

Kenan Oudin

Master 2 Médiation Numérique de la Culture et des Patrimoines

kenan.oudin@etu-univ-tours.fr

Mémoire de recherche

Université de Tours - Centre d'Études Supérieures de la Renaissance

Les interfaces de programmation d'application (API) en humanités
numériques - étude de cas de l'API d'Images IIF

2021 - 2022

Tuteur pédagogique : Elena Pierrazzo

INTRODUCTION	4
CONTEXTUALISATION	7
1. Définir une API	7
<i>1.1 Origine</i>	7
<i>1.2 Fonctionnement des API</i>	8
<i>1.3 La structure des api de twitter</i>	12
DÉVELOPPEMENT	15
1. Find your own place : enjeux et problématiques des API en humanités numériques	15
<i>1.1 « Do APIs Have a Place in the Digital Humanities? »</i>	16
<i>1.2 Les enjeux</i>	17
<i>1.3 l'impact des données ouvertes</i>	19
2. Le protocole IIIF	22
<i>2.1 la communauté IIIF</i>	22
<i>2.2 les fonctionnalités de l'api Image</i>	24
<i>2.3 Projet de recherche Righettino</i>	26
3. Cas d'études d'utilisation des API Images	27
<i>3.1 BnF - Gallica</i>	27
<i>3.2 Europeana</i>	29
OUVERTURE	31
La valorisation des données patrimoniales au sein des institutions publiques	31
CONCLUSION	33
BIBLIOGRAPHIE /	34

SITOGRAFIE	34
Sitographie	34
Bibliographie	36

INTRODUCTION

« As a community of practice, digital humanists deal with data and metadata not as static artifacts, but rather as complex, multi-dimensional and multi-layered datasets that can be analyzed, annotated and manipulated in order to produce new knowledge. »¹

Les humanités numériques (HN) ont modifié en profondeur la vision des données provenant des sciences humaines. L'idée, maintenant lointaine, de sources « statiques » n'interagissant pas entre elles est le résultat de plusieurs problématiques. On peut citer, entre autre, les difficultés des chercheurs à partager leurs résultats ou l'accessibilité des sources. Cela a pu être largement résolu avec l'apparition du numérique et la mise à disposition des données à travers le monde. Les humanités numériques, traduit de l'anglais *Digital humanities* (DH), se développent dès les années 50², jusqu'à aujourd'hui de nombreux chercheurs des sciences humaines se forment pour adapter leurs données. Ainsi, les données se développent, interagissent entre elles et amènent à évoluer. Dans ce contexte d'échange de données, qui n'est pas exclusif aux humanités numériques, les acteurs de l'informatique développent de nouvelles technologies d'interactions entre les applications. Les interfaces de programmation d'application (API), au début du numérique sont principalement utilisées en tant que bibliothèques pour les systèmes d'exploitation. C'est au début des années 2000 qu'elles deviennent importantes pour l'intégration des données à distance. La numérisation des données patrimoniales et des demandes constantes des chercheurs à l'accès aux données questionnent les environnements numériques. Cette confrontation de problématiques et de développement numérique va amener les chercheurs à faire appel aux API. Dès 2005, Dan Cohen

¹ Tasovac, Toma, Adrien Barbaresi, Thibault Clérice, Jennifer Edmond, Natalia Ermolaev, Vicky Garnett, et Clifford Wulfman. « APIs in Digital Humanities: The Infrastructural Turn ». In *Digital Humanities 2016*, 93-96. Digital Humanities 2016 Conference Abstracts. Cracovie, Poland, 2016. <https://hal.archives-ouvertes.fr/hal-01348706>.

² Création d'un index de l'oeuvre de Thomas D'Aquin par Roberto Busa - Sinatra, Michael E., et Marcello Vitali Rosati. « Histoire des humanités numériques ». In *Pratiques de l'édition numérique*. Les Ateliers de [sens public], 2014. <http://www.parcoursnumeriques-pum.ca/1-pratiques/> <https://www.parcoursnumeriques-pum.ca/1-pratiques/chapitre3.html>.

doyen et professeur à la Northeastern University se demande si « Les API ont une place dans les humanités numériques ? »³. Ce développement précoce a toujours sa place dans l'évolution des humanités numériques. En observant les API présentes dans l'échange des données patrimoniales, nous ne pouvons pas répondre négativement à la question de Dan Cohen. Nous l'aborderons par la suite à l'aide de cas d'étude, les API rassemblent les données, les organisent et permettent aux chercheurs ou aux professionnels en sciences humaines de questionner ces données. L'objectif premier est de créer une bibliothèque de données interrogeable. L'infrastructure des données change, il faut trouver l'outil pouvant faciliter ce travail. À l'image d'un fonctionnement dans une bibliothèque, ou tout autres lieux de conservation de données, l'API correspond au personnel répondant à nos demandes et nous aiguillant vers les données que nous souhaitons.

Dans ce contexte de transformations des données, les médias patrimoniaux (représentations 3D, images numérisées, peintures,...) doivent s'adapter. Grâce à la collaboration de nombreuses institutions⁴, en 2011, le protocole International Image Interoperability Framework (IIIF) apparaît. Ce protocole, qui s'organisera en communauté IIIF, développe une API Image dès 2012. L'objectif de ce développement est de proposer une communication entre plusieurs systèmes autour de la diffusion d'images en haute résolution sur le Web. La plupart des institutions culturelles et patrimoniales vont mettre en place ce protocole permettant une interopérabilité des données, mais aussi offrant de nouveaux outils d'interactions avec ces données. Les médias ne sont plus seulement des objets de présentation et d'illustration, mais des données vivantes et explicables. Notre travail de recherche va présenter les enjeux des API dans les humanités numériques et leur impact sur les données en sciences humaines. L'objectif n'est pas de refaire cette historiographie, mais en prenant le cas d'étude de l'API Image, d'expliquer les protocoles mis en œuvre pour répondre aux problématiques des chercheurs sur l'interaction des données et de proposer une solution pérenne et commune à cette réflexion.

Pour étayer ce propos, notre première partie va contextualiser et définir ce qu'est une API, son développement rapide, ainsi que son utilisation dans les humanités numériques. Après cet état de l'art, nous développerons notre propos en expliquant le fonctionnement du protocole IIIF et

³ « Do APIs Have a Place in the Digital Humanities? – Dan Cohen ». Consulté le 6 septembre 2022. <https://dancohen.org/2005/11/21/do-apis-have-a-place-in-the-digital-humanities/>.

⁴ British Library, l'Université de Stanford, la Bibliothèque Bodléienne (Université d'Oxford), la Bibliothèque nationale de France, la Nasjonalbiblioteket (Bibliothèque Nationale de Norvège), le Laboratoire National de Los Alamos de la Bibliothèque de Recherche, et l'Université de Cornell.

l'utilisation de l'API Image. Pour cela, nous travaillerons autour d'exemple d'intégration de cette API dans une structure culturelle française et une structure européenne.

CONTEXTUALISATION

1. Définir une API

« La mise en œuvre d'interfaces de programmation applicatives ou API (Applications Programming Interface) est une méthode d'ouverture par les données visant le développement de services annexes et ainsi la croissance d'un écosystème périphérique à la plateforme. »⁵

Avec l'avènement du numérique, les flux de données, quelques soient les domaines, sont massifs et nécessitent d'être exploités et interprétés. On remarque à travers la place des API dans le monde numérique qu'elles ont permis de répondre aux problématiques émergentes. Avant d'expliquer comment les humanités se sont appropriées les API pour leurs données, il faut retracer leur création et leur démocratisation dans le Web.

1.1 ORIGINE

Lors de l'apparition des API dans les années 50, l'objectif était de simplifier la programmation, et l'interface était destiné à l'utilisateur final. C'était donc un programme d'application ayant pour logique d'effectuer une unique tâche. La dénomination de bibliothèque était récurrente pour expliquer le fonctionnement des API. Le développeur pouvait questionner localement ses données et interagir avec. Un programme d'application ou *application software*, comme le sont les API, facilité l'usage des utilisateurs pour exercer une activité : manipulation textuelle, utilisation de médias,... À partir des années 70, la notion d'interface entre en jeu pour décrire l'interaction possible entre une application et le système informatique. Si l'API est suffisamment développée, elle peut interagir avec d'autres interfaces. L'essor de cette technologie prend sens avec le développement des API Web. Les requêtes des développeurs peuvent donc interroger des bibliothèques disponibles en ligne et non plus localement. Les données deviennent donc interconnectées, pluridisciplinaires et questionnables. Roy Fielding définit dans sa thèse⁶ cette idée des API fonctionnant en réseau. Il ne faut donc plus penser la donnée comme une ressource

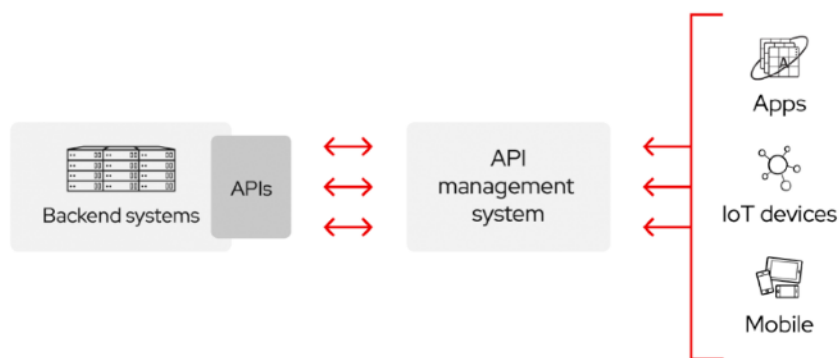
⁵ Francony, Jean-Marc. « L'éditorialisation des données aux bornes des API : enjeux et perspectives pour une analyse empirique ». *Les Enjeux de l'information et de la communication* 19/2, n° 2 (2018): 69-79. <https://doi.org/10.3917/enic.025.0069>.

⁶ Fielding, Roy Thomas. *Architectural Styles and the Design of Network-based Software Architectures*. Doctoral dissertation, University of California, Irvine, 2000.

« locale » que l'on diffuse à une plus grande échelle, mais l'intégrer lors de sa création dans une logique d'interopérabilité.

Ainsi, les API vont permettre aux développeurs d'écrire des programmes informatiques pour « accéder à des jeux de données dans l'objectif de remplir un certain nombre de fonctions »⁷. Ces outils vont induire de nouvelles problématiques dans leur développement et leur conception. La compréhension de l'architecture web et numérique n'est pas accessible à tout le monde de par la disparité d'accès des API (libre accès, accès sur demande ou accès restreint), mais aussi l'uniformité des données. Pour comprendre comment ont évolué les API et leur intégration dans les humanités numériques, il faut expliquer leur fonctionnement.

1.2 FONCTIONNEMENT DES API



Nous avons donc vu précédemment que les API permettent de communiquer avec d'autres produits et services sans connaître le développement de chaque service questionné. Les données sont ainsi partagées, librement ou non, avec des utilisateurs externes. Pour développer une API, deux éléments vont être importants. Tout d'abord avoir un serveur qui va fournir et exécuter le programme défini dans l'API. Ensuite, le client qui est un programme distinct qui va recevoir les données. Ce dernier est le plus en contact avec l'utilisateur final se matérialisant par un site web, une application,... Comme symbolisé dans le schéma ci-dessus, l'API peut être considérée comme un outil de médiation des données. La spécification des API va se concrétiser au moment du développement du projet. Il va tout d'abord falloir définir le type d'API souhaitée en fonction des

⁷ Samuel Goyet. De briques et de blocs. La fonction éditoriale des interfaces de programmation (API) web : entre science combinatoire et industrie du texte.. Sciences de l'information et de la communication. Paris IV Sorbonne, 2017. Français

besoin, API SOAP ou API REST. Il faut également convenir de l'accessibilité des données pour les utilisateurs.

Typologie des API

Le développement d'Internet a transformé en profondeur le fonctionnement des API. Les environnements locaux se font plus rares pour laisser place aux API distantes⁸. Les API distantes vont interagir au sein « d'un réseau de communication ». Les données que l'on interroge ne vont pas se trouver sur la machine qui formule la requête. Dans ce cas, des technologies diverses, des plateformes ou des langages différents peuvent fonctionner. Les API web, majoritairement présentes, utilisent le protocole HTTP pour les requêtes. Cette interface est accessible via le client (smartphone, tablettes, ...) en direction d'un serveur web. La réponse de l'API se convertit en format JSON (JavaScript Object Notation) ou en XML (Extensible Markup Language). Ainsi, les API Web vont venir progressivement remplacer les API système gérant le fonctionnement internes d'un système (paramètres de connexion, instructions d'une application ou d'un matériel physique). L'intégration en réseau des API permet à plusieurs API d'être combinées. L'exemple le plus pertinent que nous développerons par la suite concerne l'API de Twitter permettant un accès, pour les développeurs, aux données de Twitter grâce à son API REST, et son API de recherche pour interagir avec les données et les tendances. En sus, deux protocoles se développent au sein des API, les API SOAP et les API REST.

Protocole des API

Ces deux développements, SOAP ou REST vont définir le développement des API pour les échanges des données entre plusieurs applications web. La différence significative entre les deux, c'est que SOAP est un protocole et REST non⁹. Ce n'est pas cette maigre affirmation qui va nous convaincre dans l'utilisation de l'un ou de l'autre. *Simple Object Access Protocol* ou SOAP est un protocole standard permettant la communication sur une multitude de langages et de plateformes. Une API SOAP va être privilégiée par les entreprises pour intégrer un niveau important de sécurité,

⁸ « Qu'est-ce qu'une API ? ». <https://www.redhat.com/fr/topics/api/what-are-application-programming-interfaces>.

⁹ « REST et SOAP : quelle est la différence ? ». <https://www.redhat.com/fr/topics/integration/whats-the-difference-between-soap-rest>.

mettre à jour ou supprimer des comptes et des mots de passe utilisateurs. La requête provenant du client peut être faite via un protocole HTTP (navigateur web) ou SMTP (protocole de communication pour les courriers électroniques). Cependant, la réponse se fera sous la forme d'un document XML, langage utilisant un balisage simple et lisible par un utilisateur mécanique ou humain. L'un des exemples d'entreprise utilisant le protocole SOAP est PayPal¹⁰. Ce protocole plus contraignant, mais largement utilisé dans un premier temps, est remplacé par les API REST. *Representational State transfer* est une architecture d'API plus flexible que le protocole SOAP et principalement adapté aux services web. Comme le protocole SOAP, les requêtes sont envoyées en HTTP, mais la réponse peut se faire sous différents formats, JSON, HTML ou XML. L'objectif de cette architecture est de simplifier les API et la réponse se fera souvent au format JSON. Par ailleurs, les API REST tendent à se standardiser pour faciliter le travail des développeurs. Loin d'être un protocole aussi strict que SOAP, certaines API peuvent caractériser d'API RESTful en suivant un certains nombres de normes. Dans la thèse de Roy Fielding¹¹, il définit six contraintes pour obtenir ce qualificatif. Tout d'abord, elles doivent reposer sur une architecture client-serveur traitant les requêtes via le protocole HTTP. Les informations du client ne sont pas stockées sur le serveur dit *stateless*. Par exemple, si vous effectuez une recherche internet et que votre connexion est interrompue vous devez recommencer. Les données doivent ensuite être mises en mémoire cache pour éviter les interactions client-serveur. La médiation des interactions entre le client et le serveur par un système de couches pouvant posséder une fonction unique. Un code adaptable qui permet au serveur d'avoir de nouvelles fonctionnalités. Enfin, la plus importante définie par Roy Fielding, c'est d'avoir une interface uniforme permettant un transfert uniformisé des échanges. Les ressources sont identifiées lors des requêtes, ces ressources vont avoir des représentations pouvant être modifiées ou supprimées. La façon dont les informations doivent être traitées par le client est lisible par des « messages autodescriptifs ». Enfin, une fois les ressources acquises, le client REST doit pouvoir accéder aux actions disponibles par des hyperliens.

¹⁰ « Get Started ». Consulté le 20 septembre 2022. <https://developer.paypal.com/api/nvp-soap/get-started/>.

¹¹ Fielding, Roy Thomas. *Architectural Styles and the Design of Network-based Software Architectures*. Doctoral dissertation, University of California, Irvine, 2000.

Disponibilité des API

Pour ces deux protocoles API, il faut identifier si elles sont publiques, privées ou partenaires. La disponibilité des API va avoir une importance dans les humanités numériques pour faciliter ou non l'accès aux données. Cela va dépendre notamment de la nature des données mises à disposition. Une entreprise collectant des données sensibles sur ses clients ne va pas avoir le même niveau de sécurité qu'une institution publique culturelle. Les API privées vont être réservées à une utilisation interne à la structure qui l'a développée. Elle va favoriser la collaboration pour les développeurs et la protection des données. Dans un esprit d'ouverture et de développement en réseau, certaines API vont être partenaires. Elles ne sont pas publiques, mais requièrent une autorisation grâce à l'obtention d'une identification. Enfin, nous avons les API publiques, disponibles par tous. Les API REST sont notamment développées si on souhaite la rendre publique. L'idée prédominante est d'avoir une interface *open source* permettant de faciliter le travail des développeurs et de fournir des ressources réutilisables. Cependant, les API publiques nécessitent une veille constante pour s'assurer de la stabilité de l'interface.

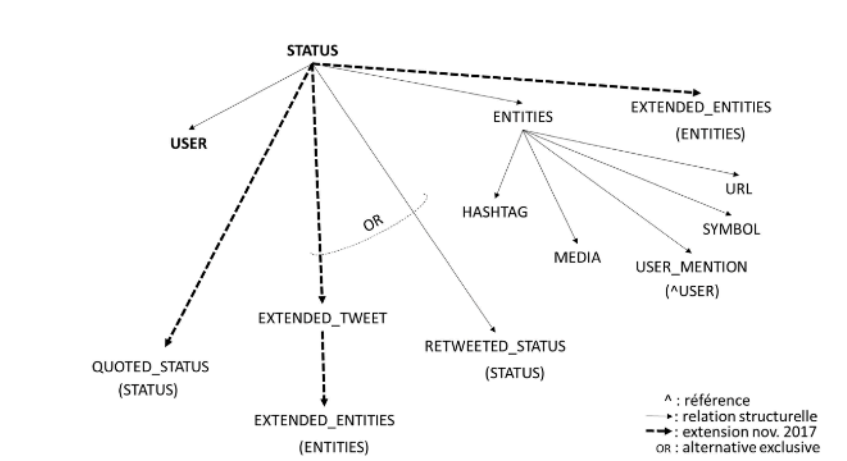
L'architecture d'une API tend à se standardiser pour faciliter le travail de développement. Elle reste malgré tout adaptable et sa structure est dépendante des objectifs et des fonctions de l'API. Par ailleurs, la concurrence importante des entreprises sur le web va définir leur accessibilité. La spécification OpenAPI¹² tend à devenir la norme, notamment pour les API RESTful. C'est ce modèle qui est mis en avant par les institutions culturelles valorisant leurs données par l'utilisation API. L'objectif est de mutualiser les données culturelles, en respectant des normes précises et en facilitant l'utilisation de leur API par des *guidelines*. Pour finir notre introduction sur les API et comprendre leur fonctionnement, nous allons prendre comme cas d'étude les API du réseau social Twitter.

¹² « Spécification OpenAPI - Version 3.0.3 | fanfaronnade ». https://swagger.io/specification/?utm_source=thenewstack&utm_medium=website&utm_content=inline-mention&utm_campaign=platform.

1.3 LA STRUCTURE DES API DE TWITTER

« L'instantanéité de ses publications fait de Twitter un outil privilégié d'alerte lors de la survenue d'événements, ou de signalement lorsque de nouvelles publications sont en ligne. Cette caractéristique se double de l'avantage qu'offrent les API de Twitter publiques d'accéder en masse aux représentations numériques très structurées de tweets qui concentrent pour chacune d'elles une information qu'il serait fastidieux de collecter depuis la page Web de comptes Twitter. »¹³

Twitter est l'exemple parfait des enjeux et des problématiques des API, et notamment sur les avantages et les inconvénients des API publiques. Pour le cas de Twitter, il ne faut pas parler singulièrement d'une API propre à ce réseau social, mais d'un « ensemble d'API » ayant des spécificités propre à chacune. Les données matérialisées dans l'objet numérique « tweet » s'organisent dans un schéma structurel précis :



Twitter met à disposition son API publiquement pour afficher des tweets sur son site web ou extraire des données (possibilités de sélectionner des tweets ayant le même hashtag par exemple). Ce n'est pas l'ensemble de ces API que Twitter offre généreusement. Tout d'abord, les données pouvant être collectées se limitent à 1 % du flux mondial des tweets. Ce taux ne peut pas être fixe et quantifié précisément étant dépendant de l'horaire et de la zone géographique lors de la réception des données. Les données transmissent exploitent ce que Jean-Marc Francony appelle « les relations éditoriales ». Grâce à cela, il est possible de mettre en évidence des publications communes, un flux de données identiques, ou des relations entre des acteurs. En mettant son API publique, comme

¹³ Francony, Jean-Marc. « L'éditorialisation des données aux bornes des API : enjeux et perspectives pour une analyse empirique ». *Les Enjeux de l'information et de la communication* 19/2, n° 2 (2018): 69-79. <https://doi.org/10.3917/enic.025.0069>.

a pu le faire Ebay, Twitter a permis à des développeurs de concevoir des API et d'extraire des données facilement. Répondant à des CGU précises et documentées¹⁴, le réseau social met à disposition des développeurs plusieurs groupes de données personnelles d'un compte ou ses messages privés. Comme dit précédemment, les tweets et réponses sont publics et peuvent publiés via l'API Twitter. Malgré de nombreux soucis de sécurité, avec la fuite de compte de personnalités publiques, ce modèle d'OpenAPI est un exemple dans le domaine.

Cette entreprise offrant un service totalement gratuit à ses utilisateurs, a réussi à mettre en place une économie autour de ses API. Nous la définissons comme API publique depuis le début de notre propos, cependant nous nous devons de contraster cette notion. Tout d'abord, pour avoir accès à l'API ouverte, il faut remplir plusieurs conditions : ouvrir un compte développeur Twitter, créer une clé API et autoriser l'application et le compte Twitter à communiquer. Cette procédure d'identification permet de contrôler les flux et les modifications de données. Par ailleurs, en 2017 Twitter a mit en place une API premium, bâtit sur un modèle économique capitalisme en opposition avec les attentes de l'*Opendata*¹⁵. Cette offre permet aux développeurs ou aux entreprises (plusieurs formules payantes) de rechercher dans les données de Twitter sur les 30 jours qui précèdent. Nous comprenons donc que la société, pionnière dans l'OpenAPI tend à commercialiser ses données. Il est important de préciser que les données exploitées par l'API payante ou l'API gratuite ne concernent pas les données personnelles des utilisateurs, mais les données publiques.

En conséquent, le cas d'étude de Twitter permet de comprendre la matérialité que peut avoir les données circulant par les API entre le client et le serveur. Toutes les API Twitter ne sont pas des API RESTful au vu de leurs fonctionnalités. Par ailleurs, la volonté de mettre en place une standardisation de l'OpenAPI semble acceptée par les entreprises, au regard des données qu'elles centralisent. Depuis 2006, Twitter a pu modifier sa politique de partage des données via les API, se situant aujourd'hui à une frontière que nous ne pouvons nommer. *A l'instar* de notre cas d'étude, Amazon, Google Maps ou encore Facebook publient leur API au début des années 2000 permettant la création d'écosystèmes autour de leurs données. Dans sa thèse publiée en 2017, Samuel Goyet¹⁶ définit les API comme « des points de passage névralgiques » des données. Ces carrefours

¹⁴ « À propos des API Twitter ». <https://help.twitter.com/fr/rules-and-policies/twitter-api>.

¹⁵ Dans la thèse de Samuel Goyet (voir note 16), il questionne les API Web comme un « capitalisme textuel contemporain? ».

¹⁶ Goyet, Samuel. « De briques et de blocs. La fonction éditoriale des interfaces de programmation (API) web: entre science combinatoire et industrie du texte. », s. d., 755.

numériques ont permis d'entrer pleinement dans le Web 2.0, avec l'abandon des pages web statiques servant de vitrines aux données. Les données sont aujourd'hui au centre de nombreux enjeux politiques ou encore économiques. Les humanités numériques vont devenir des acteurs du développement d'API en prônant majoritairement le partage des données et donc la constitution d'API publiques. Pour cela, il va falloir questionner l'utilité des API dans la valorisation des données patrimoniales. Se former et adapter ces interfaces aux données propres aux humanités numériques ayant des problématiques différentes, mais aussi des données culturelles sensibles.

DÉVELOPPEMENT

1. *Find your own place* : enjeux et problématiques des API en humanités numériques

« Plus largement, les API en tant qu'elles permettent la création de «petites formes» participent à un mouvement de transformation, d'hybridation d'objets culturels fondamentaux par le biais du numérique. »¹⁷

Samuel Goyet voit en l'objet API, « un lieu de définition de la culture numérique » aboutissant à la réutilisation et à la mutation des données. Les données patrimoniales recensées et organisées depuis les prémices d'Internet par les humanités numériques, sont elles aussi amenées à évoluer. Le Web statique a permis aux chercheurs en sciences humaines de promouvoir leurs recherches et d'exposer leurs données. La possibilité de pouvoir interroger ces données, les mettre au cœur d'un réseau international grâce aux fonctionnalités du web 2.0 ne peut pas se faire sans la mise en place des API. L'évolution des humanités numériques se fait en parallèle du développement web et n'est pas, outre mesure, en retard vis à vis des autres domaines scientifiques. Cependant, le développement d'API nécessite des connaissances informatiques répondant à des besoins précis.

L'objectif est de comprendre la mise en place des API au sein des humanités numériques. Nous soulèverons les premiers questionnements qui émergent de la communauté scientifique, mais également quels sont les enjeux apportés par l'implémentation des API pour les chercheurs. Enfin, nous montrerons ce que les API impliquent comme modifications structurelles pour l'organisation des données patrimoniales.

¹⁷ Goyet, Samuel. « De briques et de blocs. La fonction éditoriale des interfaces de programmation (API) web: entre science combinatoire et industrie du texte. », s. d., 755, p 41.

1.1 « DO APIS HAVE A PLACE IN THE DIGITAL HUMANITIES? »¹⁸

Dan Cohen, vice-recteur, doyen et professeur à la Northeastern University, questionne dès 2005 la place que pourrait prendre les API dans les humanités numériques. Nous avons vu précédemment, que les API étaient des interfaces d'échanges de données entre un client et un serveur. Le déploiement des données produites par les chercheurs sur le web amène de nouvelles problématiques. L'objectif n'est pas seulement de présenter les résultats des années de recherches, mais de permettre d'interroger les données et de les inscrire dans un réseau multidisciplinaire et international. Pour cela, les chercheurs en humanités numériques ont dû prendre en main les nouveaux outils numériques à leur disposition. L'impact des API locales, mais surtout les API distantes apparaissent rapidement comme l'outil parfait pour répondre à ces problématiques.

Avant de décrire les enjeux évidemment des API pour les humanités numériques, nous allons présenter les questionnements permettant la mise en place de ce processus de transformation. En partant de l'objectif principal selon lequel les API permettent d'interagir de multiples façons avec les données des chercheurs, et donc de rendre ces données publiques, cette technologie pourrait « dévaloriser le travail et la réflexion » des scientifiques. Depuis la création d'Internet, les chercheurs choisissent de partager leurs résultats de recherches librement. Cependant, la possibilité d'interagir avec ces données et de les réutiliser librement induit, hypothétiquement, la banalisation d'un travail de recherche considérable. Il serait alors envisageable qu'un chercheur refuse de publier ces données. La mise en place de données *Open source* n'est pas une évidence. En admettant que l'idée d'un réseau de données culturelles pleinement ouvert et interopérable soit partagée par l'ensemble de la communauté, rêve idyllique, comment financer un tel projet. Les financements publics ou privés dans les humanités numériques et plus largement dans les sciences humaines ne permettent pas le développement d'un outil numérique puissant comme les API. La plupart des API présentent proviennent des sciences dures ou des entreprises privées. Le coût de développement est pris en charge par des prestataires les vendant ou récupérant des bénéfices par la suite. Il paraît donc compliqué d'imaginer des chercheurs utiliser la quasi-totalité de leur budget dans une technologie, souvent inconnue pour eux dans la conception, réutiliser par la suite gratuitement par d'autres structures. En 2022, l'opportunité d'être pionnier dans la création d'API en humanités numériques serait une aubaine pour de nombreux chercheurs. Avec la temporalité de notre article de Dan Cohen, le recul n'est pas suffisant.

¹⁸ « Do APIs Have a Place in the Digital Humanities? – Dan Cohen ». Consulté le 6 septembre 2022. <https://dancohen.org/2005/11/21/do-apis-have-a-place-in-the-digital-humanities/>.

Nous avons vu précédemment que les API n'étaient pas uniformes dans leurs architectures et dans leurs disponibilités. Des entreprises peuvent se permettre de développer une multitude d'API afin de proposer des services divers à de potentiels clients. Dans le cadre institutionnel des humanités numériques, il faudrait réussir à produire rapidement un outil uniforme et compréhensible par tous. En omettant de notre réflexion la difficulté d'un consensus sur un tel outil, la formation du personnel à l'utilisation des langages de programmation complexe suffit à compromettre l'intégration des API en humanités numériques.

Après avoir dépeint un tableau bien sombre de l'avenir des API dans les humanités numériques, il semblerait évident que cet outil ne trouve pas sa place dans ce domaine de recherche. Cependant, leur utilité indispensable pour la mutation des données collectées par les chercheurs en sciences humaines a permis leur développement progressif. Il a donc fallu présenter aux chercheurs quelles étaient les enjeux pour leurs données d'utiliser des API. Cela demande également de former les chercheurs à l'utilisation des API déjà existantes sur le web. Nous allons donc maintenant expliquer l'intérêt concret des API dans les humanités numériques.

1.2 LES ENJEUX

L'interaction des données et leur réutilisation par le plus grand nombre est l'objectif premier de l'intégration des API dans les humanités numériques. Lorsque nous aborderons dans la suite de notre recherche des cas d'étude, nous montrerons comment la communauté scientifique met en place cela.

Concrètement, les chercheurs implémentent leurs données dans des bases de données pour un projet. Ces données doivent être publiques ou l'accès peut se faire par une identification. Une fois cette mise en ligne, l'API développée va permettre aux utilisateurs d'interagir avec ces données. Pour comprendre le fonctionnement d'une API fonctionnelle dans les humanités numériques, nous prenons comme exemple le projet *The Proceedings of the OLD BAILEY - London's Central Criminal Court, 1674 to 1913*¹⁹. Ce site internet met à disposition une collection numérisée de toutes les éditions des *Old Bailey Proceedings* de 1674 à 1913. Une base de données de plus de 197 000 procès est disponible dans un usage non commercial. Une fois numérisés, les textes ont été balisés en XML selon diverses catégories que nous n'exposerons pas ici. Le projet commencé en 2003 et actualisé jusqu'en 2018 selon les informations du site, a reçu une aide financière de nombreuses institutions. On peut tout d'abord relever le fait que ces données soient

¹⁹ « Old Bailey Online - Old Bailey API - Central Criminal Court ». <https://www.oldbaileyonline.org/static/API.jsp>.

entièrement accessibles pour tous. Elles s'inscrivent donc dans un réseau de données de recherche pertinent et répondant aux besoins techniques des humanités numériques (format de données adéquat, utilisation de langage de balisage adapté). Ces données peuvent être téléchargées dans leur ensemble sous format XML, mais il est possible pour l'utilisateur de travailler directement dans le texte. Ce travail de fouille est permis grâce à la création de l'API Old Bailey. Grâce aux balisages et aux requêtes, il est possible d'exporter les résultats.

Deux options sont possibles en fonction de votre profil utilisateur. Soit vous utilisez le « démonstrateur d'API Old Bailey » en sélectionnant des filtres multiples et ainsi faire apparaître des résultats souhaités. Soit, pour un profil développeur, une *guideline* explique entièrement comment fonctionne l'API et les comment faire des requêtes. Nous avons fait un test pour générer des fichiers à partir du démonstrateur. Il faut tout d'abord savoir qu'il y a deux contraintes, les données qui sont questionnées concernent les épreuves, l'exclusion des avant-projets, des publicités ou des *Comptes de L'ordinaire*. Par ailleurs, il n'est pas possible de télécharger plus de 10 essais complets avec une seule requête.

Résultat après avoir fait une requête

Une fois la recherche faite, un fichier zip est téléchargeable avec 10 fichiers XML sur notre demande. Cet outil précoce et fonctionnel est extrêmement documenté et montre tous les avantages à la mise en place d'une API. Nous avons vu l'importance d'avoir des données publiques permettant leur réutilisation grâce à la *Licence Creative commons*. Ce choix de libre publication des données est évidemment un choix politique, mais aussi dépendant de la nature des données traitées. L'accès

public des données des chercheurs au sein des humanités numériques, dans des formatsinteropérables permet de créer un réseau de données viables. Par ailleurs, ce projet est un exemple à suivre de documentation pour comprendre le fonctionnement de son API. Nous allons voir que le principal problème pour les chercheurs est de se former au développement d'API. La communauté en humanités numériques effectue une veille permanente sur l'évolution des nouvelles technologies et le moyen de les adapter à leurs données. Aujourd'hui, le regard se porte vers une valorisation du patrimoniale avec l'utilisation de technologies 3D, de réalité virtuelle ou d'intelligence artificielle. La médiation numérique de ces données a été permise grâce à la volonté des chercheurs de mettre à disposition les protocoles techniques de leurs outils.

La majorité des structures rendent disponibles leurs documentations sur leur API Web publique permettant aux autres chercheurs de se former et de comprendre comment fonctionne les interfaces. Nous avons vu cela avec le projet Old Bailey, avec des *guidelines* adaptées aux différents profils. Cependant, nous l'avons vu lors de notre travail de recherche sur le sujet, comprendre comment fonctionne les API semble évident dans la théorie, mais la prise en main n'est pas si simple.

1.3 L'IMPACT DES DONNÉES OUVERTES

Selon notre point de vue de recherche, l'impact des API dans les humanités numériques prend son sens si elles sont publiques et normalisées. L'objectif n'est pas de créer une API unique réutilisable, cela est impossible, mais de standardiser les données en humanités numériques grâce aux langages de balisage XML ou XML-TEI. Pour réussir cela, les projets de recherche ont mis en place des documentations complètes pour questionner leur API, mais aussi pour comprendre leur fonctionnement et les réutiliser. Cette vision *open source* sur les données et sur les technologies à utiliser est en rupture avec la vision capitaliste des données. Nous ne cherchons pas à comparer des données d'entreprises privées en les fustigeant de ne pas partager leurs compétences, mais les humanités numériques auraient pu développer une économie de la donnée culturelle. Nous avons présenté précédemment la difficulté d'obtenir des financements pour les sciences humaines et donc de se former aux nouvelles technologies. Dans ce cadre économique, les premières universités, notamment américaines, ont réussi à développer rapidement des API fonctionnelles pour interagir avec leurs données. Il n'aurait pas été étonnant de voir leurs technologies sécurisées pour être

commercialisé à d'autres universités. Cependant, les humanités numériques ont toujours eu comme objectifs de faciliter le travail des chercheurs dans un but « non lucratif ».

La politique d'accès ouvert vers les données patrimoniales est donc le facteur premier d'utilisation efficace des API en humanités numériques. Cela se retrouve également dans les stratégies des institutions publiques, musées ou bibliothèques, en diffusant numériquement leurs connaissances. Il a fallu plusieurs années pour voir apparaître de nouveaux métiers au sein des institutions publiques culturelles avec des compétences numériques développées.

« Le choix d'une politique d'ouverture des données fondée sur des API est une évolution récente dans les modalités d'accès aux ressources. La priorité donnée à l'accès à l'information sans prise en compte de sa conceptualisation correspond à une culture d'ingénierie informatique. »²⁰

Les modes d'accès aux informations grâce au API dans les institutions culturelles viennent de l'implication chercheurs en humanités numériques. Ces institutions, avec des financements plus importants que des laboratoires de recherches ont reconnu l'importance d'ouvrir leurs données via une API. Pour se former, les chercheurs vont accéder aux données en ligne présentées par des institutions et notamment aux documentations sur leurs API. Cependant, ce monde ouvert vers l'information culturelle possède ces limites. Dans le billet de blog paru en 2018, Clarisse Bardiot²¹ essaye de montrer les limites de cette technologie. L'objectif de son article n'est pas de remettre en cause l'intégration des API en humanités numériques, mais de faire comprendre que l'accessibilité des données par cette technologie à ces limites. Son propos présente les avantages des API pour aller questionner des données précises sans avoir à consulter une base de données entièrement, notamment grâce aux API ouvertes, mais également que cette technologie n'est pas facile à prendre en main pour les chercheurs. Le langage de programmation étant complexe et demandant une compréhension des interactions entre des applications précises, certaines données ne sont pas accessibles. Par ailleurs, les données dites ouvertes requièrent généralement des étapes d'identification et des demandes auprès des institutions pouvant ralentir le processus de recherches.

²⁰ Juanals, Brigitte, et Jean-Luc Minel. « Stratégies éditoriales des musées. Une approche de la médiation par l'accès ouvert aux données numérisées ». *Culture & Musées. Muséologie et recherches sur la culture*, n° 35 (1 juin 2020): 49-75. <https://doi.org/10.4000/culturemusees.4427>.

²¹ Bardiot, Clarisse. « Happy APIs : Débridons les APIS pour développer les humanités numériques ». Billet. *DORRA-DH* (blog). Consulté le 6 septembre 2022. <https://dorradh.hypotheses.org/66>.

Les documentations disponibles sont principalement tournées vers un public de développeurs et non de chercheurs en sciences humaines. Nous l'avons vu avec l'API Old Bailey fournissant les étapes précises à suivre pour utiliser son API, visant un public avec de lourdes connaissances en informatique. Avec cet exemple, nous essayons de montrer que les données ouvertes sont la solution à l'intégration systémique des API dans les humanités numériques. Cependant, elles doivent être adaptées au public qui va être amené à utiliser ces données. En reprenant la citation ci-dessus sur les *Stratégies éditoriales des musées [...] par l'accès aux données ouvertes numérisées*, nous observons que ces institutions font appel à des développeurs et non à des professionnels en humanités numériques. Par ailleurs, les données ouvertes ne le sont pas dans leur entièreté limitant la compréhension de celle-ci, empêchant « la construction d'un espace culturel intermuséal ».

La place des API dans les humanités numériques n'est donc plus à justifier. Lors de leur expansion dans le web, les chercheurs ont su comprendre l'intérêt de les utiliser pour créer un réseau de données fonctionnelles et interopérable. La prise de conscience de l'impact du numérique dans les sciences humaines amène les structures publiques à ouvrir leurs collections en libre accès en ligne. Cette récupération des connaissances numériques pour traiter des données culturelles et efficaces permet le développement de réseaux de recherches puissant européen et mondial. Pour répondre aux problématiques propres aux humanités numériques, les chercheurs et les développeurs (peuvent être la même personne) ont adapté la technologie des API aux besoins spécifiques. Sans perdre de vue l'objectif premier d'un ensemble de données *open source*, les universités vont créer un protocole performant pour la visualisation d'images. Lors de la création des humanités numériques, les principales données étaient textuelles, cependant les ressources ont évolué ainsi que les problématiques. Les interactions avec les données doivent donc s'adapter et de nouveaux protocoles sont développés par les institutions de recherches. L'enjeu majeur dans le traitement des données est la possibilité d'interagir avec les médias. Une représentation graphique ne vient plus seulement étayer un propos, mais devient le cœur du corpus de données. L'intégration d'API focalisé sur l'interaction avec les médias apparaît et semble être la première véritable standardisation d'une technologie adaptée aux sciences humaines.

2. Le protocole IIIF

« *Le cadre normatif de IIIF définit des interfaces de programmation applicative (API) communes qui fournissent une méthode standardisée de description et d'accès aux images sur le Web, ainsi que l'exposition de métadonnées descriptives et structurelles nécessaires à la présentation d'un document numérique dans une interface.* »²²

À partir des années 2010, les API ont le monopole dans l'interaction des données en ligne produites dans les humanités numériques. En réponse à ce développement, les universités et les centres de recherches ont développé en communauté l'initiative *International Image Interoperability Framework* (IIIF). Comme cité précédemment, les objectifs de cette communauté consistent à normaliser des méthodes de description d'images, dans un premier temps, grâce aux API. Ce consortium, regroupant des spécifications techniques est aujourd'hui l'une des API les plus utilisés dans les humanités numériques. Elle a réussi à créer un ensemble standardisé de données *open source* réutilisables par la communauté de chercheurs. Son utilisation dans les plus grandes institutions publiques mondiales, a permis d'interagir avec les données images en se spécifiant au fur et à mesure des années.

Nous allons expliquer le processus de création permis par ce protocole innovant pour la gestion des nouvelles données grâce à l'API Image. Après la décomposition du processus, nous entrerons en détails sur les interactions possibles sur ces données. Cela nous amène à présenter l'utilisation de ce protocole dans le projet de recherche *Righettino* développé au sein du laboratoire Ouvroir à l'Université de Montréal.

2.1 LA COMMUNAUTÉ IIIF

IIIF, apparu en 2011 est une communauté et un ensemble de spécifications techniques permettant l'intégration et l'interaction avec des images sur le web. Elle est le résultat d'une

²² « International Image Interoperability Framework ». In *Wikipédia*, 3 novembre 2020. https://fr.wikipedia.org/w/index.php?title=International_Image_Interoperability_Framework&oldid=176197307.

collaboration étroite entre l'université de Stanford²³ et les plus grandes institutions publiques de nombreux pays (Bibliothèque National de France, British Library, ou encore la Bibliothèque Nationale de Norvège). Les trois axes définissant ce protocole d'interaction avec les images sont la mise à disposition des données libre accès (*open source*), la création d'un protocole standardisé répondant aux besoins des utilisateurs, et l'inscription de cette technologie dans un réseau international. Il ne faut donc pas imaginer IIIF comme une entité technique numérique évoluant indépendamment en fonction des besoins, mais comme un réseau de travail collaboratif.

« C'est un effort de base entre de nombreuses institutions différentes pour résoudre leurs problèmes communs de livraison, de gestion, de partage et de travail avec leurs ressources. »²⁴

Il n'est pas question ici de créer l'API répondant à tous les besoins, mais de se concentrer sur l'objet image. La numérisation grandissante des archives papiers et des images dans le but de les rendre disponibles en ligne a induit l'adaptation des visionneuses entre les données. Les utilisateurs vont donc être limités à des usages définis par les structures mettant à disposition leurs données et leurs moyens technologiques. Il a fallu adapter les protocoles informatiques à la réception d'image de tailles différentes et en haute résolution. Les premières fonctionnalités se sont concentrées sur le déplacement de l'utilisateur dans l'image et permettre un zoom rapide sur les médias publiés. L'exemple le plus parlant est la navigation dans l'image *Omi Kuni-ezu* (345 x 504 cm) photographiée par Wayne Vanderkuil. L'API Image va donc permettre une liberté de circulation dans la ressource visuelle. Les techniques de balisages des documents textuelles permettaient depuis longtemps de travailler sur chaque élément de ces données. Cette nouveauté a permis d'élargir le travail d'études sur les images en les présentant comme des sources aussi importantes que les ressources textuelles.

Dans l'idée de créer une technologie standardisée dans l'affichage des images, IIIF initie les visionneuses interopérables. Cela va permettre aux bibliothèques, mais aussi aux chercheurs voulant réutiliser la visualisation de ces images, d'adapter leurs visionneuses à leurs besoins. L'utilisateur peut donc utiliser la visionneuse *OpenSea Dragon* ou *Universal Viewer* pour afficher l'image et ses métadonnées, mais il peut également se servir de *Mirador* pour annoter l'image. La partie suivante

²³ Documentation sur API IIIF de l'université de Stanford : « IIIF · API Documentation ». <https://example.com/>.

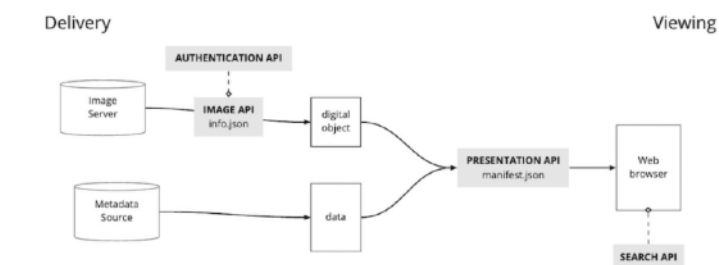
²⁴ Google My Maps. « IIIF Implementations and Consortium Map ». https://www.google.com/maps/d/viewer?mid=1faJRKJpj2Vau__RDwt8af040x0GTVOzp.

est consacrée à la présentation en détails des fonctionnalités disponibles dans les visionneuses. De plus, en rendant les images disponibles pour tous, elles permettent la récupération des données visuelles (et des métadonnées) par les institutions en humanités numériques pour construire des bases de données complètes. Le travail de recherche impliquant de nombreux déplacements à travers le monde semble bien loin grâce à ce protocole. Pour réussir ce travail d'*open data* standardisé IIIF globalise deux API structurelles : l'API Image et l'API Présentation. L'API image permet d'obtenir des fonctionnalités visuelles de déplacement et de zoom dans l'image. L'API Présentation vient compléter ces interactions en permettant d'ajouter des annotations dans l'image, de créer des collections navigables entre plusieurs données.

La création de cette communauté autour d'une API interopérable, *open source* et standardisé a permis aux structures culturelles d'adapter la présentation de leurs données. Il a également prouvé que les humanités numériques ont su développer les technologies du web en fonction de leurs besoins. Il est donc important préciser quelles sont ses besoins et comment l'API Image permet d'y répondre.

2.2 LES FONCTIONNALITÉS DE L'API IMAGE

Notre propos n'aborde pas toutes les fonctionnalités disponibles par les API développées par IIIF. L'intérêt suscité par le développement de l'API Image du consortium IIIF nous permet suffisamment de comprendre pourquoi les API sont devenues aussi prédominantes dans les humanités numériques.



Un diagramme de la façon dont les principales API IIIF fonctionnent ensemble.

Les navigateurs web comprennent aisément l’affichage d’images dans des formats standards comme .jpg et .mp4 dans des tailles définies. En s’alignant sur les normes de Web, IIIF a pu apporter de nouvelles fonctionnalités, comme vu précédemment, dans l’interaction avec l’image. L’API d’image IIIF renvoie une réponse à une requête HTTP ou HTTPS standard. L’URI (*Uniform Resource Identifier*), permettant d’identifier une ressource, peut être modifié pour spécifier la taille, la région, la rotation et les caractéristiques de qualité et du format de l’image. En créant un manifeste autour de l’image on peut interagir dedans grâce à un langage simplifié, JSON-LD rendant la lecture des données plus facile. L’API Image n’impose pas de restrictions sur la composition des identifiants qu’un serveur peut utiliser.

```
{scheme}://{server}
{/prefix}/{identifiant}/{region}/{size}/{rotation}/{quality}.{format}
```

```
https://example.org/image-service/abcd1234/full/max/0/default.jpg
```

Le travail sur les zones internes aux images permet de travailler en profondeur sur cette donnée comme on le fait dans un corpus textuel. Reprenons l’exemple de l’API Old Bailey vu précédemment. Le texte ayant été balisé en XML, à la suite de notre requête l’API, nous donne des résultats textuels comportant nos mots-clés. Le fonctionnement est identique avec l’API Image. En définissant des zones de zoom, la taille de l’image et même la rotation on va pouvoir naviguer à l’intérieur. L’API de présentation va permettre de compléter ces fonctionnalités en annotant des zones précises de la donnée.

Les visionneuses IIIF vont ensuite permettre d’adapter les images selon les besoins des utilisateurs. Plusieurs images vont pouvoir être comparées grâce à *Mirador*²⁵, ou de combiner plusieurs numérisations pour créer un corpus et naviguer dedans librement. Le format de programmation des images et réponses entre client et serveur géré par l’API va suivre un protocole précis. Le protocole est pensé pour deux utilisations distinctes des données images. Tout d’abord, la médiation autour de données ouvertes présentées par les chercheurs. Par les visionneuses et les

interactions avec les images, les utilisateurs, différents des chercheurs en humanités numériques peuvent apprécier une œuvre dans son ensemble. De plus, les chercheurs en humanités numériques peuvent réutiliser ces fonctionnalités pour agrémenter leurs ressources numériques. En s'inscrivant dans la communauté IIIF, le réseau s'agrandit et se standardise. C'est grâce à ce travail de documentation et de standardisation des données que les laboratoires de recherche font évoluer leur projet.

Pour agrémenter notre propos sur les API en humanités numériques et plus particulièrement sur l'API Image, nous allons expliquer le processus de création d'une API IIIF au sein d'un projet de recherche en histoire de l'art. Nous ne pourrions pas aborder ce propos sans utiliser l'API Presentation, mais nous montrerons comment IIIF a permis de mettre en place ce travail.

2.3 PROJET DE RECHERCHE RIGHETTINO

L'Ouvroir, laboratoire de recherche en histoire de l'art et muséologie numérique de l'université de Montréal a décidé d'utiliser l'API IIIF pour l'un de ces projets. Le projet Righettino dirigé par Emmanuel Château-Dutier enseignant chercheur au sein de l'université de Montréal confronte les difficultés d'interactions entre une image de la renaissance et une notice textuelle. L'incapacité du format textuelle pour comprendre la construction du dessin d'une vue de Turin de 1583 réalisé par le chanoine Girolamo Righettino au regard d'une notice construite par l'enseignant chercheur Denis Ribouillault, impose une réflexion sur l'utilisation du numérique. La création d'un manifeste reprenant les API IIIF semble tout trouver et questionne sur les interactions possibles dans cette image. Pour ce projet, deux sources de données sont présentes : la notice textuelle provenant d'un travail de recherche sur un manuscrit produit par Righettino sur une vue de Rome (perdu aujourd'hui), et le dessin de la vue de Turin. L'objectif est de pouvoir articuler ces deux formats de données pour comprendre comment l'autre a construit cette oeuvre. L'API Image va permettre de définir le format de l'image, sa résolution et les déplacements possibles. Grâce à l'API Presentation, des zones sont définies et interagissent avec les données textuelles expliquant ces représentations.

La pensée en humanités numériques est donc la suivante. Nous partons d'un processus de recherche sur un corpus varié. Une fois les données définies, nous recherchons à produire la technologie adéquate pour partager ce travail et distribuer librement les données. Sans le protocole IIIF, l'interaction entre ces deux ressources n'aurait pas été possible. Ainsi, le prototype développé à

l'été 2022 permet à l'utilisateur de naviguer dans le texte ou dans l'image et les données suivent un protocoleinteropérable. Pour réussir ce travail, la documentation *open source* IIF est une ressource complète et fiable.

Notre questionnement de départ a pour objectif de présenter les réels enjeux des API dans les humanités numériques. Nous avons vu durant notre propos que cette technologie largement utilisée dans le développement web, questionne les interactions possibles avec les données et leur libre accès. La création d'une technologie IIF et d'une communauté pour répondre à une problématique précise confirme l'efficacité des API dans le développement des recherches en humanités numériques. L'API Images a également permis de valoriser les collections et de créer des parcours de collections au sein des institutions publiques culturelles. Le numérique a favorisé le partage des informations au plus grand monde en forçant les structures, notamment culturelles, à repenser la présentation de leurs données. Pour comprendre l'importance de l'API Image dans le monde culturel, nous avons pris deux institutions utilisant ce protocole.

3. Cas d'études d'utilisation des API Images

La création d'une communauté internationale autour du consortium IIF résulte de l'implication des institutions culturelles dans le développement de cette technologie. La bibliothèque nationale de France (BnF) a été un acteur majeur de ce travail en repensant la structuration de ces données. L'utilisation des API IIF et notamment de l'API Image s'est accompagné d'un travail documentaire important. Les données ouvertes et les documentations expliquant comment les manipuler sont le rouage indispensable au bon fonctionnement de l'intégration des API dans les humanités numériques. Pour terminer notre travail de recherche, nous avons choisi deux cas d'études utilisant à grande échelle l'API Image pour leurs données. Tout d'abord, la BnF documentant toutes les API utilisées par l'institution et les données présentent. Puis, pour s'insérer dans un réseau international, le cas d'Europeana.

3.1 BNF - GALLICA

« Depuis une vingtaine d'années, l'avènement des humanités numériques bouleverse la recherche en sciences humaines et sociales (SHS) et interroge le rôle des bibliothèques. [...] De l'autre côté, les bibliothèques, qui sont au cœur de l'activité de recherche, se trouvent interrogées

*sur l'évolution de leurs services avec le développement de l'Internet et l'omniprésence du Web dans les pratiques documentaires. »*²⁶

Les humanités numériques ont été vectrices des modifications structurelles pour les institutions publiques. La demande fréquente des corps de recherches pour accéder aux données culturelles qu'elles conservent obligent ces structures à faciliter l'accès de leurs données. Pour réussir ce travail, la BnF s'inscrit dans un « programme à dimension internationale ». Les données ouvertes doivent être interopérables et réutilisables dans une même interface. Le consortium IIF permet de développer les connaissances techniques viables pour réussir ce virage numérique important. Gallica, la bibliothèque numérique de la BnF est l'exemple parfait du développement numérique dans les institutions. Les données de Gallica représentent 7 millions de documents et plus de 1000 millions d'images. C'est grâce à IIF qu'autant d'images ont pu être numérisées et utilisables. Toutes ces données sont entièrement interopérables et leur manipulation est « indépendante de la localisation physique et des établissements qui les hébergent ». Il faut évoquer le fait que les API utilisées par la BnF, et pas seulement les API IIF sont entièrement ouvertes et mises à disposition des institutions. Leur fonctionnement est disponible dans une documentation détaillée²⁷. La stratégie d'ouverture des données de la BnF est développée depuis 2014, peu de temps après la création du consortium IIF. Les fonctionnalités disponibles sur les données de Gallica répondent à des usages définis par l'institution : l'extraction de données et les interroger. L'utilisation de la visionneuse *Mirador* va permettre à l'utilisateur de comparer des corpus hébergés dans des lieux différents. Nous avons donc un travail facilité pour la présentation des données, leurs interactions et leur réutilisation. Ces trois processus sont puissants et entièrement fonctionnels sur les corpus présentés dans Gallica. L'autre objectif qui résulte de ce protocole, que nous avons peu abordé pour l'instant, concerne la médiation du contenu culturel. La médiation numérique²⁸ va être la finalité de l'utilisation des API dans les humanités numériques. Rendre disponibles ses travaux, ses données, pouvoir les réutiliser au sein de sa recherche a été permis par les API. Les API Images viennent compléter ce travail de diffusion et vont progressivement ouvrir la documentation

²⁶ Carlin, Marie, et Arnaud Laborderie. « Le BnF DataLab, un service aux chercheurs en humanités numériques ». *Humanités numériques*, n° 4 (1 décembre 2021). <https://doi.org/10.4000/revuehn.2684>.

²⁷ « Accueil | Api ». Consulté le 6 septembre 2022. <https://api.bnf.fr/>.

²⁸ Dailymotion. « Les API ouvertes de la BnF - Vidéo Dailymotion », 1 juin 2021. <https://www.dailymotion.com/video/x81o823>.

culturelle à tous les publics. Il faut donc adapter les données et la navigation dans l'image par L'API Image facilite cet usage. Nous allons retrouver cette réflexion autour de la médiation culturelle numérique dans la fin de notre travail de recherche.

Nous comprenons donc que le protocole IIIF et notamment l'API Image qui a été développé est une innovation technologique primordiale pour la diffusion des données. Les humanités numériques n'ont plus comme seul public les chercheurs de ce domaine, mais la donnée culturelle devient un outil d'apprentissage. Facilite l'observation et l'utilisation de ces données pour tous les publics interroge de nouveaux enjeux structurels. Par ailleurs, la BnF s'inscrit dans une politique européenne de numérisations des ressources. La plateforme Europeana²⁹ lancée en 2008 donne accès numériquement aux données des institutions culturelles de l'Union européenne. En 2020, la plateforme compte plus de 58 millions d'objets numériques. Pour la diffusion de ses données et leur interopérabilité, la plateforme utilise les API IIIF.

3.2 EUROPEANA

*« Europeana offre aux amateurs de patrimoine culturel, aux professionnels, aux enseignants et aux chercheurs un accès numérique aux ressources du patrimoine culturel européen. »*³⁰

Dans cette description proposée par Europeana sur leur site internet, nous observons une ouverture des données patrimoniales à tous les publics. L'accessibilité des données doit donc répondre à de nouveaux enjeux de diffusion et utilisable facilement sans un apport de compétences numériques important. L'objectif de la plateforme est de centraliser les données des institutions du patrimoine culturel (ICH) en respectant le protocole IIIF. Les ressources collectées obtiennent une meilleure visualisation et sont accessibles grâce aux API. L'implication de ce réseau international va contribuer au développement du protocole IIIF, notamment au Etats-Unis et la standardisé des formats de données³¹. Grâce à cette structure, les institutions ont pu favoriser le développement de logiciels *Open Source* en utilisant les API IIIF.

²⁹ « À propos d'Europeana ». <https://www.europeana.eu/fr/about-us>.

³⁰ *Ibid.*

³¹ Europeana Pro. « IIIF datasets in Europeana: A scholar's delight ». <https://pro.europeana.eu/post/iiif-datasets-in-europeana-a-scholar-s-delight>.

Pour répondre aux besoins des publics, la plateforme est une API unique utilisant le protocole IIIF. L'API European IIIF répond aux standards imposés par le protocole en améliorant l'utilisation. L'élément commun à ces deux cas d'études, et évidemment l'utilisation et la participation active au consortium IIIF, mais également de documenter finement l'utilisation API utilisés.

La structuration des données patrimoniales au sein des institutions publiques s'est transformée grâce à l'implication des groupes de pensées des humanités numériques autour du protocole IIIF. La standardisation des données utilisées et transférées par les API au sein des réseaux *open source*, permet aux humanités numériques de devenir une science d'utilité publique. Par cette notion, nous sous-entendons que l'intérêt des chercheurs en sciences humaines de partager et d'utiliser les données dans ce domaine impact les outils de médiation des institutions patrimoniales publiques. L'accès aux données culturelles ne devrait pas être réservé aux élites intellectuelles, mais pouvoir se diffuser dans les publics. Cependant, elle induit la mise en place de protocoles d'utilisation précis des nouvelles technologies et l'adaptabilité de ces dernières aux données traitées. Souvent en retard par rapport aux sciences brutes, les humanités numériques ont su s'approprier les API pour générer un outil répondant à leur besoin. Cet écosystème patrimoniale(chercheurs en universités, institutions publiques, enseignants,...) est devenu un acteur actif de l'évolution numérique. L'objectif final des données est d'être assimilé par le plus grand nombre, alors un travail de médiation numérique est mis en place, sur les sites internet, mais aussi dans les lieux culturels pour faciliter la compréhension de ces données. L'évolution des pratiques numériques dans les musées est le résultat de réflexions et de productions documentaires sur l'utilité des protocoles informatiques pour les données patrimoniales. La dernière partie de notre recherche va donc présenter comment les institutions publiques culturelles, notamment muséales, ont réussi à se former face à ce changement de paradigme. Le musée n'est plus seulement un espace de conservation et de préservation de l'information, mais également un espace de diffusion mouvant.

OUVERTURE

La valorisation des données patrimoniales au sein des institutions publiques

« Il ne s'agit donc pas tant d'évoquer ici les catalogues de musées en ligne, les applications pour terminaux mobiles, ou les déploiements d'expositions virtuelles, mais plutôt des formes plus spécifiques à l'informatique comme la mise à disposition de jeux de données, ou la création d'interfaces programmables (API) qui peuvent directement intéresser les historiens d'art. L'aménagement de ce type d'accès oblige en quelque sorte à penser aujourd'hui le musée comme un service d'information. »³²

L'accès aux données patrimoniales dans les musées a longtemps était réservé à un public privilégié. Les données accessibles par la matérialité des expositions ne pouvaient plus suffire à la conception de cette institution patrimoniale. Lieu de conservation et d'exposition, l'institution muséale se devait de partager son information brute. L'un des musées pionnier dans cette ouverture des données en accès libre grâce au numérique est le *British Museum*. Le travail de numérisation important de leur collection et la diffusion de ces informations en ligne ont permis d'ouvrir la voie vers une « interface muséales » de la donnée ouverte. Les musées ont mis en place une nouvelle « approche de la médiation culturelle par l'accès ouvert aux données numérisées »³³. La médiation culturelle n'est pas seulement un concept destiné à un public néophyte face aux données patrimoniales. La production documentaire sur l'utilisation des API pour l'utilisation des données en libre accès, est un service de médiation culturelle nouveau. Les musées deviennent donc des acteurs principaux de la diffusion des ressources patrimoniales numérisées permettant la production de

³² Chateau-Dutier, Emmanuel. « Le musée comme service d'information. Pour une politique des interfaces muséales ». In *DHNord 2019 : Corpus et archives numériques*. Lille, France: Maison Européenne des Sciences de l'Homme et de la Société (MESHs), 2019. <https://hal.archives-ouvertes.fr/hal-03351734>.

³³ Juanals, Brigitte, et Jean-Luc Minel. « Stratégies éditoriales des musées. Une approche de la médiation par l'accès ouvert aux données numérisées ». *Culture & Musées. Muséologie et recherches sur la culture*, n° 35 (1 juin 2020): 49-75. <https://doi.org/10.4000/culturemusees.4427>.

recherches en humanités numériques. Cela induit donc la création d'un discours patrimoniales permettant de comprendre les données conservées.

La vision de *l'open data* n'est pas neutre d'enjeux politiques et économiques. Les musées doivent donc composer avec les nouvelles normes concernant les droits d'auteur et la diffusion des données en ligne. Par ailleurs, la donnée a toujours été sujette à l'évolution de la pensée politique sur la libre circulation des idées et l'accès pour tous à ce savoir. Levier de pression culturelle et d'identification nationale, les institutions culturelles avancent progressivement dans le partage de leur collection.

La valorisation patrimoniale est un enjeu de compréhension de la donnée et les outils numériques cherchent à adapter leurs protocoles pour répondre au mieux à ces besoins.

CONCLUSION

Les humanités numériques développées depuis les prémices de l'informatique semblent prendre un tournant important dans l'utilisation des données qu'elles produisent. Encore peu développées dans les universités, elles sont pourtant présentes dès les origines du numérique. Ce questionnant régulièrement sur l'utilisation des nouvelles technologies pour faciliter la production de recherches dans leur domaine, elles ont su s'approprier les outils numériques pour répondre à leurs besoins. Les interfaces de programmation d'application (API), ont permis l'avènement du web 2.0 et l'accessibilité à l'information au plus grand nombre. Les données produites dans les humanités numériques ne sont pas restées statiques face à ce changement et la création de communautés ayant les mêmes besoins permettent de diffuser largement les données patrimoniales. L'apparition du numérique dans les institutions patrimoniales est le résultat du pouvoir de la donnée *open source*.

L'évolution constante du protocole IIIF et les financements grandissants dans les laboratoires de recherches en humanités numériques encouragent le développement du travail collaboratif. La recherche personnelle d'un chercheur sur un sujet donnée et sa diffusion au plus grand nombre permet la propagation du savoir culturel commun. L'objectif de notre recherche a été de montrer comment les humanités numériques se sont appropriés un outil technologique et la place de cet outil dans le traitement des données. Le concept guidant notre réflexion de pensée autour de ce questionnement a été de montrer qu'une standardisation des protocoles informatiques et la pensée de l'information en Web des données, permet le développement d'un savoir universel et accessible.

BIBLIOGRAPHIE / SITOGRAPHIE

Sitographie

« À propos des API Twitter ». <https://help.twitter.com/fr/rules-and-policies/twitter-api>.

« À propos d'Europeana ». <https://www.europeana.eu/fr/about-us>.

« Accueil | Api ». <https://api.bnf.fr/>.

« API ». In *Wikipédia*, <https://en.wikipedia.org/w/index.php?title=API&oldid=1110431511#Usage>.

APIs, Nordic. « The Bezos API Mandate: Amazon's Manifesto For Externalization | Nordic APIs | ». Nordic APIs, 19 janvier 2021. <https://nordicapis.com/the-bezos-api-mandate-amazons-manifesto-for-externalization/>.

« Architectural Styles and the Design of Network-based Software Architectures ». <https://www.ics.uci.edu/~fielding/pubs/dissertation/top.htm>.

Bardiot, Clarisse. « Happy APIs : Débridons les APIS pour développer les humanités numériques ». Billet. *DORRA-DH* (blog). <https://dorradh.hypotheses.org/66>.

Digital Humanities at MIT. « Digital Humanities at MIT ». <https://digitalhumanities.mit.edu/>.

« Do APIs Have a Place in the Digital Humanities? – Dan Cohen ». <https://dancohen.org/2005/11/21/do-apis-have-a-place-in-the-digital-humanities/>.

« FromThePage - transcribe handwritten documents ». <https://fromthepage.com/>.

« Get Started ». <https://developer.paypal.com/api/nvp-soap/get-started/>.

« Getting Data for Digital Humanities with APIs: A Gentle Introduction – Digital Humanities @ Pratt School of Information ». <https://studentwork.prattsi.org/dh/2019/05/13/getting-data-for-digital-humanities-with-apis/>.

Hall, Mark. « Introduction to APIs ». *DARIAH-Campus*, 12 octobre 2021. <https://campus.dariah.eu/en/resource/posts/introduction-to-apis>.

« Historique des API ». <https://history.apievangelist.com/>.

« IIIF · API Documentation ». <https://example.com/>.

Europeana Pro. « IIIF datasets in Europeana: A scholar's delight ». <https://pro.europeana.eu/post/iiif-datasets-in-europeana-a-scholar-s-delight>.

Google My Maps. « IIIF Implementations and Consortium Map ». https://www.google.com/maps/d/viewer?mid=1faJRKJpj2Vau__RDwt8af040x0GTVozp.

« International Image Interoperability Framework ». In *Wikipédia*, 3 novembre 2020. https://fr.wikipedia.org/w/index.php?title=International_Image_Interoperability_Framework&oldid=176197307.

Dailymotion. « Les API ouvertes de la BnF - Vidéo Dailymotion », 1 juin 2021. <https://www.dailymotion.com/video/x81o823>.

« Mirador — Home ». <https://projectmirador.org/>.

Naeem, Tehreem. « Un guide sur la façon d'utiliser les API pour la récupération de données ». *Astera* (blog), 27 avril 2020. <https://www.astera.com/fr/type/Blog/comment-utiliser-api-pour-la-r%C3%A9cup%C3%A9ration-de-donn%C3%A9es/>.

« Old Bailey Online - Old Bailey API - Central Criminal Court ». <https://www.oldbaileyonline.org/static/API.jsp>.

BnF - Site institutionnel. « Portail BnF API et jeux de données ». <https://www.bnf.fr/fr/portail-bnf-api-et-jeux-de-donnees>.

« Qu'est-ce qu'une API ? » <https://www.redhat.com/fr/topics/api/what-are-application-programming-interfaces>.

« REST et SOAP : quelle est la différence ? » <https://www.redhat.com/fr/topics/integration/whats-the-difference-between-soap-rest>.

Rittenhouse, Brad, Ximin Mi, et Courtney Allen. « Beginner's Guide to Twitter Data ». *Programming Historian*, 16 octobre 2019. <https://programminghistorian.org/en/lessons/beginners-guide-to-twitter-data>.

Sinatra, Michael E., et Marcello Vitali Rosati. « Histoire des humanités numériques ». In *Pratiques de l'édition numérique*. Les Ateliers de [sens public], 2014. <http://www.parcoursnumeriques-pum.ca/1-pratiques/> <https://www.parcoursnumeriques-pum.ca/1-pratiques/chapitre3.html>.

« Spécification OpenAPI - Version 3.0.3 | fanfaronnade ». https://swagger.io/specification/?utm_source=thenewstack&utm_medium=website&utm_content=inline-mention&utm_campaign=platform.

Bibliographie

Carlin, Marie, et Arnaud Laborderie. « Le BnF DataLab, un service aux chercheurs en humanités numériques ». *Humanités numériques*, n° 4 (1 décembre 2021). <https://doi.org/10.4000/revuehn.2684>.

Chateau-Dutier, Emmanuel. « Le musée comme service d'information. Pour une politique des interfaces muséales ». In *DHNord 2019 : Corpus et archives numériques*. Lille, France: Maison Européenne des Sciences de l'Homme et de la Société (MESHS), 2019. <https://hal.archives-ouvertes.fr/hal-03351734>.

Citton, Yves. « Manifeste pour des humanités numériques 2.0 ». Traduit par Quentin Julien-Saavedra. *Multitudes* 59, n° 2 (2015): 181-95. <https://doi.org/10.3917/mult.059.0181>.

Francony, Jean-Marc. « L'éditorialisation des données aux bornes des API : enjeux et perspectives pour une analyse empirique ». *Les Enjeux de l'information et de la communication* 19/2, n° 2 (2018): 69-79. <https://doi.org/10.3917/enic.025.0069>.

Goyet, Samuel. « De briques et de blocs. La fonction éditoriale des interfaces de programmation (API) web: entre science combinatoire et industrie du texte. », s. d., 755.

Hamilton, Gill, et Fred Saunderson. *Open Licensing for Cultural Heritage*. Facet Publishing, 2017.

Iliadis, Andrew, et Federica Russo. « Critical Data Studies: An Introduction ». *Big Data & Society* 3, n° 2 (1 décembre 2016): 2053951716674238. <https://doi.org/10.1177/2053951716674238>.

Juanals, Brigitte, et Jean-Luc Minel. « Stratégies éditoriales des musées. Une approche de la médiation par l'accès ouvert aux données numérisées ». *Culture & Musées. Muséologie et recherches sur la culture*, n° 35 (1 juin 2020): 49-75. <https://doi.org/10.4000/culturemusees.4427>.

Tasovac, Toma, Adrien Barbaresi, Thibault Clérice, Jennifer Edmond, Natalia Ermolaev, Vicky Garnett, et Clifford Wulfman. « APIs in Digital Humanities: The Infrastructural Turn ». In *Digital Humanities 2016*, 93-96. Digital Humanities 2016 Conference Abstracts. Cracovie, Poland, 2016. <https://hal.archives-ouvertes.fr/hal-01348706>.

Weinard, Chad, et Anne Luther. « Cultural Collections as Data: Aiming for Digital Data Literacy and Tool Development ». MW19 | Boston, 1 janvier 2019. https://www.academia.edu/38278177/Cultural_Collections_as_Data_Aiming_for_Digital_Data_Literacy_and_Tool_Development.