

MEMORIA

SI en ligne pour décrire des protocoles de recherche

Iwona Dudek*, Jean-Yves Blaise*, Miora Rabefandroana*

*UMR CNRS/MC 3495 MAP 31 chemin Joseph Aiguier 13402, Marseille, France
iwona.dudek@map.cnrs.fr, jean-yves.blaise@map.cnrs.fr,
miora.rabefandroana@map.cnrs.fr
<http://memoria-dev.gamsau.archi.fr/is/enter.php>

Résumé. Le prototype MEMORIA est un système d'informations en ligne dont le but est d'apporter aux acteurs d'opérations scientifiques une solution pratique pour formaliser et décrire des flux de recherche. L'initiative repose sur l'idée qu'au-delà de métadonnées décrivant des résultats finaux, des productions, ces acteurs attendent des moyens pour assurer leur vérifiabilité, leur reproductibilité et leur comparabilité en leur associant un '*historique de production*' partageable. Cette contribution synthétise les objectifs du système puis en présente les différentes facettes (service attendu, fonctionnement, choix majeurs en matière de modalités d'interactions) et les limites intrinsèques.

1 Introduction

Depuis plusieurs années déjà la notion de science ouverte fait son chemin dans différentes disciplines scientifiques, et avec elle l'établissement de principes directeurs, et en premier lieu les principes dits « FAIR » (cf. Wilkinson et al., 2016). Mais au-delà de l'adoption de ces principes, faite bien volontiers en théorie par les acteurs de la science, comment les traduire concrètement dans le suivi des opérations scientifiques ? Et ces principes sont-ils à appliquer aux **données** de la recherche *stricto sensu*, ou plus largement aux **opérations et protocoles** de recherche ayant été mobilisés pour produire ces données ?

Le projet MEMORIA (et le Système d'Informations en ligne qui en est la partie émergée), est une tentative de réponse concrète au besoin des acteurs de la science - et en particulier en SHS - de donner une traduction exemplifiée aux notions de vérifiabilité, de reproductibilité et de traçabilité. Ceci dit le concept (et les outils) de « gestion des connaissances »¹ n'est pas propre à l'activité des scientifiques, loin de là (cf. Antoine et Blum, 2014). MEMORIA s'intéresse à cette question dans le contexte spécifique de l'activité scientifique (cf. Bouhedi, 2017), mais plus précisément encore dans la communauté des sciences patrimoniales, où les flux de recherche comportent une part significative de décisions humaines subjectives, de protocoles de recherche non explicites, de connaissances mal formalisées, d'interprétations intuitives non documentées, non reproductibles, *etc.* Pourtant, cette communauté a assisté au cours des dernières décennies à l'émergence de très grandes quantités de « produits » numériques, soit à

1. Autrement dit le besoin de partager, d'utiliser, d'organiser et de maintenir les connaissances et les informations au sein d'un collectif.

la suite d'efforts massifs de numérisation, soit en raison de la capacité croissante des acteurs à produire du matériel d'origine numérique. *Comment donc accompagner cette démarche en termes de reproductibilité, de réutilisabilité et de contre-examen des résultats si les protocoles de recherche restent des « événements » ponctuels non formalisés ?*

L'approche présentée part du constat que transmettre des connaissances ne se limite pas à transmettre des conclusions, mais doit inclure la volonté de transmettre un raisonnement, un ensemble des choix - volontaires ou subis - concernant des méthodes, des outils, des techniques et des instruments. Elle part également du constat que l'article scientifique (souvent présenté comme l'objectif principal de l'activité du chercheur), est pour des questions de format rarement l'outil propice à l'inscription de travaux dans cette logique de transmission.

Ce document constitue l'accompagnement écrit d'une démonstration du prototype MEMORIA, au travers duquel est expérimentée une solution pratique de formalisation et de description intersubjective des flux de recherche. Le prototype a été conçu et développé au sein de l'UMR MAP², et doit permettre de décrire, d'enregistrer et d'archiver de manière structurée les résultats de notre travail dans leur profondeur. Il ne s'agit donc pas de sauvegarder ou décrire des fichiers contenant des résultats, mais d'identifier, de structurer et de préserver des informations sur les objectifs, choix et contexte de création de ces résultats.

Le terrain d'expérimentation déclencheur de l'expérience a été à l'origine les pratiques scientifiques de son laboratoire d'émergence, dont les problématiques de recherche sont centrées sur un renouvellement des méthodes d'acquisition et d'analyse de données sur le patrimoine architectural et culturel à l'heure du « tout numérique ».

Sur le plan de son ancrage disciplinaire l'initiative MEMORIA se situe à l'intersection de :

- travaux portant sur la notion de provenance des données (cf. Lebo et al., 2013) et plus particulièrement sur l'application des principes « FAIR » dans les sciences du patrimoine bâti (cf. Marlet et al., 2019; Ronzino et al., 2016) ;
- recherches issues du champ InfoVis portant sur l'assistance par des moyens non verbaux au raisonnement sur des protocoles à composante temporelle (cf. Tominski et al., 2017; Mclachlan et Webley, 2021). L'approche MEMORIA relève plus précisément de la sous-discipline visualisation de connaissances telle que définie par Kienreich (2006) - solutions se concentrant sur le transfert de connaissances entre personnes, et opérant le plus souvent sur des jeux d'information relativement petits mais fortement structurés. La visualisation des protocoles s'appuie sur le modèle *temps ordinal* décrit dans Aigner et al. (2011) permettant de travailler sur l'ordre (*A* avant *B*, *A* en parallèle de *C*) sans pour autant devoir inscrire chaque activité dans un calendrier.

Il faut préciser que l'initiative ne vise clairement pas à instrumenter les connexions entre les outils et plateformes existants construits autour des questions telles que l'interopérabilité, la standardisation, l'archivage à long terme - questions relevées par exemple dans la plateforme ISIDORE (SHS) - mais plutôt à aider les acteurs à mieux se comprendre, à échanger sur les méthodologies, à réfléchir sur les pratiques - enjeu relevant d'une forme *d'éthique* (cf. Dudek et Blaise, 2018; Letterier, 2016).

2. MEMORIA est un projet à long terme (initié en 2014), visant à développer un système d'information permettant la description, la structuration et l'archivage des ressources produites au travers des différentes activités de l'UMR 3495 CNRS/MC MAP. Il se déploie en s'appuyant sur les ressources propres du laboratoire, avec le soutien du Département de la Recherche, de l'enseignement Supérieur et de la Technologie du Ministère de la Culture, et de l'ANR (projet SESAMES ANR-18-CE38-0009-01).

Le présent document est organisé en 3 parties. La section 2 synthétise les objectifs du SI MEMORIA puis en présente les facettes principales : rôle, fonctionnement, modalités d'interactions, *etc.*. La section 3 revient sur quelques verrous scientifiques et technologiques majeurs, puis la section 4 fait état des limites intrinsèques du système, et résume les étapes à venir.

Nous utiliserons un exemple très simple (un processus de relevé architectural manuel se traduisant par des plans et coupes portant indications sur les dimensions de l'édifice), comme fil rouge pour donner une idée concrète de l'application du prototype sur des cas réels.

2 Le SI MEMORIA : principes et développement

Le SI MEMORIA a pour objectif de permettre aux acteurs concernés de décrire des résultats de recherche (*i.e.* les extrants), et de les associer à des indicateurs permettant de mieux retracer la façon dont ces résultats ont été obtenus, pour :

- assurer la transmissibilité intersubjective de ces processus, sur le long terme ;
- permettre l'interprétabilité, la vérifiabilité et la reproductibilité des résultats ;
- favoriser une lecture comparative et cumulative facilitant le raisonnement sur nos méthodes de travail, leur évolution, les biais disciplinaires, ... ;
- encourager des pratiques intègres en matière de reconnaissance et d'attribution des résultats scientifiques.

La présentation du SI dans cette section est organisée en trois sections. La première est consacrée aux services attendus, et aux notions clés à comprendre pour en déchiffrer le fonctionnement. La seconde commente et illustre le choix fait de privilégier des interfaces à forte composante visuelle. La troisième situe les principales caractéristiques techniques du système.

2.1 Rôle, notions clés, principes d'interfaçage du SI MEMORIA

L'usage attendu du SI MEMORIA peut être résumé ainsi :

- documenter *les extrants* (numériques ou non) issus de protocoles de recherche ou de développements technologiques ;
- décrire les processus (méthodes, choix, outils, ...) mobilisés pour produire ces extrants *d'une manière qui permette à terme une mise en comparaison* ;
- assurer un accès structuré aux données au sein d'un système cohérent (*le quoi, où, quand et comment d'un extrant, où puis-je le trouver*) ;
- proposer un mode visuel d'analyse des données cumulées pour par exemple assoir un travail d'analyse réflexive sur le cumul *ou* la dispersion des connaissances et de savoir-faire (par exemple dans des projets construits autour de la participation de personnels en CDD).

Le prototype est construit sur une série de notions clés qu'il convient de définir pour en expliquer le fonctionnement concret, et le service qui en est attendu. Nous en mentionnons ici quatre essentiels, et renvoyons le lecteur au système lui-même pour plus de détails. **L'output** est le composant de base, c'est un *résultat* considéré significatif au sein d'un flux de recherche, quelle que soit sa position chronologique dans ce flux (ex. des données brutes de terrain, une ontologie, une restitution 3D). L'output sera dans notre exemple fil rouge *relevé architectural manuel* un jeu de plans, coupes et schémas préparés et remplis à la main *in situ* face à l'édifice, puis numérisés une fois de retour en laboratoire.

MEMORIA. SI en ligne pour décrire des protocoles de recherche

L'output est décrit par une liste de métadonnées classiques permettant de spécifier son contenant (ex. titre, auteurs, type, format, ou date de production) et son contenu (ex. objet d'étude, couverture temporelle d'analyse). On décrira notre output *relevé* comme une collection de fichiers (plans, coupes, schémas) numérisés avec un format de sortie donné (typiquement JPG, PDF, *etc.*), produit pendant une période allant du jour de la visite sur site à celui de la numérisation, lié à une ou plusieurs personnes, portant sur tel ou tel édifice, *etc.*

Il est associé à un *processus*, constitué d'activités dont l'ordre d'exécution peut être spécifié ou non. Si l'on a assisté à (ou conduit) un relevé architectural manuel on va pouvoir spécifier cet ordre en distinguant notamment les phases préparatoires (dessin à la main du plan et de la coupe, puis identification sur ces dessins des dimensions à relever), les phases d'exécution (relevé de telle ou telle dimension, prise de notes, vérification en comparant la dimension relevée avec celles déjà notées, *etc.* - le tout de manière répétitive pour les différentes parties de l'édifice), et enfin les phases de numérisation. Mais si en revanche on doit renseigner un processus de relevé architectural manuel *a posteriori*, processus qui aurait été exécuté par d'autres à un moment que personne n'est capable d'identifier clairement, alors la description du processus peut se faire sous une forme moins assertive et ordonnée. Un processus peut être lié à des processus « précédents », *i.e.* des processus antérieurs dont il est dépendant. Si, par exemple, une phase de fouille de données en laboratoire (moissonnage web ou consultation d'archives) a eu lieu avant le relevé architectural manuel lui-même, afin de préparer un premier plan approximatif de l'édifice, ce processus préparatoire peut être décrit indépendamment et agir comme processus précédent le processus de relevé de l'édifice *in situ*.

L'élément atomique d'un processus est l'*activité*. Chaque activité est définie, exemplifiée et documentée, et dans la majorité des cas caractérisée par une série de descripteurs spécifiques (285 activités différentes identifiées à ce stade). Les activités sont réparties en 5 groupes (acquisition de données, filtrage et traitement de données, analyse des données, protocoles d'exploitation, finalisation). À l'intérieur de chaque groupe, elles sont organisées de manière hiérarchique, des plus générales aux plus spécifiques et représentées dans ce que nous appelons des *roues d'activités*. Cette hiérarchie est conçue pour apporter une forme de flexibilité à l'usage : cela permet à l'utilisateur au besoin de sélectionner une activité relativement générique en cas d'informations imprécises ou lacunaires. Notre relevé manuel aura typiquement inclus (en phase d'exécution) des activités individuelles telles que prise de mesures (*manual distance acquisition*), report écrit sur les plans coupes et schémas préparés en amont (*recording and documenting*), annotation (*annotation*) le cas échéant, analyse (*analysis and interpretation*) par exemple dans le cas de surfaces irrégulières, contrôle (*verification*) en croisant avec d'autres mesures. Chacune de ses activités est susceptible d'être caractérisée par des descripteurs spécifiques : dans le cas d'une prise de mesure on a tout intérêt à préciser notamment avec quel instrument elle a été effectuée. Dans l'étape de spécification d'un processus, des activités, ou des séquences d'activités, peuvent par ailleurs être déclarées comme étant répétitives ou itératives. Dans notre exemple fil rouge la séquence comprenant prise de mesure, report écrit, annotation, puis analyse et vérification est répétitive : elle est reproduite pour chaque mesure individuelle. Mais elle est également déclarée comme itérative : la phase d'analyse et vérification peut conduire à recommencer la mesure, si on a un doute sur sa fiabilité.

Le prototype est aussi construit sur des éléments de méthode *de fond* :

- une phase d'élicitation chronophage mais indispensable comprenant identification collective des activités et recherche bibliographique débouchant sur des jeux de défini-

tions alternatives, reformulées et exemplifiées, chaque définition renvoyant aux sources d'origine ;

- des interfaces visuelles, qui doivent par exemple donner accès aux résultats de requêtes sur des outputs, mais aussi montrer l'évolution des méthodes de travail ou des techniques et outils utilisées au cours du temps ;
- un module d'analyse comparative qui doit notamment rendre lisibles par exemple les interactions entre acteurs au sein d'un processus, ou les rapports de dépendance entre processus au sein d'un flux de recherche.

Deux modalités d'interactions utilisateur-système en accès restreint ont été implémentées, les modes *indexation* et *modification*. Le système est par ailleurs ouvert au monde extérieur dans ses parties *navigation* et *analyse*.

2.2 Ergonomie et esthétique

Une des caractéristiques fondamentales du SI MEMORIA est la volonté d'offrir de la souplesse en mode indexation - ex. permettre de varier l'ordre dans lequel les informations sont renseignées. Mais plus largement il se caractérise par la volonté de faire appel de façon soutenue à des langages visuels comme moyens de synthétiser et de partager des informations à un niveau *non verbal* (Fig 1).

Cette stratégie répond à quatre priorités :

- aider l'utilisateur à se repérer facilement, par des moyens visuels (ex. lui permettre de savoir à quel stade du travail il en est, différencier éléments obligatoire et facultatifs) ;
- maximiser la quantité d'informations exprimées sans faire appel aux langues ethniques (ex. couleurs des groupes d'activités, icônes des activités) ;
- encourager l'utilisateur dans ses efforts en lui proposant un environnement visuel simple et que l'on veut agréable ;
- s'appuyer sur les enseignements de la discipline InfoVis (visualisation d'informations) au sens large pour ce qui est des modules d'analyse.

2.3 Implémentation et caractéristiques 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 (système de gestion de bases de données relationnelles MySQL, serveur Apache, scripts PHP, composants graphiques SVG, méthode AJAX, bibliothèques JavaScript JQuery / JQuery UI, HTML2Canvas, ...). MEMORIA est donc un système d'Information pour le Web relativement classique sur le plan technique. Ceci dit le prototype porte un choix technique un peu particulier, un mode d'indexation « hybride » - l'idée est de combiner des indexations utilisateur classique via formulaires HTML et des indexations via des formulaires PDF téléversables en temps décalé pour compatibilité avec des situations de travail hors ligne.

Le SI n'a pas été conçu autour d'exigences de performance au sens temps de réponse ou masse de données, mais au sens service potentiel rendu, à titre exploratoire. Le passage de ce prototype à un système déployé à grande échelle nécessiterait des moyens bien évidemment largement supérieurs aux nôtres, et une reprise de certains des choix techniques très pragmatiques qui ont été faits. A ce stade nous avons privilégié une démarche de questionnement du

MEMORIA. SI en ligne pour décrire des protocoles de recherche

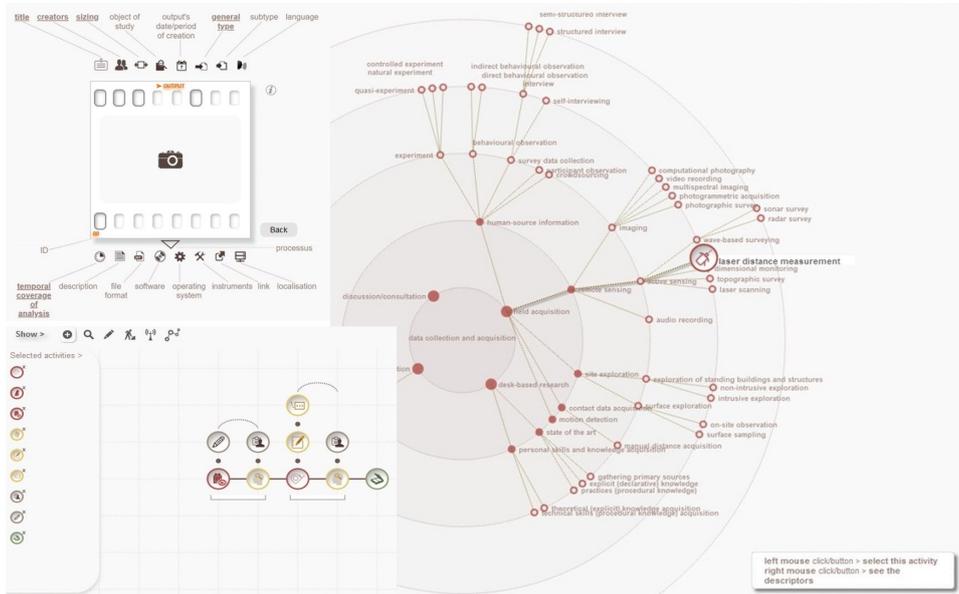


FIG. 1 – Trois exemples des choix de composition graphique du SI MEMORIA. Haut gauche, la métaphore visuelle « pellicule de film » utilisée pour informer l'utilisateur sur les jeux d'informations à renseigner. Bas gauche, l'espace de composition des processus avec sa grille systématique sur laquelle vont « s'accrocher » les activités que l'utilisateur va glisser-déposer depuis la liste située en bandeau gauche. Est représenté ici notre processus fil rouge « relevé architectural manuel », caractérisé par deux séquences à la fois répétitives et itératives (préparation des plans et coupes sur lesquels reporter les mesures, prises de mesures elles-mêmes). Le processus se termine sur une activité de numérisation (une fois de retour en laboratoire) des documents produits in situ. En fond, capture d'écran d'une roue d'activités (catégorie activités d'acquisition) au moment où l'utilisateur amène le curseur sur un des noeuds.

prototype, une démarche de « tests d'acceptabilité » avec la mise en place d'une 'copie' partagée du système d'information en ligne servant d'outil d'évaluation du système et de dialogue entre partenaires (UMR 3495 MAP et UMR 7324 CITERES - LAT, voir section 4).

3 Verrous scientifiques et technologiques

Ce travail se heurte d'abord à un verrou scientifique majeur : parvenir à identifier et organiser les activités propres à une discipline ou à une unité de recherche, et ce dans un contexte où des vocabulaires de spécialité divergents existent. Cela se traduit par la nécessité de mettre en place une longue phase d'élicitation de connaissances (ou de savoir-faire). Il faut donc évidemment préciser que la liste structurée des activités présente dans le SI MEMORIA à ce stade n'est pas universelle - elle ne correspond qu'aux besoins de ceux qui l'ont créée puis étendue.

Un autre verrou de fond à mentionner est la question de la granularité d'analyse : autrement dit celle du choix utilisateur, difficile à opérer, de ne décrire que l'indispensable pour éviter une surcharge d'informations.

Enfin il faut mentionner également des verrous plutôt technologiques ou organisationnels : stratégie de maintenance et de gestion des accès distants, choix d'une stratégie de vérification de nouveaux contenus, exploration de modalités d'interrogation et de mise à jour du système originales, et surtout développement de dispositifs de visualisation d'informations et d'analyse permettant réellement d'aller vers un travail de lecture comparée des évolutions dans les pratiques de recherche .

Plus largement le SI MEMORIA pose la question de l'acceptabilité, par les acteurs visés, d'un investissement dans le recueil et l'indexation de données *non valorisables directement* (par un article scientifique par exemple). Renseigner des outputs ou des processus est « seulement » répondre à une exigence déontologique (que l'on parle de MEMORIA ou non), et faire de la réponse à cette exigence un travail valorisable sur le plan individuel n'est pas trivial. Mais ce dernier point va évidemment bien au-delà du contexte de ce document, centré sur la présentation du prototype lui-même.

4 Limites et perspectives

À ce stade, MEMORIA reste une initiative de recherche et développement en cours. Le SI est fonctionnel mais il doit être considéré comme un prototype en cours du développement et d'évaluation. En effet le problème ici n'est pas seulement de finaliser un système d'information Web robuste, mais d'atteindre un point où son acceptabilité, la charge de travail pour les utilisateurs, et finalement son efficacité et son utilité peuvent être clairement établis et évalués.

L'expérimentation en cours dans le cadre du projet SESAMES (ANR-18-CE38-0009-01) met l'accent sur une question fréquente en sciences du patrimoine. Il s'agit de l'impératif de formaliser et de garder trace des différentes étapes qui mènent des *lieux* (reliques archéologiques, sources) à une reconstruction virtuelle 3D. Ce produit « final » demeure encore aujourd'hui bien souvent sans lien pérenne avec les processus d'interprétation qui le sous-tendent. Cette expérience doit permettre de questionner et d'évaluer la démarche d'une double manière : enrichissement ou adaptation des activités, et retour d'expérience des chercheurs de différentes disciplines notamment sur l'utilisabilité du système.

Ce qui ressort de l'utilisation faite dans ce cadre du système MEMORIA, est qu'un tel système force à la réflexion et à l'analyse critique des méthodes de travail. Plusieurs améliorations ou réorientations ont d'ores et déjà pu être apportées via cette expérimentation (ex. notion d'output cumulatif, de processus théorique, ou de processus de planification). Ainsi, bien que conçu au départ comme une solution concrète pour rendre compte de protocoles de recherche *a posteriori*, le SI s'ouvre aujourd'hui sur une autre forme d'utilisation. L'idée est d'assister l'individu ou le groupe à planifier une opération de recherche, à identifier les compétences nécessaires, les organiser, et visualiser les dépendances à anticiper entre activités comme entre acteurs.

Enfin il est important pour conclure de préciser que MEMORIA n'a pas vocation à assurer le suivi des activités quotidiennes d'individus. Il se veut seulement un moyen offert aux acteurs de la recherche de structurer des indices permettant de comprendre et de réutiliser la production scientifique d'aujourd'hui, pour eux-mêmes et pour les générations futures.

MEMORIA. SI en ligne pour décrire des protocoles de recherche

Références

- Aigner, W., S. Miksch, H. Schumann, et C. Tominski (2011). *Visualization of Time-Oriented Data* (1st ed.). Springer Publishing Company, Incorporated.
- Antoine, A. et G. Blum (2014). La gestion des connaissances ou le knowledge management (km). *Revue internationale de psychosociologie et de gestion des comportements organisationnels XX(49)*, 23–31.
- Bouhedi, M.-C. (2017). Les pratiques de partage des connaissances d’une unité de recherche pluridisciplinaire en interne et externe. *Communication & Management, Eska 14(1)*, 69–86.
- Dudek, I. et J.-Y. Blaise (2018). MEMORIA-la préservation des processus d’étude comme enjeu éthique. In V. Ginouvès et I. Gras (Eds.), *La diffusion numérique des données en SHS, Guide des bonnes pratiques éthiques et juridique*, pp. 231–240. Aix-en-Provence : PUP.
- Kienreich, W. (2006). Information and knowledge visualisation : an oblique view. *MiaJournal, MAP CNRS 0(1)*, 7–17.
- Lebo, T., S. Sahoo, et D. McGuinness (2013). Prov-o: The prov ontology. W3c recommendation, CNRS-CPU (<https://www.w3.org/TR/prov-o/>). 30 April 2013.
- Letterier, L. (2016). Integrity and responsibility in research practices a guide. Guide, CNRS-CPU (<http://www.cpu.fr/wp-content/uploads/2016/12/integrity-and-responsibility-in-research-practices-a-guide-19.11.16.pdf>).
- Marlet, O., E. Zadora-Rio, P.-Y. Buard, B. Markhoff, et X. Rodier (2019). The archaeological excavation report of rigny: An example of an interoperable logicist publication. *Heritage 2(1)*, 761–773.
- Mclachlan, S. et L. C. Webley (2021). Visualisation of law and legal process: An opportunity missed. *Information Visualization 20(2-3)*, 192–204.
- Ronzino, P., F. Niccolucci, A. Felicetti, et M. Doerr (2016). Crmba a crm extension for the documentation of standing buildings. *International Journal on Digital Libraries 17*, 71–78.
- Tominski, C., W. Aigner, et S. Miksch (2017). Images of time visual representations of time-oriented data. *Information design: research and practice, Routledge*, 23–42.
- Wilkinson, M., M. Dumontier, et I. Aalbersberg (2016). The fair guiding principles for scientific data management and stewardship. *Sci Data 3*, art. 160018.

Summary

The MEMORIA prototype is an online information system offering a practical solution to formalize and describe research workflows. The initiative is based on the idea that beyond metadata describing outputs themselves, scientific results - especially in historical and heritage sciences - should be depicted with indicators that would allow for a better understanding of the process through which these results were achieved. The contribution summarizes the objectives of the system and briefly presents its main principles and its intrinsic limits.