

### Valoriser le patrimoine culturel pyrénéen à l'aide d'une plateforme et d'une application mobile

*Valorising the Tangible Cultural Heritage of the Pyrenees via a Platform and a Mobile Application*

**Landy Rajaonarivo, Marie-Noëlle Bessagnet, Christian Sallaberry, Annig Le Parc Lacayrelle et Philippe Roose**

---



#### Édition électronique

URL : <https://journals.openedition.org/revuehn/1343>

DOI : 10.4000/revuehn.1343

ISSN : 2736-2337

#### Éditeur

Humanistica

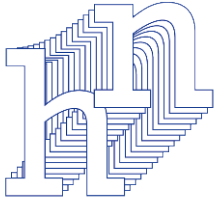
#### Référence électronique

Landy Rajaonarivo, Marie-Noëlle Bessagnet, Christian Sallaberry, Annig Le Parc Lacayrelle et Philippe Roose, « Valoriser le patrimoine culturel pyrénéen à l'aide d'une plateforme et d'une application mobile », *Humanités numériques* [En ligne], 3 | 2021, mis en ligne le 01 mai 2021, consulté le 12 mai 2021. URL : <http://journals.openedition.org/revuehn/1343> ; DOI : <https://doi.org/10.4000/revuehn.1343>

---



Les contenus de la revue *Humanités numériques* sont mis à disposition selon les termes de la Licence Creative Commons Attribution 4.0 International.



## Valoriser le patrimoine culturel pyrénéen à l'aide d'une plateforme et d'une application mobile

### *Valorising the Tangible Cultural Heritage of the Pyrenees via a Platform and a Mobile Application*

Landy Rajaonarivo, Marie-Noëlle Bessagnet, Christian Sallaberry, Annig Le Parc Lacayrelle et Philippe Roose

#### Résumés

Le projet *TCVPYR* est un projet européen FEDER dont le but est de promouvoir le tourisme dans la région française des Pyrénées en exploitant son patrimoine culturel. Il implique des chercheurs de différents domaines : géographes, historiens, anthropologues et informaticiens. Cet article présente deux approches pour tirer parti des points d'intérêts (POI) liés au patrimoine culturel, toutes deux intégrant des concepts du Web de données. La première s'intéresse à la publication automatique des données dans des plateformes *open data* connues comme *Wikipédia*. La seconde concerne une application mobile *open source* dédiée à des itinéraires touristiques. La première approche propose un processus automatique complet permettant de publier n'importe quel jeu de données sur *Wikipédia*. Grâce à une preuve de concepts (POC), nous avons validé ce processus sur un jeu de données contenant des données géoréférencées du patrimoine culturel, collectées par les chercheurs du projet *TCVPYR* dans différentes régions des Pyrénées. Ces mêmes données sont également mises à disposition du grand public dans la seconde approche originale de valorisation : une application mobile *open source* dédiée à la valorisation du patrimoine pyrénéen via la génération personnalisée d'itinéraires touristiques.

*TCVPYR* is a European FEDER project which aims to promote tourism in the French Pyrenees region by leveraging its cultural heritage. It involves scientists from various domains: geographers, historians, anthropologists, and computer scientists. This paper presents two approaches to exploit cultural heritage points of interest (POI), both using Web of data concepts. One consists in publishing these data into the database of well-known open data platforms such as *Wikipedia*. The other one consists in an open source mobile application dedicated to touristic itineraries. The first approach involves a fully automated process to publish any dataset on *Wikipedia*. Thanks to a proof of concept (POC), we validate this process on a sample of geo-referenced cultural heritage data collected by *TCVPYR* researchers in different regions of the Pyrenees. These same data are also highlighted in the second approach via a mobile application recommending personalised touristic itineraries.

## Entrées d'index

MOTS-CLÉS : humanités numériques spatialisées, Web sémantique, patrimoine culturel, valorisation, données ouvertes, système de recommandation

KEYWORDS: spatial digital humanities, semantic Web, cultural heritage, valorisation, recommendation system, open data

*Ce projet est réalisé dans le cadre du programme de recherche européen TCVPYR (2017-2020), financé par l'Union européenne (FEDER) en partenariat avec les régions Occitanie et Nouvelle-Aquitaine.*

*Nous remercions Lucas Levêque pour sa relecture et ses conseils avisés sur la partie Wiki\*.*

## Introduction

<sup>1</sup> Définir le patrimoine culturel est une entreprise difficile. Selon l'UNESCO (2020), « [...] *Cultural heritage does not end at monuments and collections of objects. It also includes traditions or living expressions inherited from our ancestors and passed on to our descendants, such as oral traditions, performing arts, social practices, rituals, festive events, knowledge and practices concerning nature and the universe or the knowledge and skills to produce traditional crafts [...]* ». Il y a bien lieu de distinguer le patrimoine bâti et le patrimoine culturel immatériel (respectivement, *tangible cultural heritage* et *intangible cultural heritage*, en anglais).

<sup>2</sup> Le projet de recherche européen FEDER *TCVPYR* (2017-2020) contribue à un inventaire du patrimoine bâti et du patrimoine culturel immatériel de la villégiature et du thermalisme dans le massif pyrénéen français. Ainsi, à travers ses atouts patrimoniaux, il renforce le développement de la connaissance et de la fréquentation des Pyrénées.

<sup>3</sup> Dans le cadre de ce projet pluridisciplinaire mêlant des chercheurs historiens, géographes, anthropologues, l'objectif, pour les chercheurs en informatique, était double : constituer une base de données fédérant des points d'intérêts (POI) patrimoniaux référencés par ces experts ;

proposer une visualisation de ces POI dans le cadre de différents scénarios de valorisation. L'objectif de cet article est de détailler le deuxième objectif. Concernant le premier objectif du projet, le lecteur intéressé pourra se reporter aux références bibliographiques suivantes : Bessagnet *et al.* (2018), Fonteles *et al.* (2018), Rajaonarivo *et al.* (2019a), Rajaonarivo *et al.* (2019b), ainsi qu'au [site Web du projet](#).

4 En matière de valorisation de ce patrimoine, nous avons donc identifié deux principaux défis : publier automatiquement des données sur des plateformes *open data* connues comme *Wikipédia* ; développer une application mobile *open source* capable de recommander des itinéraires touristiques personnalisés et contextualisés. Le choix de ces deux canaux de diffusion n'est pas anodin : la notoriété de *Wikipédia* n'est plus à faire. Wikidata est un exemple de dépôt de données ouvertes (*open data*). Hébergé par la Wikimedia Foundation et étroitement associé à Wikidata, *Wikipédia* compte actuellement un total de 2,1 millions d'articles en langue française, qui augmente chaque jour (300 articles supplémentaires en moyenne quotidiennement). Les données sont publiées manuellement sur *Wikipédia* selon un processus collaboratif. Par exemple, les données du site *PCI Lab*<sup>1</sup> sur le patrimoine culturel immatériel ont été publiées de manière manuelle (Casteret et Larché 2018).

5 Un article de Marden *et al.* (2013) souligne les avantages de la diffusion des données sur le patrimoine culturel par le biais des *Linked Open Data* (LOD) : « *Linked Open Data offers a new way for cultural heritage institutions to share their holdings with a wider audience, and to change the traditional relationship between the holder of knowledge, the interpreter of knowledge, and the consumer of knowledge.* » Ainsi, dans la lignée des préconisations de cet article et de Freire *et al.* (2018), nous souhaitons joindre nos efforts à ceux des institutions qui s'efforcent d'améliorer la visibilité, la facilité d'utilisation et la durabilité des collections du patrimoine culturel conservées par les bibliothèques, les archives, les musées et autres. Actuellement, le ministère français de la Culture lance un grand projet appelé *POP*<sup>2</sup> (il s'agit d'une plateforme ouverte dédiée aux données patrimoniales). L'objectif de ce projet est de rendre les données sur le patrimoine culturel français librement accessibles. Cependant, tous les POI patrimoniaux d'ampleur régionale ne seront pas publiés dans *POP* et nous pensons que, pour l'instant, *Wikipédia* est encore un standard de fait pour la publication en *open data*. Ainsi, pour notre premier défi, le principal problème que nous nous proposons de traiter ici est la conception et l'expérimentation d'un processus de publication automatisé dédié à l'encyclopédie *Wikipédia*, au dépôt d'images dans Wikimedia Commons et à la base de données Wikidata. Un tel processus implique la gestion des barrières sociales et technologiques auxquelles sont confrontés tous les membres du projet. Par exemple, les partenaires experts qui collectent des données sur le patrimoine local souhaitent que leur nom apparaisse sur l'article *Wikipédia* qui sera publié. Ou encore, ils veulent que personne d'autre ne puisse modifier ces articles.

6 Le développement et l'utilisation d'applications mobiles ont cru de manière exponentielle ces dernières années grâce aux deux écosystèmes que sont iOS et Android. Sur le site de statistiques Statista<sup>3</sup>, on relève

par exemple que 3,85 % des applications mobiles actives sur l'App Store d'Apple en 2019 concernent le tourisme, occupant ainsi la septième position parmi les types d'applications les plus populaires.

<sup>7</sup> Ainsi, pour conseiller des POI, de nombreuses applications mobiles telle que la renommée TripAdvisor<sup>4</sup> (voir Amorim *et al.* 2018), mais aussi de nombreuses plateformes telles que TRIPBUILDER (voir Brillhante *et al.* 2014 ; Brillhante *et al.* 2015), utilisent le Web de données. L'UNESCO tente de trouver des fonds pour financer également un grand projet pour développer « *an official UNESCO World Heritage Mobile application*<sup>5</sup> ». Des applications mobiles ou des plateformes intègrent des POI du patrimoine culturel afin de les promouvoir dans la recommandation d'itinéraires de visite (Corallo *et al.* 2017). Aussi, pour notre second défi, nous avons conçu et développé un algorithme de recommandation d'itinéraires personnalisés et contextualisés : il prend en compte le profil de l'utilisateur (ses préférences thématiques et historiques) et le contexte (sa position, les caractéristiques de son appareil ainsi que d'autres paramètres tels que le temps disponible pour effectuer la visite des POI ou le moyen de transport). Chaque recommandation est également enrichie par des informations extraites de l'*Open Data*.

<sup>8</sup> Les sections suivantes sont organisées comme suit. Dans la section « Travaux connexes », nous abordons les travaux liés à la publication en *Linked Open Data* et aux applications mobiles promouvant le patrimoine culturel. Dans la section « De la collecte à la valorisation de données patrimoniales », nous détaillons notre processus général en détaillant l'architecture logicielle globale mise en place. Dans les sections « Valorisation via *Wikipédia* » et « Valorisation via l'application mobile *TCVPYR* », nous aborderons les deux défis de valorisation du patrimoine du projet *TCVPYR*. Tout d'abord, nous verrons le processus générique de publication automatique vers *Wikipédia*, Wikimedia Commons et Wikidata que nous avons conçu et formalisé par un algorithme. Nous y présentons l'expérimentation que nous avons menée sur un ensemble spécifique de données patrimoniales ainsi que les principaux retours d'expérience sur ces premiers essais. Ensuite, nous présenterons le processus de recommandation d'itinéraires patrimoniaux, au cœur du développement de l'application mobile. Enfin, dans une dernière section, nous concluons et esquissons des perspectives.

## Travaux connexes

<sup>9</sup> Sur le site Web d'Ontotext<sup>6</sup>, les « données liées » sont définies comme l'un des principaux piliers du Web sémantique, également connu sous le nom de « Web de données ». Grâce aux technologies du Web sémantique, il est possible d'établir des liens entre des ensembles de données qui sont compréhensibles non seulement pour les humains, mais aussi pour les machines. Les données liées peuvent être considérées comme un ensemble de principes de conception pour le partage de données liées lisibles par les machines sur le Web. « Données liées » est un terme créé par Tim Berners-Lee en 2006. Celui-ci, dans le contexte de la sémantique du Web, a promulgué quatre principes (Berners-Lee 2009), notamment celui d'utiliser les adresses URI pour identifier les choses. Lorsque les données sont ouvertes, on utilise le concept de *Linked Open Data* (LOD).

## Publication en *Linked Open Data*

<sup>10</sup> La publication de données liées et ouvertes est un domaine actif. Selon les travaux actuels (Roman *et al.* 2018<sup>7</sup> ; Abella *et al.* 2019), le processus général de publication peut être résumé en trois étapes : (1) la préparation des données, (2) le stockage des données et (3) l'accès aux données selon différentes approches et technologies. L'étape de préparation permet de construire, nettoyer, modifier et partager les données. Les données sont, par exemple, converties en graphes RDF<sup>8</sup> (Resource Description Format) mais toute autre représentation de données pourrait être pertinente, par exemple le format CSV (*comma-separated values*). L'étape de stockage prend en charge l'entreposage et le partage des données sur une plateforme privée ou publique fiable (par exemple, un *triple store* accessible gratuitement à tous les utilisateurs). L'étape d'accès fournit des fonctions d'exploration et de visualisation : par exemple, interroger les données stockées dans un SPARQL *endpoint*<sup>9</sup>, partager et réutiliser les requêtes SPARQL ou utiliser des API pour accéder aux données à travers un programme.

<sup>11</sup> En ce qui concerne l'étape 1, un travail récent (Vetrò *et al.* 2016) visait à surmonter les difficultés de mesure de la qualité des *Open Government Data* (OGD). En ce qui concerne à la fois l'étape 1 et l'étape 2, le projet *Zeri et LODE* (Daquino *et al.* 2017) s'est concentré sur la création d'un ensemble de données LOD répertoriant environ 50 000 entrées des archives photographiques de *Zeri*. Il s'appuie sur la mise en correspondance du vocabulaire des métadonnées de *Zeri* avec celui du CIDOC-CRM<sup>10</sup> et la création de l'ensemble de données ouvertes correspondant en RDF. De plus, Abella *et al.* (2019) ont étudié et catégorisé les producteurs de données ouvertes (par exemple, l'administration publique) et les utilisateurs (par exemple, les citoyens). Ce travail concerne les étapes 1 et 3 du processus de publication. Du point de vue des données relatives au patrimoine culturel, le projet présenté dans Candela *et al.* (2018) vise principalement à améliorer l'accès aux corpus du patrimoine culturel (par exemple, la Biblioteca Virtual Miguel de Cervantes) grâce à l'enrichissement et à la désambiguïsation des lieux dans les requêtes textuelles, en utilisant les LOD (par exemple, Wikidata, GeoNames).

## Publication sur *Wikipédia*

<sup>12</sup> La publication sur *Wikipédia* se scinde en deux catégories principales : le processus de publication manuel standard et le processus automatisé (Agirre *et al.* 2012).

<sup>13</sup> La majorité des articles *Wikipédia* sont créés<sup>11</sup> manuellement par des contributeurs. La procédure classique part d'une section déjà existante et la complète avec les informations trouvées dans des ressources de qualité. La création n'étant pas un acte individuel, les personnes se coordonnent souvent pour partager la charge de rédaction et de vérification des sources. Dès que l'admissibilité d'un article est actée, il intègre l'encyclopédie et y demeure, quel que soit son état d'avancement. Ainsi, sur un total de 168 millions de modifications de l'encyclopédie, seuls 2 ou 3 millions concernent la création de nouveaux articles. La contribution



est donc en grande partie l'ajout d'informations aux articles existants. Le projet *PCI Lab* (Casteret et Larché 2018) décrit une première expérience de publication de données patrimoniales françaises sur *Wikipédia* où environ 300 articles sur le patrimoine culturel immatériel ont été publiés. Cependant, le processus de saisie sur *Wikipédia* est entièrement manuel et les textes, rédigés à l'origine par des experts, ont été réécrits pour mieux s'adapter à un public plus large. Ce processus manuel et collaboratif a pris beaucoup de temps et d'énergie.

14 La publication automatisée sur *Wikipédia*, quant à elle, est plus rare. Elle consiste à importer automatiquement des articles à l'aide de robots. Elle commence néanmoins par une étape de validation des articles et d'échanges avec la communauté. En effet, comme pour la publication manuelle, la publication automatisée d'articles déclenche des contrôles *a posteriori*. Pour simplifier, la publication sur *Wikipédia* comprend trois étapes principales : (1) la préparation, la rédaction et le formatage des textes, (2) la « wikification », c'est-à-dire la publication, (3) la révision et la finalisation d'un article. Concernant la publication automatisée, nous nous concentrerons sur cette deuxième étape. C'est ainsi que dans le cadre du projet *Valdensia*<sup>12</sup>, la bibliothèque cantonale et universitaire de Lausanne (BCU Lausanne), en collaboration avec l'association Wikimedia.ch, a importé 1 010 pages de notes biographiques d'un jeu de données du canton de Vaud. En ce qui concerne les images, d'autres projets visent la publication sur Wikimedia Commons, qui nécessite un processus de vérification beaucoup plus simple. Différents scénarios existent : les institutions qui téléchargent les images ont des droits d'auteur et publient leurs œuvres sous licence Creative Commons BY-SA, par exemple, ou bien les œuvres sont déjà dans le domaine public et ne posent aucun problème. Ainsi, le projet du Metropolitan Museum of New York<sup>13</sup> a téléversé sur Wikimedia Commons environ 375 000 images en libre accès d'œuvres d'art du domaine public de la collection du musée, et a invité des wikipédiens du monde entier à lier ces images à des articles *Wikipédia* nouveaux ou déjà existants.

15 À notre connaissance, aucun projet n'a encore publié automatiquement un jeu de données sur le patrimoine culturel simultanément sur *Wikipédia*, Wikidata et Wikimedia Commons, en établissant des liens entre descriptions textuelles, données structurées et images. Nous présentons un tel processus générique dans la section « Valorisation via *Wikipédia* ».

## Applications mobiles, données patrimoniales et recommandation d'itinéraires

16 Le développement rapide des technologies informatiques mobiles a permis la création exponentielle de divers types d'applications mobiles. Les applications mobiles pour le tourisme sont devenues très populaires et couvrent un ensemble de catégories. En accord avec Amorim *et al.* (2018), ces applications concernent cinq étapes liées à l'expérience du touriste, allant du choix de la destination jusqu'au partage de l'expérience. « *The experience phase, which is fundamental for the tourist, aims to provide information about the public transport network, the main points of interest and static information about these places [...].* »

17 Les applications mobiles dédiées aux touristes (voir Emmanouilidis, Koutsiamanis et Tasidou 2013 ; Smirnov, Kashevnik et Ponomarev 2017) doivent fournir des services de visite en rapport avec le contexte et riches en information multimédia. Nous nous intéressons plus particulièrement aux applications mobiles de recommandation du patrimoine culturel proche du touriste, liées à ses préférences et à sa situation dans la zone géographique considérée. Le système que nous avons développé entre dans cette catégorie.

18 Au-delà des méthodes de recommandation, des interfaces et de la manière de représenter les préférences des utilisateurs (Smirnov, Kashevnik et Ponomarev 2017), prendre en considération les sources d'information utilisées pour fournir aux touristes des données sur les endroits intéressants est important. De nombreuses plateformes telles que TRIPBUILDER (Brilhante *et al.* 2014) ou encore Tourist Assistant (TAIS, voir Smirnov, Kashevnik et Ponomarev 2017) utilisent le Web de données pour la recommandation d'itinéraires de visite. Par exemple, le système TRIPBUILDER collecte des attractions touristiques (c'est-à-dire des points d'intérêt, ou POI) dans les ressources de *Wikipédia* et fournit aux touristes des informations telles qu'une vignette descriptive du POI, ses coordonnées géographiques, les catégories thématiques auxquelles le POI appartient. En outre, de nombreux touristes effectuent directement des recherches dans l'encyclopédie *Wikipédia* afin de construire eux-mêmes leur prochain circuit touristique.

19 Dans notre cas, la source principale est notre propre base de données, qui contient des informations sur le patrimoine culturel des Pyrénées. Aussi, recommander des POI patrimoniaux peu connus par le grand public est un atout dans notre travail de recherche.

## De la collecte à la valorisation de données patrimoniales

20 La figure 1 décrit le processus de collecte et de validation des données patrimoniales dans le cadre du projet TCVPYR

FIGURE 1. PROCESSUS GLOBAL AVANT VALORISATION

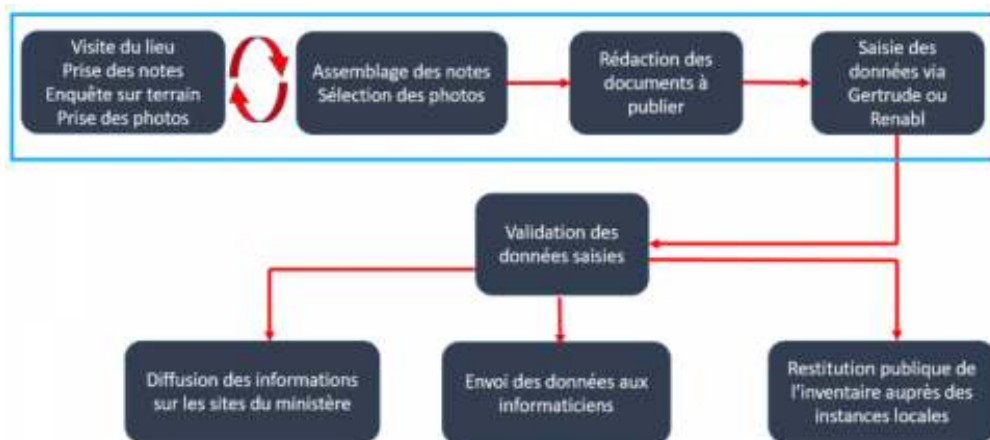


Image produite par les auteurs



21 Pour collecter les informations patrimoniales, les chercheurs en sciences humaines et sociales recueillent sur le terrain les informations, qu'ils enregistrent par le biais de deux logiciels : RenabLP2 et Gertrude<sup>14</sup>. Ces deux logiciels permettent d'exporter les données collectées au format XML. Cependant, les exports de RenabLP2 et de Gertrude sont très différents, aussi bien au niveau de la structuration et de l'organisation que du nommage des informations.

22 Pour exploiter ces données hétérogènes dans un cadre dédié au tourisme, nous avons conçu un modèle de données unifié et homogène : voir Bessagnet *et al.* (2018) ; Fonteles *et al.* (2018).

23 Nous avons développé un premier prototype d'application fédérant ces données, qui permet de visualiser les POI et quelques informations sur un fond de carte<sup>15</sup>. Ce prototype présente un double avantage. Il permet aux chercheurs de vérifier l'exactitude de la localisation des POI mais également de vérifier le contenu des fiches descriptives. Afin de valoriser au mieux ces informations patrimoniales, nous avons développé deux approches plus complexes de valorisation de POI patrimoniaux : d'une part, une application permettant d'automatiser le processus de publication dans l'univers Wikimédia ; d'autre part, une application mobile dont l'objectif est de proposer des itinéraires de visite personnalisés et contextualisés, à destination d'un large public. L'une des particularités de nos approches réside dans l'utilisation de connaissances *a priori* : un thésaurus de classification de concepts patrimoniaux nous permet de catégoriser aussi bien les POI (point de vue des experts) que les préférences des utilisateurs (point de vue de l'utilisateur ou de l'application de recommandation). Nous disposons d'un thésaurus expert du domaine patrimonial qui comporte 4 007 concepts et 7 946 liens.

24 Une fois le travail de collecte et validation opéré, nous pouvons passer à l'étape de valorisation. La figure 2 détaille l'architecture logicielle globale.

FIGURE 2. ARCHITECTURE DES DEUX APPROCHES POUR LA VALORISATION

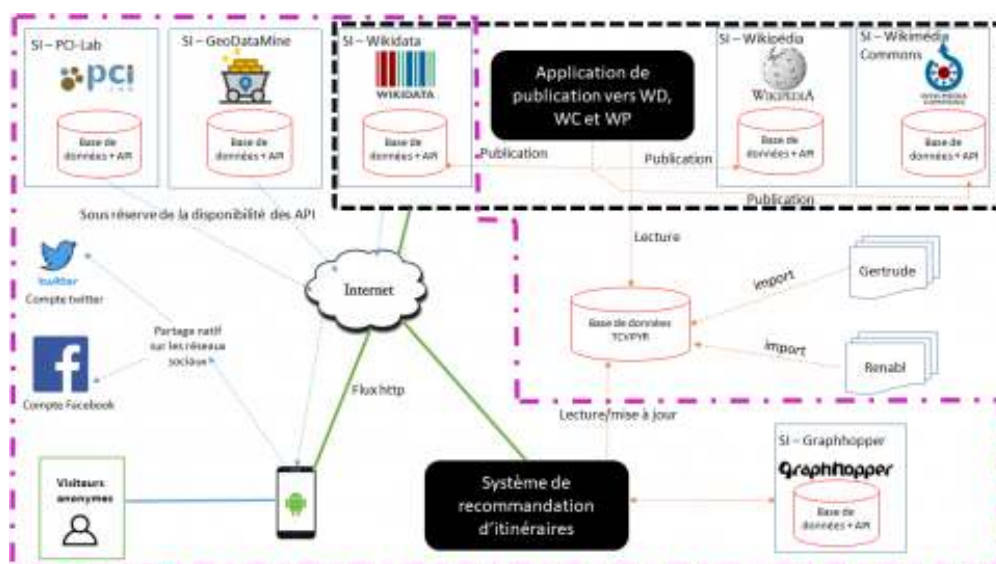


Image produite par les auteurs

25 Les environnements logiciels que nous avons développés accèdent à la base de données commune TCVPYR afin d'extraire les données patrimoniales utiles. Ils interagissent avec le Web de données soit pour enrichir

l'expérience de l'utilisateur (en plus des données patrimoniales), soit pour les besoins de la publication d'articles sur *Wikipédia*.

26 Nous allons détailler dans « Valorisation via *Wikipédia* » et « Valorisation via l'application mobile *TCVPYR* » les algorithmes, les mises en œuvre et expérimentations que nous avons pu mener sur les deux volets.

## Valorisation via *Wikipédia*

### Un processus générique pour la publication automatique sur *Wikipédia*

27 Toute publication d'un jeu de données dans des articles de *Wikipédia* doit se conformer aux règles de publication attendues par la communauté des wikipédiens<sup>16</sup>. Par exemple, tout article doit être vérifiable : il doit donc contenir une bibliographie. De même, un article doit être unique : deux articles différents ne peuvent donc avoir ni le même titre, ni le même contenu. Le contenu d'un article doit être écrit dans un style encyclopédique (viser à la neutralité). Un article peut être associé à un ou plusieurs portails *Wikipédia* et à une ou plusieurs catégories *Wikipédia*. Un portail permet d'associer un article à un regroupement thématique, alors qu'une catégorie permet d'accéder à un article via un index thématique. Ainsi, durant la constitution du jeu de données à publier, il est nécessaire d'identifier les portails et catégories relatifs aux données.

28 Un article *Wikipédia* peut intégrer des données structurées provenant de Wikidata et des images provenant de Wikimedia Commons. Une donnée structurée s'appelle un *item* dans Wikidata et est identifiée de façon unique par un code commençant par la lettre Q. Pour publier un *item* dans Wikidata, il est nécessaire d'avoir, au minimum, un *label* et une description. Pour publier une image dans Wikimedia Commons, il est nécessaire d'avoir une localisation (quand cela est possible), le type d'élément décrit par l'image, son auteur et sa source de provenance. Toutes les données doivent être libres de droit, par exemple soumises à une licence Creative Commons BY-SA.

29 Ainsi, la publication d'une donnée  $d$  dans un article  $art_d$  de *Wikipédia* peut amener à l'ajout d'un *item* correspondant à  $d$  ( $item_d$ ) dans Wikidata et à l'ajout d'une image  $img_d$  (quand elle existe dans le jeu de données) dans Wikimedia Commons.  $art_d$ ,  $item_d$  et  $img_d$  sont reliés de la manière suivante :

- $item_d$  référence  $img_d$  et  $art_d$
- $img_d$  référence  $item_d$
- $art_d$  référence  $img_d$

30 Nous avons conçu l'algorithme WikiAlgo permettant la publication automatique de n'importe quel jeu de données dans *Wikipédia*. Afin de respecter la règle d'unicité dans *Wikipédia*, il est nécessaire de vérifier, avant toute publication, qu'une donnée  $d$  n'est pas déjà le sujet d'un article  $art_d$  dans *Wikipédia* ou d'un  $item_d$  dans Wikidata. Ainsi, quatre cas de figure sont distingués dans notre algorithme :

- cas 1 :  $art_d$  et  $item_d$  existent déjà (lignes 12 à 16 dans l'algorithme)
- cas 2 :  $art_d$  existe et  $item_d$  n'existe pas (lignes 18 à 23 dans l'algorithme)
- cas 3 :  $art_d$  n'existe pas et  $item_d$  existe (lignes 27 à 31 dans l'algorithme)
- cas 4 :  $art_d$  et  $item_d$  n'existent pas (lignes 25 à 31 dans l'algorithme)

ALGORITHME 1. L'ALGORITHME WIKIALGO

```

1  procedure LierImage (imgd, itemd, artd)
2    // imgd, itemd, artd sont des paramètres inout
3  début
4    | Lier imgd avec itemd
5    | Lier itemd avec imgd
6    | Lier artd avec imgd
7
8  Données :  $C_{DON}$  // Corpus à importer dans Wikipedia
9  début
10 | pour  $d \in C_{DON}$  faire
11 |   Rechercher  $d$  dans Wikipedia
12 |   Rechercher  $d$  dans Wikidata
13 |   si  $art_d \in Wikipedia \wedge item_d \in Wikidata$  alors
14 |     Modifier le contenu de  $art_d$  en le complétant avec  $d$ 
15 |     si  $d$  possède une image  $img_d$  alors
16 |       | Importer  $img_d$  dans Wikimedia Commons
17 |       | LierImage (imgd, itemd, artd)
18 |   sinon
19 |     si  $art_d \in Wikipedia$  alors
20 |       Modifier le contenu de  $art_d$  en le complétant avec  $d$ 
21 |       Créer  $item_d$  dans Wikidata
22 |       Lier  $item_d$  avec  $art_d$ 
23 |       si  $d$  possède une image  $img_d$  alors
24 |         | Importer  $img_d$  dans Wikimedia Commons
25 |         | LierImage (imgd, itemd, artd)
26 |     sinon
27 |       si  $item_d \notin Wikidata$  alors
28 |         | Créer  $item_d$  dans Wikidata
29 |         Créer  $art_d$  dans Wikipedia
30 |         Lier  $item_d$  avec  $art_d$ 
31 |         si  $d$  possède une image  $img_d$  alors
32 |           | Importer  $img_d$  dans Wikimedia Commons
33 |           | LierImage (imgd, itemd, artd)

```

Image produite par les auteurs

Avant de lancer le processus, il faut informer la communauté et les portails concernés par le jeu de données qu'un téléchargement automatique va être effectué.

## Expérimentation du processus de publication automatique vers Wikimedia

<sup>32</sup> Rappelons que les données sur le patrimoine culturel des Pyrénées utilisées dans ce travail de recherche viennent du projet européen FEDER *TCVPYR* et sont stockées dans la base de données du projet (figure 2). Aussi, cette section présente une première expérimentation de notre processus générique sur des données patrimoniales du projet *TCVPYR*.

### Constitution du corpus de données

<sup>33</sup> Notre corpus de données  $C_{\text{DON}}$  contient un ensemble de données patrimoniales issues du projet *TCVPYR*. Chaque donnée  $d$  correspond à un fichier JSON contenant un titre, une description, un historique, un ou plusieurs *labels* thématiques et historiques (voir la section « Processus générique pour la publication automatique sur *Wikipédia* »), une position géographique, le nom de la ville dans laquelle se trouve le patrimoine culturel  $d$ . Une image est aussi associée à  $d$ . Ces données sont extraites via un service d'extraction d'information que nous avons développé et qui pointe vers la base de données du projet *TCVPYR*. Actuellement, cette base de données contient approximativement 600 données patrimoniales culturelles et couvre une dizaine de villes.

<sup>34</sup> La figure 3 et le tableau 1 illustrent un exemple de données de la base de données *TCVPYR*.



FIGURE 3. UN EXEMPLE DE *IMG<sub>D</sub>*



Image produite par les auteurs

<i>Parameter</i>	<i>Value</i>
Title	voie ferrée du transpyrénéen
Type	gare
City	Ax-les-Thermes
Region	Occitanie
Reseachers	Anaïs Comet
Bibliography	Garau, M., « Les transpyrénéens, ... », Revue géographique des Pyrénées ... Hagimont, S., « Un essor touristique ... », Archives ariégeoises ...

Image produite par les auteurs

## Processus de publication

35 Lors de leur collecte, les données du projet *TCVPYR* sont indexées par les experts à l'aide d'un thésaurus métier. Afin de conserver cette indexation lors de la publication sur *Wikipedia*, nous avons, d'une part, élaboré un fichier de correspondance entre ce thésaurus métier et des *items* de Wikidata. Ce fichier se nomme *TH*. Chaque instance de *TH* est un tuple constitué par un *label* thématique, le titre et le code (*Qxx*) de l'*item* correspondant dans Wikidata. D'autre part, afin de lier les données de notre corpus aux portails et catégories *Wikipedia*, nous avons créé un fichier *CT* dans lequel chaque ville pyrénéenne est associée au code d'un *item* Wikidata (*Qxx*) et à des portails *Wikipedia* tels le portail de la région pyrénéenne et celui de son département. Ainsi, chaque instance de *CT* est un tuple qui est constitué par le nom d'une ville, le code de l'*item* correspondant dans Wikidata et le portail correspondant dans *Wikipedia*. *TH* et *CT* sont constitués en une seule fois pour l'ensemble du corpus.

36 Nous avons développé une application Java qui implémente le processus de publication vers Wikimédia (voir la section « Un processus générique pour la publication automatique sur *Wikipedia* »). À titre de preuve de concept, nous avons d'abord établi un corpus de données qui ne sont ni dans *Wikipedia* ni dans Wikidata. Ainsi, ce processus de publication est divisé en cinq étapes implémentant le cas 4 de WikiAlgo :

1. Publication des données vers Wikidata. Cette phase a pour but de créer un *item* dans Wikidata (ligne 27 dans WikiAlgo). Notre programme appelle une API Wikidata dédiée à la création et à l'édition d'un *item*. Pour créer *item<sub>d</sub>*, l'API Wikidata a besoin des éléments suivants : titre de *item<sub>d</sub>*, position géographique, nom de la ville, *label* thématique (information disponible dans *TH*). Cette phase prend en entrée les données dans *TH*, dans *d* et dans *CT* et retourne *item<sub>d</sub>*.



2. Publication des données dans *Wikipédia*. L'objet de cette phase est de créer  $art_d$  dans *Wikipédia*, qui correspond à la donnée  $d$  dans  $C_{DON}$  (ligne 28 dans WikiAlgo). Notre programme appelle une API *Wikipédia* dédiée à la création et à l'édition d'un article dans *Wikipédia*. L'API *Wikipédia* a besoin des éléments suivants : titre de  $art_d$ , description, historique, position géographique, nom de la ville, code de l'*item* Wikidata correspondant à son *label* thématique (disponible dans *TH*), noms des chercheurs, références bibliographiques, portails (disponible dans *CT*). Cette phase prend en entrée les informations dans *TH*, dans  $d$  et dans *CT* et retourne  $art_d$ .

3. Lien entre l'*item* Wikidata et l'article *Wikipédia*. Une fois que la création de  $item_d$  et  $art_d$  a réussi, une étape consiste à lier automatiquement  $item_d$  et  $art_d$  (ligne 29 dans WikiAlgo). Pour cela, l'attribut « Wikipédia » dans  $item_d$  est rempli (par une référence à  $art_d$ ) en utilisant l'API Wikidata.

4. Publication d'images vers Wikimedia Commons. Cette phase consiste à publier  $img_d$  (ligne 31 dans WikiAlgo). Cette phase de publication d'image utilise une API appelée Pattypan, qui importe une image vers Wikimedia Commons via un fichier Excel. Ce fichier contient deux feuilles : une première détaillant les informations concernant l'image (nom, titre, auteurs, source, position géographique, etc.) et la seconde décrivant le modèle de l'image (œuvre d'art, article, livre, carte, etc.). Plusieurs API permettent d'importer des images vers Wikimedia Commons telles que Communist, Vicuna Uploader, ComeOn. Nous avons choisi d'utiliser Pattypan parce que le fichier en entrée de cet outil peut être rempli de façon automatique. Cette phase prend en entrée la donnée  $d$  et retourne  $img_d$ .

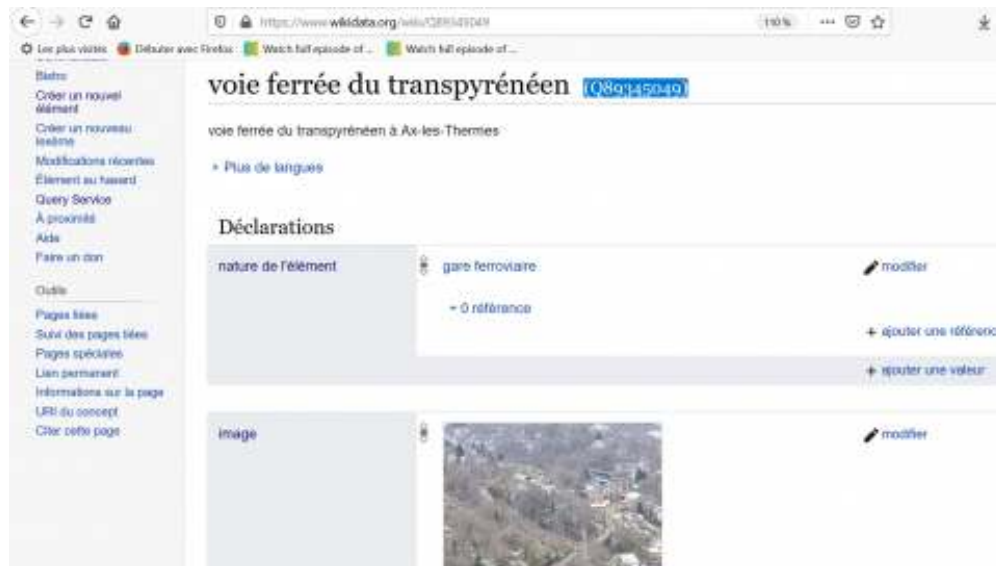
5. Lien entre l'*item* Wikidata, l'article *Wikipédia* et l'image Wikimedia Commons. Une fois que la phase de publication d'image a réussi, cette phase consiste à mettre à jour  $img_d$ ,  $item_d$  et  $art_d$  (ligne 32 dans WikiAlgo). Tout d'abord, une API de Wikimedia Commons est appelée pour renseigner l'attribut « données structurées » de  $img_d$  par  $item_d$  (ou Qxx). Ensuite, deux appels de l'API Wikidata et de l'API *Wikipédia* mettent à jour l'attribut « image » de  $item_d$  et de  $art_d$ .

## Résultats

37

La figure 4 présente l'exemple d'un *item* publié dans Wikidata. L'image correspondante a été publiée dans Wikimedia Commons ( $img_d$ ). Le lien entre  $item_d$  et un autre *item* correspondant à son *label* thématique (par exemple, « gare ferroviaire ») est représenté par l'attribut nommé « nature de l'élément ». Le lien entre  $item_d$  et  $img_d$  est représenté par l'attribut nommé « image ». Sa valeur correspond au nom de  $img_d$  dans Wikimedia Commons.

FIGURE 4. ITEM WIKIDATA CORRESPONDANT ( $ITEM_D$ )



Source : Wikidata

38 La figure 5 illustre l'« infobox » d' $art_d$  (image, type, pays et ville...) et sa description. L'infobox présente le lien entre  $art_d$  et  $img_d$  ainsi que le lien entre  $art_d$  et un autre article *Wikipédia* qui n'est autre que celui de la ville où se trouve  $art_d$ . Ce lien a été réalisé via l'attribut nommé « ville ».

FIGURE 5. ARTICLE WIKIPÉDIA, PARTIE 1 ( $ART_D$ )



Source : Wikipédia

39 La figure 6 illustre les références bibliographiques et plus spécifiquement les portails correspondants à  $art_d$ . Chaque portail a son article *Wikipédia*. De plus, le nom du chercheur qui a réalisé l'inventaire et qui a écrit le contenu de l'article  $art_d$  est indiqué dans la section « sources et références » de l'article.

**Annexes** [modifier | modifier le code]

**Bibliographie** [modifier | modifier le code]

- GARAU, M., « Les trans pyrénéens, La pénétration de la voie d'écartement européen en Espagne », *Revue géographique des Pyrénées et du Sud-Ouest*, n°4-4, 1993, p. 449-471.
- HAGIMONT, Steve, « Un essor touristique et thermal contrainé au XXe siècle - Ax-les-Thermes », *Archives ariégeoises*, n°10, 2016, p. 87-110.
- AD Ariège, 220 EDT O 25, Chemin de fer à la frontière d'Espagne, plan parcellaire, par Rougé, ingénieur à Toulouse, 1927.
- AD Ariège, 220 EDT T 2, Plaquette de communication pour le lotissement d'En Castel, Société immobilière d'Ax-les-Thermes, 1924.
- AD Haute-Garonne, Fonds Labouche, 26 FI PYR09 567, carte postale, « Les Pyrénées ariégeoises – 814 – Ax-les-Thermes – DMNÉ du transpyrénéen et montagne de Mérens », 1916-1937.
- AD Haute-Garonne, Fonds Labouche, 26 FI PYR09 847, carte postale, « Les Pyrénées ariégeoises – 1103 – Ax-les-Thermes – Le grand viaduc du transpyrénéen », 1916-1937.
- AD Haute-Garonne, Fonds Labouche, 26 FI PYR09 580, carte postale, « Les Pyrénées ariégeoises – 827 – Ax-les-Thermes – Le petit viaduc du transpyrénéen », 1930-1937.

**Sources et références** [modifier | modifier le code]

Tout ou partie de cet article provient du travail d'inventaire effectué par **Ariège Conn** dans le cadre du projet européen Feder Projet FEDER TCVPYR. La région Occitanie (membre du projet TCVPYR) a opté pour une licence semi-ouverte : la licence créative common CC-BY-NC-ND 2.0.

Source : *Wikipédia*

40 Ces liens entre les *items* Wikidata, les articles *Wikipédia* et les images Wikimedia Commons ont été réalisés via le processus de publication automatique.

## Feedback

41 Selon la classification de Kucera *et al.* (2015), nous pouvons mettre en avant des défis techniques, mais également des défis politiques et sociaux. Concernant les défis techniques, nous avons intégré les trois API suivantes : l'API Wikidata dédiée à la création et à l'édition d'un *item*, l'API *Wikipédia* dédiée à la création et à l'édition d'un article dans *Wikipédia* et une API dénommée Pattypan, qui importe une image dans Wikimedia Commons. Au final, ce ne sont pas les problèmes techniques qui posent problème dans un tel processus, mais plutôt les problèmes politiques et sociaux.

42 Pour affronter ces derniers, nous avons deux catégories d'experts : (1) les producteurs de données, c'est-à-dire les experts impliqués dans le projet TCVPYR et chargés du travail d'inventaire ; (2) les contrôleurs des données, c'est-à-dire les experts de *Wikipédia* appelés « la patrouille *Wikipédia* ».

43 Ainsi, les producteurs de données sont les auteurs des fiches d'inventaire (données du patrimoine culturel). Naturellement, ils souhaitent que leur nom apparaisse sur l'article *Wikipédia*. Le processus le mentionne dans la section « Références » de l'article et dans un onglet spécial relatif au projet TCVPYR<sup>17</sup>. De plus, les producteurs de données souhaitaient que personne ne puisse modifier leur production. Cela n'étant pas possible, ils ont accepté cette contrainte. Aussi, le plus grand problème rencontré a été celui des droits d'auteur. Les données du patrimoine culturel sont accessibles sur plusieurs plateformes qui n'ont pas nécessairement de licence ouverte : la plateforme nationale *POP* et les sites régionaux. En conséquence, durant notre expérimentation, quelques articles ont été bloqués par les contrôleurs à cause de la violation du droit d'auteur. En effet, ils ont trouvé les mêmes données du patrimoine culturel pyrénéen sur des sites Web<sup>18</sup>. Nous avons cherché le meilleur moyen de résoudre ce problème : une référence à ces publications dans les articles *Wikipédia* n'est pas suffisante. Afin de résoudre ces pro-

blèmes, nos partenaires des inventaires régionaux auraient dû attribuer à toutes les productions relatives au projet *TCVPYR* une licence ouverte, par exemple la licence Creative Commons BY-SA, reconnue par *Wikipédia*. Hélas, les inventaires régionaux ont adopté seulement la licence BY-NC-ND 2.0, qui n'est pas une licence complètement libre. Nous étions encore confrontés à une limite politique.

44 Nous avons résolu tous les défis techniques. Notre POC démontre qu'un processus de publication automatique dans le monde Wikimedia fonctionne. Ce processus peut désormais être utilisé pour la publication de différents corpus *TCVPYR*. Nous avons connu des échecs sur les défis politiques et sociaux : par exemple, lors de notre première expérimentation, sur dix articles publiés, neuf ont été rejetés par les contrôleurs (« violation de droits d'auteur », « article existant dans d'autres sources », « article ne respectant pas les règles de publication (ton neutre...) », etc.).

## Valorisation via l'application mobile *TCVPYR*

45 L'application mobile *TCVPYR* présente six grandes fonctionnalités : la configuration du profil utilisateur, la visite libre (hors itinéraire généré), la recherche, la consultation des POI, la création ou visualisation d'itinéraires et la création ou visualisation de carnets de voyage contenant les différents itinéraires précédemment suivis.

46 Dans le cadre de cette application mobile proposant des itinéraires de visite personnalisés, nous avons conçu un algorithme dédié à la génération d'itinéraires « contextualisés » adaptés aux utilisateurs d'une application mobile. Cette recommandation d'itinéraires prend en compte le profil de l'utilisateur, sa position, les caractéristiques de son appareil, tels que l'état de sa batterie, ainsi que d'autres paramètres tels que le temps disponible pour effectuer la visite des POI ou le moyen de transport. Une recommandation est enrichie par des informations extraites de l'*Open Data* (GeoDataMine, *PCI Lab*). Nous proposons une approche basée sur la notion de recommandation hybride (Rajaonarivo *et al.* 2019a).

47 Nous avons détaillé ailleurs (Rajaonarivo *et al.* 2019a ; Rajaonarivo *et al.* 2019b) :

- un ensemble de modèles : un modèle utilisateur (définissant le profil de l'utilisateur caractérisé par le genre, la catégorie d'âge et les préférences), un modèle de contexte (composé des trois facettes « utilisateur », « physique » et « ressource ») et un modèle d'itinéraire (un itinéraire étant une succession d'étapes dans lesquelles figurent les POI à visiter)
- un système de recommandation hybride : l'originalité de notre approche consiste à rapprocher les POI déjà parcourus par le touriste de ceux également parcourus par d'autres touristes afin de lui proposer de nouveaux POI intéressants
- un système de génération d'itinéraires prenant en compte une fonction de *scoring* originale basée sur trois aspects : la pertinence de chaque POI de l'itinéraire proposé par rapport (1) aux préférences de

l'utilisateur, (2) au trajet et à la durée de visite et (3) au comportement des autres utilisateurs

48

Il s'agit de croiser les profils (position géographique, mode de transport, fenêtre de disponibilité, centres d'intérêt, etc.) des utilisateurs et la base de données patrimoniales pour mettre en place un système de recommandation d'itinéraires de visite comprenant des POI pertinents. Grâce au thésaurus de classification de concepts patrimoniaux, nous avons des connaissances pour déterminer des liens sémantiques entre POI, entre utilisateurs et entre POI et utilisateurs. Nous pouvons, par exemple, calculer la proximité sémantique entre un POI et des préférences utilisateur en mesurant leur similarité sur la base des connaissances modélisées sur ce thésaurus patrimonial.

## Un algorithme original pour la recommandation de POI pour la génération d'itinéraires

49

La majorité des systèmes de génération d'itinéraires se basent sur la résolution de l'*Orienteering Problem* (OP), qui est un système de *scoring* trouvant son origine dans la littérature de recherche opérationnelle. Dans notre cas, l'OP consiste pour un touriste à collecter des points qui sont assignés à des POI. Notre approche de génération d'itinéraires fait partie des approches TDOP (*Time-dependent Orienteering Problem*) car nous considérons le temps de déplacement entre les POI. Nous ne pouvons développer dans le cadre de cet article l'intégralité de notre approche.

50

L'originalité de notre approche hybride consiste à rapprocher des POI déjà parcourus par le touriste de ceux parcourus par d'autres touristes afin de mettre en exergue et d'évaluer des POI candidats selon le principe qu'« un autre touriste qui a visité les mêmes POI que vous a aussi visité les suivants ». Ce principe a été mis en œuvre dans notre approche grâce à l'intégration d'un algorithme d'optimisation de colonies de fourmis (*Ant Colony Optimization*, ACO), type d'algorithme largement utilisé dans le commerce électronique (voir Lu et Guo 2016 ; Zhang et Pang 2015 ; Minjing *et al.* 2017), mais peu courant pour la recommandation de POI. Le lecteur intéressé pourra se reporter à nos articles (Rajaonarivo *et al.* 2019a ; Rajaonarivo *et al.* 2019b).

### ALGORITHME 2. L'ALGORITHME POUR LA GÉNÉRATION D'ITINÉRAIRES

```
Données :  $E_{POI}$ ,  $temps_{visiteTot}$ ,  $zone_{Geo}$  //  $E_{POI}$  : Ensemble des POI
Résultat :  $It_{POI}$  // Itinéraire composé d'une liste de POI
1 début
2    $It_{POI} \leftarrow \{\}$ 
3    $temps_{Restant} \leftarrow temps_{visiteTot}$ 
4   Construire  $C_{POI}$  en sélectionnant les POI de  $E_{POI}$  situés dans  $zone_{Geo}$ 
5   Calculer un score pour tous les POI de  $C_{POI}$ 
6    $pt_{depart} \leftarrow$  POI de  $C_{POI}$  ayant le score le plus élevé
7   Ajouter  $pt_{depart}$  dans  $It_{POI}$ 
8   Supprimer  $pt_{depart}$  dans  $C_{POI}$ 
9    $temps_{Restant} \leftarrow temps_{Restant} -$  temps de visite de  $pt_{depart}$ 
10  tant que  $C_{POI} \neq \{\} \wedge temps_{Restant} > 0$  faire
11    Calculer un score pour tous les POI de  $C_{POI}$ 
12     $pt_{suiv} \leftarrow$  POI de  $C_{POI}$  ayant le score le plus élevé
13    Ajouter  $pt_{suiv}$  dans  $It_{POI}$ 
14    Supprimer  $pt_{suiv}$  dans  $C_{POI}$ 
15     $temps_{Restant} \leftarrow temps_{Restant} -$  (temps de déplacement vers  $pt_{suiv}$  + temps de visite de  $pt_{suiv}$ )
```

Image produite par les auteurs



51 La première étape de notre approche (algorithme 2, ligne 4) filtre les POI figurant dans la zone de visite de l'utilisateur. La deuxième étape calcule le score de l'ensemble des POI candidats obtenus lors du filtre (la fonction de calcul de score est décrite dans Rajaonarivo *et al.* 2019a). Elle nécessite les informations relatives aux POI et les informations utilisateur (profil et contexte). La troisième étape, quant à elle, vise la sélection du POI de départ parmi cet ensemble : par défaut, le POI de score le plus élevé (ligne 6). Le POI de départ est ensuite ajouté dans l'itinéraire en cours de construction. Les itérations suivantes (ligne 10) visent le calcul de score des POI restants suivi de l'ajout du POI de score le plus élevé dans l'itinéraire. Nous relançons ce calcul de score de POI candidats à chaque nouvelle itération afin d'intégrer le paramètre de temps de trajet et celui, plus original, de popularité par rapport au dernier POI de l'itinéraire en cours de construction. Cette forme de popularité prendra notamment en compte le fait que plusieurs utilisateurs qui ont visité le dernier POI ont aussi visité le POI candidat. L'itération s'arrête quand il n'y a plus de POI candidat ou bien quand il ne reste plus assez de temps pour poursuivre la visite. À la fin de ces étapes, nous obtenons un itinéraire qui est défini par une liste ordonnée de POI avec les durées estimatives de visite et de trajet entre les POI.

52 Cet algorithme a été implémenté dans un premier temps sur un prototype fonctionnel (en JavaScript et PHP) puis intégré avec d'autres dans l'implémentation de l'application finale par la société Mazedia.

## Expérimentation du processus de recommandation dans une application mobile

53 Pour des raisons de portabilité ultérieure (sur iOS, en Progressive Web App, ou PWA), l'application mobile finale a été réalisée en s'appuyant sur le SDK Flutter<sup>19</sup> de Google et l'IDE Android Studio<sup>20</sup>. Le tracé GPS des itinéraires est réalisé en s'appuyant sur le moteur *open source* GraphHopper Directions API<sup>21</sup>. L'écran principal est illustré dans la figure 7a. L'application mobile comprend un ensemble de fonctionnalités déclinées sur une succession d'écrans respectant les règles d'interaction dans un environnement mobile.

54 Lors du premier démarrage de l'application, une fenêtre *pop-up* s'ouvre pour indiquer à l'utilisateur qu'il peut créer un profil. Ainsi, l'écran « Mon profil » permet de configurer le profil utilisateur selon divers critères : le genre, les thèmes favoris (préférences thématiques), les périodes favorites (préférences historiques), la composition du groupe (enfants, séniors) et enfin une préférence d'itinéraire qui permettra de sélectionner des POI soit parce qu'ils sont proches thématiquement, soit à cause de la popularité du POI (paramètre qui servira lors de la recommandation et de la génération des itinéraires).

55 Lors d'une visite libre (figures 7b et 7c), l'application affiche tous les POI recensés dans la ville (ici Eaux-Bonnes) ou autour de l'utilisateur (figure 7c).



FIGURE 7A. APPLICATION TCVPY-AT : ACCUEIL

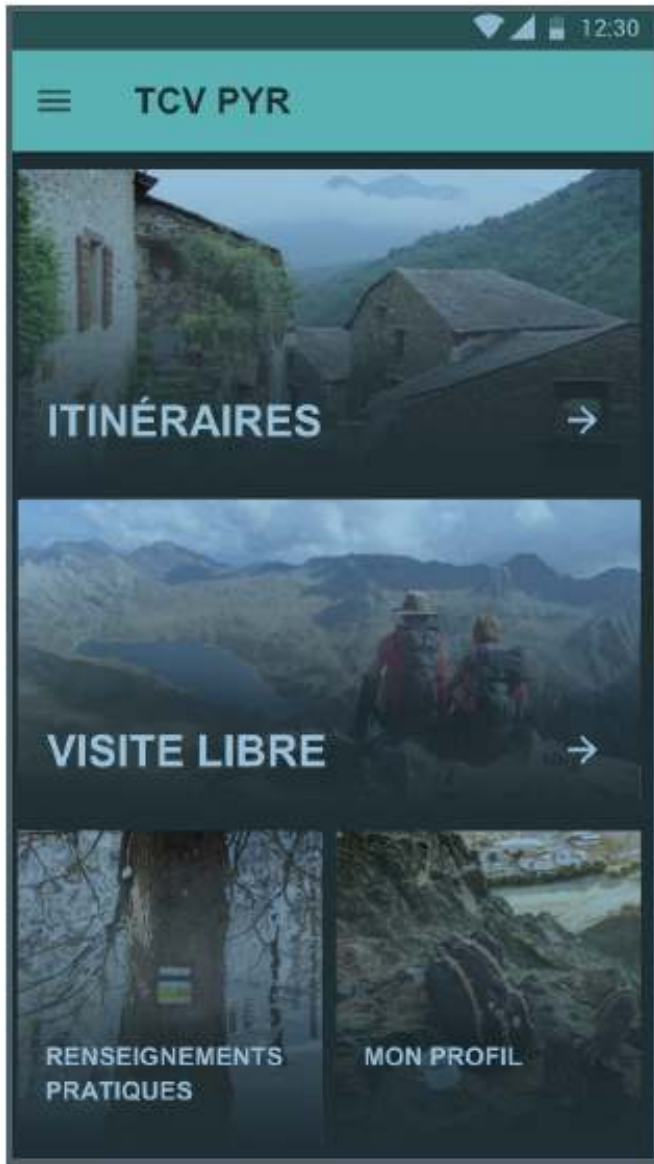


Image produite par les auteurs

FIGURE 7B. VISITE LIBRE DES EAUX-BONNES



Image produite par les auteurs



Image produite par les auteurs

Une fois la génération d'itinéraires faite (figures 8a, 8b et 8c), l'utilisateur peut accéder à son historique d'itinéraires dans son carnet de voyage. Chaque itinéraire est caractérisé par la date de demande de création de l'itinéraire, le nom de la ville de départ, la date et l'heure de départ de l'itinéraire. Un clic sur une ligne de l'historique permet d'accéder à l'écran « Détail itinéraire » et d'afficher ce dernier (figure 8a). Sur site, la trace GPS permettra à l'utilisateur de se situer. En effet, l'utilisateur est géocentré et la carte se déplace en fonction de ses mouvements. L'utilisateur peut obtenir les détails d'un POI (figure 8b) mais également le noter s'il est détecté au moins une minute sur site, dans un rayon de 200 mètres autour du POI. Le bouton « Service » en haut de la carte permet de basculer sur une carte avec uniquement les POI représentant les toilettes publiques, les réseaux wifi publics, les parkings, les hôtels et les restaurants (figure 8c).

FIGURE 8A. EXEMPLE DE RECOMMANDATION D'UN ITINÉRAIRE



Image produite par les auteurs



FIGURE 8B. VISITE VIRTUELLE D'UN POI (DÉTAIL)



Image produite par les auteurs



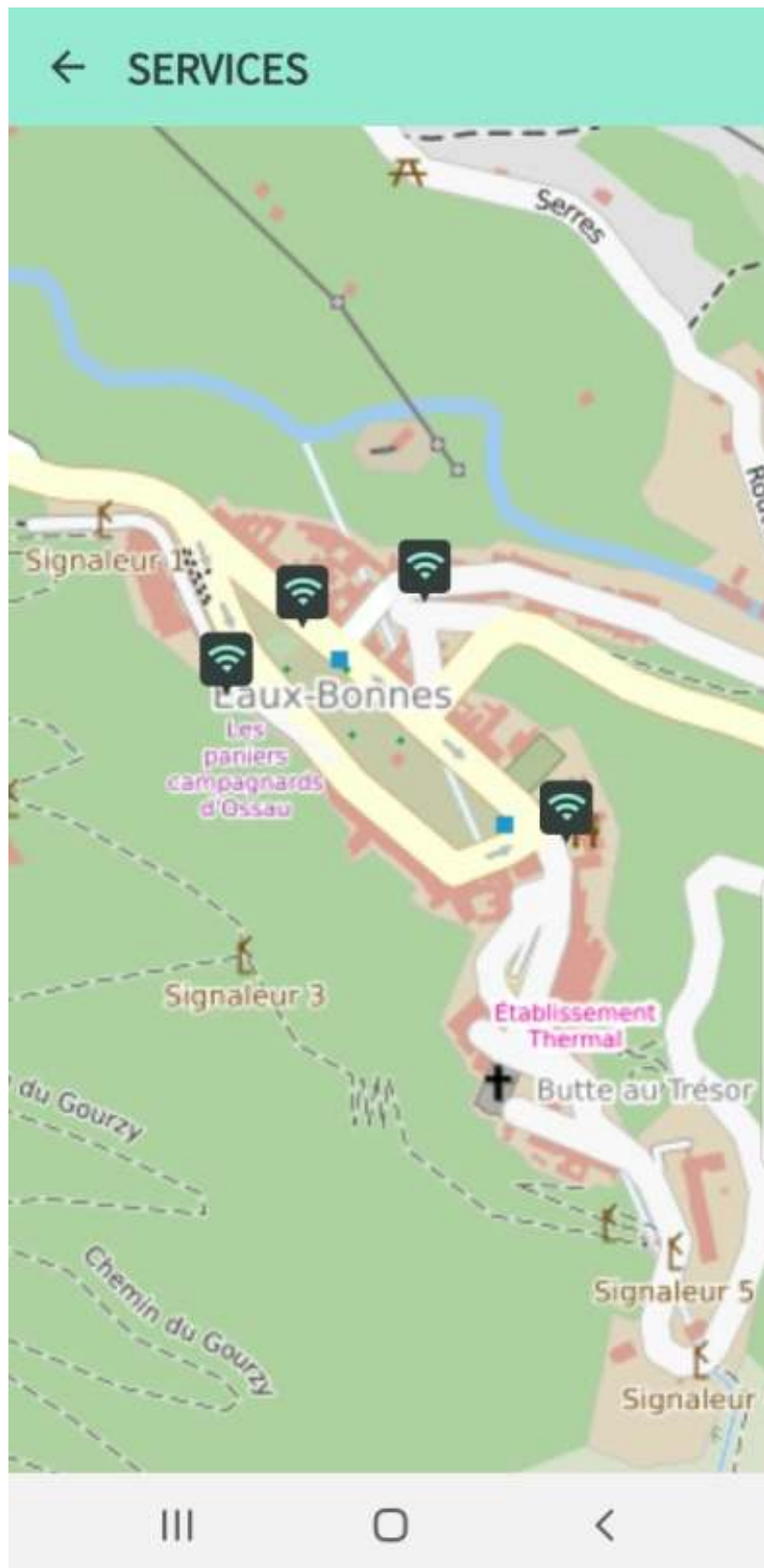


Image produite par les auteurs

57 De plus, les POI de type « Patrimoine culturel immatériel » recensés dans *PCI Lab* sont importés et associés à la ville la plus proche (située à moins de 20 kilomètres). Une gestion de l'état de la batterie du téléphone est réalisée de manière à privilégier l'itinérance. De plus, l'itinéraire tient compte de l'identification des zones blanches pour anticiper les téléchargements des POI à venir sur le tracé et de l'état de la batterie : si la batterie est chargée à moins de 30 %, arrêt des affichages de vidéos et images ; à moins de 20 %, arrêt des téléchargements de contenus relatifs aux POI ; à moins de 15 %, recalcul de l'itinéraire pour arriver au plus vite à la fin du parcours ; à moins de 10 %, arrêt du rafraîchissement de la position GPS sur la carte.

## Feedback

58 Les POI recensés dans le cadre du projet *TCVPYR* vont permettre aux touristes de découvrir des lieux nouveaux. Cette application est gratuite et téléchargeable sur Google Play depuis mai 2020.

59 Cette application présente aux touristes des POI qui sont peu connus dans une ville mais qui présentent une histoire intéressante. Les agents de l'office du tourisme avec qui nous avons discuté ont montré un intérêt pour l'application, car elle recommande des POI qui sont complémentaires de leur recommandation. Cela permet de bien valoriser les sites et objets patrimoniaux de la ville.

60 En raison du Covid-19, nous n'avons à ce jour pu faire des expérimentations sur site qu'avec le prototype. Nous les avons détaillées dans Rajaonarivo *et al.* (2019a). Ces expérimentations préliminaires nous ont permis d'ajuster les valeurs de certains paramètres, de vérifier le niveau de satisfaction des contraintes (par exemple, la durée de visite et les moyens de transport). Les recommandations correspondent à la réalité.

## Conclusion

61 Nous avons proposé des services de promotion de POI patrimoniaux auprès du grand public, illustrés avec différents scénarios de valorisation. À cette fin, nous avons conçu un processus complet de diffusion de ces connaissances à travers deux canaux de publication : la publication automatique d'articles et d'images dans l'environnement Wikimedia ; le développement d'une application mobile en *open source* de recommandation d'itinéraires touristiques personnalisés et contextualisés.

62 Concernant le processus de publication d'articles, nous avons démontré que nous pouvons publier automatiquement des données sur *Wikipédia* ainsi que les données liées correspondantes sur Wikidata et Wikimedia Commons. Une expérimentation a été menée sur un échantillon de données patrimoniales (des rapports d'inventaires et des photos réalisés par des chercheurs géographes et historiens) du projet européen FEDER *TCVPYR*, qui vise la promotion des villes thermales des Pyrénées. Ainsi, *Wikipédia* devient un vecteur intéressant pour la promotion automatisée du patrimoine culturel et donc du tourisme dans la région. Les conditions technologiques sont désormais remplies pour la construction de chaînes d'alimentation automatique de *Wikipédia*. La principale limite reste cependant celle de la gestion des droits d'auteur.

Étant donné que les géographes, les historiens, mais aussi le personnel des autorités locales responsables des inventaires ont l'habitude de publier en priorité leurs données dans les bases de données régionales et nationales, ces problèmes d'antériorité des publications et de droit d'auteur restent, à l'heure actuelle, difficiles à résoudre.

63 Nous sommes convaincus que *Wikipédia* est un canal particulièrement intéressant pour la promotion de données patrimoniales. C'est pourquoi nous voyons de nombreuses perspectives pour ce travail. Tout d'abord, nous pensons que les autorités nationales françaises en charge du patrimoine culturel devraient collaborer avec les personnes en charge du contrôle sur *Wikipédia* afin d'élaborer une chaîne de publication française mieux organisée aux niveaux régional et national, ainsi que pour choisir des licences adaptées à *Wikipédia*. Deuxièmement, nous prévoyons d'étendre notre expérimentation à un ensemble de données plus large ainsi qu'à de nouvelles données (par exemple, les données sur le patrimoine culturel immatériel du projet *PCI Lab*) provenant de différents domaines d'application. Enfin, pour en revenir à l'aspect technique, les cas 1, 2 et 3 de WikiAlgo ne sont pas encore mis en œuvre. Ils concernent la modification automatique du contenu d'articles *Wikipédia* existants. C'est un défi considérable, que nous voulons relever. Il nécessite notamment l'utilisation de techniques de traitement automatique du langage naturel (TALN).

64 Concernant l'application mobile, nous avons, dans un premier temps, développé un prototype testé sur site. Les premiers résultats obtenus ont permis notamment d'ajuster les paramètres liés à la gestion des préférences des utilisateurs, mais également, d'un point de vue technique, de proposer une API REST mettant en application les différents algorithmes précédemment présentés. L'application mobile et l'API ont ensuite été industrialisées. L'application est disponible sur l'Android Store. Nous comptons poursuivre le travail d'analyse des données utilisateurs pour conforter ou adapter nos paramétrages (par exemple, les valeurs initiales des coefficients des formules de mesure de pertinence) en nous fondant sur un panel d'utilisateurs varié et plus important. Enfin, nous prévoyons de généraliser notre approche en alimentant le système de recommandations avec des POI patrimoniaux publiés sur *Wikipédia*, sans passer par nos bases de données locales. Cette modification de la chaîne de traitement *TCVPYR* aura un double avantage : nous libérer de la maintenance d'une base de données locale de gestion des POI ; bâtir dynamiquement nos itinéraires à partir de nos POI publiés sur *Wikipédia* puis enrichis par de nouveaux POI patrimoniaux publiés collaborativement par la communauté. De plus, nous allons exploiter des POI issus du LOD (Datatourisme, GeoDataMine, ou encore, *PCI Lab*), spatialement et thématiquement proches des préférences associées à une demande d'itinéraire.

## Bibliographie

Abella, Alberto, Marta Ortiz-de Urbina-Criado et Carmen De Pablos-Heredero. 2019. « The Process of Open Data Publication and Reuse ». *Journal of the Association for Information Science and Technology* 70 : 296-300.

Agirre, Eneko, Ander Barrena, Oier Lopez de Lacalle, Aitor Soroa, Samuel Fernando, et Mark Stevenson. 2012. « Matching Cultural Heritage Items to Wikipedia ». Dans *Proceedings of the Eighth International Conference on Language Resources and Evaluation. LREC 2012*,

Istanbul, Turkey, May 23-25, édité par Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Mehmet Uğur Doğan, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk et Stelios Piperidis, 1729-1735. European Language Resources Association (ELRA). <http://www.lrec-conf.org/proceedings/lrec2012/summaries/1021.html>.

Amorim, Mario, Adriana Mar, Fernando Monteiro, Stella Sylaiou, Pedro Pereira et João Martins. 2018. « Smart Tourism Routes Based on Real Time Data and Evolutionary Algorithms ». Dans *Digital Heritage. Progress in Cultural Heritage : Documentation, Preservation, and Protection*, édité par Marinos Ioannides, Eleanor Fink, Raffaella Brumana, Petros Patias, Anastasios Doulamis, João Martins et Manolis Wallace, 417-426. Cham : Springer. [https://doi.org/10.1007/978-3-030-01762-0\\_36](https://doi.org/10.1007/978-3-030-01762-0_36).

Berners-Lee, Tim. 2009. « Linked Data ». *World Wide Web Consortium*. <https://www.w3.org/DesignIssues/LinkedData.html>.

Bessagnet, Marie-Noëlle, Patrick Etcheverry, Annig Le Parc Lacayrelle, Christophe Marquesuzaà, Landy Rajaonarivo, Philippe Roose, André Fonteles et Christian Sallaberry. 2018. « Leveraging Heterogeneous Cultural Heritage Data to Promote Tourism ». Communication présentée à *Open Source Geospatial Research and Education Symposiums (OGRS)*, Lugano, octobre. <https://hal.archives-ouvertes.fr/hal-01976405>.

Brilhante, Igo, Jose Antonio de Macêdo, Franco Maria Nardini, Raffaele Perego et Chiara Renso. 2014. « Tripbuilder : A Tool for Recommending Sightseeing Tours ». Dans *Advances in Information Retrieval. European Conference on Information Retrieval (ECIR) 2014*, édité par Maarten de Rijke, Tom Kenter, Arjen P. de Vries, ChengXiang Zhai, Franciska de Jong, Kira Radinsky et Katja Hofmann, 771-774. Cham : Springer. [https://doi.org/10.1007/978-3-319-06028-6\\_93](https://doi.org/10.1007/978-3-319-06028-6_93).

Brilhante, Igo, Jose Antonio de Macêdo, Franco Maria Nardini, Raffaele Perego et Chiara Renso. 2015. « On Planning Sightseeing Tours with Tripbuilder ». *Information Processing and Management* 51 (2) : 1-15. <https://doi.org/https://doi.org/10.1016/j.ipm.2014.10.003>.

Candela, Gustavo, Pilar Escobar, Rafael C. Carrasco et Manuel Marco-Such. 2018. « A Linked Open Data Framework to Enhance the Discoverability and Impact of Culture Heritage ». *Journal of Information Science* 45 (6) : 756-766. <https://doi.org/10.1177/0165551518812658>.

Casteret, Jean-Jacques et Mélanie Larché. 2018. « Le projet PciLab pour la valorisation numérique de l'inventaire français du PCI ». *Les cahiers du CFPCI* 5 : 102-112. <https://www.maisondesculturesdumonde.org/archives/patrimoine-culturel-immateriel-et-numerique>.

Corallo, Angelo, Laura Fortunato, Clara Renna, Marco Sarcinella et Alessandra Spennato. 2017. « Mobile App for Promoting Cultural Heritage : Geostatistic and Textual Analysis ». Dans *Proceedings of IMEKO International Conference on Metrology for Archaeology and Cultural Heritage*, 194-201. IMEKO

Daquino, Marinela, Francesca Mambelli, Silvio Peroni, Francesca Tomasi et Fabio Vitali. 2017. « Enhancing Semantic Expressivity in the Cultural Heritage Domain : Exposing the Zeri Photo Archive as Linked Open Data ». *Journal on Computing and Cultural Heritage (JOCCH)* 10 (4) : 21. <https://doi.org/10.1145/3051487>.

Emmanouilidis, Christos, Remous-Aris Koutsiamanis et Aimilia Tasidou. 2013. « Mobile Guides : Taxonomy of Architectures, Context Awareness, Technologies and Applications ». *Journal of Network and Computer Applications* 36 (1) : 103-125. <https://doi.org/10.1016/j.jn-ca.2012.04.007>.

Fonteles, André, Marie-Noëlle Bessagnet, Annig Le Parc Lacayrelle et Christian Sallaberry. 2018. « Un environnement pour la valorisation de données patrimoniales hétérogènes ». Dans *Actes de conférence Spatial Analysis and GEomatics (SAGEO 2018)*, édité par Maguelonne Teisseire et Mathieu Roche, 156-161. SAGEO.

Freire, Nuno, Enno Meijers, René Voorburg et Antoine Isaac. 2018. « Aggregation of Cultural Heritage Datasets Through the Web of Data ». *Procedia Computer Science* 137 : 120-126. <https://doi.org/10.1016/j.procs.2018.09.012>.

Kucera, Jan, Dusan Chlapek, Jakub Klímek et Martin Necaský. 2015. « Methodologies and Best Practices for Open Data Publication ». Dans *Proceedings of the DATESO 2015 Workshop* : 52-64. DATESO.

Lu, Qibei et Guo Feipeng. 2016. « A Novel e-Commerce Customer Continuous Purchase Recommendation Model Research Based on Colony Clustering ». *International Journal of Wireless and Mobile Computing* 11 (4) : 309-317. <https://doi.org/10.1504/IJWMC.2016.082288>

Marden, Julia, Carolyn Li-Madeo, Noreen Whysel et Jeffrey Edelstein. 2013. « Linked Open Data for Cultural Heritage : Evolution of an Information Technology ». Dans *Proceedings of 31st ACM International Conference on Design of Communication*, 107-112. New York : Association for Computing Machinery. <https://doi.org/10.1145/2507065.2507103>.

Minjing, Peng, Liu Xinglin, Luo Ximing, Zhu Mingliang, Zhang Xianyong, Deng Xiangming et Wu Mingfen. 2017. « Recognizing Intentions of e-Commerce Consumers Based on Ant Colony Optimization Simulation ». *Journal of Intelligent and Fuzzy Systems* 33 (5) : 2687-2697. <https://doi.org/10.3233/JIFS-169318>.

Rajaonarivo, Landy, André Fonteles, Christian Sallaberry, Philippe Roose, Marie-Noëlle Bessagnet, Patrick Etcheverry, Annig Le Parc Lacayrelle, Christophe Marquesuzaà, Cécile Cayère et Quentin Coudert. 2019a. « Recommandation et valorisation d'objets patrimoniaux hétérogènes ». Dans *Actes de la conférence Inforsid*, 171-186. Inforsid. [http://inforsid.fr/actes/2019/Actes\\_INFORSID2019.pdf](http://inforsid.fr/actes/2019/Actes_INFORSID2019.pdf).

Rajaonarivo, Landy, André Fonteles, Christian Sallaberry, Marie-Noëlle Bessagnet, Philippe Roose, Patrick Etcheverry, Christophe Marquesuzaà, Annig Le Parc Lacayrelle, Cécile Cayère et Quentin Coudert. 2019b. « Recommendation of Heterogeneous Cultural Heritage Objects for the Promotion of Tourism ». *ISPRS Int. J. Geo-Inf.* 8 (5) : 230. <https://doi.org/10.3390/ijgi8050230>.

Roman, Dumitru, Nikolay Nikolov, Antoine Putlier, Dina Sukhobok, Brian Elvesæter, Arne Berre, Xianglin Ye, Marin Dimitrov, Alex Simov, Momchill Zarev, Rick Moynihan, Bill Roberts, Ivan Berlocher, Seonho Kim, Tony Lee, Amanda Smith et Tom Heath. 2018. « Datagraft : One-Stop-Shop for Open Data Management ». *Semantic Web* 9 (4) : 393-411. <https://doi.org/10.3233/SW-170263>.

Smirnov, Alexander V., Alexey M. Kashevnik et Andrew Ponomarev. 2017. « Context-Based Infomobility System for Cultural Heritage Recommendation : Tourist Assistant-TAIS ». *Personal and Ubiquitous Computing* 21 : 297-311. <https://doi.org/10.1007/s00779-016-0990-0>.

UNESCO. 2020. « What is Intangible Cultural Heritage ? » UNESCO. <https://ich.unesco.org/en/what-is-intangible-heritage-00003>.

Vetrò, Antonio, Lorenzo Canova, Marco Torchiano, Camilo Orozco Minotas, Raimondo Iemma et Federico Morando. 2016. « Open Data Quality Measurement Framework : Definition and Application to Open Government Data ». *Government Information Quarterly* 33 (2) : 325-337. <https://doi.org/10.1016/j.giq.2016.02.001>.

Zhang, Xin et Xiao Pang. 2015. « Analysis on the Mobile Electronic Commerce Recommendation Model Based on the Ant Colony Algorithm ». Dans *Proceedings of the 2015 International Conference on Industrial Technology and Management Science*, édité par Junjiao Zhao. Atlantis Press. <https://doi.org/10.2991/itms-15.2015.423>.

## Notes

1 <http://www.pci-lab.fr>.

2 <http://www.pop.culture.gouv.fr>.

3 <https://www.statista.com/statistics/270291/popular-categories-in-the-app-store/>.

4 <https://www.tripadvisor.fr>.

5 <http://whc.unesco.org/fr/activites/913/>.

6 <https://www.ontotext.com/knowledgehub/fundamentals/linked-data-linked-open-data/>.

7 <https://datagraft.io>.

8 <http://www.w3.org/RDF/>.

9 <http://www.w3.org/TR/rdf-sparql-query/>.

10 <http://www.cidoc-crm.org>.

11 [https://en.wikipedia.org/wiki/Help:Your\\_first\\_article/](https://en.wikipedia.org/wiki/Help:Your_first_article/).

12 <https://fr.wikipedia.org/wiki/Projet:Valdensia/>.

13 <http://www.metmuseum.org/blogs/now-at-the-met/2018/open-access-at-the-met-year-one/>.



14 RenabLP2 et Gertrude sont des outils de création et de gestion de dossiers électroniques appliqués à la documentation de l'inventaire du patrimoine culturel. RenabLP2 est utilisé en région Occitanie et Gertrude l'est en région Nouvelle-Aquitaine et Occitanie.

15 Ce prototype est disponible à l'adresse <http://tcvpyr.iutbayonne.univ-pau.fr>.

16 [https://en.wikipedia.org/wiki/Wikipedia:Policies\\_and\\_guidelines/](https://en.wikipedia.org/wiki/Wikipedia:Policies_and_guidelines/).

17 <https://fr.wikipedia.org/wiki/Projet:TCVPYR/>.

18 <https://www.pop.culture.gouv.fr/notice/palissy/IM31000912/> ; <http://dossiers-inventaire.aquitaine.fr/dossier/station-thermale-des-eaux-bonnes/83a90a4b-51a1-49fa-936c-d56395643c1e/> ; [https://ressourcespatrimoines.laregion.fr/ark:/46855/inventaire\\_IA31012369?posInSet=1&queryId=6fc139dc-75b5-4743-ae8a-1c6b5c7d3524/](https://ressourcespatrimoines.laregion.fr/ark:/46855/inventaire_IA31012369?posInSet=1&queryId=6fc139dc-75b5-4743-ae8a-1c6b5c7d3524/).

19 <https://flutter.dev>.

20 <https://developer.android.com/studio/>.

21 <https://www.graphhopper.com>.

## Auteurs

### Landy Rajaonarivo

EA 3000 LIUPPA, université de Pau et des pays de l'Adour, E2S UPPA, Pau, France

Landy Rajaonarivo est actuellement en post-doctorat au Japon à l'université de Kyushu ; elle a obtenu son doctorat en 2018 à l'université de Brest et a ensuite travaillé durant trois ans sur le projet *TCVPYR* à l'université de Pau et des pays de l'Adour en tant que post-doc-  
torante ; ses travaux de recherche visent notamment la recommandation.

ORCID 0000-0003-1831-270X

[landy.rajaonarivo@univ-pau.fr](mailto:landy.rajaonarivo@univ-pau.fr)

### Marie-Noëlle Bessagnet

EA 3000 LIUPPA, université de Pau et des pays de l'Adour, E2S UPPA, Pau, France

Marie-Noëlle Bessagnet est maître de conférences à l'université de Pau et des pays de l'Adour depuis 1992 ; elle a obtenu son doctorat en 1991 ; ses principaux axes de recherche se situent autour du traitement sémantique de l'information dans des documents textuels ou dans des traces numériques, par la conception de modèles ou plateformes d'extraction, annotation et indexation spatiales, temporelles et thématiques.

ORCID 0000-0002-2538-1298

[marie-noelle.bessagnet@univ-pau.fr](mailto:marie-noelle.bessagnet@univ-pau.fr)

### Christian Sallaberry

EA 3000 LIUPPA, université de Pau et des pays de l'Adour, E2S UPPA, Pau, France

Christian Sallaberry est maître de conférences en informatique à l'université de Pau et des pays de l'Adour depuis 1994 ; ses travaux de recherche portent sur l'analyse de corpus textuels et de trajectoires : extraction, indexation et recherche d'informations spatiales, temporelles et thématiques, dans des domaines d'application tels que le tourisme et les villes intelligentes.

ORCID 0000-0002-3605-3927

[christian.sallaberry@univ-pau.fr](mailto:christian.sallaberry@univ-pau.fr)

### Annig Le Parc Lacayrelle

EA 3000 LIUPPA, université de Pau et des pays de l'Adour, E2S UPPA, Pau, France

Annig Le Parc Lacayrelle, maître de conférences à l'université de Pau et des pays de l'Adour, a obtenu un doctorat de l'université de Toulouse III en 1997 ; ses recherches actuelles concernent la modélisation de données hétérogènes ainsi que leur valorisation.

ORCID 0000-0002-2716-6663

[annig.lacayrelle@univ-pau.fr](mailto:annig.lacayrelle@univ-pau.fr)

### Philippe Roose

EA 3000 LIUPPA, université de Pau et des pays de l'Adour, E2S UPPA, Pau, France

Philippe Roose, dont les principaux axes de recherche se situent autour des *middleware*, de la prise en compte du contexte et de l'informatique écoresponsable, est maître de conférences à l'université de Pau et des pays de l'Adour depuis 2001 et a obtenu une habili-

tation à diriger les recherches en 2008 ; co-inventeur de la plateforme Kalimucho (<https://kalimucho.univ-pau.fr/>), qui a donné lieu au dépôt d'une marque, et de trois brevets internationaux, il crée actuellement une startup.

ORCID [0000-0002-2227-3283](https://orcid.org/0000-0002-2227-3283)

[philippe.roose@univ-pau.fr](mailto:philippe.roose@univ-pau.fr)

## Droits d'auteur



Les contenus de la revue *Humanités numériques* sont mis à disposition selon les termes de la [Licence Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/).