
Métamorphoses et hybridations d'un système d'information pour la valorisation des archives numériques

Vers des écosystèmes de connaissances

Samuel Szoniecky

*Laboratoire Paragraphe, Université Paris 8 Vincennes Saint Denis
2 rue de la liberté, 93526 Saint Denis, France
samuel.szoniecky@univ-paris8.fr*

RÉSUMÉ.

Dans cet article, nous proposons une méthode de conception et d'analyse de systèmes d'information pour les humanités numériques. Pour illustrer cette méthode, nous présentons des dispositifs numériques dédiés à l'enrichissement collectif des connaissances par l'expérimentation de la réflexivité et l'analyse comparative des interprétations récoltées. Le système d'information que nous avons développé dans le cadre d'une recherche-action a pour objectif de valoriser les reportages photographiques des présidences de la République française en accès libre sur le site des Archives Nationales. Nous montrons comment la numérisation des documents et leur mise à disposition sont analysables en termes de métamorphose documentaire et d'hybridation des données. Puis, nous expliquons comment ces transformations de l'information peuvent s'automatiser pour obtenir une culture intensive de l'information et une consommation standard des documents. Enfin, nous présentons un dispositif d'enrichissement collectif des connaissances pour montrer comment stimuler l'interprétation et rendre interopérable l'expression d'un point de vue tout en quantifiant sa complexité.

ABSTRACT.

In this paper, we propose a method of designing and analyzing information systems for digital humanities. To illustrate this method, we present web applications dedicated to the collective enrichment of knowledge by the experimentation of the reflexivity and the comparative analysis of the interpretations harvested. The information system that we developed as part of an action research aims to enhance the photographic reports of the presidencies of the French Republic in free access on the site of the french National Archives. We show how the digitization of documents and their availability can be analyzed in terms of document metamorphosis and hybridization of data. Then, we explain how these transformations of information can be automatized to obtain an intensive information culture and a standard consumption of documents. Finally, we present a system of collective

enrichment of knowledge to show how to stimulate the interpretation and make interoperable the expression of a point of view while quantifying its complexity.

Mots-clés : archives numériques, métamorphose, hybridation, intelligence collective, écosystème de connaissances

KEYWORDS: digital archives, metamorphosis, hybridization, collective intelligence, knowledge ecosystem

1. Introduction

Dans le contexte de la prolifération des données, des outils numériques et surtout de leur accessibilité grâce à des portails dédiés à leurs partages, les questions de pourquoi et comment utiliser ces nouvelles ressources dans des projets de recherches en sciences humaines et sociales (SHS) sont au cœur de nombreux débats. Nous n'analyserons pas ici les points de vue des uns et des autres, préférant alimenter ces discussions par l'exposé d'une recherche-action (Mantran, 2017) d'où émergent des problématiques auxquelles nous répondrons par des propositions concrètes. Pour mener nos réflexions sur les humanités numériques (HN), nous sommes partis d'un constat simple. Juste une rue sépare l'université Paris 8 et les nouveaux bâtiments des Archives Nationales (AN) à Pierrefitte mais il n'existe que peu d'interaction entre ces deux mondes. Il nous est apparu qu'un des objectifs des HN pouvait être de faciliter et valoriser la rencontre entre d'un côté des milliers d'étudiants qui viennent apprendre, et de l'autre des millions de documents qui attendent qu'on les interprète. Motivés par l'ambition de transformer une proximité géographique en opportunités pour créer des collaborations intellectuelles entre ces deux institutions, nous nous sommes interrogés sur les moyens de mener à bien ce projet. Comment développer les compétences des étudiants tout en enrichissant nos connaissances sur les archives ? En quoi des systèmes d'informations numériques peuvent-ils faciliter l'émergence d'une intelligence collective ?

Pour contribuer à l'exploration de ces questions et avant d'envisager des expérimentations à plus grande échelle, nous avons conduit en compagnie de Catherine Nyeki et de Jean-Marc Meunier un atelier laboratoire porté par l'Initiative D'Excellence en Formations Innovantes CreaTIC¹. À partir d'une journée d'étude sur les questions de valorisation pédagogique des archives numériques², nous avons travaillé avec les étudiants à des scénarios innovants qui feront l'objet d'autres publications. Dans cet article, nous présenterons le prototype d'un système d'information dédiée à la valorisation des reportages photographiques des présidences de la République française³ que nous avons conçu pour faire émerger des problématiques, les analyser et proposer des réponses concrètes aux questions de valorisation collectives des informations publiques et de leur communication.

Le champ scientifique que nous cultivons à travers ces recherches s'étend des interrogations posées par les chercheurs concernant la calculabilité du sens (Bachimont *et al.* 2011 ; Lévy, 2011), la culture de l'interprétation, l'économie de la connaissance et l'écologie de l'attention (Citton, 2014), la sémiologie graphique (Groupe μ , 2015) pour une représentation dynamique et interactive des connaissances (Reyes, 2017), la modélisation dans les sciences humaines à l'ère du numérique (Latour, 2012 ; Meunier, 2017 ; Sauret, 2017), l'exploration et

¹ <http://idefi-creatic.net/>

² <http://valarnum.univ-paris8.fr/>

³ <https://goo.gl/1c4hR8>, <https://goo.gl/XZP7C1>, <https://goo.gl/LFfPwY>

l'indexation de corpus (Moiragui, 2018). À partir de ces cadres théoriques, nous proposons une méthode de conception et d'analyse des systèmes d'information par modélisation d'existences informationnelles dans des écosystèmes de connaissances (Szoniecky, 2017). Cet article explique l'utilisation de cette méthode pour concevoir des dispositifs numériques d'intelligence collective dédiés à l'enrichissement des connaissances par l'expérimentation de la réflexivité et l'analyse comparative des interprétations récoltées par le dispositif.

Nous montrerons tout d'abord comment la numérisation des documents et leur mise à disposition sont analysables sur le plan des métamorphoses documentaires et des hybridations des données. Puis, nous expliquerons comment ces transformations de l'information peuvent s'automatiser pour obtenir une culture intensive de l'information et une consommation standard des documents. Enfin, nous présenterons un dispositif de valorisation collective des connaissances pour montrer comment stimuler l'interprétation et rendre interopérable l'expression d'un point de vue.

2. Métamorphoses documentaires et hybridation des données

En tant qu'artefacts physiques, les documents sont soumis à de multiples transformations qu'elles soient liées aux dégradations matérielles causées par le temps ou aux actions des individus qui se les approprient pour une simple consultation ou pour les utiliser comme partie d'une composition. Dans le cas des documents numériques, ces transformations sont pratiquement illimitées puisque la matière numérique est décomposable et recomposable à l'infini. Plus encore, ces potentialités de transformation donnent à cette matière la capacité d'un développement autonome (Parisi, 2016) tant dans la forme extérieure de ses « physicalités » que dans le fond conceptuel de ces « intériorités » (Descola, 2005). Les documents numériques développent ainsi un mode d'existence (Latour, 2012) « analogiste » (Serres, 2009) dont nous analyserons l'absolue plasticité en suivant les transformations d'un fonds d'archive.

Une des premières actions que nous avons menée dans le cadre de cette recherche nous a conduits à la salle de consultation des AN pour expérimenter les conditions d'accès aux documents, la numérisation de ceux-ci et leur partage à partir d'un smartphone. Notre ambition était de questionner la viabilité d'une annotation collective des documents par « crowdsourcing » (Andro, 2017). Pour mener cette expérience, nous avons examiné les archives de l'exposition universelle de 1937 à Paris⁴. Cette expérience nous a montré la richesse et la diversité des documents conservés aux AN⁵ mais aussi la difficulté de bien organiser la consultation des documents pour garder la cohérence de l'inventaire et comment la prise de vue avec un smartphone pourrait être améliorée avec une application d'aide à l'annotation.

⁴ <https://goo.gl/Z4SMn7>

⁵ <https://goo.gl/aFcddz>

Nous avons aussi constaté le problème de destruction d'une partie des archives⁶, même en prenant grand soin des documents.

Cette première expérience nous a confirmé que le développement d'une intelligence collective pour la valorisation des archives passe par une numérisation des documents afin de les rendre accessibles et modifiables tout en les préservant de la destruction. Mais aussi et surtout, la transformation numérique doit absolument définir une référence unique et pérenne, une clef d'indexation, entre le document original et ses numérisations afin de conserver ce qu'on pourrait appeler son ADN, c'est-à-dire les attributs qui lui sont propres en tant qu'artefact physique. Lors de la consultation d'un dossier d'archives, cette clef de référence permet de retrouver un document parmi les milliards qui sont conservés. Or s'il existe bien une clef pour le dossier, voire pour les chemises que les archivistes ont créées pour organiser l'inventaire ; à l'intérieur de ces chemises, les documents n'ont pas de référence unique et pérenne. La numérisation des documents que nous avons faite avec notre smartphone devient difficilement exploitable dans un projet d'ingénierie des connaissances et pratiquement impossible pour un travail d'intelligence collective en l'absence de lien entre le document original et sa numérisation. Il n'existe aucune garantie que telle photo est bien la numérisation de tel document provenant de tel fonds d'archive. Un projet d'annotation collective des archives par « crowdsourcing » n'est pas si évident à mettre en place ne serait-ce qu'au niveau de cette première étape de transformation numérique du document qui consiste à l'indexer pour qu'il soit compatible avec des systèmes d'information ouverts et interopérables comme le Linked Open Data (LOD).

Pour différencier un système d'information opérationnel dans le LOD et la simple numérisation d'un document sans indexation, comme ce que nous avons fait avec notre smartphone, nous appelons « métamorphose » une transformation numérique du document qui préserve un lien unique et garanti entre le document source et sa transformation numérique, tout comme l'ADN du papillon est conservé au fil de ses métamorphoses en larve, nymphe et imago.

Pour continuer d'avancer dans notre projet de valorisation collective des archives, nous avons travaillé sur des fonds d'archives pour lesquels des « métamorphoses » étaient possibles. Notre choix s'est porté sur les reportages photographiques autour des présidences de la République française, car ces fonds offrent les avantages d'être libre de droits, de posséder une matière multimédia de qualité et de présenter une description détaillée sous une forme numérique. Ces descriptions en XML sont transformables par des algorithmiques tout en préservant une référence sous la forme d'une URL unique et pérenne qui permet à tout moment de retrouver le document original, quelles que soient ces « métamorphoses ». On peut envisager un autre type de transformation numérique qui donne aux documents de nouvelles potentialités de transformations en le plaçant dans un système d'information particulier. Nous appelons « hybridation » cette transformation qui modifie le « pouvoir d'agir » (Brun, 2017) du document en l'enrichissant de nouvelles données tout en préservant le lien vers un document source.

⁶ <https://goo.gl/7SbcW3>

Entre « métamorphoses » et « hybridations », nous vous proposons de suivre maintenant les transformations des fonds d'archives pour mieux comprendre les enjeux de la valorisation des archives numériques et comment développer des dispositifs d'enrichissement par l'expression d'une interprétation.

3. Cultures intensives de l'information

Un des principaux intérêts des archives numériques tient dans la possibilité d'automatiser les tâches du cycle de vie des documents en le soumettant à des algorithmes. On peut alors industrialiser le traitement documentaire, par exemple lors de campagne de numérisation où l'on crée des documents numériques en les décrivant suivant des normes internationales. Nous ne détaillerons pas les difficultés d'un tel chantier et les problèmes d'interopérabilité qu'il occasionne (Juanals et Minel, 2016), notons en revanche que ces traitements automatiques produisent une croissance rapide de nombre de données et de leurs complexités, ce que nous appelons une culture intensive de l'information par analogie avec l'agriculture intensive dont un des objectifs est de produire le plus possible sur une surface réduite.

La modélisation de ces systèmes d'information sous forme d'écosystème de connaissances permet de les représenter et de calculer précisément la complexité qu'il occasionne en quantifiant les existences informationnelles qui le composent suivant quatre dimensions : des documents, des acteurs, des concepts et des rapports. Le diagramme suivant illustre l'organisation de ces dimensions que nous avons décrites en détail dans d'autres publications (Szoniecky, 2017 ; 2018).

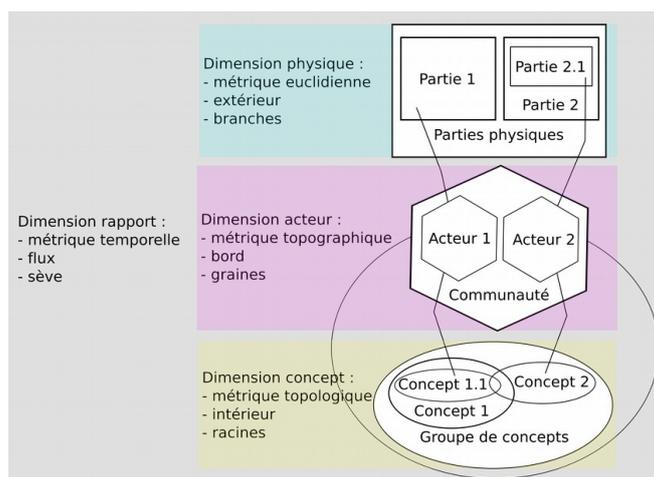


Figure 1. Modèle générique d'existence informationnelle

3.1. État initial de l'écosystème

L'état initial du système d'information sur lequel nous travaillons est le résultat de cette première « métamorphose » qui transforme les inventaires des AN en fichiers. Avec la méthode de modélisation des écosystèmes de connaissances, nous représentons cet état sous la forme du diagramme suivant :

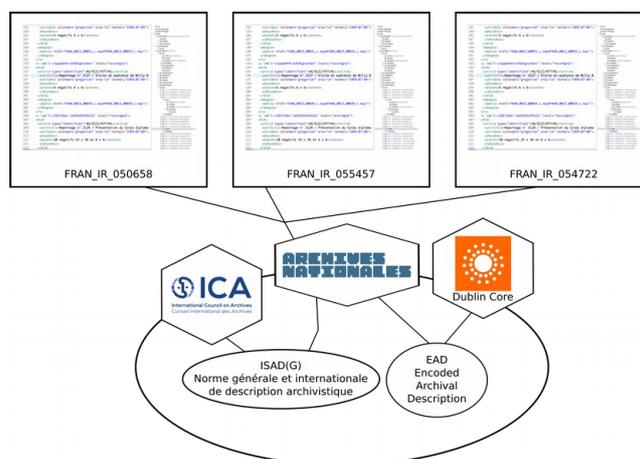


Figure 2. État initial du système d'information

L'état initial est composé :

- de trois documents correspondant chacun au descriptif XML d'un inventaire,
- de l'acteur AN qui produit les documents,
- des acteurs ICA et Dublin Core qui sont les éditeurs des normes utilisées
- de deux concepts correspondant aux deux normes : ISAD(G)⁷ et EAD⁸
- de sept rapports représentant les liens entre acteurs, document et concepts.

À partir de cette première modélisation, nous calculons la complexité du système d'information en comptant le nombre d'éléments composant ce modèle. Concrètement, le calcul indique (cf. tableau ci-dessous) que nous ne rentrons pas dans les détails du système d'information par exemple en modélisant les éléments qui composent chaque document, ou au niveau des acteurs en indiquant les services et personnes responsables. De même au niveau des concepts que nous pourrions détailler pour montrer les articulations sémantiques que les normes mettent en avant. Le niveau de détail que nous avons choisi indique juste le fait que les fichiers XML sont mis à disposition du public sur le site des AN et qu'ils respectent des normes archivistiques. Ce qui se traduit par une complexité faible : 60

⁷ <https://goo.gl/P1Y8G1>

⁸ <http://www.loc.gov/ead/>

Tableau 1. Complexité de l'état initial

<i>Dimensions</i>	<i>Niveau</i>	<i>Nombre d'éléments</i>	<i>Complexité</i>
Document	1	3	3
Acteurs	1	3	3
Concept	1	2	2
Rapport	1	7	7
TOTAL	4	15	60

3.2. Deuxième métamorphose, première hybridation

Pour développer notre système d'information que nous appelons « Jardin des connaissances » (JDC), nous partons de cet état initial très simple que nous métamorphosons en transposant les fichiers XML dans une base de données MySQL dédiée aux écosystèmes de connaissances⁹. Nous effectuons une première hybridation pour que les données obtenues se développent de façon autonome par rapport à l'état initial. Ainsi, nous n'interférons pas avec le système d'information des AN et pouvons enrichir librement les données tout en gardant la référence originale. Le diagramme suivant illustre cette première hybridation :

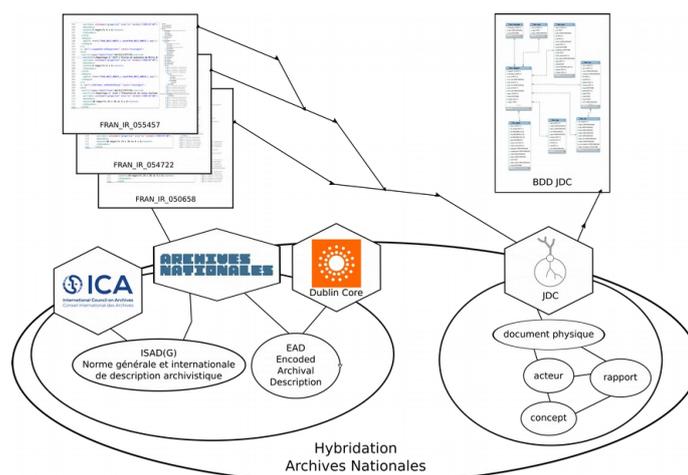


Figure 3. Hybridation du système d'information des AN

⁹ Le modèle de la base de données est ici : <https://goo.gl/qbFv6w>

L'algorithme¹⁰ que nous avons développé pour effectuer cette première hybridation, analyse le document XML pour en extraire les métadonnées et le fragmenter (Kembelec et Bottini, 2017) en autant de dimensions que celles définies dans notre modèle d'écosystème de connaissances : documents, acteurs, concepts, rapports.

3.2.1. Des documents

Pour modéliser la dimension des documents, nous avons extrait avec un algorithme récursif les séries et sous-séries des inventaires. Nous n'avons pas pris en compte les références à des documents annexes comme les références précisées dans les balises <relatedmaterial>, <separatedmaterial>, <otherfindaid>, <bibliography>. De même, nous n'avons pas traité les attributs 'sources' des balises <controlaccess> qui donnent la référence de l'archive d'où provient l'information.

L'arborescence documentaire obtenue par la fragmentation comporte un nombre différent de niveaux suivant les inventaires¹¹. Par exemple, l'inventaire « Reportages photographiques de la présidence de Charles de Gaulle (1959-1969) » (FRAN_IR_054722) ne possède que trois niveaux : le fichier XML, les reportages et les photos. En revanche, l'inventaire « Reportages photographiques autour des chefs de l'État et des présidents de la République (1938-1959) » (FRAN_IR_050658) possède quatre niveaux : le XML, la présidence, le reportage, les photos. Excepté pour la présidence de Vincent Auriol qui possède un niveau supplémentaire correspondant au fournisseur de l'archive : « Reportages du service central photographique du ministère de l'Intérieur et du service central de l'Identité judiciaire » ou « Reportages du service photographique de l'Élysée ».

La difficulté algorithmique concernant l'extraction des documents vient du fait que suivant les inventaires, le lien vers les documents photo n'est pas formalisé de la même manière. L'algorithme doit donc prendre en compte plusieurs modes d'extraction. De plus, il existe des photos présentes sur le site des AN mais qui ne sont pas décrites dans l'inventaire. En analysant le numéro d'ordre présent dans le nom des fichiers photo, nous avons trouvé 224 photos qui ne sont pas décrites dans les inventaires, mais que nous avons trouvées sur le serveur des AN¹².

3.2.2. Des acteurs

Concernant la dimension des acteurs, nous avons récupéré dans les inventaires trois types d'acteurs. Premièrement, nous avons extrait les auteurs des inventaires pour obtenir une liste de 13 noms dont certains sont explicites comme « Jean-Pierre Bat » d'autres moins comme « agents techniques » et d'autres qui demandent une correction de l'algorithme pour affiner l'extraction par exemple : « Nicole Even et Isabelle Chave » ou « par Isabelle Chave ». Cette multiplication des formes « Isabelle Chave » entraîne une perte d'information puisque cette personne n'aura pas automatiquement une importance correspondant au fait qu'elle est la seule

¹⁰ Le code PHP de l'algorithme est accessible ici : <https://goo.gl/gJ4mXy>

¹¹ Les données de répartition des séries sont accessibles ici : <https://goo.gl/oXGC2p>

¹² Liste des photos accessibles ici : <https://goo.gl/N3Euxs>

personne qui apparaît dans les trois inventaires en tant qu'auteur. C'est pourquoi il est nécessaire de compléter la métamorphose automatique par un processus semi-automatique couplant des algorithmes de traitement automatique de la langue (TAL) avec des corrections faites par des experts du domaine. Ces traitements font émerger des « entités nommées » pour créer des rapprochements entre des formes équivalentes évitant ainsi des pertes d'information importantes.

Le deuxième type d'acteur que nous avons extrait concerne les personnes que les inventaires mentionnent explicitement dans une balise <persname>. La valeur de cette balise se décompose ainsi « Nom, Prénom (date de naissance-date de mort) ». Malheureusement, seul l'inventaire FRAN_IR_050658 contient cette balise, nous n'avons donc récupéré par ce moyen que 39 acteurs.

Nous avons extrait le troisième type d'acteur à partir de la balise <geogname> qui elle aussi n'est présente que dans un seul inventaire. Nous considérons les 145 entités géographiques comme des acteurs en accord avec notre modèle d'existence (cf. Figure 1.) qui fait de l'acteur le porteur d'une métrique topographique en représentant la dimension du lieu, l'« ici » de l'existence informationnelle. Nous considérons le lieu dans sa double dimension à la fois comme « topos » c'est-à-dire le nom d'un espace géographique, mais aussi comme « chôra » c'est-à-dire la dimension générative du lieu, ce en quoi il est fertile et par là même créateur d'une potentialité de rapports. Par exemple, l'acteur « Le Havre (Seine-Maritime) » ne désigne pas uniquement un territoire aux frontières administratives bien définies et modélisées par une liste de longitudes et de latitudes (topos), mais aussi l'ensemble des institutions, des élus, des personnes, des vivants qui jouent tous un rôle dans l'activité du territoire (chôra) (Berque, 2012).

3.2.3. Des concepts

Concernant les concepts, il n'y a pas dans les inventaires de balise pour définir un vocabulaire formalisé faisant référence à un thésaurus comme SKOS¹³ ou à une URI dans une ontologie comme SyMoGIH (Beretta, 2012). La sémantique des séries se retrouve principalement dans les balises <unittitle> qui définissent en langage naturel le titre de la série. On trouve aussi des balises <scopecontent> dans lesquelles une liste de balise <item> précise le titre des photos. Dans le cas où cette balise est absente, nous avons intitulé les documents : « photo » concaténée avec le n° d'ordre de la photo. Mis à part le titre, les deux seuls concepts que nous avons récupérés de cette métamorphose sont « début » et « fin » lorsqu'elles sont définies dans les balises <unitdate>.

Des analyses textométriques très simples faites sur les titres des documents avec le logiciel Voyant Tools¹⁴ révèlent des mots-clefs que l'on peut considérer comme les concepts importants de notre corpus. À cause de la règle sur les photos sans titre que nous avons précisée ci-dessus, le mot-clef qui écrase statistiquement les autres est « photo » car plus de 90% des photos n'ont pas de titre (Figure 4). En revanche, si on prend en compte le descriptif complet de la photo à savoir le titre de chaque

¹³ <https://www.w3.org/2004/02/skos/>

¹⁴ Les analyses sont accessibles ici : <https://goo.gl/M49eLq> et <https://goo.gl/JBiLrR>

série dont elle fait partie, l'importance des mots-clés est proportionnelle au nombre de photos dans chaque série (Figure 5). D'autres concepts apparaissent comme celui de « voyage » ou de « chasse ». Il faudrait sans doute approfondir l'interprétation de ces analyses, mais ce n'est pas notre propos ici.



Figure 4 : mots-clés des titres



Figure 5 : mots-clés des titres proportionnellement au nombre de photo

3.2.4. Des rapports

Concernant la dimension des rapports, ils sont composés d'une source, d'une destination et d'un prédicat qui peut être nul. Le lien entre ces trois éléments est fonction d'une actualité précise c'est-à-dire : une date qui définit son « maintenant ». Nous avons pris en compte les indications temporelles de chaque série pour définir cette actualité qui est soit une date ponctuelle soit une période pour presque 8% des cas. Dans le cas des inventaires que nous avons traités, la date fournie dans la balise <unitdate> de chaque série correspond à la date de l'événement pendant lequel des photos ont été prises. Avec cette date, nous avons créé des rapports entre les séries comme document source, les concepts « début » et/ou « fin » comme concept prédicat et la valeur de la date comme destination. Nous avons aussi créé des rapports entre les séries et les photos qui sont décrites dans chaque série ; ainsi que des liens entre les documents et les acteurs (auteurs, personnes, entités géographiques). Le diagramme ci-dessous montre une répartition temporelle de ces rapports suivant le nombre de documents pour chaque date. Nous ne rentrerons pas dans l'interprétation du diagramme pour expliquer les fluctuations historiques, ce travail reste à faire.

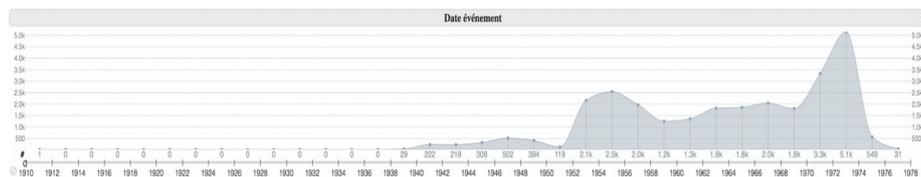


Figure 6 : Répartition historique du nombre de documents

3.2.5. *Complexité de la métamorphose*

Le système d'information ainsi transformé voit son niveau de complexité grandement augmenté, il passe de 60 à 175 712. Au cours de la métamorphose, de nombreux fragments ont été créés dans les quatre dimensions de notre modèle que le tableau ci-dessous détaille en présentant pour chaque dimension le nombre d'éléments par niveau.

Tableau 2. *Complexité de la première hybridation*

<i>Dimensions</i>	<i>Niveau</i>			<i>Nombre d'éléments</i>				<i>Complexité</i>
Document	1			3				3
	2			1999				3998
	3			18262				547886
	4			2453				9812
	5			8645				43225
	5			31362				111824
Acteur	1			52				52
	1			52				52
Concept	1			3				3
	1			3				3
Rapport	<i>Nb. de rapport</i>	<i>Nb. source</i>	<i>Type source</i>	<i>Nb. destination</i>	<i>Type destination</i>	<i>Nb. prédicat</i>	<i>Type prédicat</i>	<i>Complexité</i>
	27539	3301	document	27539	document	11	document	58390
	139	83	document	39	acteur (personne)	7	document	268
	13	3	document	13	acteur (auteur)	1	concept	30
	179	83	document	145	acteur (géographie)	7	document	414
	3566	3303	document	2	concept	14	document	6885
	31436	6773		27738		40		65987
	1			31436				65987
TOTAL	8			62377				175712

L'algorithme que nous avons développé fragmente automatiquement les données initiales suivant la structuration de chaque inventaire. L'objectif est qu'il soit le plus générique possible pour traiter automatiquement tous les inventaires des AN. Toutefois, cette fragmentation a mis en évidence les particularités des XML originaux en relevant des incohérences formelles qui occasionnent des « bugs » dans le déroulement du processus de métamorphose ou des approximations lorsque les données présentes ne sont pas suffisamment structurées. Par exemple, nous avons

trouvé dans l'inventaire « FRAN_IR_050658 » un titre de photo qui possède un espace en trop dans le n° de l'item ce qui occasionne la création de doublons si l'algorithme ne gère pas ce type d'erreur :

<item>n° 79 1: le général de Gaulle salue des blessés de guerre.</item>

Dans le même inventaire, c'est cette fois le n° de l'item qui ne correspond pas à l'ordre puisque le n°1334 est entre le n°1323 et le n°1325

<item>n° 1323 : le président Vincent Auriol reçoit un bouquet de fleurs lors de son accueil en gare d'Évian.</item>

<item>n° 1334 : honneurs militaires place de la gare.</item>

<item>n° 1325 : dépôt d'une gerbe devant le monument aux morts.</item>

Autre exemple avec des doublons dans le n° de l'item¹⁵ et des items n°522 et n°524 qui ne sont pas mentionnés alors qu'ils existent bien¹⁶ :

<item>n° 521 : visite du château de Pau (chambre natale d'Henri IV et ancien cabinet du roi).</item>

<item>n° 521-n° 523 : pose de la première pierre de l'internat du lycée Louis Barthou.</item>

<item>n° 525-n° 529 : le cortège présidentiel dans les rues de Pau.</item>

Nous avons trouvé d'autres erreurs qui ont été corrigées dans les fichiers originaux. Nous avons averti les personnes responsables aux AN de ces incohérences que le développement de notre algorithme a révélées. Attention toutefois à ne pas faire de l'algorithme le moyen d'évaluer la qualité des inventaires en se basant sur sa complexité, c'est-à-dire le nombre de spécificités que l'algorithme prend en compte. Cette complexité algorithmique est tout au plus un moyen de révéler une autre complexité, celle du travail des archivistes. Comme l'ont montré Langlais (2017) avec Gallica et Glorieux (2016) avec DataBNF, les versions successives d'un algorithme construisent un point de vue clair sur les données mises à disposition et leurs éventuels manques. La modélisation que nous proposons donne une vision précise et comparable de ces points de vue notamment pour évaluer les différences entre plusieurs versions d'un algorithme d'extraction ou les choix stratégiques de ces auteurs. Dans notre perspective de développement d'un système d'information générique pour la valorisation des archives numériques, cette modélisation nous donne le moyen de mesurer le pouvoir d'agir de l'écosystème et des existences qui le composent.

3.3. Troisième métamorphose : enrichissement Google

À partir de la première hybridation, nous avons augmenté le pouvoir d'agir de l'écosystème grâce à une nouvelle métamorphose qui consiste à utiliser les AIaS

¹⁵ cf. id="c-21r6uib53-1dypx4p4a5zd4" et id="c-20j9xe576--ul2yzbbr3oev"

¹⁶ cf. <https://goo.gl/AVq3bk>

(Artificial Intelligence as Service) de Google (Guillaud, 2018) pour extraire des formes signifiantes à partir des titres présents dans les inventaires (Google Language) et extraire des visages et leurs caractéristiques émotionnelles et colorimétriques à partir des photos (Google Vision). Ces services donnent accès à des technologies d'intelligence artificielle et plus particulièrement de « deep learning » sans avoir des compétences poussées dans ces domaines. Nous avons juste interrogé ces services via un algorithme qui transforme les réponses de Google en existence dans l'écosystème. Ce processus de métamorphose est illustré par le diagramme suivant :

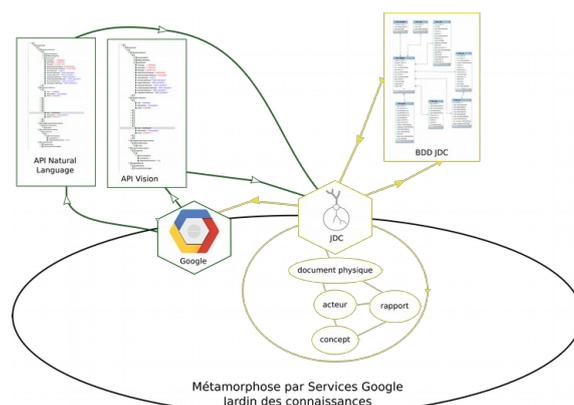


Figure 7. Métamorphose Google

Nous ne rentrons pas dans le détail des fragmentations effectuées à chaque étape de la transformation en document, acteurs, concepts et rapports¹⁷. Grâce aux services Google d'extraction des formes signifiantes et d'analyse des photos, nous avons augmenté la complexité de notre écosystème de plus de 7%. Mais il reste encore de nombreux « nettoyages » à mener pour par exemple éliminer les doublons d'acteurs ou regrouper les concepts dans des catégories plus opératoires. Concernant les entités géographiques, il faudrait mener un travail de géolocalisation pour traduire les noms de lieux en coordonnées géographiques afin de réaliser des cartographies de répartition spatiale des événements ou tracer les itinéraires des voyages présidentiels.

Cet enrichissement de l'écosystème a aussi profité à la société Google qui fait payer l'usage de ces services au-delà d'un certain nombre d'appels¹⁸. Concrètement, l'utilisation des services Google avec l'algorithme que nous avons utilisé nous a coûté 219,49 \$US pour 15 685 appels de l'API Natural Language (0,58 \$US) et 129 010 appels de L'API Vision (218,91 \$US). Ces appels ont été faits avec une version

¹⁷ Algorithmes pour Google Natural Language : <https://goo.gl/MV4szK>

Pour Google Vision : <https://goo.gl/t6M5tS>

¹⁸ Pour calculer le prix de l'usage : <https://cloud.google.com/products/calculator>

intermédiaire de la base de données ce qui fait que des photos présentes dans la base actuelle n'ont pas été prises en compte.

3.4. 4ème métamorphose 2ème hybridation : du JDC vers OMEKA S

Suite à l'enrichissement du système d'information avec les analyses de Google, nous avons effectué une nouvelle métamorphose de celui-ci pour importer les données dans un CMS (Content Management System) dédié à la gestion sémantique des archives numériques : Omeka S¹⁹. Nous avons choisi ce CMS, car il offre de nombreux avantages notamment pour gérer des descriptions de contenu à partir d'ontologie, mais aussi parce qu'il propose une extension pour importer des fichiers multimédias²⁰ à partir d'un fichier CSV²¹ et surtout parce qu'une extension implémente un serveur IIIF²² qui permet notamment de contrôler l'affichage d'une image en paramétrant les dimensions, les tailles, les rotations et les filtres colorimétriques.

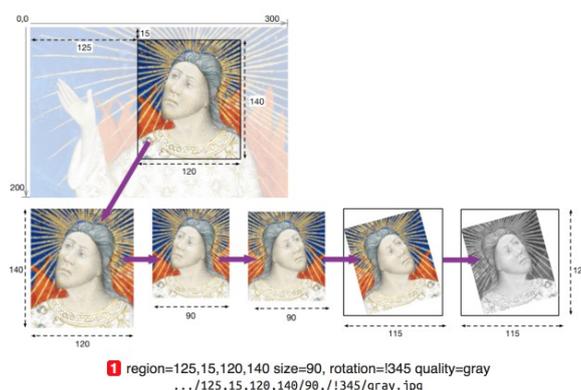


Figure 1. Exemple de contrôle d'une image avec IIIF²³

Pour rendre accessibles les données de notre système d'information dans Omeka S nous avons développé un algorithme qui crée un fichier CSV à partir de la base de données MySQL²⁴. Puis nous avons importé ce fichier dans Omeka S en préservant la relation avec le JDC et les inventaires des AN grâce à un identifiant unique. Ainsi nous bénéficions des avantages de ce CMS tout en gardant le lien vers les données originales. Ces données sont dès lors autonomes, elles suivent leurs propres cycles

¹⁹ <https://omeka.org/s/>

²⁰ <https://omeka.org/s/modules/FileSideload>

²¹ <https://omeka.org/s/modules/CSVImport>

²² Présentation de IIIF cf. <http://iiif.io/>, module Omeka S cf. <https://goo.gl/LpqUtz>

²³ <http://iiif.io/api/image/2.1/#order-of-implementation>

²⁴ Algorithme getCsvToOmeka : <https://goo.gl/6fgSRo>

de vie qui les amèneront à subir de nouvelles transformations. Le diagramme suivant illustre ce nouveau processus de métamorphose et d'hybridation :

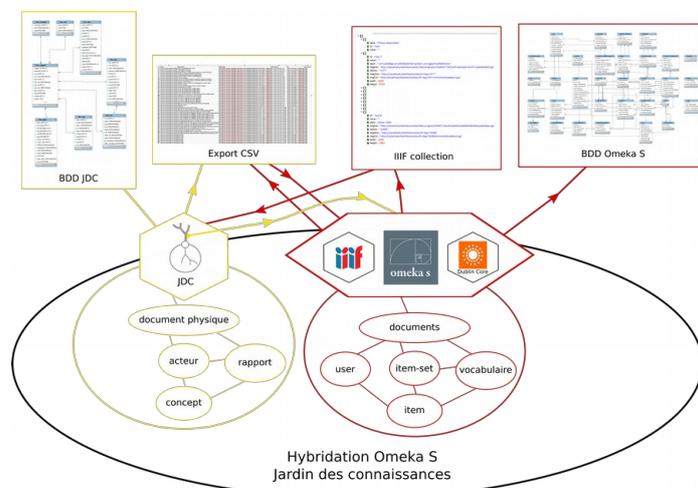


Figure 2. Hybridation Omeka S

3. Consommation standard des documents

À partir des cultures intensives de l'information que nous avons menées, nous disposons de deux hybridations des inventaires XML : l'une sous la forme d'une base de données dédiée aux écosystèmes de connaissances, l'autre sous la forme du CMS Omeka S. Jusqu'à cette étape de développement de notre système d'information, les données ne sont pas accessibles aux publics, mais uniquement aux personnes qui disposent des droits et des compétences pour intervenir dans l'écosystème. Examinons maintenant comment créer des existences informationnelles autonomes par un simple paramétrage du système d'information et montrons en quoi ces existences entraînent une consommation standard des documents numériques que nous avons récoltés.

3.1. 5ème métamorphose, 3ème hybridation : gabarit Omeka S

La première existence que nous présentons est celle proposée par Omeka S pour rendre visible les données. Comme de nombreux CMS, Omeka S rend accessible les informations avec des gabarits qui ont été spécifiquement développés pour afficher les différents objets gérés dans le CMS sans avoir à créer la moindre ligne de code. Il suffit de remplir un formulaire pour automatiquement créer un site Web possédant les navigations nécessaires pour explorer et rendre visible les documents.

Plusieurs gabarits sont disponibles chacun possédant un style particulier qu'on peut enrichir de modules supplémentaires pour ajouter au site Web de nouvelles fonctionnalités. Une connaissance du CMS et des modules disponibles rend sans doute la création du site Web plus optimale, mais après quelques essais et avec l'aide de la documentation, la création d'un site Web est à la portée du plus grand nombre. Le résultat sera un site Web mettant en jeu un système de communication standard qui apporte ses contraintes et inconvénients (Farge, 2014). Nous pourrions modéliser ce système suivant le modèle écosystémique pour montrer les limites du pouvoir d'agir des sites standard et les comparer avec les potentialités d'autres sites, mais cela dépasse l'ambition de cet article.

En remplissant quelques champs, en moins d'une heure, nous avons créé un site et rendu accessibles plusieurs dizaines de milliers de photos et les visages qui les composent. Voici par exemple une des pages créées automatiquement qui présente une des photos récoltées lors de la phase de culture intensive :



Figure 10. Omeka S page item²⁵

Nous avons ajouté un module de Folksonomie²⁶ pour donner aux internautes la possibilité de commenter les photos en ajoutant des tags. L'existence obtenue après cette métamorphose est donc une hybridation puisqu'elle possède ses propres capacités d'évolution.

²⁵ Consultable ici : <https://goo.gl/7tjsYw>

²⁶ Lien vers le module : <https://omeka.org/s/modules/Folksonomy/>

3.2. 6ème métamorphose : exploration à facettes Keshif

La deuxième existence standard que nous présentons est celle d'une navigation à facette (Desfriche Doria, 2013) développée à partir de la librairie JavaScript Keshif²⁷. Cet exemple demande des connaissances techniques un peu plus poussées que le précédent parce que le paramétrage se fait en modélisant un objet informatique avec la syntaxe JSON et en développant des lignes de code HTML et JavaScript²⁸. Toutefois, cette librairie est aussi accessible via une application Web qui propose des formulaires simples pour paramétrer les facettes et leur format²⁹. Pour une analyse de ce type d'outil, nous renvoyons le lecteur au chapitre de l'ouvrage que nous avons dédié à cette question (Louapre et Szoniecky, 2017)

Le principal avantage de cette librairie est de fournir les outils d'exploration générique d'un corpus de données. Les facettes sont autant de filtres pour sélectionner les documents qui forment un sous-corpus explorable avec des facettes automatiquement recalculées. Chaque sélection d'une des facettes crée un état particulier du corpus qui devient dès lors une nouvelle potentialité d'exploration. En partant du principe que chaque état correspond à un écosystème particulier, la complexité du système de navigation est proportionnel au nombre de facettes, d'éléments à l'intérieur de chaque facette et des relations entre ces éléments.

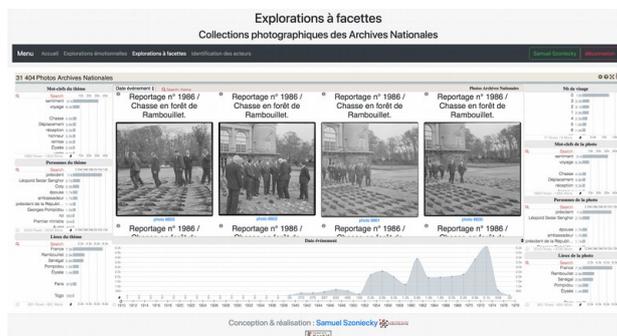


Figure 11 : Navigateur à facettes Keshif³⁰

Avec ce dispositif, nous ne parlons pas d'hybridation, mais uniquement de métamorphose parce que rien dans l'interface n'est disponible pour modifier les données. Il n'y a pas de développement autonome des existences informationnelles dans cette interface, elles n'évolueront pas par l'action des utilisateurs. Plus exactement, les informations présentes sur la page Web pourront toujours être récupérées par un utilisateur pour les transformer selon ses besoins, mais ces modifications échapperont complètement au dispositif initial.

²⁷ Lien vers la librairie : <https://github.com/adilyalcin/Keshif>

²⁸ Lien vers le code de la page : <https://goo.gl/BVMgq3>

²⁹ Lien vers l'application : <https://keshif.me/>

³⁰ Consultable ici : <http://jardindesconnaissances.univ-paris8.fr/public/valarnum/photofacettes>

4. Jardinage collectif des connaissances

Parallèlement à la consommation standard des documents que nous venons d'illustrer par deux exemples, nous avons conçu des existences informationnelles dédiées au jardinage collectif des connaissances (Szoniecky 2017). Nous cherchons à concevoir des expériences utilisateurs (UX) à la fois conviviales (Illich, 1975) et respectueuses des impératifs de pérennité et d'interopérabilité du LOD. Les enjeux sont essentiellement ergonomiques tant au niveau de l'interface utilisateur que de la gestion informatique des données en lecture, écriture, mise à jour et suppression (CRUD).

4.1. 7ème métamorphose, 4ème hybridation : Cartographie émotionnelle

La première existence a pour objectif de récolter les émotions des utilisateurs face aux photos décrites dans les inventaires. Ce travail alimente une réflexion concernant la place des émotions dans les écosystèmes numériques (Pierre et Alloing, 2017) plus particulièrement leurs rôles dans le processus d'apprentissage (Habib, 2015) et leurs utilités pour la construction d'un point de vue réflexif et complexe. Notre hypothèse est que la stimulation des émotions et leur récolte font entrer les utilisateurs dans un processus de réflexivité d'autant plus complexe que la palette des émotions est grande et le support de l'émotion surprenant. En quoi le dispositif permet-il de capter les impressions les plus sensibles (Boutaud, 2015) de l'utilisateur ?

Pour stimuler les émotions, nous présentons aux utilisateurs un flux ininterrompu de photos sous la forme d'un diaporama que nous avons fragmenté aléatoirement en utilisant les fonctionnalités de découpage du serveur IIF. En ne montrant qu'une sélection fragmentaire de documents, le dispositif réduit la complexité de l'écosystème en focalisant l'attention uniquement sur une petite niche de potentialité. En décontextualisant la photo par un resserrement du cadre, les documents semblent ne plus avoir d'intérêt : un bout de ciel, un visage tronqué, un bout de tapisserie, une touffe d'herbe... Mais face à ce non-sens apparent, l'utilisateur en tant que générateur de pensées analogiques en flux continu (Hofstadter et Sander, 2013), va parcourir les images à la recherche d'une signification sur laquelle faire rebondir ces pensées. Pour certains ce sera l'expression d'un visage, pour d'autres un détail de décoration intérieure ou un fragment de paysage. Nous retrouvons ici des principes très anciens (Vernant, 1974) de stimulation de la pensée par confrontation à l'aléatoire (Cage, 1973) afin de provoquer une synchronicité entre l'événement et son interprétation (Jung, 1988).

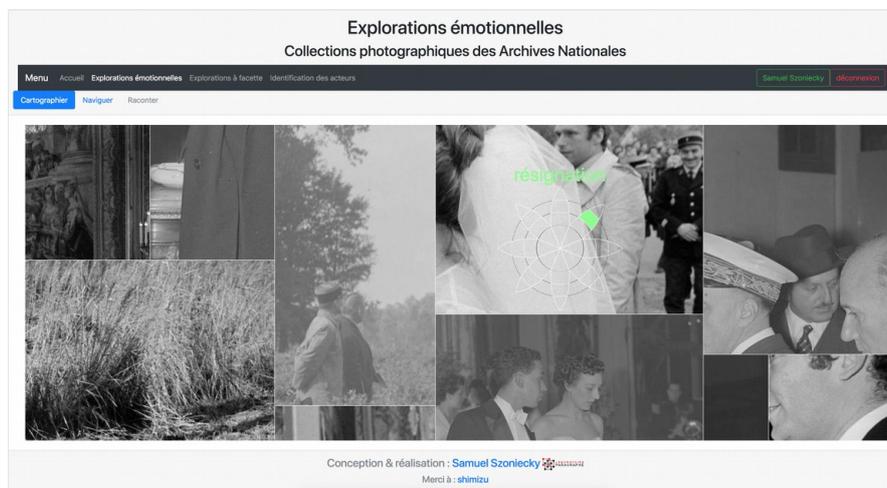


Figure 12. Diaporama pour l'exploration émotionnelle³¹

Une fois provoquée cette stimulation sensorielle, il faut ensuite récolter son expression de la manière la plus fluide possible, mais aussi de façon à pouvoir la comparer avec d'autres expressions pour construire une économie des affects (Citton, 2008). Face à ce chaos infini des potentialités interprétatives, il est nécessaire de contraindre l'expression dans une grille analogique pour faire émerger une cohérence sémantique (Descola, 2005). Parmi toutes les possibilités de représentation des émotions (Tayari-Meftah, 2013), la grille que nous avons choisie est la « roue des émotions » proposée par Plutchick (1991) parce qu'elle offre à la fois les qualités graphiques d'une « boussole », une échelle colorimétrique cohérente, une diversité et une graduation des émotions.

Le processus de récolte est très simple. Sur la mosaïque des photos, le curseur de la souris est transformé en une « roue des émotions » qui aléatoirement affiche une des émotions et son nom. Quand l'utilisateur clique sur une photo, la roue se centre sur la position du clic et affiche la totalité des émotions. L'utilisateur choisit ensuite l'émotion qu'il souhaite exprimer en cliquant sur la couleur correspondante. Tant que l'utilisateur garde enfoncé le bouton de la souris sur la couleur, le rayon du cercle coloré augmente afin d'augmenter l'intensité de l'émotion. Une fois le bouton relâché, l'émotion et son intensité sont enregistrées dans la base de données en lien avec le fragment photographique et l'utilisateur qui l'exprime.

³¹ Consultable ici : <http://jardindesconnaissances.univ-paris8.fr/public/valarnum/exploemo>



Figure 13 : Processus d'expression des émotions

Une page dédiée à la visualisation des émotions permet de naviguer dans le corpus des photos émotionnellement catégorisées. La gauche de l'écran est occupée par un diagramme « stream » qui montre l'évolution temporelle des émotions. La partie droite de l'écran présente un diaporama des photos dans leur taille originale, des cadres montrent le découpage des fragments et les cercles de couleurs indiquent l'intensité émotionnelle.



Figure 14 : navigation dans les expressions

Ce dispositif exploite les possibilités de transformation de l'écosystème pour offrir aux utilisateurs une expérience de catégorisation extrêmement simple tout en ne négligeant pas l'interopérabilité des expressions. En effet, le passage par une cartographie sémantique permet de traduire un simple clic en expression parfaitement structurée pour le LOD par exemple en associant chaque émotion à une URI dans une ontologie dédiée à l'expression des émotions. D'autre part, enregistrer les expressions comme des existences à part entière permet de multiplier les analyses en mettant en jeu les éléments constitutifs de l'écosystème de connaissance, par exemple entre photos (documents), entre utilisateurs (acteurs), entre émotions (concepts). De nouvelles métamorphoses sont alors envisageables pour présenter ces analyses ou de nouvelles hybridations pour les créer.

4.2. 8ème métamorphose, 5ème hybridation : Identification des acteurs

À l’opposé de l’existence précédente, celle que nous présentons maintenant ne cherche pas à récolter une expression subjective pouvant varier suivant les personnes et pour une même personne suivant le moment de l’expression, mais au contraire cette nouvelle expérience a pour but de faire émerger un consensus concernant l’identification des personnes présentes sur une photo afin de faire correspondre un nom à une personne.

Nous avons présenté plus haut comment l’extraction des informations de l’inventaire et leurs analyses par le service « Natural Language » de Google ont permis d’identifier plus de 2 000 acteurs. Parallèlement, grâce au service d’extraction de visages, nous avons récolté plus de 60 000 visages pour lesquels nous n’avons pas d’identification précise. Au mieux, nous pouvons dire que tel visage est peut-être un des acteurs présents sur telle photo. Pour valider cette identification entre un acteur et un visage, un traitement automatique de reconnaissance de personne trouverait les visages similaires. Toutefois, cela ne permet pas de dire que ces visages correspondent à telle personne portant tel nom, encore moins de dire que ces visages correspondent à la personne identifier dans le LOD par telle URI.

Pour répondre à ce besoin, l’expérience que nous avons développée place l’utilisateur face à un mur d’image qu’il peut explorer soit de façon chronologique ou suivant un choix aléatoire. Pour chaque visage le passage de la souris révèle le titre de la photo originale et les évaluations émotionnelles faites automatiquement par Google. Un clic sur l’image affiche un formulaire pour l’identification de la personne (Figure 16 : IHM pour choisir l’identification d’une personne) qui met à disposition de l’utilisateur deux moyens d’identifier le visage.



Figure 15. Mur des visages à identifier³²

Premièrement, en haut de la fenêtre est présentée une liste de personnages potentiellement en lien avec la photo originale. Soit parce que les analyses

³² Consultable ici : <http://jardindesconnaissances.univ-paris8.fr/public/valarnum/identacteurs>

automatiques ont fait un lien entre cette photo est un personnage, soit parce qu'un utilisateur a déjà interprété cette photo en y reconnaissant la présence de personnes. Il suffit alors de sélectionner dans cette liste la personne qui correspond au visage et ensuite de préciser la confiance que l'on accorde à son choix et finalement de cliquer sur le bouton « nommer le visage » pour enregistrer dans la base de données son interprétation.

The screenshot shows a web application interface for identifying actors. At the top, there is a search bar and a table titled 'Acteurs du contexte'. Below the table, there is a section for 'photo 04943' with a date of '1972-06-02'. A photo of Léon Zitronne is displayed. To the right of the photo, there is a list of references with columns for 'URI', 'Type', and 'Confiance'. The interface also includes a sidebar with 'WIKIDATA' and 'Léon Zitronne' information, and a bottom section with 'Liens de références'.

Figure 16 : IHM pour choisir l'identification d'une personne

Deuxièmement, au milieu de la fenêtre, le bouton « chercher dans le LOD » va utiliser l'API Knowledge Graph³³ pour trouver des références du LOD correspondant à une interrogation. Par exemple, l'utilisateur qui reconnaît le visage de Léon Zitronne trouve avec la requête « Léon Zitronne » une liste de références dont il peut consulter les détails afin de choisir celle qui correspond à la personne qu'il a reconnue. On peut envisager d'affiner automatiquement les réponses du moteur de recherche en développant un algorithme de filtre qui exclurait les réponses ne correspondant pas à un type de réponse (humain, chose...), à une période historique, à un espace géographique ou tout autre critère pertinent présent dans le LOD. Toutefois, il nous semble qu'utiliser les capacités humaines de discernement des incohérences pour effectuer ce choix reste encore un moyen très efficace. Surtout, un tel dispositif permet de constituer un réseau d'experts dans un domaine particulier en recoupant les choix des utilisateurs afin de faire émerger des intérêts communs ou des différences d'interprétations. Plus encore, en exploitant le degré de confiance que l'utilisateur a exprimé pour chacun de ces choix, on peut extraire de la base de données les interprétations les plus propices à une controverse pour les proposer aux autres utilisateurs et ainsi stimuler un débat entre des individus, ce qui

³³ <https://developers.google.com/knowledge-graph/>

en dernière instance reste pour nous l'objectif principal d'un dispositif d'intelligence collective.

On le voit dans le cas d'une interprétation visant un consensus sur une réponse unique, tel visage correspond à telle personne, le processus d'expression de l'interprétation est plus complexe que dans le cas précédent où l'expression subjective passe simplement par un clic sur une carte sémantique. Nous avons développé ces premières existences pour valider les traitements avec la base de données, nous envisageons maintenant de les hybrider pour rendre l'expérience utilisateur beaucoup plus conviviale et accessible sur des smartphones.

5. À suivre

Au terme de cet article, nous avons montré la nécessité d'une identification unique des documents composant une archive numérique afin de rendre fiables les transformations de ces documents. Ajoutons en conclusion que les technologies de « Blockchain » sont sans doute une solution intéressante pour assurer une pérennité et une non-falsification de ces identifications. Nous avons ensuite décrit comment ces documents numériques de par leur caractère absolument plastique pouvaient subir une infinité de transformations que nous appelons métamorphose quand un lien avec le document original est maintenu et hybridation quand le résultat de la métamorphose possède des fonctionnalités de développement autonome.

La méthode générique de modélisation des écosystèmes de connaissances que nous utilisons nous a permis de quantifier l'évolution des complexités après chaque transformation et surtout de développer des applications Web pour une navigation standard dans les archives numériques (Omeka S et Keshif) ou plus spécifiquement pour l'expression d'une interprétation. À travers ces exemples d'hybridation, nous souhaitons rendre interopérables et pérennes ces interprétations en les liant au LOD. Nous voulions aussi insister sur les différences entre une existence conçue pour l'expression d'une subjectivité et une autre dédiée à l'émergence d'un consensus. Un des enjeux principaux de l'ingénierie des connaissances et de l'intelligence collective pour ces prochaines années sera de concilier les deux approches que nous avons exposées pour proposer de nouvelles expériences utilisateurs basées sur des dispositifs ludiques utilisant des cartographies sémantiques pour stimuler l'interprétation. Cet enjeu est d'autant plus importants qu'il a des répercussions dans les méthodes d'apprentissages utilisant les technologies numériques pour enseigner à nos enfants la complexité de l'écosystème dans lequel ils grandiront.

Le « work in progress » que nous présentons dans cet article demande encore de nouvelles étapes de développement notamment pour mener des tests utilisateurs afin d'analyser la pertinence de nos hypothèses ergonomiques et la faisabilité d'un projet d'intelligence collective entre les étudiants de l'Université Paris 8 et les Archives Nationales. À travers ces tests, nous espérons récolter des données permettant d'alimenter un moteur de recommandation qui calculera les relations entre des personnes pour lesquelles la potentialité de controverse est importante et donc le dialogue entre elles nécessaire. Un autre point important qu'il nous faudra

approfondir, concerne la gestion éthique des données récoltées par un paramétrage de la « vie privée » où chaque utilisateur pourra le plus simplement possible choisir quelles données seront préservées et celles qui sont librement diffusables.

L'ouverture des données numérisées et leur valorisation offrent aux chercheurs en sciences humaines l'opportunité d'explorer de nouvelles pistes en menant des expériences inédites qui mettent en jeu des problématiques épistémologiques, ergonomiques, sociales et politiques. Si l'accessibilité de ces recherches est aujourd'hui facilitée par la mise à disposition de portails ouverts et gratuits, la question de leur interopérabilité est en revanche loin d'être résolue, sans parler de la calculabilité des raisonnements qu'elles expriment qui relève encore pour certains de l'utopie. Au-delà de la formalisation du sens n'y a-t-il pas là un défi technologique et humain que nous nous devons de relever ?

Bibliographie

- Abbott, A., Bouvard, H., Ollion, É., (2016). L'avenir des sciences sociales. *Annales. Histoire, Sciences Sociales* 71e année, 575–596.
- Andro, M. (2017). *Bibliothèques numériques et crowdsourcing*, ISTE éditions, Londres.
- Bachimont, B., Gandon, F., Poupeau, G., Vatant, B., Troncy, R., Pouyllau, S., Martinez, R., Battisti, M., Zacklad, M., (2011). Enjeux et technologies : des données au sens. *Documentaliste-Sciences de l'Information*, n°48, p. 24–41.
- Beretta, F., Vernus, P., (2012). Le projet SyMoGIH et la modélisation de l'information : une opération scientifique au service de l'histoire. *Les Carnets du LARHRA*, p. 81–107.
- Berque, A. (2012). La chôra chez Platon, *Espace et Lieu Dans La Pensée Occidentale de Platon à Nietzsche*. La Découverte, Paris.
- Berque, A. (2009). *Ecumène : Introduction à l'étude des milieux humains*. Belin, Paris.
- Boutaud, J.-J. (2015). *Sensible et communication, Sciences cognitives*. ISTE éditions, Londres.
- Brun, G. (2017). Pouvoir d'agir, en analyse de l'activité. *Activités*, n°14. 2017, <https://doi.org/10.4000/activites.2957>
- Cage, J. (1973). *Silence : lectures and writings*. Wesleyan University Press, Middletown, Conn.
- Citton, Y. (2014). *Pour une écologie de l'attention*. Seuil, Paris.
- Citton, Y. (2010) *L'avenir des humanités : Economie de la connaissance ou cultures de l'interprétation ?*, Editions La Découverte, Paris.
- Citton, Y., Lordon, F., (2008). *Spinoza et les sciences sociales : de la puissance de la multitude à l'économie des affects*. Éd. Amsterdam, Paris.
- Descola, P. (2005). *Par-delà nature et culture*. NRF : Gallimard, Paris.
- Desfriches Doria, O. (2013). *La classification à facettes pour la gestion des connaissances métier : méthodologie d'élaboration de FolkClassifications à facettes*. Thèse en Sciences de l'Information et de la Communication, CNAM.

- Farge, O. (2014). *“Rhétorique de la conception” : pour une prise de conscience des stratégies de l’outil de création : proposition d’une typologie de postures d’auteurs*. Thèse en Sciences de l’Information et de la Communication, Paris 8.
- Glorieux, F. (2016). *Data.bnf.fr, les documents. J’attends des résultats*. <https://resultats.hypotheses.org/795>
- Guillaud, H. (2018). *Vers des applications d’intelligence artificielle accessibles à tous ?* <http://www.internetactu.net/2018/02/16/vers-des-applications-dintelligence-artificielle-accessibles-a-tous/>
- Habib, M., Borst, G., Poirel, N., Houdé, O., Moutier, S., Cassotti, M., (2015). Socio-Emotional Context and Adolescents’ Decision Making: The Experience of Regret and Relief After Social Comparison. *Journal of Research on Adolescence*, n°25, p. 81–91.
- Hofstadter, D., Sander, E., (2013). *L’analogie : Coeur de la pensée*. Odile Jacob, Paris.
- Illich, I. (1975). *La convivialité*, Points. Éd. du Seuil, Paris.
- Juanals, B., Minel, J.-L. (2016). La construction d’un espace patrimonial partagé dans le Web de données ouvert. *Communication*, vol. 34/1 | 2016, <http://journals.openedition.org/communication/6650>
- Jung, C.G. (1988). *Synchronicité et Paracelsica*, Albin Michel, Paris.
- Kembellec, G., Bottini, T., (2017). Réflexions sur le fragment dans les pratiques scientifiques en ligne : entre matérialité documentaire et péricope, *20e Colloque International Sur Le Document Numérique (CiDE.19)*, Lyon, p. 71–86.
- Langlais, P.-C. (2017). Les bibliothèques numériques sont-elles représentatives ? , *Sciences communes*, <https://scoms.hypotheses.org/799>.
- Latour, B. (2012). *Enquêtes sur les modes d’existence : Une anthropologie des Modernes*. Éditions La Découverte, Paris.
- Lévy, P. (2011). *La sphère sémantique : Tome I, Computation, cognition, économie de l’information*. Hermes Science Publications, Paris.
- Louâpre, M., Szoniecky, S., (2017). Outils pour une modélisation de l’interprétation des archives numériques, *Intelligence Collective et Archives Numériques*. ISTE éditions, Londres.
- Meunier, J.-G. (2017). Humanités numériques et modélisation scientifique. *Questions de communication*, n° 31, 2017, p. 19–48.
- Moiraghi, E. (2018) « Explorer des corpus d’images. L’IA au service du patrimoine », *Carnet de la recherche à la Bibliothèque nationale de France*. <https://bnf.hypotheses.org/2809> (accessed 7.5.18).
- Parisi, L. (2016). La raison instrumentale, le capitalisme algorithmique et l’incomputable. *Multitudes*, n°62, 98–109.
- Pierre, J., Alloing, C., (2017). Le design du web affectif : entre empathie et universalité. Retour sur les phases de conception de l’affectivité numérique, *Actes de la conférence H2PTM’17*. ISTE éditions, Londres, p. 112–130.
- Plutchik, R. (1991). *The Emotions*. University Press of America.

- Prié, Y. (2011). *Vers une phénoménologie des inscriptions numériques. Dynamique de l'activité et des structures informationnelles dans les systèmes d'interprétation*. HDR Université Claude Bernard - Lyon I, Lyon.
- Reyes-Garcia, E. (2017). *The Image-Interface: Graphical Supports for Visual Information*. ISTE Ltd, Hoboken, NJ.
- Sauret, N. (2017). Epistémologie du modèle. *Sens Public*, <http://sens-public.org/article1287.html>
- Serres, M. (2009). *Ecrivains, savants et philosophes font le tour du monde*, Les essais du Pommier. Pommier, Paris.
- Szoniecky, S. (2017). *Écosystème de connaissances : méthode de modélisation et d'analyse de l'information et de la communication*. ISTE Editions, Londres.
- Szoniecky, S. (2018). Modeling of power to act for an ethics of the Internet of Things, *Challenges of the Internet of Things: Technique, Use, Ethics*, Hoboken, NJ: ISTE Ltd, 2018, p. 215- 243.
- Tayari-Meftah, I. (2013). *Modélisation, détection et annotation des états émotionnels à l'aide d'un espace vectoriel multidimensionnel*, Thèse en informatique, Université Nice Sophia Antipolis.
- Vernant, J.-P. (1974). *Divination et rationalité*. Seuil, Paris.
- μ, G., Edeline, F., Klinkenberg, J.-M., (2015). *Principia semiotica : Aux sources du sens*. Les impressions nouvelles éditions, Bruxelles.