexes p.28)*

- Conception des modalités d'interrogation et mise à jour du système. *(cf. Annexes pp.19-28, 31)*

Développement du Système d'Informations

- Implémentation, débogage et tests d'un protocole d'alimentation de la base de données MEMORIA à partir de formulaires PDF téléversables en temps décalé (compatibilité avec des situations de travail en ligne / hors ligne). *(cf. Annexes p. 31)*

  - ✓ mise à jour des formulaires PDF téléversables en temps décalé,
  - ✓ élaboration et implémentation des stratégies de validation/rejet de nouveaux contenus,
  - ✓ interface pour la 'control zone' du système d'information MEMORIA --- Interface dédiée au contrôle du protocole d'alimentation et de mise à jour du système d'information à partir de formulaires PDF téléversables en temps décalé.

- Implémentation d'interfaces visuelles et de modalités d'interrogation du Système d'information MEMORIA (interactions en ligne).

  - ✓ implémentation et évaluation des formulaires HTML en-ligne, mise en œuvre et chainage des formulaires de mise à jour du système exploitant la métaphore visuelle «pellicule de film» *(cf. Annexes pp. 19-23),*
  - ✓ introduction d'un mode d'accès aux activités par une « roue d'activités » interactive *(cf. Annexes pp. 15-16),*
  - ✓ mise en place d'une grille de composition des processus (structuration et description), outil visuel favorisant la formalisation et la description de flux d'activités (*i.e.* visualisation de processus sous la forme d'un graphe soulignant les caractéristiques et les relations temporelles des activités particulières engagées pour réaliser le processus) *(cf. Annexes pp. 24-26),*
  - ✓ caractérisation des activités atomiques formant un processus, assurant la traçabilité des différentes étapes mises en œuvre dans des projets scientifiques *(cf. Annexes pp. 27-28),*
  - ✓ outils visuels d'analyse, *(cf. Annexes pp. 29-30, 35).*

- Conception, implémentation et débogage d'un ensemble de formulaires dynamiques pour alimenter une partie de la base de données dédiée au référencement (caractérisation des références bibliographiques). L'objectif principal de cet effort de catégorisation était de permettre une visualisation comparative du caractère de ces ressources.

  - ✓ implémentation et évaluation de la série de sous-formulaires qui se succèdent (chacun correspondant à une couche d'information spécifique). L'utilisateur est guidé par un fil d'images expliquant les étapes à suivre et les fonctionnalités du système. *(cf. Annexes p. 33)*

- développement d'interfaces en indexation et requête, assistance utilisateurs

Alimentation de la base de données

- Description des ressources produites (outputs) et des processus (notion à comprendre comme chaînes d'activités) mobilisés pour les produire dans une sélection d'expériences hétérogènes.

  UMR 3495 – MAP-Gamsau *(cf. Annexes pp. 20, 23, 25)*

  - ✓ ensemble de processus d'acquisition et de traitement de données métriques (projet SESAMES),

- ✓ ensemble des étapes allant de la collecte de données historiques aux phases de production des chronographes – visualisation de l'état de nos connaissances sur l'évolution de 15 chapelles rurales (projet SESAMES),
- ✓ processus liés au développement du système d'informations MEMORIA (projet MEMORIA),
- ✓ processus d'analyse et de catégorisation des sources électroniques et alimentation de la base de données MEMORIA-ref (projet MEMORIA),
- ✓ ensemble des étapes allant de la collecte de données historiques aux phases de reconstruction virtuelle sur le cas de restitutions de l'ancien hôtel de ville de Cracovie (projet ARKIW).

UMR 7324 - CItés, TERritoires, Environnement et Sociétés, Laboratoire Archéologie et Territoires *(cf. Annexes pp. 20-21, 24, 26)*

- ✓ ensemble des étapes allant de la collecte de données historiques aux phases de reconstruction virtuelle en passant par toutes les étapes d'interprétation et de traitement des données dans le cadre de restitutions alternatives de l'hôtellerie de l'Abbaye de Marmoutier (Tours)

UMR 7061 - P.R.I.S.M. Perception Représentation Image Son Musique *(cf. Annexes p. 21)*

- ✓ processus d'acquisition et de traitement de données sonores (projet SESAMES)

Vérification, validation et développement de métriques d'évaluation du système

- Mise en place d'une 'copie' du système d'information en ligne (memoria-lat) - une plateforme partagée servant d'outil d'évaluation du système existant et de dialogue entre partenaires (UMR 3495 MAP et UMR 7324 CITERES - LAT).

  - ✓ évaluation du système existant, réorganisation et raffinement du modèle de description des extrants (outputs), d'activités et de chaînes d'activités dans une démarche pragmatique et exhaustive d'application au cas des inférences mobilisées dans le passage observé - restitué.

Collaboration scientifique et encadrement du personnel technique

- collaboration scientifique dans le cadre du projet Sesames
  - ✓ UMR 7324 - CItés, TERritoires, Environnement et Sociétés, Laboratoire Archéologie et Territoires,
  - ✓ UMR 7061 - P.R.I.S.M. Perception Représentation Image Son Musique

- Encadrement CDD Ingénieur et Technicien

  Accompagnement de la recherche : Informatique, statistique et calcul scientifique > Études, développement et déploiement (12 mois) employé comme un renfort d'équipe dans le cadre de deux projets incitatifs - MEMORIA (financement MC/DREST) et SESAMES (financement ANR)

- Sélection et encadrement de stagiaires (développement de sous parties du système)
  Stages de fin de parcours (télétravail en raison des restrictions COVID-19)

  - ✓ Implémentation des interfaces d'entrée-sortie pour une base de données
    3 étudiants du DUT Informatique (IUT d'Aix-Marseille)
  - ✓ Profilage de e-ressources > implémentation des outils visuels pour l'analyse de ressources électroniques et bibliographiques
    1 étudiant, Licence Pro Métiers de l'informatique : « Agile: de l'analyse au déploiement multi-supports » (ADMS AMU)

Valorisation

- Présentation du projet et discussion avec des spécialistes et du public dans le cadre de la *Table ronde - "Diffuser des données aujourd'hui : enjeux juridiques et éthiques",* Le droit et l'éthique comme alliés, Maison méditerranéenne des sciences de l'homme, Aix-en-Provence, 03/10/2019 – *invitation suite à la publication.*

  MEMORIA – la préservation des processus d'étude comme enjeu éthique, La diffusion numérique des données en SHS, [dans) Guide des bonnes pratiques éthiques et juridiques sous la direction de V. Ginouvès, I. Gras, collection DIGITALES, Presses Universitaires de Provence, Aix-Marseille Université 2018, ISBN: 9791032001790, pp. 231-240
  < https://halshs.archives-ouvertes.fr/halshs-01902564>

- Présentation du projet dans un atelier - public internationale

  2nd workshop about 3D objects within the scope of PARTHENOS Work Package 4 on standardization, SHARE - PUBLISH - STORE – PRESERVE, Methodologies, Tools and Challenges for 3D Use in Social Sciences and Humanities

- Présentation du projet pendant la visite inaugurale des locaux de l'UMR 3495 MAP – journée porte ouverte, public scientifique.

- Brochure de présentation du projet

  Lien flyer < http://memoria.gamsau.archi.fr/projet/pdf/Flyer_MEMORIA2019_fin.pdf >

2.2 État d'avancement

Le projet MEMORIA comprend différents volets, implique différents développements (dont certains prospectifs), s'accompagne de différentes expérimentations au fil de l'eau : le schéma ci-dessous permet de récapituler d'une façon synthétique où se situent les niveaux d'aboutissement à ce stade, et l'avancement par rapport à 2019.

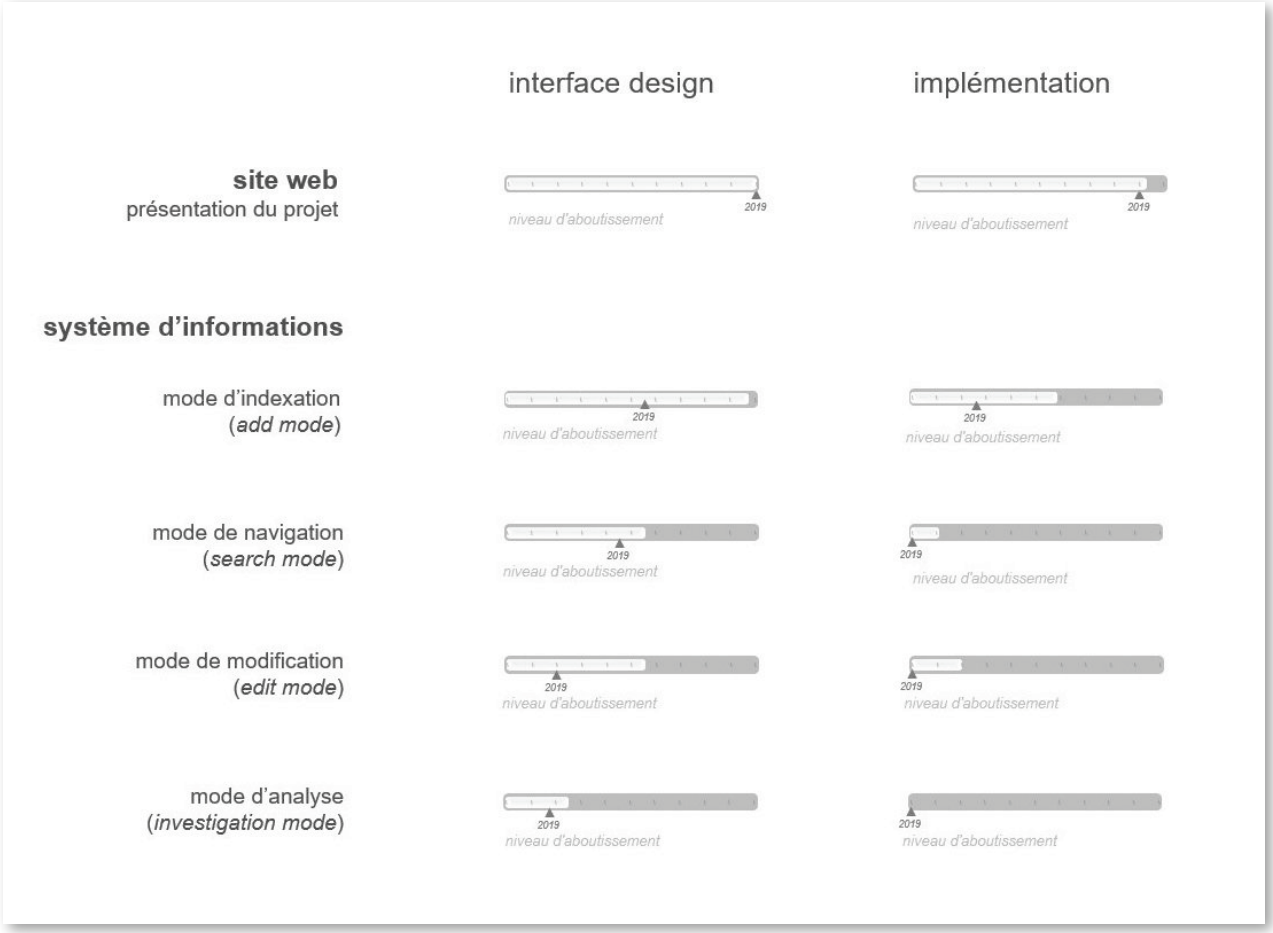
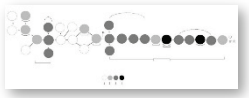


Fig.2 Schématisation de l'état d'accomplissement du système MEMORIA en septembre 2020.

Le schéma souligne les niveaux de réalisations accomplis par rapport aux objectifs du projet : *système d'information permettant l'élicitation, la structuration, la pérennisation et l'analyse comparative des modalités d'obtention de résultats de recherche (couvrant à la fois des données, des inférences, des sources primaires utilisées en phase d'analyse, comme des techniques et outils utilisés pour produire les ressources numériques ou non).*

3 Liste des publications de l'année

Journaux/revues



- I. Dudek, J.Y. Blaise, "Enabling the comparability of research workflows: a case study ", Proceedings - CAA 2019 (in press)

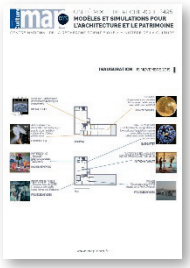
HAL-SHS <https://halshs.archives-ouvertes.fr/halshs-02927631>

Colloques/séminaires

- Table ronde - "Diffuser des données aujourd'hui : enjeux juridiques et éthiques", Le droit et l'éthique comme alliés, Maison méditerranéenne des sciences de l'homme, Aix-en-Provence, 03/10/2019

  contribution: « Store Preserve Visualize - The MEMORIA project»,

  programme <https://sygefor.reseau-urfist.fr/#/training/8112/9067>
  URL <http://memoria.gamsau.archi.fr/projet/pdf/Memoria_MapGamsau_IDudek.pdf>



- Visite inaugurale des locaux de l'UMR 3495 MAP - porte ouverte, Marseille, 19/01/2019

  < http://www.gamsau.map.cnrs.fr/?p=5391 >

Liste des produits ou outils informatiques

- Site du projet
  volets : objectives, équipe, collaborations, vocabulaire, publications et présentations

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

- Système d'Informations en ligne

  < http://memoria-dev.gamsau.archi.fr/is/enter.php>

- Système d'information en ligne secondaire (memoria-lat) – pour notre partenaire UMR 7324 – CITERES-LAT.

  < https://memoria-lat.map.cnrs.fr/is/enter.php>

Listes des autres produits (dont créations artistiques)

- Brochure de présentation du projet
  lien flyer < http://memoria.gamsau.archi.fr/projet/pdf/Flyer_MEMORIA2019_fin.pdf >

- Le projet MEMORIA ne relève clairement pas d'une ambition de création artistique mais d''une démarche scientifique pragmatique. Il n'en reste pas moins, et cela dit avec l'humilité qui s'impose, qu'il comprend une part de création graphique largement illustrée dans les annexes qui suivent. Sans donc parler de 'création artistique' au sens propre il faut souligner que la mise en œuvre du projet s'est accompagnée d'un effort volontaire et spécifique de création graphique, effort très chronophage. Cela est vrai tant dans l'interfaçage général du SI (métaphores visuelles pour faciliter les interactions utilisateur – système, , qu'au niveau de composants graphiques individuels comme par exemple les icones activités, formant un langage visuel cohérent, *antonomases* ou *synecdoques graphiques* qui par le soin mis à leur composition, et par leur capacité d'expressivité, siéent bien à un travail initié à l'articulation science et culture.

4 Moyens mobilisés :

Le projet MEMORIA se déploie en s'appuyant sur les ressources propres du laboratoire (le matériel et le personnel) avec le soutien financier du Département de la Recherche, de l'Enseignement Supérieur et de la Technologie du Ministère de la Culture et de la Communication et celui de l'Agence Nationale de la Recherche (ANR) au travers du projet SESAMES (ANR-18-CE38-0009-01).

Responsabilités et rôles attribués aux membres de l'équipe en 2019-2020

direction scientifique du projet - I. Dudek-Blaise
comité de pilotage - J.Y. Blaise, L. De Luca, I. Dudek-Blaise
modélisation des connaissances - J.Y. Blaise, I. Dudek-Blaise
explicitation des connaissances - J.Y. Blaise, I. Dudek-Blaise, G. Simon (LAT), A. Vidal (PRISM), S. Fargeot (PRISM)
saisie des données - J.Y. Blaise, I. Dudek-Blaise, G. Simon (LAT)
encadrement des stagiaires et personnel CDD - P. Bénistant, I. Dudek-Blaise
conception des interfaces - I. Dudek-Blaise
administration système - P. Bénistant
implémentation de l'interface SI - M. Rabefandroana, V. Samg-Mouit (AMU),
test d'implémentation d'une plateforme de visualisation - N. Guyomarc'h (IUT), A. Salles (IUT), Zaid-Mohamed (IUT)
production des icônes - N. Renaudin

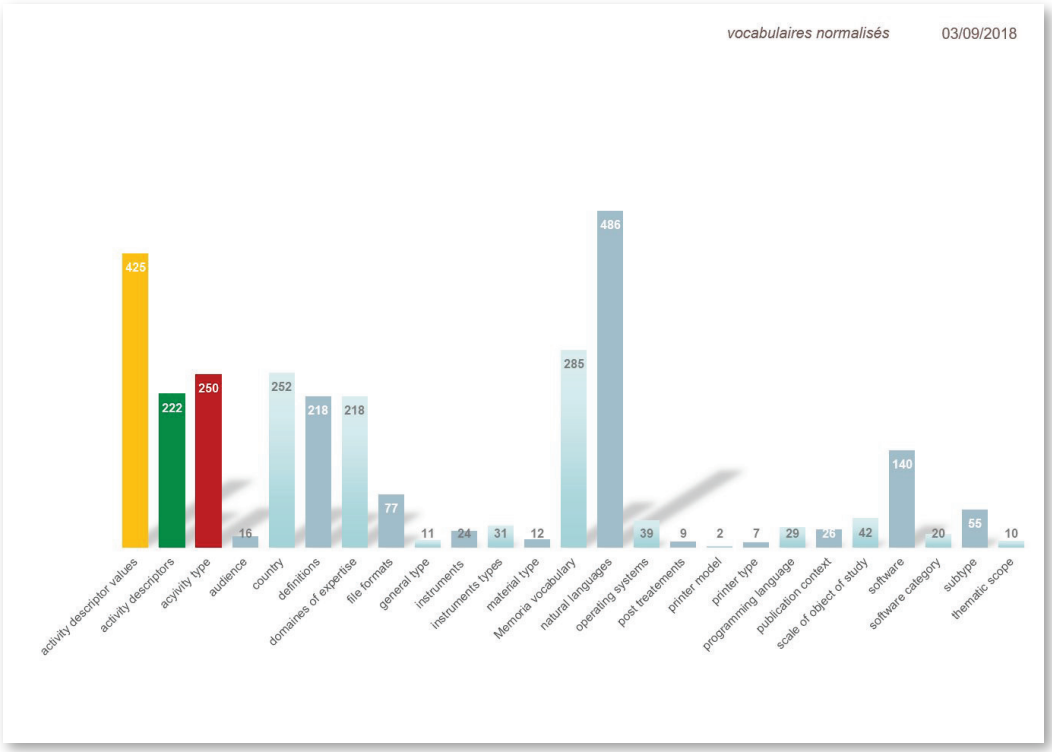
Moyens techniques

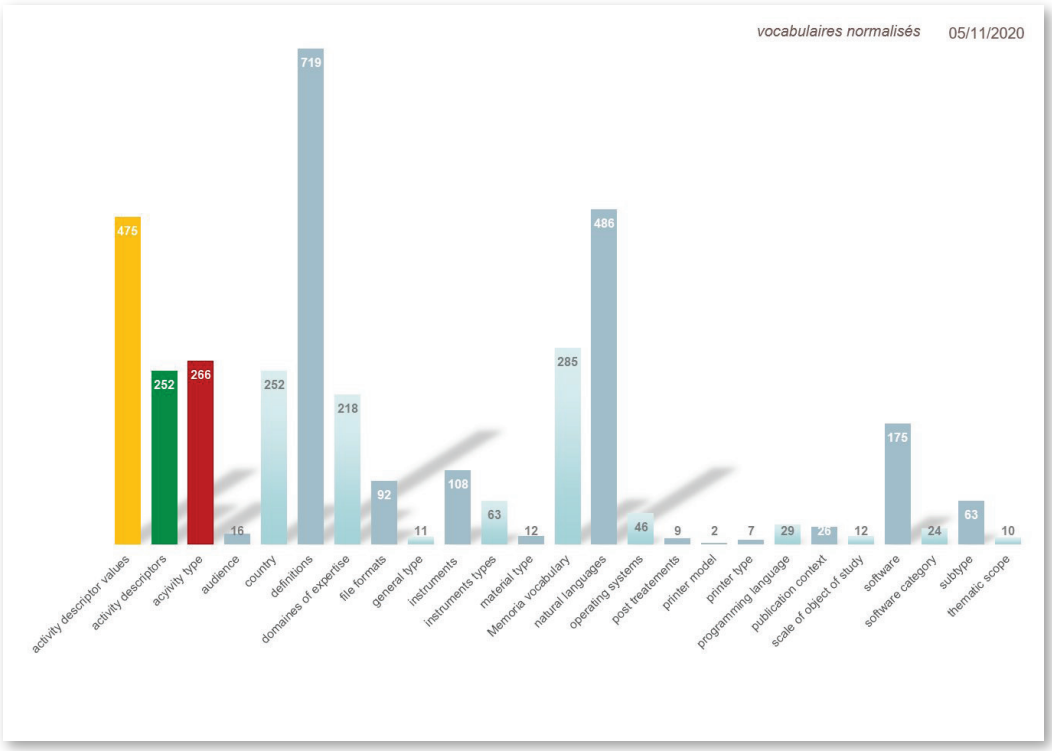
Du point de vue des choix technologiques, nous nous appuyons sur des solutions, des standards et des formalismes issus du monde du logiciel libre. Ces technologies sont par ailleurs conçues pour le développement d'applications sur Internet (Base de données MySQL, Serveur Apache, Scripts PHP, SVG, Librairie 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. Annexes p. 14)*

ANNEXES

Bilan 2019/2020

2020





Élargissement (sur la période 2018-2020) des listes de vocabulaire contrôlé.

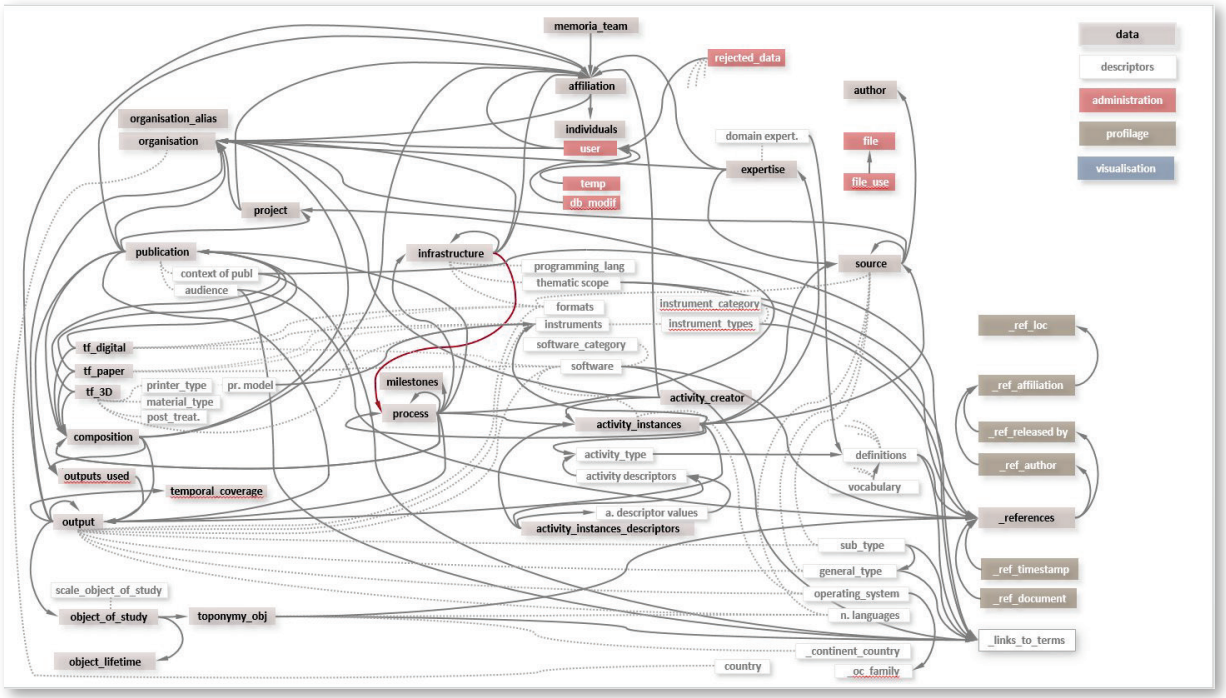


Schéma actuel de la structure de la base de données.

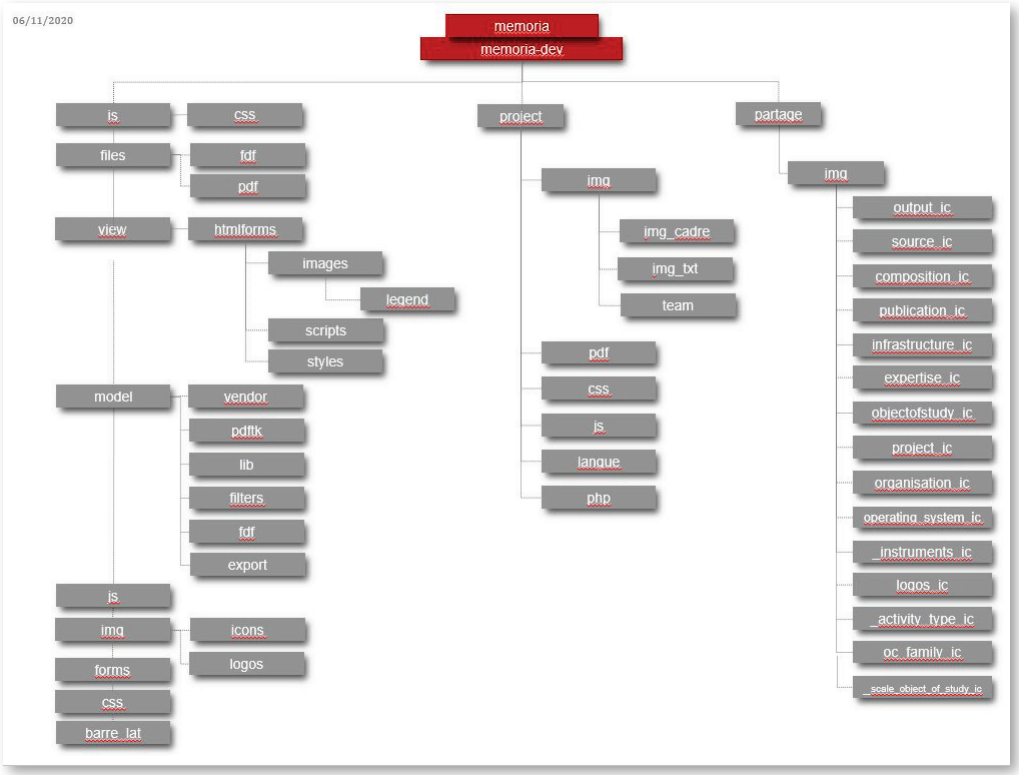
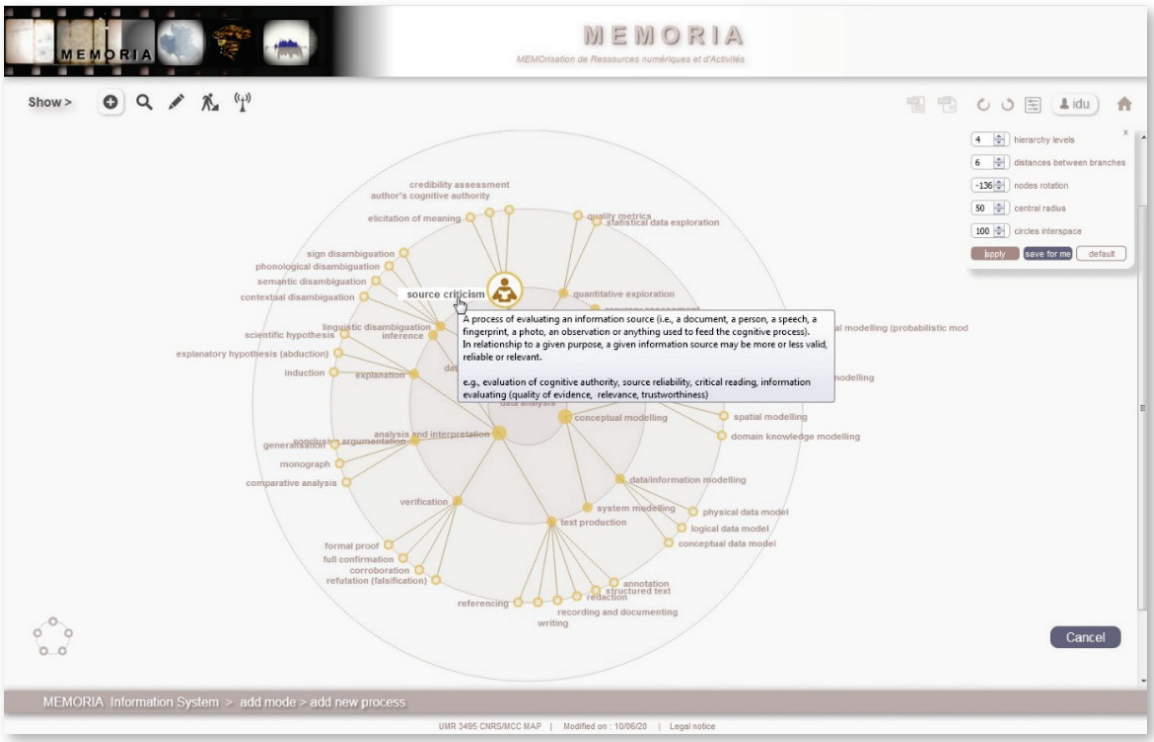
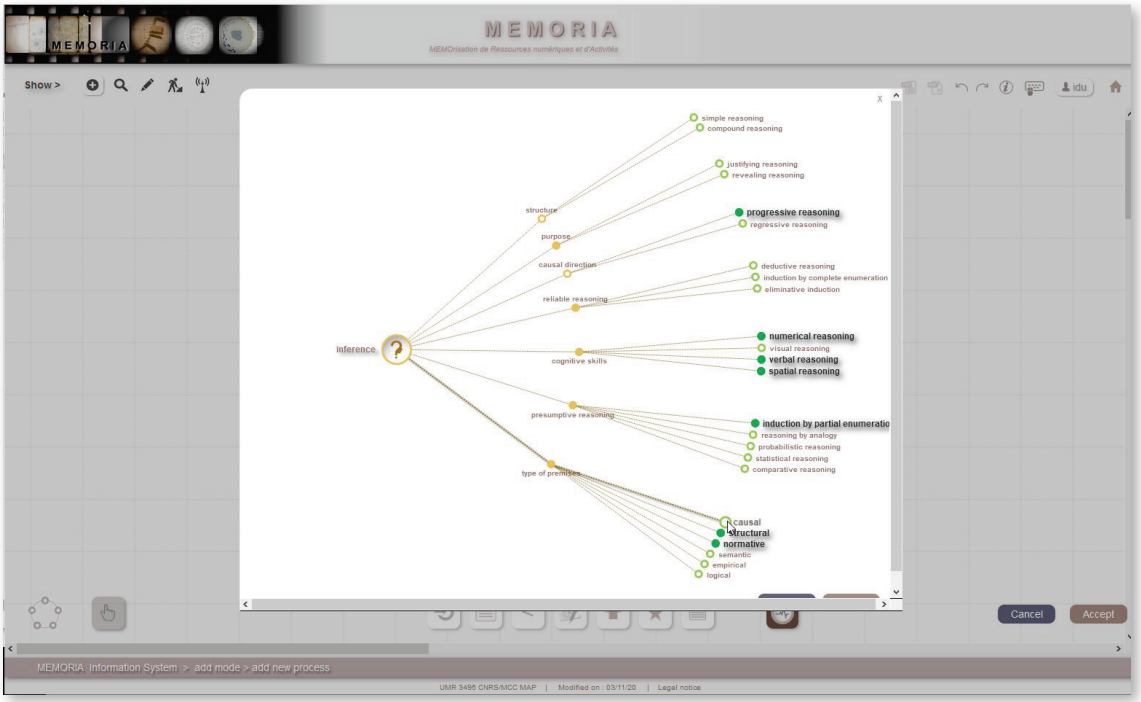


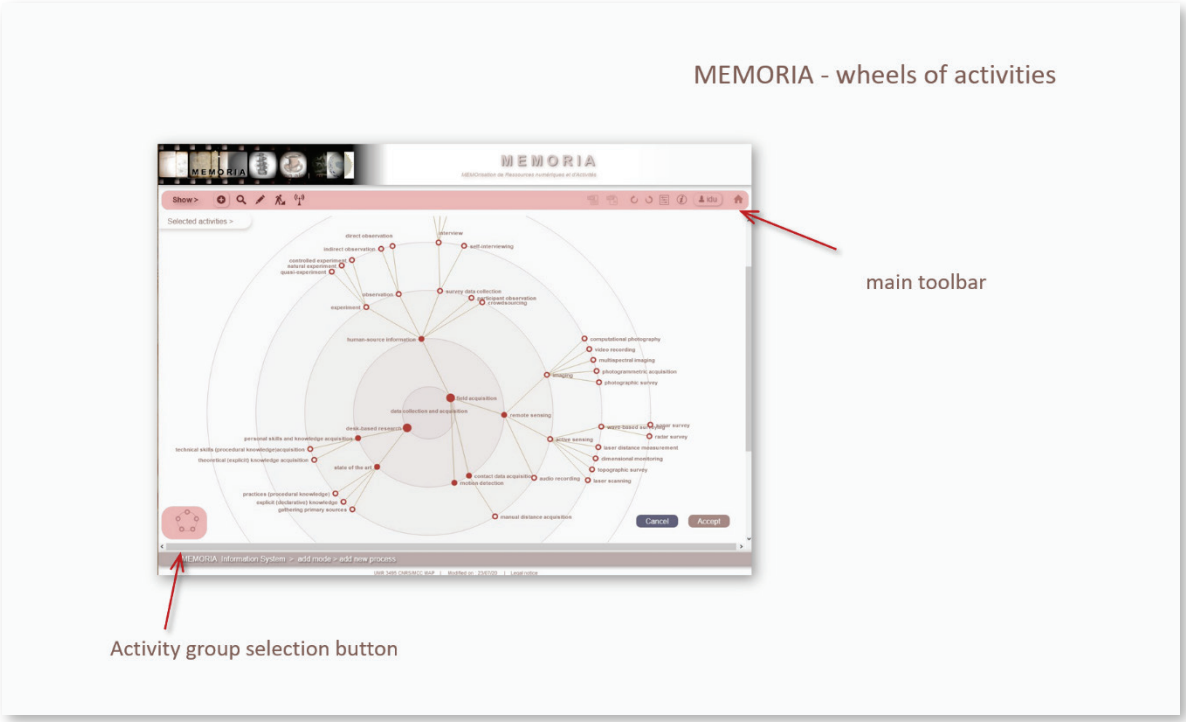
Schéma actuel de la structure des fichiers.

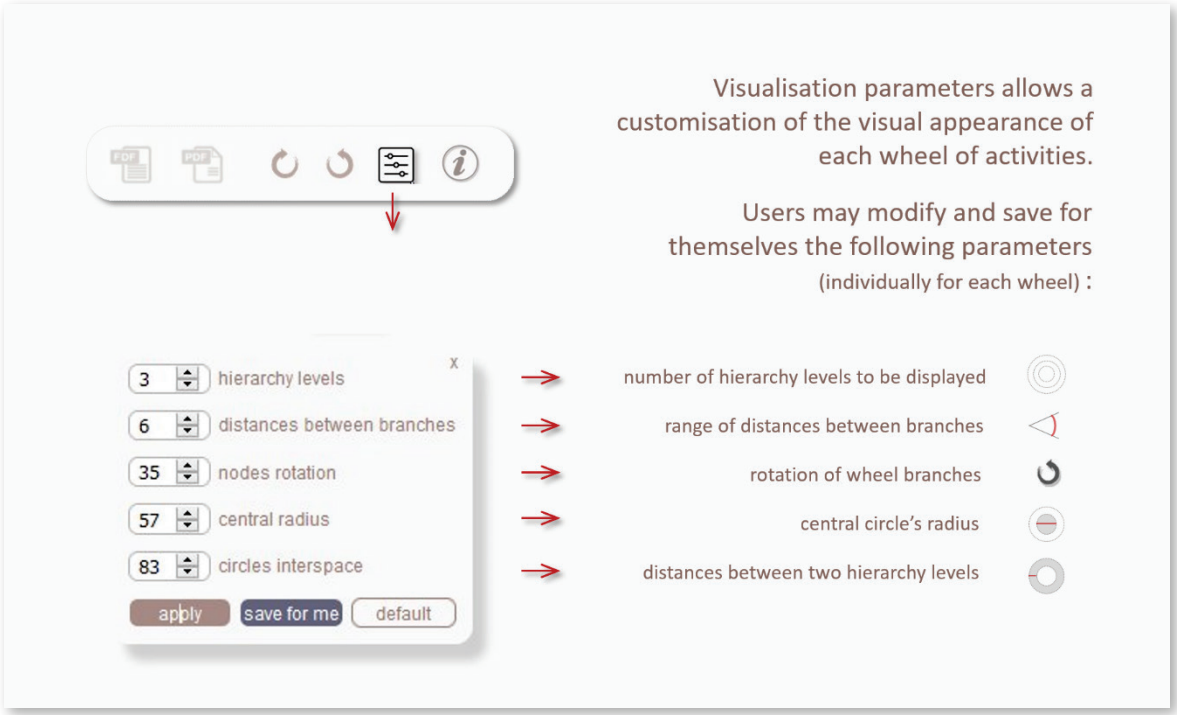


Implémentation du SI MEMORIA : accès à la sélection d'activités via un formalisme visuel appelé « roue d'activités interactives » (illustrée ici sur un des cinq groupes d'activités, groupe « analyse de données »).

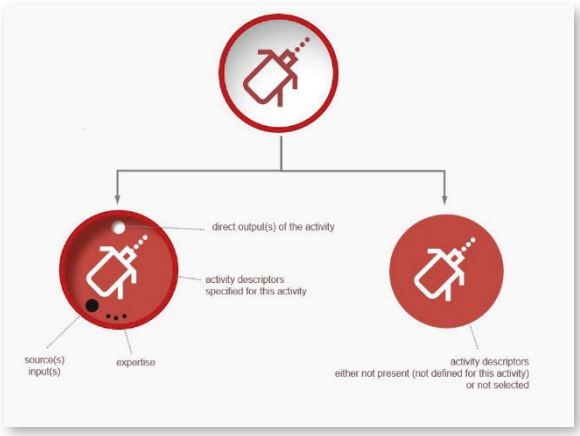


Implémentation du SI MEMORIA : accès – ici dans l'interface de composition des processus - à l'ensemble des descripteurs (cercles jaunes) et à leurs valeurs pour une activité spécifique, l'activité « inférence » (groupe « analyse de données »).



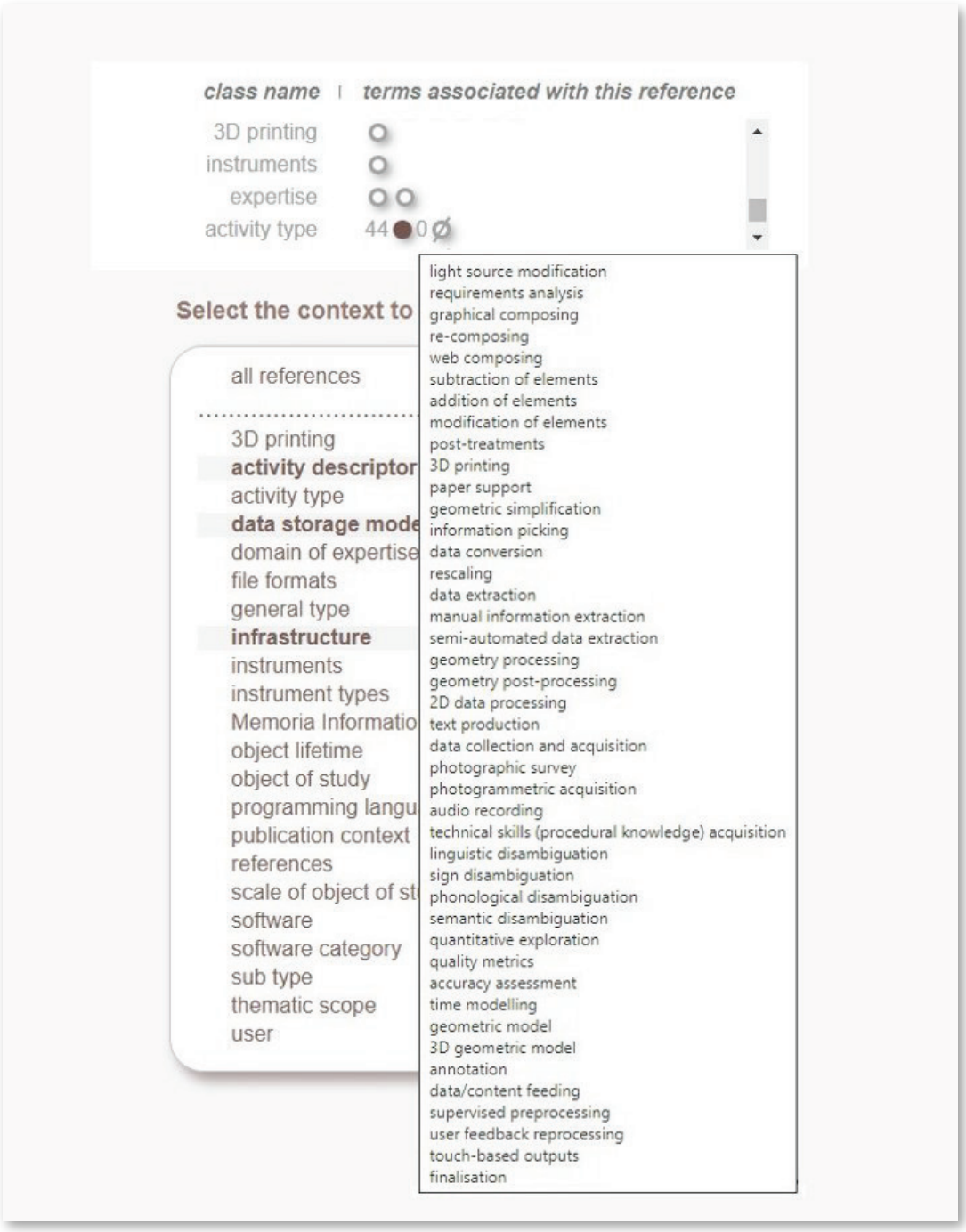


Implémentation du SI MEMORIA : la sélection d'activités se fait via un formalisme visuel appelé « roue d'activités interactives », l'interface autorise l'utilisateur à configurer la forme qu'il souhaite donner à cette « roue » : écartement des branches, rotation des branches, nombre de niveaux hiérarchiques affiches, *etc*.

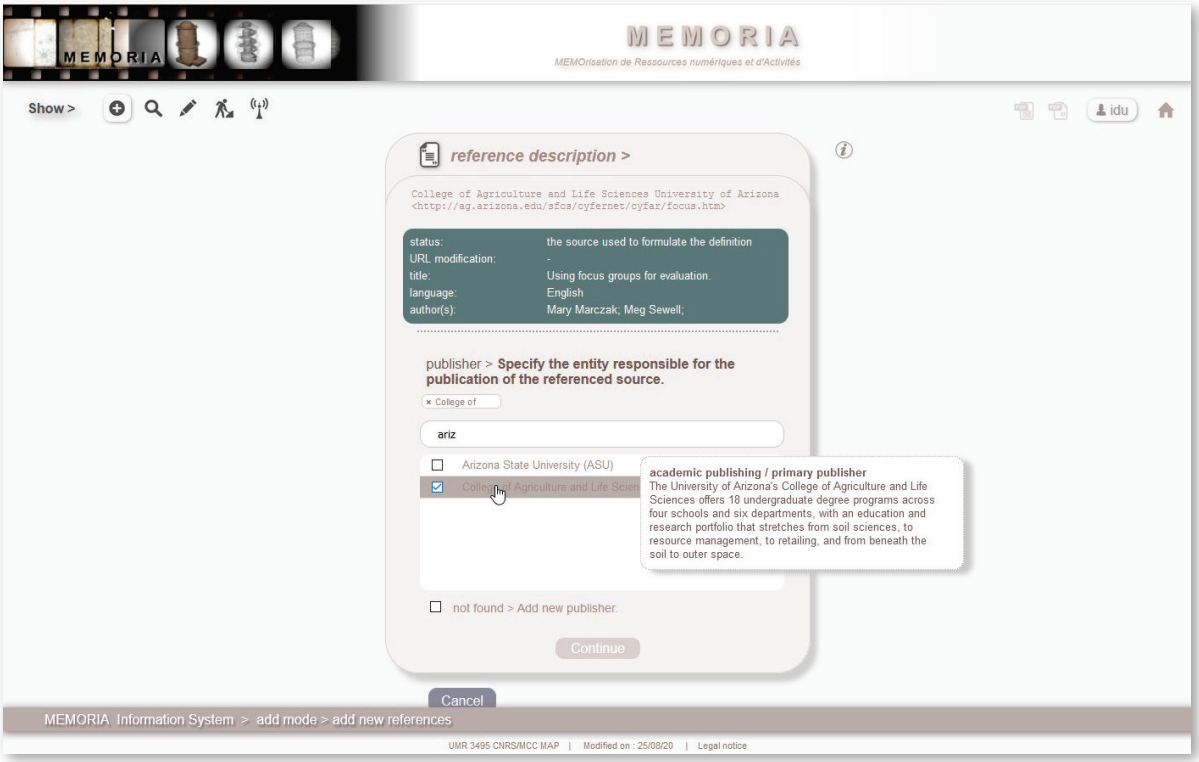


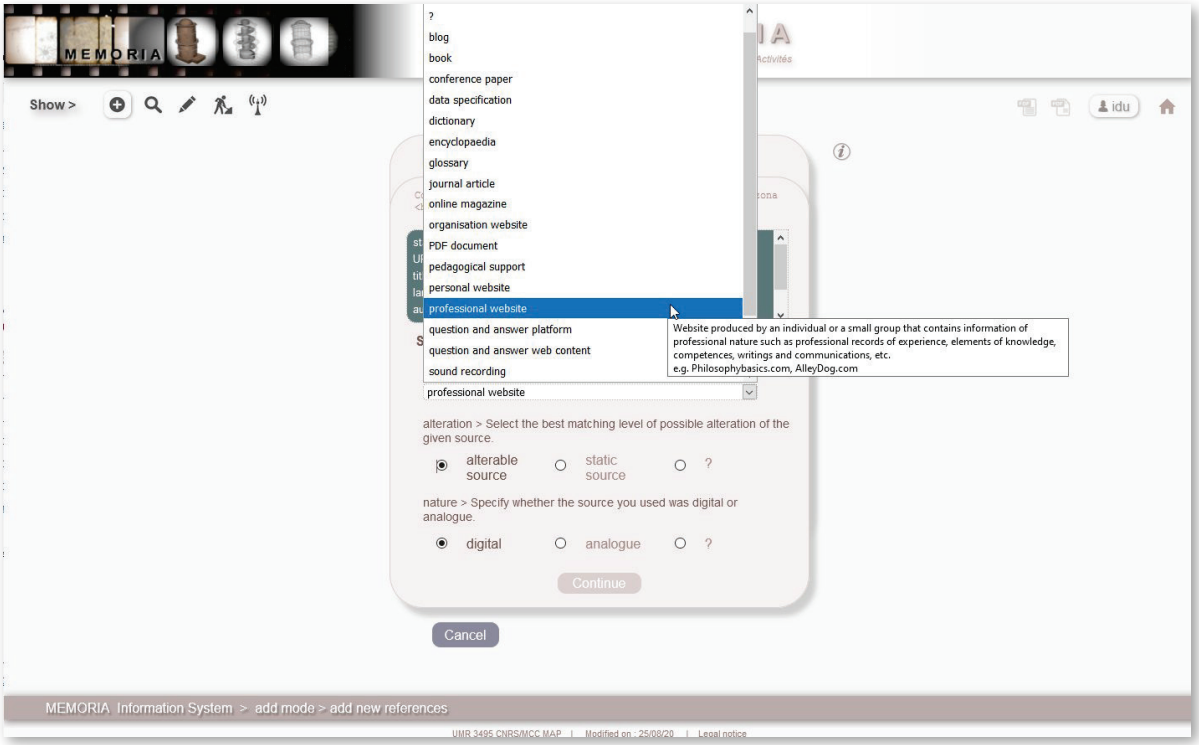
Un exemple de la façon dont les icônes '*déclaratifs'* sont modifiés pour rendre compte de l'état de complétude dans l'effort *de* renseignement d'une activité – ici trois *niveaux de renseignement* basique sont illustrés pour une même activité - « mesure par télémètre laser ». (*Haut: l'utilisateur a juste fait état de cette activité, sans aucun détail. Gauche: L'utilisateur a renseigné l'activité et ses descripteurs. Droite: L'utilisateur a renseigné l'activité mais pas les descripteurs - ici par exemple qui et quand, mais rien sur l'instrumentation*).

Sélection d'icônes correspondant à deux groupes d'activités (groupes « acquisition de données » – haut, et « analyse de données » - bas).

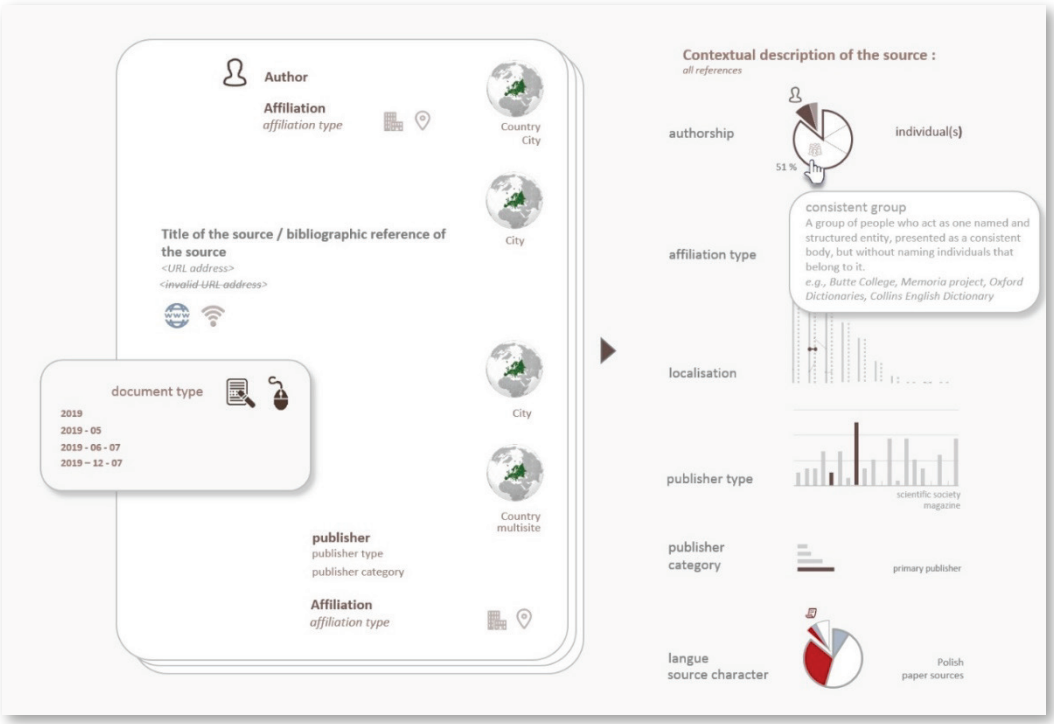


Travail de requêtage et de visualisation mené en 2020 sur la partie « sources documentaires » du projet MEMORIA (i.e. ancrage biblio-webliographique des termes et définitions mobilisés par le projet). Sélection des groupes des termes (memoria-ref).



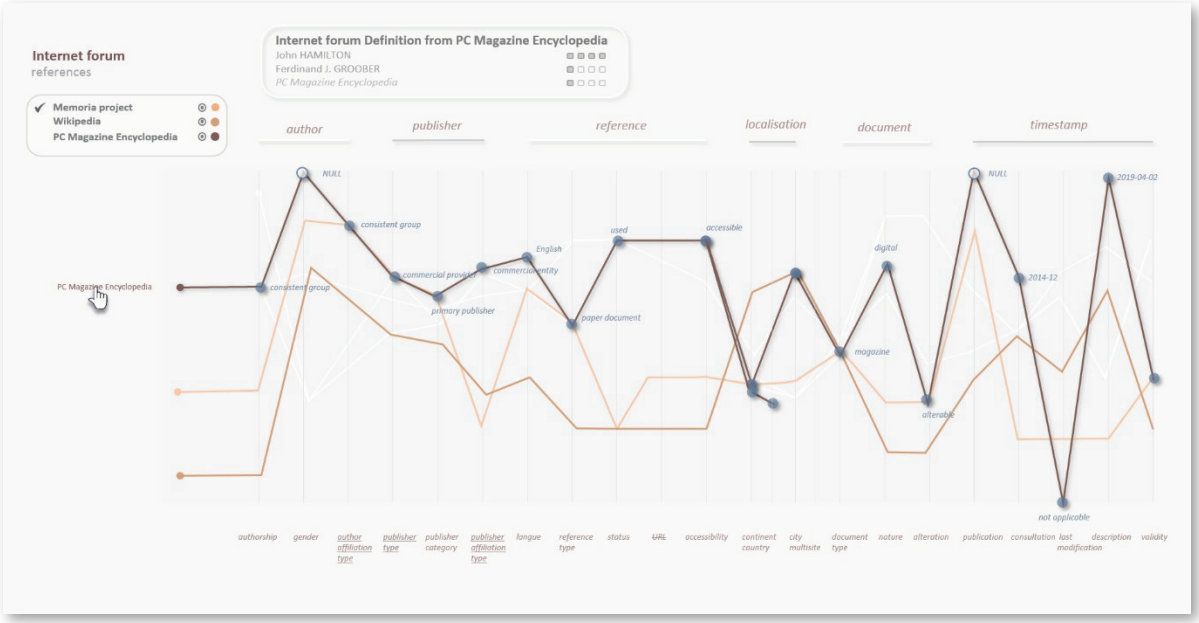


Partie « sources documentaires » du projet MEMORIA (i.e. ancrage biblio-webliographique des termes et définitions mobilisés par le projet) - mode *indexation*. Formulaire permettant la catégorisation des sources électroniques (memoria-ref).



Projet d'interface du système d'information MEMORIA en mode *analyse* – partie « sources documentaires » (langage visuel et interactions). Préfiguration d'un dispositif de visualisation comparative facilitant l'analyse des ressources (ressources électroniques et bibliographiques) en fonction des critères de catégorisation définis dans la base de données memoria-ref.

Ces outils visuels combinent différents types de visualisations donnant accès à des informations détaillées sur une ressource spécifique (en haut), ou sur la particularité d'une ressource par rapport à l'ensemble des références, ou par rapport à un groupe de références spécifique (différents indicateurs sont représentés par différents formalismes visuels).



Projet d'interface du système d'information MEMORIA en mode analyse – partie « sources documentaires » (langage visuel et interactions). Préfiguration d'un dispositif de visualisation comparative facilitant l'analyse des ressources (ressources électroniques et bibliographiques) en fonction des critères de catégorisation définis dans la base de données memoria-ref.
Visualisation comparative des particularités de différentes ressources (ou groupes des ressources) par rapport à l'ensemble des références, ou par rapport à un groupe de références sélectionné - inspirée par le formalisme visuel *parallel cooordinates* (diagramme à coordonnées parallèles).