

ÉCOLE NATIONALE DES CHARTES

Clara GROMETTO

licenciée ès lettres

Le partage des outils de la recherche

**Élaboration d'une plateforme extensive
pour le traitement des données visuelles**

Mémoire pour le diplôme de master
« Technologies numériques appliquées à l'histoire »

2024

Résumé

Ce mémoire porte sur l'intégration des outils numériques dans la recherche en SHS (*Sciences Humaines et Sociales*), notamment l'utilisation de la *computer vision* pour l'enrichissement et la sémantification des sources historiques. Le projet central étudié est EIDA (*Editing and analysing hIstorical astronomical Diagrams with Artificial intelligence*), porteur du développement de la plateforme AIKON, cette dernière mettant à disposition des outils de *deep learning* pour l'extraction et l'analyse des diagrammes astronomiques de tradition ptoléméenne.

Le mémoire pose la question du niveau de spécificité ou de généralité à prévoir dans le développement des outils numériques : comment construire une chaîne de traitement à la fois flexible, adaptée à des données variées, et capable de répondre à des besoins spécifiques ? Via cette question, ce travail explore les opportunités liés à l'intégration des technologies numériques dans la recherche en SHS, et relève les défis tenant au partage des outils, impactant le partage des pratiques et des méthodes.

Mots-clés : histoire des science ; humanités numérique ; diagrammes astronomiques ; IA (*Intelligence Artificielle*) ; *deep learning* ; Python ; modularité ; standardisation technique ; interopérabilité.

Informations bibliographiques : Clara GROMETTO, *La standardisation des outils numériques dans la recherche, Élaboration d'une plateforme extensive pour le traitement des données visuelles basé sur la vision artificielle*, mémoire de master « Technologies numériques appliquées à l'histoire », dir. Ségolène Albouy, École nationale des chartes, 2024.

Remerciements

Avant tout, je tiens à exprimer ma profonde gratitude à toutes les personnes ayant contribué à faire de ce stage une expérience profondément enrichissante sur le plan intellectuel comme personnel.

Avant tout, je tiens à remercier ma directrice de mémoire Ségolène Albouy et ma tutrice Jade Norindr pour leur soutien, leurs conseils avisés, leur patience, leur disponibilité et leur générosité. Je me sens extrêmement chanceuse d'avoir pu profiter de leur expertise et découvrir leurs qualités humaines, elles forcent l'admiration. Un grand merci à Matthieu Husson ainsi qu'à tous les chercheur.ses d'EIDA : Scott, Divna, Samuel, Chen... Merci à Éleonora pour son délicieux cours de philologie. J'ai trouvé à l'Observatoire un environnement valorisant et encourageant, ce qui a été extrêmement précieux.

Merci à tous les étudiants de la promotion 2024 du master TNAH pour leur soutien et leur amitié. L'émulation intellectuelle qui règne au sein de notre groupe a été un moteur essentiel et a rendu cette année de master particulièrement stimulante.

Enfin, merci à mes parents, mes plus fidèles relecteurs, pour leur intérêt curieux et leur soutien indéfectible pendant la douloureuse rédaction de ce mémoire. Et merci à Jingwei pour son enthousiasme à toute épreuve, sa douceur et son oreille attentive.

Bibliographie

Histoire de l'astronomie

- COSTABEL (Pierre), *Biographie de Claude Ptolémée* (90 env.-env. 168), fr-FR, URL : <https://www.universalis.fr/encyclopedie/claudie-ptolemee/> (visité le 03/03/2024).
- DASTON (Lorraine), « On Scientific Observation », *Isis*, 99–1 (mars 2008), Publisher : The University of Chicago Press, p. 97-110, DOI : 10.1086/587535.
- EIDA Conference 2024, en-US, URL : <https://eida.hypotheses.org/conferences/conference-2024> (visité le 26/05/2024).
- EVANS (James), *Astronomy - Ancient Greece, Stars, Planets*, URL : <https://www.britannica.com/science/astronomy/Ancient-Greece> (visité le 03/03/2024).
- KUPFER (Marcia), COHEN (Adam S.) et CHAJES (Howard), *The Visualization of Knowledge in Medieval and Early Modern Europe - Marcia Kupfer, Adam S. Cohen, Jeffrey Howard Chajes*, 2020, URL : <https://www.decitre.fr/livres/the-visualization-of-knowledge-in-medieval-and-early-modern-europe-9782503583037.html> (visité le 27/05/2024).
- Le Monde en Sphères*, fr, URL : <http://expositions.bnf.fr/monde-en-spheres/> (visité le 26/05/2024).
- LEQUEUX (James), *Système géocentrique de Ptolémée*, fr-FR, URL : <https://www.universalis.fr/encyclopedie/systeme-geocentrique-de-ptolemee/> (visité le 03/03/2024).
- MERCIER (Raymond), *Studies on the transmission of medieval mathematical astronomy*, 2004, URL : <https://data.bnf.fr/temp-work/10f87cf325908d90792f9f269c04ee41/> (visité le 06/04/2024).
- OBRIST (Barbara), « Visual Representation and Science : Visual Figures of the Universe between Antiquity and the Early Thirteenth Century », *Spontaneous Generations : A Journal for the History and Philosophy of Science*, 6–1 (oct. 2012), Number : 1, p. 15-23, DOI : 10.4245/sponge.v6i1.17158.
- RAYMOND JONES (Alexander), *Ptolemy*, en, févr. 2024, URL : <https://www.britannica.com/biography/Ptolemy> (visité le 03/03/2024).
- ROUSSEAU (Jean-Jacques), *Epicycles de Ptolémée*, URL : <https://ressources.univ-lemans.fr/AccessLibre/UM/Pedago/physique/02/divers/ptolemee.html> (visité le 03/03/2024).

SAFRAN (Linda), HAMBURGER (Jeffrey) et ROXBURGH (David), *The Diagram as Paradigm : Cross-Cultural Approaches*, 2022.

IA : généralités

A Quick Introduction to Neural Networks, en, août 2016, URL : <https://ujjwalkarn.me/2016/08/09/quick-intro-neural-networks/> (visité le 15/04/2024).

An Intuitive Explanation of Convolutional Neural Networks, en, août 2016, URL : <https://ujjwalkarn.me/2016/08/11/intuitive-explanation-convnets/> (visité le 11/04/2024).

AZENCOTT (Chloé-Agathe), *Introduction au machine learning*, 2e édition, Malakoff, 2022 (Info Sup).

BERMÈS (Emmanuelle) et MOIRAGHI (Eleonora), « Le patrimoine numérique national à l'heure de l'intelligence artificielle », *Revue Ouverte d'Intelligence Artificielle*, 1–1 (juill. 2020), p. 89-109, DOI : 10.5802/roia.5.

BOURDELOIE (Hélène), « Ce que le numérique fait aux sciences humaines et sociales. » *tic&société*–Vol. 7, N° 2 (mai 2014), Number : Vol. 7, N° 2 Publisher : ARTIC, DOI : 10.4000/ticetsociete.1500.

CARREMANS (Bert), *Handling overfitting in deep learning models*, en, janv. 2019, URL : <https://towardsdatascience.com/handling-overfitting-in-deep-learning-models-c760ee047c6e> (visité le 05/03/2024).

CASTELLANO (Giovanna) et VESSIO (Gennaro), « A Deep Learning Approach to Clustering Visual Arts », *International Journal of Computer Vision*, 130–11 (nov. 2022), p. 2590-2605, DOI : 10.1007/s11263-022-01664-y.

CHARNIAK (Eugene), *Introduction au Deep Learning*, 2021.

CHOLLET (François) et FORIEN (Jacqueline Isabelle), *L'apprentissage profond avec Python*, Saint-Cyr-sur-Loire, 2020 (Les essentiels de l'IA).

CLAVERT (Frédéric), *2DH — Design & Digital Humanities : le design comme méthode pour les Humanités Numériques ?*, fr-FR, Billet, ISSN : 2261-2971, mai 2015, DOI : 10.58079/u02u.

CLAVERT (Frédéric), BERRA (Aurelien) et HEIMBURGER (Franziska), « L'historien programmeur ? », dans *THATCamp Paris 2012 : Non-actes de la non-conférence des humanités numériques*, Code : THATCamp Paris 2012 : Non-actes de la non-conférence des humanités numériques, Paris, 2012 (La Non-Collection), DOI : 10.4000/books.editionsmsh.305.

- CRAWFORD (Kate) et PAGLEN (Trevor), *Excavating AI : The Politics of Training Sets for Machine Learning*, en-us, sept. 2019, URL : <https://perma.cc/NE8D-P6AW> (visité le 03/03/2024).
- DELUA (Julianna), *Supervised vs. Unsupervised Learning : What's the Difference ?*, en-US, mars 2021, URL : <https://www.ibm.com/blog/supervised-vs-unsupervised-learning/www.ibm.com/blog/supervised-vs-unsupervised-learning> (visité le 11/03/2024).
- DI LEONARDO (Isabella), SEGUIN (Benoit) et KAPLAN (Frédéric), « Visual Patterns Discovery in Large Databases of Paintings. » Dans *Digital Humanities 2016 : Conference Abstracts*, Kraków, 2016, p. 169-172, URL : <https://dh2016.adho.org/abstracts/348> (visité le 10/03/2024).
- FARLEY, *Multimodal embeddings concepts - Image Analysis 4.0 - Azure AI services*, en-us, févr. 2024, URL : <https://learn.microsoft.com/en-us/azure/ai-services/computer-vision/concept-image-retrieval> (visité le 10/03/2024).
- GRØNNE (Mathias), *Introduction to Embedding, Clustering, and Similarity*, en, oct. 2022, URL : <https://towardsdatascience.com/introduction-to-embedding-clustering-and-similarity-11dd80b00061> (visité le 10/03/2024).
- GUICHARD (Éric), « L'internet et les épistémologies des sciences humaines et sociales », *Revue Sciences/Lettres*-2 (févr. 2014), Number : 2 Publisher : École normale supérieure, DOI : 10.4000/rsl.389.
- GUICHARD (Éric) et POIBEAU (Thierry), « « Éditorial », Épistémologies digitales des sciences humaines et sociales », *Revue Sciences/Lettres*-2 (févr. 2014), Number : 2 Publisher : École normale supérieure, DOI : 10.4000/rsl.358.
- INDOLIA (Sakshi), GOSWAMI (Anil Kumar), MISHRA (S. P.) et ASOPA (Pooja), « Conceptual Understanding of Convolutional Neural Network- A Deep Learning Approach », *Procedia Computer Science*, International Conference on Computational Intelligence and Data Science 132 (janv. 2018), p. 679-688, DOI : 10.1016/j.procs.2018.05.069.
- JUNEJA (Sahil), *Deep Learning vs Machine Learning — The Difference Explained!*, en-US, mars 2023, URL : <https://www.dataquest.io/blog/deep-learning-vs-machine-learning-the-difference-explained/> (visité le 05/07/2024).
- KERMORVANT (Christopher), *Teklia - Automatic Text Recognition - The convergence between OCR and HTR technologies*, URL : <https://teklia.com/blog/202212-atr/> (visité le 17/08/2024).
- KLINKE (Harald), « Big Image Data within the Big Picture of Art History », *International Journal for Digital Art History*-2 (oct. 2016), Number : 2, DOI : 10.11588/dah.2016.2.33527.
- « L'Open Data a-t-il tenu ses promesses ? », dans ADEMEC, Paris, France, 2023.

- MALLORY (Gavin), *IIIF for museums, explained*, en, juill. 2019, URL : <https://blog.cogapp.com/iiif-for-museums-explained-49fd0560e1ba> (visité le 07/08/2024).
- MOIRAGHI (Eleonora), *Explorer des corpus d'images. L'IA au service du patrimoine*, fr-FR, Billet, ISSN : 2493-4437, avr. 2018, DOI : 10.58079/m3cp.
- NORINDR (Jade), *Le traitement des sources historiques par la vision artificielle : l'exemple des manuscrits d'astronomie de tradition ptoléméenne*, fr, mémoire de master, École Nationale des Chartes, 2023, URL : <https://dumas.ccsd.cnrs.fr/dumas-04255677> (visité le 02/02/2024).
- STRIEN (Daniel van), BEELEN (Kaspar), WEVERS (Melvin), SMITS (Thomas) et McDONOUGH (Katherine), « Computer Vision for the Humanities : An Introduction to Deep Learning for Image Classification (Part 2) », *Programming Historian* (, août 2022), URL : <https://programminghistorian.org/en/lessons/computer-vision-deep-learning-pt2> (visité le 13/03/2024).
- WEVERS (Melvin) et SMITS (Thomas), « The visual digital turn : Using neural networks to study historical images », *Digital Scholarship in the Humanities*, 35–1 (avr. 2020), p. 194-207, DOI : 10.1093/linc/fqy085.
- What Is Deep Learning and How Does It Work?*, en, URL : <https://builtin.com/machine-learning/deep-learning> (visité le 05/07/2024).

Documentation technique, méthodes, projets annexes

About — ALFA / Digital Alfonsine Corpus, URL : <https://dishes.obspm.fr/about> (visité le 25/05/2024).

ALBOUY (Ségolène), *Médiation des données de la recherche : Élaboration d'une plateforme en ligne pour une base de tables astronomiques anciennes*, fr, mémoire de master, École Nationale des Chartes, 2019.

APPLEBY (Micheal), CRANE (Tom), SANDERSON (Robert), STROOP (Jon) et WARNER (Simeon), *Presentation API 3.0*, en, URL : <https://iiif.io/api/presentation/3.0/> (visité le 01/09/2024).

BEAUDOUIN (Valérie) et PEHLIVAN (Zeynep), *Cartographie de la Grande Guerre sur le Web*, Research Report, Bibliothèque nationale de France ; Bibliothèque de documentation internationale contemporaine ; Télécom ParisTech, 2017, URL : <https://hal.science/hal-01425600> (visité le 29/07/2024).

BOUCHARD (Ariane), *Présentation du projet CORPUS à la BnF*, fr-FR, Billet, ISSN : 2608-1393, mai 2017, DOI : 10.58079/va8e.

BÜTTNER (Jochen), MARTINETZ (Julius), EL-HAJJ (Hassan) et VALLERIANI (Matteo), « CorDeep and the Sacrobosco Dataset : Detection of Visual Elements in Historical Documents », *Journal of Imaging*, 8–10 (oct. 2022), Number : 10 Publisher : Multi-disciplinary Digital Publishing Institute, p. 285, DOI : 10.3390/jimaging8100285.

CANNY (John), « A Computational Approach to Edge Detection », *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-8–6 (nov. 1986), Conference Name : IEEE Transactions on Pattern Analysis and Machine Intelligence, p. 679-698, DOI : 10.1109/TPAMI.1986.4767851.

CARBONI (Nicola), BARRAS (Marie) et JOYEUX-PRUNEL (Béatrice), « Pister des circulations visuelles à l'échelle mondiale », dans *Humanistica 2023*, Genève, Switzerland, 2023 (Images), URL : <https://hal.science/hal-04094170> (visité le 15/05/2024).

CARON (Mathilde), TOUVRON (Hugo), MISRA (Ishan), JEGOU (Hervé), MAIRAL (Julien), BOJANOWSKI (Piotr) et JOULIN (Armand), « Emerging Properties in Self-Supervised Vision Transformers », dans *2021 IEEE/CVF International Conference*

on Computer Vision (ICCV), ISSN : 2380-7504, 2021, p. 9630-9640, DOI : 10.1109/ICCV48922.2021.00951.

CHAGUÉ (Alix), CLÉRICE (Thibault) et ROMARY (Laurent), « HTR-United : Mutualisons la vérité de terrain ! », dans *DHNord2021 - Publier, partager, réutiliser les données de la recherche : les data papers et leurs enjeux*, Lille, France, 2021, URL : <https://hal.science/hal-03398740> (visité le 09/08/2024).

CHAMPENOIS (Robin) et JOYEUX-PRUNEL (Béatrice), « Visual Contagions : extraire et tracer la circulation d'images dans des imprimés illustrés », dans *Humanistica 2023*, Genève, Switzerland, 2023 (Circulations), URL : <https://hal.science/hal-04108205> (visité le 15/05/2024).

CHRISTENSEN (Kelly), PINCHE (Ariane) et GABAY (Simon), « Gallic(orpor)a : Traitement des sources textuelles en diachronie longue de Gallica », dans *DataLab de la BnF*, Paris, France, 2022, URL : <https://hal.science/hal-03716534> (visité le 11/05/2024).

COCO - Common Objects in Context, URL : <https://cocodataset.org/#home> (visité le 20/06/2024).

DOSOVITSKIY (Alexey), BEYER (Lucas), KOLESNIKOV (Alexander), WEISSENBORN (Dirk), ZHAI (Xiaohua), UNTERTHINER (Thomas), DEHGHANI (Mostafa), MINDERER (Matthias), HEIGOLD (Georg), GELLY (Sylvain), *et al.*, *An Image is Worth 16x16 Words : Transformers for Image Recognition at Scale*, arXiv :2010.11929 [cs], juin 2021, DOI : 10.48550/arXiv.2010.11929.

EGIAZARIAN (Vage), VOYNOV (Oleg), ARTEMOV (Alexey), VOLKHONSKIY (Denis), SAFIN (Aleksandr), TAKTASHEVA (Maria), ZORIN (Denis) et BURNAEV (Evgeny), « Deep Vectorization of Technical Drawings », 12358 (2020), arXiv :2003.05471 [cs], p. 582-598, DOI : 10.1007/978-3-030-58601-0_35.

EIDA, URL : <https://imagine.enpc.fr/~kallelis/icdar2024/> (visité le 23/05/2024). *Enhancing Heritage Image Databases*, en, URL : <https://anr.fr/Project-ANR-17-CE23-0008> (visité le 11/05/2024).

eScriptorium - Homepage, URL : <https://test2.fondue.unige.ch/> (visité le 15/06/2024). FOUAD (AOUINTI), SONAT (BALTAZI Zeynep), MATHIEU (AUBRY), ALEXANDRE (GUILBAUD) et STAVROS (LAZARIS), « Computer Vision and Historical Scientific Illustrations », (, 2023).

GABAY (Simon), CLÉRICE (Thibault), CAMPS (Jean-Baptiste), TANGUY (Jean-Baptiste) et GILLE-LEVENSON (Matthias), « Standardizing linguistic data : method and tools for annotating (pre-orthographic) French », dans *Proceedings of the 2nd International Digital Tools & Uses Congress (DTUC '20)*, Hammamet, Tunisia, 2020, DOI : 10.1145/3423603.3423996.

GABAY (Simon), PINCHE (Ariane) et CHRISTENSEN (Kelly), *Gallic(orpor)a : Processing Gallica's historical sources*, Publisher : Université de Genève Published : UNIGE

- Data Science Day, sept. 2022, URL : <https://hal.science/hal-03819326> (visité le 08/05/2024).
- Gallicorpora*, en, URL : <https://github.com/Gallicorpora> (visité le 08/05/2024).
- HA (David) et ECK (Douglas), *A Neural Representation of Sketch Drawings*, arXiv :1704.03477 [cs, stat], mai 2017, DOI : 10.48550/arXiv.1704.03477.
- HE (Kaiming), FAN (Haoqi), WU (Yuxin), XIE (Saining) et GIRSHICK (Ross), *Momentum Contrast for Unsupervised Visual Representation Learning*, arXiv :1911.05722 [cs], mars 2020, DOI : 10.48550/arXiv.1911.05722.
- HILAIRE (X.) et TOMBRE (K.), « Robust and accurate vectorization of line drawings », *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28–6 (juin 2006), Conference Name : IEEE Transactions on Pattern Analysis and Machine Intelligence, p. 890-904, DOI : 10.1109/TPAMI.2006.127.
- HORAE - Hours : Recognition, Analysis, Editions*, URL : https://heurist.huma-num.fr/heurist/?db=stutzmann_horae&website&id=378891 (visité le 14/08/2024).
- HUSSON (Mathieu), *EiDA - funding proposal*, 2022.
- ImageNet*, URL : <https://www.image-net.org/> (visité le 20/06/2024).
- JACQUOT (Olivier), *Décrire, transcrire et diffuser un corpus documentaire hétérogène : méthodes, formats, outils*, fr-FR, Billet, ISSN : 2493-4437, nov. 2017, DOI : 10.58079/m3ap.
- JANES (Juliette), PINCHE (Ariane), JAHAN (Claire) et GABAY (Simon), « Towards automatic TEI encoding via layout analysis », dans *Fantastic future 21, 3rd International Conference on Artificial Intelligence for Librairies, Archives and Museums*, Paris, France, 2021, URL : <https://hal.science/hal-03527287> (visité le 06/06/2024).
- JEANRENAUD (Adrien), « L'affiche de film à l'épreuve de la vision par ordinateur », dans *Humanistica 2023*, Genève, Switzerland, 2023, URL : <https://hal.science/hal-04133342> (visité le 15/05/2024).
- JOYEUX-PRUNEL (Béatrice), CARBONI (Nicola), JEANRENAUD (Adrien), VIACCOZ (Cédric), BELINA (Céline), GAUFFROY (Thomas) et BARRAS (Marie), « Un œil mondial ? La mondialisation par l'image au prisme du numérique : le cas du projet Visual Contagions », *Sociétés & Représentations*, 55–1 (2023), Place : Paris Publisher : Éditions de la Sorbonne, p. 203-226, DOI : 10.3917/sr.055.0203.
- KALLELI (Syrine), TRIGG (Scott), ALBOUY (Ségolène), HUSSON (Mathieu) et AUBRY (Mathieu), *Historical Astronomical Diagrams Decomposition in Geometric Primitives*, arXiv :2403.08721 [cs], mars 2024, DOI : 10.48550/arXiv.2403.08721.
- KAOUA (Ryad), SHEN (Xi), DURR (Alexandra), LAZARIS (Stavros), PICARD (David) et AUBRY (Mathieu), *Image Collation : Matching illustrations in manuscripts*, arXiv :2108.08109 [cs], août 2021, DOI : 10.48550/arXiv.2108.08109.
- KERMORVANT (Christopher), « Détection automatique d'objets dans les images avec YOLO » ().

KRISTENSEN (Kelly), *D'ALTO à TEI : Modélisation de transcriptions automatiques pour une pré-éditorialisant des textes*, fr, mémoire de master, École Nationale des Chartes, 2022, URL : <https://github.com/kat-kel/TNAH-Memoire>.

La numérisation à la BnF, fr, URL : <https://www.bnf.fr/fr/la-numerisation-la-bnf> (visité le 25/05/2024).

LEE (Benjamin Charles Germain), MEARS (Jaime), JAKEWAY (Eileen), FERRITER (Meghan), ADAMS (Chris), YARASAVAGE (Nathan), THOMAS (Deborah), ZWAARD (Kate) et WELD (Daniel S.), « The Newspaper Navigator Dataset : Extracting Headlines and Visual Content from 16 Million Historic Newspaper Pages in Chronicling America », dans *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, Virtual Event Ireland, 2020, p. 3055-3062, DOI : [10.1145/3340531.3412767](https://doi.org/10.1145/3340531.3412767).

LIN (Tsung-Yi), MAIRE (Michael), BELONGIE (Serge), BOURDEV (Lubomir), GIRSHICK (Ross), HAYS (James), PERONA (Pietro), RAMANAN (Deva), ZITNICK (C. Lawrence) et DOLLÁR (Piotr), *Microsoft COCO : Common Objects in Context*, arXiv :1405.0312 [cs], février 2015, DOI : [10.48550/arXiv.1405.0312](https://doi.org/10.48550/arXiv.1405.0312).

MALKOV (Yu A.) et YASHUNIN (D. A.), *Efficient and robust approximate nearest neighbor search using Hierarchical Navigable Small World graphs*, arXiv :1603.09320 [cs], août 2018, DOI : [10.48550/arXiv.1603.09320](https://doi.org/10.48550/arXiv.1603.09320).

MICHEZ (Guillaume), *GallicaPix, un nouvel outil d'exploration iconographique*, juin 2021, URL : <https://gallica.bnf.fr/blog/21062021/gallicapix-un-nouvel-outil-dexploration-iconographique?mode=desktop> (visité le 29/07/2024).

MONNIER (Tom) et AUBRY (Mathieu), « docExtractor : An off-the-shelf historical document element extraction », dans *2020 17th International Conference on Frontiers in Handwriting Recognition (ICFHR)*, arXiv :2012.08191 [cs], 2020, p. 91-96, DOI : [10.1109/ICFHR2020.2020.00027](https://doi.org/10.1109/ICFHR2020.2020.00027).

MONNIER (Tom), VINCENT (Elliot), PONCE (Jean) et AUBRY (Mathieu), *Unsupervised Layered Image Decomposition into Object Prototypes*, arXiv :2104.14575 [cs], août 2021, DOI : [10.48550/arXiv.2104.14575](https://doi.org/10.48550/arXiv.2104.14575).

PINCHE (Ariane), CHRISTENSEN (Kelly) et GABAY (Simon), « Between automatic and manual encoding », dans *TEI 2022 conference : Text as data*, Newcastle, United Kingdom, 2022, DOI : [10.5281/zenodo.7092214](https://doi.org/10.5281/zenodo.7092214).

RANSAC, en, URL : <https://www.mathworks.com/discovery/ransac.html> (visité le 14/08/2024).

REDMON (Joseph), DIVVALA (Santosh), GIRSHICK (Ross) et FARHADI (Ali), *You Only Look Once : Unified, Real-Time Object Detection*, arXiv :1506.02640 [cs], mai 2016, DOI : [10.48550/arXiv.1506.02640](https://doi.org/10.48550/arXiv.1506.02640).

RESILIENCE Tool : eScriptorium, URL : <https://www.resilience-ri.eu/blog/resilience-tool-escriptorium/> (visité le 15/06/2024).

SAGOT (Benoît), ROMARY (Laurent), BAWDEN (Rachel), ORTIZ SUÁREZ (Pedro Javier), CHRISTENSEN (Kelly), GABAY (Simon), PINCHE (Ariane) et CAMPS (Jean-Baptiste), « Gallic(orpor)a : Extraction, annotation et diffusion de l'information textuelle et visuelle en diachronie longue », dans *DataLab de la BnF : Restitution des travaux 2022*, Paris, France, 2022, URL : <https://hal.science/hal-03930542> (visité le 06/06/2024).

SHEN (Xi), EFROS (Alexei A.), JOULIN (Armand) et AUBRY (Mathieu), *Learning Co-segmentation by Segment Swapping for Retrieval and Discovery*, arXiv :2110.15904 [cs] version : 2, mars 2022, DOI : 10.48550/arXiv.2110.15904.

The Sphere, URL : <https://sphaera.mpiwg-berlin.mpg.de/> (visité le 14/08/2024).

VHS Project : Presentation, en-US, URL : <https://vhs.hypotheses.org/presentation> (visité le 12/05/2024).

Vision artificielle et analyse Historique de la circulation de l'illustration Scientifique, fr, URL : <https://anr.fr/Projet-ANR-21-CE38-0008> (visité le 12/05/2024).

WordNet, en, URL : <https://wordnet.princeton.edu/homepage> (visité le 14/08/2024).

ZHANG (Hao), LI (Feng), LIU (Shilong), ZHANG (Lei), SU (Hang), ZHU (Jun), NI (Lionel M.) et SHUM (Heung-Yeung), *DINO : DETR with Improved DeNoising Anchor Boxes for End-to-End Object Detection*, arXiv :2203.03605 [cs], juill. 2022, DOI : 10.48550/arXiv.2203.03605.

Problématiques d'édition

About the CITE architecture · The CITE Architecture, URL : <https://cite-architecture.github.io/about/> (visité le 09/07/2024).

Robert Alessi et Marcello Vitali-Rosati (éd.), *Éditions critiques numériques : entre tradition et changement de paradigme*, Montréal, 2023 (Parcours numériques), URL : <http://parcoursnumeriques-pum.ca/12-editions-critiques/>.

BLACKWELL (Christopher W.) et SMITH (Neel), « The CITE Architecture : a Conceptual and Practical Overview », dans *The CITE Architecture : a Conceptual and Practical Overview*, 2019, p. 73-94, DOI : 10.1515/9783110599572-006.

DE YOUNG (Greg), « Editing a collection of diagrams ascribed to al-Hājjāj : an initial case study », *SCIAMVS*, 15 (2014), p. 171-238.

DE YOUNG (Gregg), « Diagrams in ancient Egyptian geometry : Survey and assessment », *Historia Mathematica*, 36–4 (nov. 2009), p. 321-373, DOI : 10.1016/j.hm.2009.02.004.

DRaFT download, URL : https://www.greekmath.org/draft/draft_index.html (visité le 14/08/2024).

EPRON (Benoît) et VITALI-ROSATI (Marcello), *L'édition à l'ère numérique*, Paris, 2018 (Repères), URL : <https://papyrus.bib.umontreal.ca/xmlui/handle/1866/20642>.

HARRIS (Rich), *Svelte 3 : Rethinking reactivity*, en, avr. 2019, URL : <https://svelte.dev/blog/svelte-3-rethinking-reactivity> (visité le 16/08/2024).

Introduction • Docs • Svelte, en, URL : <https://svelte.dev/docs/introduction> (visité le 14/08/2024).

JARDINE (Boris) et JARDINE (Nicholas), « Critical Editing of Early-Modern Astronomical Diagrams », *Journal for the History of Astronomy*, 41–3 (août 2010), Publisher : SAGE Publications Ltd, p. 393-414, DOI : 10.1177/002182861004100307.

Leibniz Online - Leibniz-Edition Berlin, URL : <https://leibniz-berlin.bbaw.de/de/leibniz-online> (visité le 10/07/2024).

MOUREAU (Sébastien), « The Apparatus Criticus » (, 2015), URL : <https://dial.uclouvain.be/pr/boreal/object/boreal:155652> (visité le 14/08/2024).

RAYNAUD (Dominique), « Building the stemma codicum from geometric diagrams : A treatise on optics by Ibn al-Haytham as a test case », *Archive for History of Exact*

- Sciences*, 68–2 (2014), Publisher : Springer, p. 207-239, URL : <https://www.jstor.org/stable/24569630> (visité le 07/04/2024).
- REES (Martin), *Cosmic Origami and What We Don't Know*, en-US, nov. 2013, URL : <https://onbeing.org/programs/martin-rees-cosmic-origami-and-what-we-dont-know/> (visité le 14/07/2024).
- Référence des attributs *SVG - SVG (Scalable Vector Graphics) / MDN*, fr, juill. 2024, URL : <https://developer.mozilla.org/fr/docs/Web/SVG/Attribute> (visité le 03/08/2024).
- REYNAUD (Adeline), *Les diagrammes mathématiques paléo-babyloniens : catalogue, propriétés, matérielles, rôles dans les raisonnements*, These en préparation, Université Paris Cité, 2017, URL : <https://theses.fr/s197780> (visité le 26/08/2024).
- ROUGHAN (Christine), « Digital editions and diplomatic diagrams », dans *Proceedings of the First International Conference on Digital Access to Textual Cultural Heritage*, event-place : Madrid, Spain, New York, NY, USA, 2014 (DATECH '14), p. 77-82, DOI : 10.1145/2595188.2595189.
- SINATRA (Michaël E.) et VITALI-ROSATI (Marcello), *Pratiques de l'édition numérique*, Code : Pratiques de l'édition numérique, Montréal, 2014 (Parcours numérique), DOI : 10.4000/books.pum.317.
- telota/LeibnizVIII-LaTeX_TEI*, original-date : 2021-02-24T13:06:06Z, mars 2023, URL : https://github.com/telota/LeibnizVIII-LaTeX_TEI (visité le 10/07/2024).
- The Homer Multitext project*, URL : <https://www.homermultitext.org/> (visité le 09/07/2024).
- TROVATO (Paolo), *Everything You Always Wanted to Know about Lachmann's Method*, Google-Books-ID : VZqlBAAAQBAJ, 2014.
- Tutoriel *SVG - SVG (Scalable Vector Graphics) / MDN*, fr, juill. 2024, URL : <https://developer.mozilla.org/fr/docs/Web/SVG/Tutorial> (visité le 03/08/2024).

Introduction

“La notion de pensée spirituelle n'a pas de sens et ce que l'on croit relever d'une aptitude intellectuelle extraordinaire consiste, peut-être à 95 %, en une maîtrise de notre système de signes et de ses combinaisons, qui peut certes confiner à l'art (la possibilité d'enchaîner des centaines, voire des milliers de gestes, d'algorithmes, de recettes ou de formules) mais qui reste majoritairement technique : une somme d'apprentissages tout à fait accessibles et qui conduisent à des pratiques et des gestes enchaînés de façon de plus en plus rapide avec leur répétition. Nous retrouvons là les propos de Leibniz, de Dagognet et de Granger.”¹

Guichard souligne ici que la pensée s'ancre profondément dans la technique. À l'heure du numérique, chaque projet de recherche, avec ses sources et ses questions, a des besoins épistémologiques précis qui s'incarnent dans des structures informatiques particulières : des protocoles, des formats, des dispositifs, des visualisations de données, etc. Au cœur de cette idiosyncrasie, comment créer des outils numériques qui s'inscrivent dans un écosystème plus larges, tout en restant attentif aux spécificités de chaque objet et de chaque question de recherche ?

Cette question est centrale dans la construction d'une chaîne de traitement des sources. Le mythe du savant isolé, reclus dans son observatoire, penché sur ses grimoires, fait place à la réalité du collectif. La collaboration est désormais de mise : pour favoriser le partage et l'accès aux données, pour garantir la reproductibilité des résultats et assurer la cohérence des pratiques. Malgré cela, la prolifération des outils spécialisés, souvent conçus pour un projet particulier, persiste dans le domaine des Humanités Numériques. Non seulement nécessitent-ils la mobilisation de ressources humaines et techniques importantes, mais aussi se révèlent-ils difficilement maintenables sur le long terme. Devant ces constats, la question se pose : comment mettre en place des formes de mutualisation sur le plan technique ? Cet enjeu se joue à plusieurs niveaux : ouverture des silos de donnée, scripting, architecture applicative et infrastructure hardware, protocoles pour diffuser et maintenir ces outils, pour collaborer autour de leur développement...

Le principal défi réside dans la conciliation des exigences de généralité et de spécificité, pour créer des outils numériques qui non seulement répondent aux besoins spécifiques du projet qui les porte, mais aussi servent une communauté scientifique plus large et in-

1. Éric Guichard, « L'internet et les épistémologies des sciences humaines et sociales », *Revue Sciences/Lettres*-2 (févr. 2014), Number : 2 Publisher : École normale supérieure, DOI : 10.4000/rs1.389.

terdisciplinaire. Comment concevoir des systèmes suffisamment flexibles sans sacrifier leur pertinence et leur efficacité ?

Ce mémoire propose une exploration de ces problématiques en s'appuyant sur un cas d'étude : la construction de la plateforme AIKON, qui met à disposition des outils basés sur la *computer vision* pour l'enrichissement, la sémantification et l'analyse des données visuelles.

0.1 Mise en contexte

0.1.1 Un changement de paradigme

La numérisation massive de sources archivistiques et bibliographiques a radicalement transformé le paysage de la recherche en sciences humaines. Les ressources numériques et leurs usages se cessent de se développer et se diversifier. Les ouvrages et les manuscrits scientifiques, notamment, constituent une source essentielle pour la connaissance de l'histoire des sciences.

“The availability of large amounts of digitized historical documents opened the door to the use of computational approaches for their analysis.”²

La disponibilité des données numérisées est au fondement de la démarche basée sur des traitements automatiques des sources par des algorithmes de vision artificielle. Ces immenses corpus offrent des opportunités inédites pour comprendre les phénomènes culturels, sociaux et historiques à une échelle sans précédent, soulevant cependant des défis techniques et méthodologiques : en premier lieu l'accès à la donnée, en second lieu, sa sémantification.

Le décloisonnement des silos d'information est un enjeu majeur de la gouvernance des données, et le premier pas vers leur exploitation. Le concept d'*open-data* renvoie à une ambition d'ouverture des données pour leur libre circulation et exploitation. Il concerne à la fois la diffusion des sources numériques auprès d'un public très large³ et l'enrichissement des ressources par la communauté scientifique. Le recours à des standards communs et des mécanismes d'échange normalisés est essentiel pour garantir l'interopérabilité des données et faciliter leur intégration dans des diverses infrastructures. En effet, la capacité à intégrer des données provenant de sources multiples et hétérogènes est indispensable pour exploiter pleinement le potentiel des ressources numérisées dans la recherche en SHS. En d'autres termes, il s'agit de briser les silos applicatifs pour concrétiser une *harmonisation virtuelle*,

2. Jochen Büttner, Julius Martinetz, Hassan El-Hajj et Matteo Valleriani, « CorDeep and the Sacrobosco Dataset : Detection of Visual Elements in Historical Documents », *Journal of Imaging*, 8–10 (oct. 2022), Number : 10 Publisher : Multidisciplinary Digital Publishing Institute, p. 285, DOI : 10.3390/jimaging8100285, p.2.

3. Auprès de la communauté scientifique comme du grand public.

une unification, créant ainsi un écosystème numérique collaboratif, où les données peuvent circuler librement entre les lieux de stockage.

Au cœur des enjeux de l’open-data, la valorisation de la donnée brute reste en outre un défi à relever. Pour tirer pleinement parti de cette richesse disponible, il est nécessaire de mettre en place des infrastructures adaptées pour les mettre à disposition afin de favoriser la collaboration entre les différents acteurs (producteurs comme réutilisateur.rices). D'où la nécessité d'enrichir la donnée brute, pour la “faire parler”.

“Thus, Big Data means to acquire, store, and analyze large amount of data that are generated quickly and are not always structured [...].”⁴

Après l’étape de numérisation, le document reste à l’état d’images matricielles directement issues du scan ou de la photographie des pages. Son contenu sémantique, c'est-à-dire le texte et les illustrations, demeure illisible aux machines, et donc caché aux lecteur jusqu’à ce qu'il ouvre et lise la copie numérique. L’enjeu est alors de transformer ces images matricielles en données structurées, d’en obtenir de nouvelles représentations sémantiquement riche et manipulables, afin de construire des portails de bases de données permettant une interrogation fine du contenu des documents.

La numérisation initie donc un processus de transformation de la donnée brute en informations structurées grâce à sa segmentation. Cette étape, en conférant un niveau d’abstraction supérieur aux données, les rend aptes à supporter de nouveaux traitements algorithmiques et des analyses scientifiques. Face à l’explosion des volumes de données, l’automatisation de ces analyses devient indispensable pour appréhender de nouveaux ordre de grandeur et extraire des connaissances.

“[The] interpretation of results is still exclusive to humans, but computers can help us with the steps leading to that destination.”⁵

Des outils numériques vont permettre de décoder et d’interpréter les informations visuelles, faisant ainsi “parler” les données brutes :

0.1.2 De nouveaux outils

Les Humanités Numériques, initialement concentrées sur l’étude du texte, ont bénéficié de l’essor des technologies d’extraction du texte. La Reconnaissance Optique de Caractères (OCR (*Optical Character Recognition*)), puis la Reconnaissance d’Écriture Manuscrite (HTR (*Handwritten Text Recognition*)), ont suscité un intérêt croissant depuis les années 1950, mais c'est à partir des années 1990 que leur développement s'est véritablement accéléré. L'objectif premier est d'automatiser l'extraction de textes à partir d'images numériques de sources historiques. Confrontées à la diversité des polices, des

4. Harald Klinke, « Big Image Data within the Big Picture of Art History », *International Journal for Digital Art History*-2 (oct. 2016), Number : 2, DOI : 10.11588/dah.2016.2.33527, p.26.

5. *Ibid.*

mises en page et des orthographies rencontrées dans les documents, les méthodes se sont récemment tournées vers l'apprentissage profond (*deep learning*).

Cependant les archives et les sources contiennent aussi un grand nombre d'images. En comparaison de l'intérêt pour l'HTR et l'OCR, le développement des réseaux de neurones et du *machine learning* pour le traitement automatique du matériau visuelle accompagnant le texte est relativement récent. Pourtant, les deux champs partagent une base commune : le traitement du texte numérisé comme de ses illustrations, en apprentissage machine, revient à manipuler une image matricielle, soit une grille de pixels, et extraire puis encoder des informations sémantiques.

0.1.3 Perspectives ouvertes pour l'image

Cette convergence technologique offre aujourd'hui la possibilité d'identifier et classifier des motifs récurrents au sein de vastes collections d'images, permettant une exploration plus systématique et à plus grande échelle des éléments visuels présents dans les sources. Les réseaux de neurones profonds ouvrent de nouvelles voies d'interrogation des archives numériques au prisme de leurs illustrations.

“They [les réseaux de neurone] open up a part of the digital archive for large-scale analysis, which, until now, has been left uncovered : the millions of images in digitized books, newspapers, periodicals, and historical documents. As a result, they allow us to explore the visual side of the digital turn in historical research. Using these techniques, we can explore visual material in archives using nontextual search methods. Scholars can, for example, find visual material related to a particular topic, or, they can identify transitions in the use of a particular medium, such as illustrations and photographs.”⁶

Les représentations sémantiques extraites par les modèles de vision artificielle fournissent un niveau d'abstraction supérieur à l'image, visent à capturer le sens profond de son contenu, offrant ainsi de nouvelles perspectives pour l'analyse et l'exploitation des données de la recherche.

0.1.4 Structurer l'information

L'extraction et l'encodage de texte ouvrent de nouvelles perspectives pour mettre en œuvre des chaînes d'édition partiellement automatisées. Or les formats de l'édition numériques permettent l'introduction d'une approche sémantique de l'exploitation des textes. Au-delà de la simple mise en forme, l'édition numérique implique désormais un balisage sémantique, basé sur des langages comme XML (*eXtensible Markup Language*) et HTML (*HyperText Markup Language*), et sur des standards comme TEI (*Text Encoding*

6. Melvin Wevers et Thomas Smits, « The visual digital turn : Using neural networks to study historical images », *Digital Scholarship in the Humanities*, 35–1 (avr. 2020), p. 194–207, doi : 10.1093/11c/fqy085, p.2.

Initiative), rendant les contenus lisibles aussi bien par les humains que par les machines. Ainsi, l'édition ne se limite plus à la création d'un objet matériel, mais vise à structurer les contenus pour favoriser leur exploration et l'ouverture à de nouveaux traitements algorithmiques.

Comment étendre cette structuration aux images, de manière à les rendre exploitable par des traitements algorithmiques ? Les algorithmes de vectorisation automatique permettent de transformer des images géométriques simples, comme les diagrammes astronomiques qui constituent le corpus d'EIDA, en données structurées. Cette représentation vectorielle offre un équivalent visuel du balisage sémantique appliqué aux textes, permettant ainsi de capturer le sens mathématique inhérent à ces images.

0.1.5 Les données pour des ponts interdisciplinaires

La combinaison de deux facteurs, l'augmentation exponentielle de la puissance de calcul et de la disponibilité des données visuelles, a permis à la Computer Vision d'évoluer rapidement, permettant de créer des réseaux de neurones plus profonds, plus précis, et améliorant la vitesse et l'exactitude de tâches de plus en plus complexes⁷. Mais cet avancement des techniques est aussi catalysée par leur application concrète et en contexte, car leur développement repose sur la disponibilité de grands *datasets* annotés pour l'entraînement des modèles et l'évaluation des résultats. Ainsi si l'IA profite au domaine de la recherche historique sur du matériel visuel, la réciproque est aussi vraie :

“humanities could likewise be a boon to the development of more accurate and more sophisticated computer vision techniques. As classification algorithms have been trained on manually tagged sets, structural biases in their classification schemes will be reproduced in the results produced by computer vision techniques. In collaboration with humanities scholars, computer scientists could critically engage with these biases and rethink the way we annotate data sets and measure algorithmic accuracy.”⁸

La mise à disposition de jeux de données sélectionnés et annotés par des experts améliore les performances et l'exactitude des modèles, faisant rapidement avancer la recherche en *computer vision*. Ainsi le *deep learning*, gourmand en données, profite de la disponibilité des corpus annotés, et les historien.nes profitent du changement d'échelle de l'analyse permise par la *computer vision*. La recherche et la technologie rentrent en dialogue, l'une répondant aux besoins de l'autre, créant des collaborations interdisciplinaires bénéfiques. D'où le besoin d'autant plus prégnant dans le domaine du *machine learning* de penser les outils dans le sens de l'ouverture des systèmes afin de garantir l'interopérabilité de ces données annotées.

7. H. Klinke, « Big Image Data within the Big Picture of Art History »...

8. M. Wevers et T. Smits, « The visual digital turn... », p.12.

0.1.6 *Pipelines et workflows : spécificités de l'implémentation de l'IA*

Les problématiques liées à l'intelligence artificielle ne se limitent pas à la construction de modèles performants. Son utilisation soulève également des questions en matière d'architecture matérielle et applicative. Déjà, les modèles d'IA, particulièrement ceux basés sur l'apprentissage profond, nécessitent la puissance de calcul et les infrastructures *hard-ware* adaptées. L'intégration de ces modèles dans des systèmes applicatifs exige également des algorithmes optimisés pour la parallélisation, ainsi que des solutions pour la gestion des données à grande échelle et pour la réduction de la latence.

L'intégration de l'IA dans un processus de traitement de données implique en outre la création de *workflows* itératifs. Ces *workflows* nécessitent une récupération des données en sortie, suivie d'une intervention humaine pour leur correction et leur réintégration dans le processus d'entraînement des modèles. Ces pipelines impliquent une boucle de rétroaction où les données produites par les modèles sont régulièrement évaluées, corrigées par des experts, puis réinjectées dans le processus d'apprentissage.

Alors comment préserver le rôle des chercheur.ses dans ce processus afin de garantir la qualité des résultats ? Comment concevoir des infrastructures applicatives ou logicielles capables de gérer ces flux de données ? Et est-il possible de créer un outil qui permette d'appliquer une méthode unifiée à des ensembles de données hétérogènes, un outil suffisamment flexible et adapté à divers traitements basés sur le *machine learning* ? La réponse à ces questions nécessite une approche transversale, combinant des compétences en informatique, en développement applicatif, et en gestion de la donnée.

Les enjeux et interrogations qui entourent la mise en place de ce *workflow* m'ont occupé pendant mon stage, et constituent le cœur de ce qui est exposé dans ce mémoire.

0.2 Mission de stage

La plateforme AIKON—anciennement EIDA—est déjà dotée d'une chaîne de traitement partiellement automatisée qui intègre des fonctionnalités d'extraction et de recherche de similarités. Ces traitements permettent de repérer et de comparer des éléments visuels, en l'occurrence des diagrammes, au sein d'une base de données constituée d'images numérisées provenant de sources variées : institutions de conservation ou collections personnelles.

La prochaine étape dans le développement de la plateforme consiste en l'implémentation de la fonctionnalité de vectorisation des diagrammes extraits, visant à affiner encore le niveau de structuration de l'information. Ce processus transforme une image en un ensemble de formes géométriques élémentaires, appelées primitives. Cette représentation, particulièrement adaptée au traitement informatique, permet de traduire les informations visuelles en structures mathématiques, facilitant ainsi leur manipulation, leur analyse et

leur exploitation par des méthodes computationnelles.

Mon stage au sein de l'équipe d'histoire des sciences du laboratoire SYRTE (*Systèmes de Référence Temps-Espace*) de l'Observatoire de Paris a consisté à développer un module de vectorisation automatique dans la plateforme AIKON, tout en participant à une réflexion plus large sur l'architecture de cette plateforme. En favorisant une approche modulaire et flexible, son évolution permettra d'intégrer plus facilement de nouvelles fonctionnalités et d'ouvrir la plateforme à d'autres domaines de recherche.

0.3 Problématisation

Comment concilier la nécessité de développer des outils génériques, réutilisables et performants avec la grande variété des données et des problématiques spécifiques à chaque étude ?

La tension s'exprime dans le défi de la personnalisation : entre généralisation et spécificité. Nous explorerons comment cette dialectique se manifeste à chaque étape d'une chaîne de traitement. Au niveau de la description des données en premier lieu : comment concilier la richesse des données réelles avec la nécessité de les structurer pour une exploitation efficace ? Par ailleurs, la construction de modèles de *computer vision* s'inscrit au cœur de cette tension : ils doivent être suffisamment généraux pour pouvoir s'adapter à la complexité du réel, tout en restant assez complexes et spécialisés sur des cas particuliers. Enfin, comment penser l'implémentation de ces modèles dans une plateforme évolutive et modulable ?

Une question épistémologique sous-tendra tout au long de ce mémoire les questions techniques : comment l'outillage technique collaboratif contribue à façonner les méthodologies de la recherche ?

0.4 Annonce du plan

Le développement se déclinera en trois parties. Premièrement, nous présenterons le projet EIDA et ses objectifs, en le situant dans le paysage de la recherche en études visuelles, et nous explorerons les enjeux qui découlent de cette inscription dans un contexte plus large. Nous aborderons ensuite les outils de *computer vision* utilisés, les défis liés à leur mise en œuvre, et les perspectives ouvertes en terme d'édition numérique. Enfin, nous présenterons les contours d'une plateforme qui permettrait de rendre à la fois les méthodes et les résultats de l'analyse accessibles à la communauté scientifique, notamment à d'autres projets de recherche en études visuelles ; cette partie sera consacrée à la mise en œuvre technique d'une approche modulaire dans le développement d'une application web.

Première partie

Chaîne de traitement de la donnée visuelle : enjeux technologiques et disciplinaires

Introduction partielle

“Indeed, not only have the majority of historians of cosmology dismissed the early Middle Ages due to the fact that there is no “scientific progress” to be observed, but they also tended to disregard visual representations, limiting their inquiries to the doctrinal aspects of textual sources.”⁹

Les illustrations et leur évolution dans les sources d’ordre scientifique du Moyen Âge jusqu’aux cultures occidentales modernes n’ont été que peu étudiées. De manière plus générale, le rôle de l’image dans la construction et la diffusion du savoir scientifique soulève des questions complexes qui restent historiquement délicates à appréhender et pour lesquelles des outils d’analyse adaptés font défaut. Afin de combler cette lacune, l’objectif des projets EIDA/VHS (*Vision artificielle et analyse Historique de la circulation de l’illustration Scientifique*) est de développer des méthodes automatisées pour l’extraction et le traitement des illustrations dans les sources historiques, avec pour but ultime d’appuyer l’analyse experte des chercheur.ses en histoire des sciences.

L’enrichissement et l’exploration de vastes corpus iconographiques n’est pas l’apanage de la seule histoire des sciences. Le développement de projets en Humanités Numériques dans des disciplines annexes constituent des antécédents ou des points de comparaison riches en enseignements. Si ce mémoire porte plus spécifiquement sur les développements réalisés dans le cadre de EIDA, les différents projets s’inscrivent dans des dynamiques de continuité et de partage, créant un écosystème ouvert et fructueux, permettant d’appréhender au mieux les enjeux et les défis liés au traitement des résultats des grandes campagnes de numérisation des institutions et bibliothèques.

Cette première partie explore la complexité des relations qui se forment à différents niveaux dans le contexte de la fabrication d’outils de traitement automatique de l’image. Ces outils doivent prendre en compte les dynamiques spécifiques aux acteurs et au cadre du projet EIDA. Par ailleurs, ils s’insèrent dans un réseau plus vaste incluant des projets partenaires sur lesquels s’appuyer ou des briques fonctionnelles à intégrer. La problématique de ce chapitre réside donc dans la compréhension des liens se tissant à divers niveaux, impactant les développements, leur ouvrant des pistes comme les contraignant.

Cette première partie présente ce réseau à deux niveaux : premièrement les acteurs

9. Barbara Obrist, « Visual Representation and Science : Visual Figures of the Universe between Antiquity and the Early Thirteenth Century », *Spontaneous Generations : A Journal for the History and Philosophy of Science*, 6–1 (oct. 2012), Number : 1, p. 15-23, DOI : 10.4245/sponge.v6i1.17158, p.16.

et le cadre du projet, et deuxièmement un réseau plus vaste, celui de la recherche en générale, la communauté se regroupant autour de grands principes d'ouverture garantissant une interopérabilité technique des données. Enfin, un état de l'art sur l'intelligence artificielle et d'enrichissement de données se concentrera autour du projet *Gallic(orpor)a*. Les réflexion méthodologiques menées dans le cadre de ce projet soulignent l'importance de partager des pratiques dans le domaine du *machine learning*.

“Aller à la rencontre d’autres projets, créer des synergies avec d’autres équipes ou institutions peut être une réponse à cette difficulté de gestion de la masse. Un effort de standardisation sommé à une réflexion sur l’interopérabilité pourrait faire dialoguer les corpus et mutualiser les outils d’analyse, au-delà des clivages entre disciplines et sujets de recherches.”¹⁰

10. Olivier Jacquot, *Décrire, transcrire et diffuser un corpus documentaire hétérogène : méthodes, formats, outils*, fr-FR, Billet, ISSN : 2493-4437, nov. 2017, DOI : 10.58079/m3ap.

Chapitre 1

EIDA : Contexte institutionnel et scientifique

Le projet EIDA (Editing and analysing hIstorical astronomical Diagrams with Artificial intelligence) a pour ambition de rassembler des sources provenant de différentes traditions astronomiques et de produire des outils d'analyse et d'exploitation de la grande diversité de sources astronomiques issues d'origines géographiques et temporelles variées, les mettant à disposition de la communauté de la recherche.

Dans une perspective interdisciplinaire, EIDA vise un double objectif scientifique et technique. Exploitant les progrès récents des approches analytiques basées sur la vision artificielle le projet s'appuie sur le développement d'outils de *deep learning* capables d'automatiser l'analyse des diagrammes de l'extraction à la décomposition en composants significatifs, permettant d'envisager leur édition. Ces outils dédiés aux diagrammes servent leur étude documentaire et épistémique, diachronique ou synchronique, en s'appuyant sur des corpus à grande échelle.

Les outils développés dans ce cadre s'inscrivent dans un écosystème collaboratif qui demande un effort de normalisation des outils développés. On expliquera comment les dispositifs d'automatisation sont contraints par la pluralité des usages et des acteurs en présence.

1.1 Contexte disciplinaire

L'astronomie est issue d'une tradition continue vieille de près de 4000 ans qui transcende les cultures et les langues. Les théories, les savoirs, les méthodes, au gré de leur diffusion, se mélangent aux pratiques autochtones pour servir les usages locaux, qui bien souvent renforcent des dynamiques de pouvoir en place, que ce dernier soit politique ou religieux. Les sciences astreines revêtent ainsi une importance culturelle majeure.

Une étude approfondie des sources révèle les processus par lesquels les connaissances

se sont enrichies et transformées au contact de différentes cultures. Leur examen permet en outre de souligner les spécificités et les innovations de chaque tradition, tout en montrant les interconnexions et les influences réciproques qui ont façonné l'évolution de l'astronomie. Les schémas de circulation sont dans de nombreux cas de très grande portée géographique, chronologique et culturelle. Ils relient des contextes de production de connaissances à l'échelle afro-eurasienne et sur des périodes de siècles voire de millénaires. La trace de ces transmission supporte une vision connectée et globale de l'histoire des cultures et du savoir.

La grande diversité des sources et des approches possibles rend cependant difficile une approche globale. À ce titre, il est important de cibler un objet d'étude : EIDA se focalise ainsi essentiellement dans la transmission de la tradition ptoléméenne, et le corpus se compose donc de ressources manuscrites et imprimées relevant de cette tradition.

1.1.1 Ptolémée : modèle et transmission

Ptolémée tient une place proéminente dans l'histoire de l'astronomie et des mathématiques. Son nom reste associé à la conception d'un système astronomique qui plaçait la Terre immobile au centre du monde, et dont la mise en question, de Copernic à Newton, a commandé la révolution scientifique.

Dans sa *Syntaxe mathématique*, plus connue sous le titre d'*Almageste*, et dont la dernière observation consignée date de 141, il expose l'ensemble des connaissances astronomiques de son époque. Notamment il perfectionne le modèle élaboré par Hipparche, à qui il emprunte la découverte de l'excentricité des trajectoires apparentes du Soleil et de la Lune par rapport à la Terre, et l'idée de composer ces trajectoires à l'aide de deux mouvements distincts. Il élabore un système géocentrique au moyen d'un ensemble complexe de trajectoires circulaires des objets pris dans un mouvement uniforme : les déférents, autour de la Terre, et les épicycles, dont les centres parcouruent les déférents¹.

En effet, les sociétés anciennes attendent des corps astraux (soleil, lune, planètes et étoiles) un mouvement uniforme et le plus “parfait” possible, c'est-à-dire un cercle. Pourtant la trajectoire de ces corps, observée empiriquement, n'est pas circulaire. Le modèle de Ptolémée explique ces imperfections en postulant que les mouvement apparemment irréguliers sont dû à cette fameuse combinaison de plusieurs trajectoires circulaires régulières vues depuis la Terre, point statique. Les planètes se déplacent à vitesse uniforme sur un cercle (l'épicycle) dont le centre se déplace à vitesse uniforme sur un cercle coplanaire (le déférent), dont la Terre est le centre.

En plus de la description du mouvement des astres, Ptolémée dresse dans l'*Almageste* des tables établissant les positions de la lune et prévoyant les périodes et les caractéris-

1. James Lequeux, *Système géocentrique de Ptolémée*, fr-FR, URL : <https://www.universalis.fr/encyclopedie/systeme-geocentrique-de-ptolemee/> (visité le 03/03/2024).

tiques des éclipses avec une précision inédite², un catalogue des étoiles, un traité complet de trigonométrie plane et sphérique et une description des instruments nécessaires à un grand observatoire.

L'œuvre de Ptolémée fera référence, et en tant que synthèse des connaissances astronomiques antérieures, sa transmission correspond à celle de la vision des pratiques de l'astronomie grecque à son apogée, et sa diffusion façonnera la production astronomique ancienne pendant près de treize siècles. D'ailleurs le nom d'*Almageste* date de la transmission par les civilisations arabes à l'occident.

En effet, lors de la chute de l'Empire Romain d'occident, la majeure partie des ouvrages antiques sont perdus et la science occidentale stagnera jusqu'au XII^e siècle. Elle continuera cependant à progresser ailleurs : notamment dans le monde arabe et musulman. Dès le VIII^e et IX^e siècle, les Arabes vont traduire dans leur langue la plupart des grands textes scientifiques de l'Antiquité, en particulier les œuvres d'Aristote et l'*Almageste* de Ptolémée. Sans remettre en cause le géocentrisme et le système de Ptolémée, ils le perfectionnent et l'amènent à un très grand degré de précision. Nécessaire à la stricte observation des règles de l'islam, l'astronomie arabe se développe et se diffuse, grâce aux travaux d'al-Biruni, al-Hazen ou al-Sufi³.

À partir du XI^e et surtout du XII^e siècle, au fil des conquêtes des occidentaux en Espagne et en Sicile, les textes grecques sont traduits en latin via la traduction arabe. La transmission des savoirs gréco-arabes – notamment les traductions arabo-latines de l'*Almageste* et du Livre des étoiles fixes d'al-Sufi – ouvre la voie à un renouveau scientifique dans l'Occident chrétien, permettant ainsi l'essor des grandes universités européennes de l'époque (Paris, Oxford, Bologne, etc). On redécouvre les modèles d'Aristote et de Ptolémée en les adaptant aux conceptions chrétiennes⁴.

Avant l'avènement de l'astronomie grecque, les Babyloniens, dès le premier millénaire av. J.-C., utilisaient des calculs arithmétiques pour prévoir la position des planètes. Ces théories ont voyagé jusqu'en Perse et en Inde, où elles ont été adaptées et combinées à des méthodes autochtones. Les théories grecques de l'époque de Ptolémée et de son prédécesseur Hipparque sont également parvenues jusqu'en Inde, créant un matériel complexe dont les influences sont difficiles à démêler. Parmi les pratiques empruntées aux théories grecques, on relève l'emploi de termes – par exemple le titre du canon *Romaka Siddhanta* datant du début du V^e et qui marque les origines de la science astronomique sanskrite – ainsi que des modèles épicycliques et des méthodes de calculs requérant des paramètres numériques hérités d'Hipparque⁵.

2. Alexander Raymond Jones, *Ptolemy*, en, févr. 2024, URL : <https://www.britannica.com/biography/Ptolemy> (visité le 03/03/2024).

3. *Le Monde en Sphères*, fr, URL : <http://expositions.bnfr.fr/monde-en-spheres/> (visité le 26/05/2024).

4. “Le système géocentrique devient le modèle astronomique et théologique de l'Église, qui ne remet pas en cause la sphéricité de la Terre” *Ibid.*

5. Raymond Mercier, *Studies on the transmission of medieval mathematical astronomy*, 2004, URL :

La tradition chinoise se développe de manière relativement indépendante et les échanges ne débutent qu'autour de 200 av. J.-C.. Elle se distingue de celle des Grecs par un intérêt plus marqué pour la prédiction d'événements singuliers plutôt que pour les théories cosmologiques cherchant l'établissement d'un modèle d'organisation du ciel. En effet, en Chine impériale, l'astronomie a une fonction politique. L'empereur est considéré comme le Fils du Ciel et ainsi la régulation du calendrier, ou bien le succès (ou l'échec) de ses astronomes à prédire une éclipse, se reflétaient positivement ou négativement sur lui. L'inclusion croissante des diagrammes dans les traités après les missions jésuites à partir du XVI^e siècle révèle l'influence des pratiques d'Europe de l'ouest.

Pour conclure, l'*Almageste* se présente donc comme une sorte d'encyclopédie des connaissances d'une époque qui s'est enrichie avec le temps au point de rendre difficile l'appréciation de son état originel. Œuvre sans cesse recopiée au cours des siècles, passant du grec à l'arabe puis au latin, transmise à travers tout le bassin méditerranéen et dominant le Moyen Âge occidental après avoir conquis l'Islām, chaque traduction et chaque copie de l'*Almageste* n'ont pas seulement transmis son contenu, mais l'ont aussi adapté et enrichi en fonction des contextes culturels et scientifiques de chaque époque. L'œuvre ptoléméenne a servi de base à de nombreux commentaires et traités, intégrant progressivement des éléments de connaissance issus de diverses traditions scientifiques, et illustrant ainsi l'interconnexion des savoirs à travers les civilisations.

“Certains indices dans les manuscrits révèlent les emprunts intellectuels qui s’opèrent au fur et à mesure des copies ; les méthodes de calcul, le tracé des diagrammes, la mise en page des tables, la structuration des textes techniques, la mention d'auteurs antérieurs, la réutilisation de paramètres astronomiques ou même la récurrence de certaines erreurs sont autant de signes qui témoignent des échanges culturels qui ont façonné la pratique de l'astronomie.”⁶

Comme l'entend Albouy, les diagrammes font partie des révélateurs des échanges intellectuels.

1.1.2 Le diagramme vecteur de connaissances

L'historiographie et l'histoire des sciences n'échappent pas au récent “visual digital turn”⁷ général des humanités, montrant à quel point la production et la diffusion du savoir croisent les représentations visuelles rendant compte de ces connaissances. De fait,

<https://data.bnf.fr/temp-work/10f87cf325908d90792f9f269c04ee41/> (visité le 06/04/2024), p.6-7 in Ségolène Albouy, *Médiation des données de la recherche : Élaboration d'une plateforme en ligne pour une base de tables astronomiques anciennes*, fr, mémoire de master, École Nationale des Chartes, 2019, p.15

6. *Ibid.*, p.14.

7. “Digital humanities research has focused primarily on the analysis of texts. This emphasis stems from the availability of technology to study digitized text. Optical character recognition allows researchers to use keywords to search and analyze digitized texts. However, archives of digitized sources also contain large numbers of images.” M. Wevers et T. Smits, « The visual digital turn... ».

on ne s'intéresse plus seulement au texte. Or les astronomes, au fil de l'histoire, ont eu recours à une grande diversité de matériaux. Les sources primaires sont constituées par des instruments et des écrits, ces derniers eux-même hétérogènes. Dans les traités anciens, on trouve des descriptions détaillées, des propositions mathématiques, des tables de calcul, et des diagrammes illustrant souvent le texte qu'ils accompagnent. Les sources primaires peuvent en outre être enrichies de commentaires et de gloses, prose ou illustrations, témoignant de la manière dont les connaissances ont été transmises et interprétées. Elles révèlent également les méthodes pédagogiques employées pour enseigner ces savoirs.

Au cœur de cette diversité, le diagramme, objet hybride pour deux raisons : il combine un contenu géométrique (des lignes, arcs et cercles) et des labels, et il entretient un lien (plus ou moins étroit) avec le texte qui l'accompagne.

En tant que structure de pensée, la figuration géométrique – une forme de création de modèles associée à l'élaboration et à la résolution de problèmes dans divers domaines de la pensée humaine liés au calcul abstrait et à la modélisation des idées – est aussi ancienne que presque toute autre forme d'enregistrement des pensées et des idées. Des diagrammes utilisés pour calculer la superficie de parcelles et de terres apparaissent dans le Papyrus mathématique Rhind, la source la plus importante qui subsiste pour l'histoire des mathématiques dans l'Égypte ancienne⁸. Instruments de pensée et de démonstration, ils servent non seulement à transmettre le savoir mais aussi à le produire. Et dans un étrange mouvement métaréflexif, ils permettent aux historien.nes des sciences de produire la connaissance sur ces anciennes traditions heuristiques et leur transmission.

Les diagrammes sont, pour les astronomes, le support d'une pratique scientifique, et sont ainsi révélateurs de leurs méthodes, de leur contexte d'exercice et de leur conception de leur discipline. Ils revêtent des rôles et des aspects différents, permettant d'identifier des modes diagrammatiques⁹ spécifiques d'un lieu ou d'une époque, et traçant des lignes de diffusion des pratiques et des savoirs.

Au Moyen Âge, trois grandes cultures coexistent en Eurasie : les cultures byzantine, islamique, et d'Europe occidentale. Elles connaissent des évolutions différentes en termes linguistiques et religieux ; cette diversité est vraie également pour les usages auxquels les diagrammes astronomiques étaient destinés, pour les domaines dans lesquels ils étaient reconnus comme des instruments de pédagogie et des vecteurs de pensée, ainsi que pour la place accordée à la culture visuelle plus généralement et aux modes de représentation diagrammatiques. Pourtant cette coexistence donne lieu à des échanges intellectuels, artistiques, diplomatiques et commerciaux. Les traductions d'œuvres savantes, les transferts de manuscrits illustrent la porosité des frontières du savoir et de l'interdépendance des cultures. Les diagrammes astronomiques sont à ce titre témoins des chemins de diffusion

8. Linda Safran, Jeffrey Hamburger et David Roxburgh, *The Diagram as Paradigm : Cross-Cultural Approaches*, 2022, p.6.

9. La “diagrammatisation” désigne assez largement l’investissement des acteurs dans la complexification des représentations visuelles des propositions scientifiques présentes dans les traités.

des connaissance et des pratiques des sciences.

Comment ces diagrammes parlent-ils aux historien.nes ?

Les diagrammes peuvent être étudiés intrinsèquement (quelles conventions gouvernaient le langage visuel, quelle fonction assumaient-ils ?) ou extrinsèquement (pour comprendre la transmission de ces traditions et ces pratiques entre l'Europe et l'Asie, en passant par la péninsule arabique).

On pourrait penser que les formes et éléments visuels utilisés pour une démonstration géométrique soient universels, qu'ils restent les mêmes quelle que soit la date et la langue de l'explication textuelle, le grec, l'arabe ou le latin... Et pourtant le contexte géographique, temporel, et les aspects matériels liés aux technologies d'inscription changent profondément le fonctionnement et les objectifs des diagrammes, leurs objectifs.

L'évolution des conventions graphiques en sont un exemple frappant. Par exemple, l'axe vertical de la Terre, bien que représenté à plat sur la page, fut conventionnellement compris comme un axe perpendiculaire à la coupe du globe. Au fil du temps, de nouvelles conventions graphiques ont été adoptées, et pour les lecteurs d'aujourd'hui on représenterait sûrement le globe terrestre avec sa profondeur pour expliciter la représentation. Citons en outre la représentation des phases lunaires, qui a connu une évolution concernant l'association des couleurs claire et sombre à la pleine lune et à la nouvelle lune. Si aujourd'hui on associerait plutôt la pleine lune à un aplat de couleur claire et la nouvelle lune à une couleur sombre, les manuscrits médiévaux byzantins adoptent le référentiel contraire.

De même, le rôle du diagrammes est fluctuant, et va au delà du simple support démonstratif au service du texte. Par exemple ceux des *Traités logiques* d'Aristote ont probablement circulé indépendamment du texte, même si tout indique que les écrits les appelaient dès le départ. Cela souligne la distinction entre le diagramme en tant qu'objet de démonstration et de discussion complétant une proposition d'un côté, et le diagramme en tant qu'accompagnement des textes scolaires, qui aide à la compréhension de l'autre¹⁰. Le diagramme peut ainsi constituer la preuve et le support d'une réinterprétation de la proposition textuelle.

Par-dessus tout, les transformations subies au fil des copies et des réceptions sont éloquentes pour les chercheurs. Bien que les diagrammes soient initialement conçus pour clarifier et expliquer une proposition textuelle, ils peuvent parfois être des vecteurs de confusion (ou d'innovation). Le même diagramme d'une même œuvre soumis à un processus constant de transformation par les scribes, les artistes ou les lecteurs/commentateurs.

La recherche des erreurs transmises a ainsi un intérêt philologique important. Les cas de méprises et les malentendus sont peut-être plus nombreux que les cas de compréhension fidèle lors de la traversée des frontières géographiques, culturelles, religieuses et/ou linguistiques, et l'étude des erreurs et modifications révèle leurs aspects heuristiques,

10. *Ibid.*, p.5.

autant qu'il peut amener à l'établissement d'un stemma¹¹.

L'importance des diagrammes dans les transmissions est illustrée par l'exemple des diagrammes attribués à al-Ḥajjāj, en lien avec la transmission arabe des *Éléments* d'Euclide. Bien que la traduction originale d'al-Ḥajjāj soit perdue, les diagrammes retrouvés dans divers manuscrits montrent qu'il utilisait parfois des schémas différents de ceux adoptés dans la tradition arabe ultérieure. Ces diagrammes ont probablement joué un rôle clé dans l'élaboration d'une version alternative de la géométrie euclidienne, influençant ainsi la transmission vers l'Europe via les traductions latines et hébraïques¹².

Ainsi, au travers des variations, similarités et évolutions des diagrammes, les historien.nes peuvent reconstruire les pratiques scientifiques des astronomes et comprendre les contextes culturels et sociaux dans lesquels elles s'inscrivaient. De telles études permettent aussi de tracer la circulation des sources dans le monde entre les différentes cultures et de comprendre comment celle-ci s'approprient le contenu. En somme, l'évaluation de phénomènes diagrammatiques indéniablement disparates à travers des géographies éloignées permet d'identifier des modalités d'échanges culturels et leur impact sur la construction du savoir scientifique.

Les travaux antérieurs sur l'illustration scientifique se concentrent essentiellement sur des types spécifiques de diagrammes, situés dans des contextes déterminés chronologiquement et culturellement. Un exemple de ce paradigme est le projet précédent ALFA, qui porte sur les diagrammes de tradition alfonsine médiévaux un regard eurocentré. Cependant, à l'aune des remarques précédentes, il devient pressant de dépasser cette perspective centripète en étendant la portée géographique et temporelle des projets ; ambition rendue possible par la disponibilité des sources primaires en ligne, permettant la construction de bases de données d'images de grande envergure. Ainsi peut être mise en œuvre une analyse plus inclusive et diversifiée des sources iconographiques – notamment les diagrammes¹³.

Comme le dit Jeffrey F. Hamburger dans un plaidoyer pour une étude comparative des diagrammes astronomique : “Diagrams can thus be seen not as embodiments of eternal truths, but, rather, as culturally embedded objects”¹⁴.

“(...) to be effective, a cross-cultural comparison of diagrammatic traditions must look beyond the prima facies appearance of the diagrams under consideration to their underlying operations and the patterns of thoughts that they both codified and were intended to inculcate.”¹⁵

Les interrogations soulevées par le projet EIDA se déclinent donc comme suit : analy-

11. Dominique Raynaud, « Building the stemma codicum from geometric diagrams : A treatise on optics by Ibn al-Haytham as a test case », *Archive for History of Exact Sciences*, 68–2 (2014), Publisher : Springer, p. 207-239, URL : <https://www.jstor.org/stable/24569630> (visité le 07/04/2024).

12. Greg De Young, « Editing a collection of diagrams ascribed to al-Ḥajjāj : an initial case study », *SCIAMVS*, 15 (2014), p. 171-238.

13. Mathieu Husson, *EiDA - funding proposal*, 2022.

14. L. Safran, J. Hamburger et D. Roxburgh, *The Diagram as Paradigm...*, p.7.

15. *Ibid.*, p.3.

ser l'articulation entre les fonctions documentaires et épistémiques des diagrammes au sein de l'histoire des pratiques astronomiques, interroger l'importance des diagrammes dans la construction et la transmission des connaissances scientifiques, et enfin identifier des schémas récurrents dans les modalités de circulation de ces diagrammes. S'appuyant sur ces analyses, les chercheurs pourront à leur tour tracer des lignes : au sens figuré construire “a web of connections linking the points represented by the individual contributions together into a larger pattern”¹⁶; et au sens propre, cette vision globale sur la vie des images et des œuvres pouvant permettre l'établissement d'éditions critiques normalisées.

1.2 Genèse et écosystème du projet

Comme mentionné précédemment, les objectifs disciplinaires du projet consistent à étudier la diversité des fonctions et des modes de circulation des diagrammes astronomiques à l'échelle afro-eurasienne et sur une période de temps étendue. Cette approche trouve des assises conceptuelles et techniques chez des projets similaires (VHS) ou chez ses proches parents (DISHAS (*Digital Information System for the History of Astral Sciences*)), sans verrouiller pour autant l'émergence progressive des spécificités du projet.

1.2.1 Contributeurs et partenaires

Coordonné par Mathieu Husson, EIDA repose sur l'association fructueuse entre l'équipe d'historien.nes basée au SYRTE de l'observatoire de Paris prenant en charge la composante disciplinaire, et des chercheurs en computer vision basés au laboratoire IMAGINE (*Laboratoire d'Informatique Gaspard Monge*) de l'école des Ponts (Laboratoire d'Informatique Gaspard Monge, École des Ponts ParisTech). Ce dernier possède l'expertise en vision artificielle et apprentissage machine nécessaire à la réalisation des aspirations du projet, tandis que les chercheurs de l'équipe d'histoire des sciences développent une approche conceptuelle des sources.

Le projet VHS (Vision artificielle et analyse Historique de la circulation de l'illustration Scientifique), coordonné par l'Institut des sciences du calcul et des données, est également partenaire d'EIDA pour le développement des algorithmes de détection et de la plateforme de mise à disposition des modèles et des résultats. Mathieu Aubry, chercheur titulaire membre de l'équipe IMAGINE et PI du projet VHS, fait le lien entre les deux équipes. Les deux projets profitent de son expérience au sein de projets précurseurs comme le projet ANR EnHerit¹⁷ sur l'enrichissement des bases de données d'images du patrimoine.

16. *Ibid.*, p.10.

17. *Enhancing Heritage Image Databases*, en, URL : <https://anr.fr/Project-ANR-17-CE23-0008> (visité le 11/05/2024).

VHS s'intéresse tout comme EIDA aux illustrations et leur évolution dans les corpus scientifiques du Moyen Âge et de l'époque moderne. Les deux projets non seulement partagent certains développements (en fonction cependant de leurs objectifs spécifiques) mais se coordonnent aussi dans la production de publications académiques, de contenus pédagogiques, et l'organisation de séminaires pour la diffusion et la visibilité des résultats de la recherche.

EIDA/VHS créent alors un espace d'échange entre l'histoire des sciences et les pratiques contemporaines, concentrant l'expertise théorique des sources et celle, technique, du *deep learning* appliquée à la donnée historique.

1.2.2 Projets précédents

EIDA apparaît comme la suite logique de DISHAS, large projet de plateforme visant à rassembler les sources de la recherche en histoire des sciences astrales portant sur des corpus de traditions multiples et diverses, notamment des corpus chinois, sanskrits, arabes, latins et hébreux. Toutefois, les sources primaires astronomiques relevant d'une telle variété – textes théoriques et poétiques, diagrammes, instruments de mesures, entre autres – il n'était pas possible d'inclure tous ces objets dans une seule plateforme, les infrastructures numériques à concevoir pour leur étude étant trop différentes. Il a été décidé d'approcher ces types de documents les uns après les autres : si les tables astronomiques ont été choisies pour constituer l'unité fondamentale du projet DISHAS, EIDA porte sur les diagrammes. Fédérant plusieurs projets de recherche en astronomie ancienne, le projet pilote DISHAS considère un corpus pluriel, afin d'étudier les pratiques des astronomes et des milieux intellectuels, ainsi que la transmission et la circulation du savoir dans ces milieux. La conception d'une plateforme entend répondre à des problématiques de trouvabilité et de médiation des données de la recherche en histoire de l'astronomie.

“sans un effort de médiation la masse considérable des données produites par la communauté scientifique en SHS peut rester inaccessible et perdue dans des systèmes d'information obsolètes.”¹⁸

DISHAS et sa jeune sœur EIDA poursuivent donc un triple objectif commun : l'accès à la donnée, l'accès à des outils d'analyse et d'édition, et la valorisation des travaux des spécialistes. En élaborant des plateformes sur mesure pour exposer les données de la recherche et donner accès aux outils de traitement, des vecteurs d'analyse alternatifs deviennent disponibles pour la communauté académique. Ainsi DISHAS propose des outils d'édition, d'analyse et visualisation de tables astronomiques de traditions diverses allant de la Chine à l'Europe. EIDA poursuit des ambitions parallèles avec un autre objet : les diagrammes. L'ensemble de ces projets vise au long terme à l'élaboration d'un véritable

18. S. Albouy, *Médiation des données de la recherche...*, p.xix.

Système d'Information se voulant outil extensif pour le traitement, l'analyse et la mise à disposition des sources en histoire de l'astronomie.

Un premier programme EIDA, ALFA, examinait des sources issues du haut Moyen Âge en Europe, étudiant l'interaction entre diagrammes astronomiques, tables numériques et textes dans les manuscrits de tradition alfonsine. La révision du programme répond aux besoins, au sein de la discipline, d'étendre la typologie des diagrammes étudiés et de dépasser la perspective eurocentrée en élargissant la portée géographique des projets.

1.3 Objectifs

L'objectif commun des deux projets parallèles EIDA et VHS est de proposer une nouvelle approche de l'étude historique de la circulation des connaissances scientifiques basée sur de nouvelles méthodes d'analyse des illustrations. Le développement d'outils numériques pour l'étude des diagrammes astronomiques – notamment l'application de techniques de vision artificielle – permet d'automatiser une série de traitements et ainsi de faciliter leur analyse et exploitation. Ces outils ont pour objectif de réduire les étapes manuelles de fouille et d'annotation, optimisant ainsi l'efficacité des recherches, et permettant également de poser un regard neuf sur les objets.

1.3.1 Des outils d'exploration et d'analyse d'un large corpus

L'ouverture des données met à disposition des chercheur.ses des documents variés en masse, ce qui a des conséquences profondes sur les méthodologies de la recherche scientifique.

EIDA/VHS : des préoccupations proches

”There are tens of thousands of manuscripts and early prints in Latin, Greek, Arabic/Persian, Sanskrit and Chinese extent today which include different types of texts, numerical tables and diagrams. Indeed, astronomers made extensive and refined uses of non-discursive modes of expression, such as numerical tables and diagrams, in their practice. The analysis of the precise interactions of the different discursive (textual) and non-discursive (tables and diagrams) elements that compose these documents is key to our understanding of a history of astral sciences that goes beyond the “surface” of doctrines and astronomical models as presented in texts and opens up the potential of the history of astral sciences for global history”.¹⁹.

L'abondance d'archives et de documentation laissées par les astronomes prémodernes et modernes posent un défi majeur en termes d'exploitation en raison de la difficulté à mener des études à cette échelle. Cependant, le traitement par IA est prometteur, pour

19. M. Husson, *EiDA - funding proposal...*

accélérer l'exploration des corpus et interroger les modalités de circulation des savoirs scientifiques, ainsi que le rôle et la place de l'image dans la transmission des connaissances. Le but est de pouvoir donner du sens aux images en évitant au maximum les étapes manuelles d'annotation. Pour ce faire, EIDA et VHS développent des méthodes d'apprentissage non ou faiblement supervisées permettant d'effectuer des recherches automatiques à grande échelle dans des corpus d'envergure. L'utilisation du *deep learning* est centrale pour accomplir divers traitements analytiques qui accompagnent l'étude des modalités d'évolution et de transformation des images dans des corpus scientifiques illustrés.

La collaboration EIDA / VHS a pour but de créer un outil à la fois polyvalent (une sorte de couteau suisse) et précis, c'est pourquoi il est nécessaire de différencier et paralléliser les tâches en raison des finalités spécifiques de chaque projet. Cependant, une base commune repose sur deux aspects essentiels : la constitution des corpus numérisés et l'extraction automatique de leurs illustrations dans une base de données d'images au format IIIF (*International Image Interoperability Framework*), dotée d'une interface numérique de consultation et d'annotation partagée.

Divergences et finalités

Les objectifs de EIDA et VHS sont proches mais des divergences apparaissent en raison de la nature distincte des corpus, ce qui entraîne des finalités différentes dans la chaîne de traitement.

Côté VHS la vision artificielle sert l'analyse d'un large corpus scientifique illustré du Moyen Âge et de la période moderne, ne se limitant pas aux sciences astronomiques. La méthodologie adoptée consiste essentiellement à rechercher et identifier (dans un corpus de manuscrits témoins d'un même travail) les illustrations qui se correspondent. Cette tâche est ardue pour les ensembles de manuscrits contenant parfois des centaines d'illustrations, séparés par de nombreuses copies perdues, s'étalant sur des siècles, et qui ont pu être complètement réorganisés et fortement modifiés pour s'adapter à de nouvelles connaissances ou croyances²⁰.

"[N]otre approche consiste à concevoir des méthodes basées sur la vision artificielle pour détecter les similitudes iconographiques (c'est-à-dire des images qui ont été copiées ou partiellement inspirées les unes des autres) et textuelles (des images qui décrivent un contenu textuel similaire mais qui peuvent être visuellement différentes) entre des illustrations dans de grands corpus. L'un des principaux objectifs

20. "Most research on the automatic analysis of manuscripts and particularly their alignment, also known as collation, has focused on text. However, illustrations are a crucial part of some documents, hinting the copyist values, knowledge and beliefs and are thus of major interest to historians. One might naively think that these illustrations are much easier to align than text and that a specialist can identify them in a matter of seconds. This is only true in the simplest of cases, where the order of the illustrations is preserved and their content relatively similar. In harder cases however, the task becomes daunting and is one of the important limiting factor for a large scale analysis." Ryad Kaoua, Xi Shen, Alexandra Durr, Stavros Lazaris, David Picard et Mathieu Aubry, *Image Collation : Matching illustrations in manuscripts*, arXiv :2108.08109 [cs], août 2021, DOI : 10.48550/arXiv.2108.08109, p.1.

est d'obtenir ces associations automatiquement, en s'appuyant le moins possible sur les annotations d'experts, qui sont coûteuses et compliquées à obtenir. Nous faisons l'hypothèse que de telles méthodes d'analyse permettront d'effectuer beaucoup plus efficacement des comparaisons et des rapprochements pertinents entre images car elles fourniront, d'une part, de nouveaux regroupements d'images en fonction de leur contenu et des textes qu'elles illustrent et, d'autre part, des distinctions fines entre les différentes modalités de représentation par l'image.”²¹.

L'implémentation des méthodes de *deep learning* vise donc essentiellement la détection automatique de similarités entre les images, ce qui conduira à la constitution de séries iconographiques ; l'analyse des associations pouvant être interprétées par les historien.nes.

“VHS associe étroitement deux approches : d'une part, une approche historique qui perçoit l'image non pas comme une entité fermée et isolée, mais comme un vecteur essentiel dans la transmission des connaissances scientifiques ; d'autre part, le développement de méthodes automatisées d'analyse des similarités et des contenus dans des corpus illustrés médiévaux et modernes peu ou pas annotés.”²²

La phase d'extraction des illustrations garde une place d'importance, mais la fonctionnalité de vectorisation²³ (au cœur des objectifs d'EIDA) se trouve beaucoup moins pertinente, de par le type d'illustrations présentes dans le corpus VHS. Plus complexes, elles ne se prêtent pas – à l'inverse des diagrammes – à la décomposition en un ensemble de formes géométriques élémentaires. En outre la recherche de similarité constitue une perspective trop “haut-niveau” pour les besoins d'EIDA.

“However, these approaches typically neither enable fine-grained control of the reasons images are deemed similar, nor provide a fine-grained analysis of their content. Moreover, taking into account data specific or problem specific similarity notions will typically require the manual annotation of a large scale database, which is extremely costly and time consuming, limiting practical applications. Developing alternative solutions to these “black-box” deep learning approaches is thus a critical challenge”²⁴.

Conformément à cet objectif de produire une analyse détaillée du contenu des diagrammes, EIDA se focalise sur l'algorithme de vectorisation et extraction des labels, couleurs, et autres caractéristiques sémantiques spécifiques au diagramme astronomique. La démarche diffère alors de l'approche plus générique mais plus grossière basée sur les correspondances²⁵, centrale dans VHS, qui ne fournit pas le sens *computable* du contenu des images.

21. *Vision artificielle et analyse Historique de la circulation de l'illustration Scientifique*, fr, URL : <https://anr.fr/Projet-ANR-21-CE38-0008> (visité le 12/05/2024).

22. *Ibid.*

23. La vectorisation est le processus de conversion de données sous forme de listes de structures vectorielles, soit une série d'instructions simples, en XML (au format SVG), que l'ordinateur peut traiter très rapidement.

24. M. Husson, *EiDA - funding proposal...*

25. R. Kaoua, X. Shen, A. Durr, *et al.*, *Image Collation...*

EIDA adapte des méthodes existantes – notamment la toute récente approche “analysis-by-synthesis”²⁶ – pour identifier des primitives géométriques prédéfinies et gérer les éléments textuels (y compris les étiquettes de lettres). Les diagrammes sont suffisamment simples pour être analysés de cette manière, tout en étant suffisamment riches pour poser des défis significatifs à l’adaptation des méthodes de vision par ordinateur. Cette méthode permet une description fine et précise du contenu du diagramme, et surtout retourne un résultat lisible et compréhensible par la machine, dont l’exploitation se trouve de fait potentiellement automatisable.

La décomposition des diagrammes astronomiques en composants significatifs concourt à un double objectif : leur analyse et leur édition, en ayant recours au minimum d’annotation humaine.

“Central to EIDA is the decision to think together issues of diagram analysis and edition in the context of digital humanities enhanced with artificial intelligence.”²⁷

1.3.2 Développement d’une interface commune

La gémellité EIDA/ VHS tient à leurs objectifs très proches : la conception d’outils d’étude de la circulation des savoirs scientifiques dans l’histoire au prisme de l’illustration, en exploitant des méthodes d’analyse par apprentissage profond adaptées à des corpus anciens. La teneur des outils à concevoir varie, malgré tout il est nécessaire de développer une application unique pour les deux projets jumeaux, qui serve de pont entre les utilisateur.rices et les algorithmes de vision, et permette ainsi le dépôt et le traitement des sources par l’intermédiaire d’une interface graphique, accessible et utilisable par les chercheur.ses (dans un premier temps)²⁸.

Dans ce contexte, le développement en open-source et le partage du code entre deux projets de recherche offrent plusieurs avantages. Tout d’abord, cette approche permet de transcender les frontières disciplinaires, favorisant la transversalité et les discussions enrichissant les deux projets. La mise en commun des corpus est notamment bénéfique puisque les sujets de recherche sont connexes. Ensuite, cette stratégie implique une collaboration entre plusieurs ingénieur.es, permettant de mutualiser les expertises. Chaque

26. L’approche “analysis-by-synthesis” (analyse par synthèse) en vision par ordinateur est une méthode qui consiste à générer des prédictions d’objets, puis à comparer ces prédictions avec les données visuelles réelles pour vérifier leur précision. Cette approche intègre un cycle itératif : les différences entre les images prédites et les images réelles sont ensuite utilisées pour affiner les modèles. En itérant alternativement sur la synthèse et l’analyse, cette méthode peut gérer une grande gamme de variation dans les données visuelles, elle permet ainsi de traiter des problèmes en vision par ordinateur, y compris ceux impliquant des données complexes ou ambiguës à l’instar des diagrammes astronomiques. Cette approche s’oppose à la “sketch segmentation and labelling” (David Ha et Douglas Eck, *A Neural Representation of Sketch Drawings*, arXiv :1704.03477 [cs, stat], mai 2017, DOI : 10.48550/arXiv.1704.03477) qui gère mal la variabilité et l’ambiguïté des composants. La partie II de ce mémoire porte plus spécifiquement sur les méthodes de *computer vision* appliquées aux images du projet.

27. M. Husson, *EIDA - funding proposal...*

28. le projet d’une interface publique est en préparation

partie peut ainsi bénéficier des avancées et développements réalisés par l'autre, et alimenter un savoir-faire collectif. En outre, un problème récurrent dans le domaine des Humanités Numériques est la dépendance aux logiciels “maison” développés en interne et confrontés à des problèmes de pérennité lorsque les financements s'arrêtent. Le partage des budgets de recherche et des ressources humaines assure non seulement la durabilité des outils développés, mais permet également d'aller au-delà des objectifs initiaux en ouvrant à la continuité des développements. L'outil n'a pas seulement de meilleures chances d'être maintenu dans le temps, mais aussi d'évoluer, et le code étant ouvert, l'appui de la communauté open-source va assurer en partie ces missions. Enfin, éviter l'écriture de codes redondants s'inscrit dans une stratégie globale de sobriété, reposant sur l'interopérabilité et l'intégration des outils de recherche en général, celle-ci gagnant de fait en cohérence et en efficacité.

Pourtant, la coordination avec une autre équipe ne vient pas sans contraintes et questionnements. Le développement conjoint implique une stratégie de programmation modulaire, *a minima* pour répondre aux besoin d'EIDA et VHS, et qui sera poussée encore plus loin, comme on le détaillera plus avant²⁹. Le développement collaboratif via GitHub constitue une solution pour une implémentation *à la carte*, garantissant une interaction continue entre les équipes, tout en offrant la flexibilité de ne pas implémenter tous les développements effectués dans le cadre du projet parallèle. Les ingénieur.es des projets EIDA et VHS travaillent sur différentes branches d'un même dépôt, qui dispose d'une branche pour la mise en production de l'instance EIDA et d'une autre pour la mise en production de l'instance VHS. Alors, les développements de l'un ne sont pas nécessairement mis en production dans le cadre de l'autre, aboutissant à deux instances indépendantes de la plateforme. L'instance EIDA de la plateforme tourne sur les serveurs de l'observatoire et l'instance VHS tourne sur le serveur de la Sorbonne.

La dissociation de l'inférence des modèles de la plateforme collaborative répond à deux impératifs fondamentaux. D'une part, l'inférence des modèles requiert une puissance de calcul conséquente, typiquement fournie par des cartes graphiques GPU (*Graphics Processing Unit*). D'autre part, l'optimisation des ressources englobe le partage du *hardware*. L'Observatoire finance donc le GPU sur lequel tourne l'API (*Application Programming Interface*) dédiée à l'exécution des modèles de vision, laquelle est employée par les deux projets.

Les limites de l'approche collaborative touchent en outre à la coordination de la temporalité des deux projets. Par exemple, l'entreprise de refonte du modèle de données pour qu'il s'adapte aux sources d'EIDA, a imposé un ralentissement du côté de VHS. D'ailleurs, le défi tenant à la modélisation des données est de taille : comment conserver un degré de description suffisant en optant pour la généralisation des modèles de données ? Nous aborderons cette question dans la section suivante.

29. Voir le chapitre 3

Les difficultés tenant à la coordination des deux équipes sont par ailleurs apparues dans la mise en œuvre de l'annotation des similarités dans l'optique de *fine-tuner*³⁰ le modèle. La tâche consiste à évaluer à l'aide d'un score les résultats du modèle : pour ce faire, cinq catégories ont été définies par les chercheur.ses de VHS et le système de notation a été implémenté dans la plateforme avec les visualisations des résultats. La qualification des similarités correspond aux catégories 1, 2 et 3. Cette triple catégorisation vise à différencier les copies exactes (en autorisant des variations minimes comme celles tenant à un trait à main levée ou au compas) des copies partielles (notamment dans les imprimées : une partie d'une *type form* était souvent réutilisée) de deux images qui représentent un même contenu intellectuel sans forcément être identiques. La catégorie 4 indique un niveau de similarité non significatif, et la 5 est utilisée pour n'importe quel type d'association significative entre deux diagrammes, sans la qualifier. Cette catégorie est spécifique à l'utilisateur.rice et n'affecte ni l'annotation collective du document ni le ré-entraînement du modèle. La problématique est alors d'interpréter ces catégories dans le cadre d'EIDA.

La spécificité des diagrammes rend la tâche de définition des catégories ardue. Déjà, hors du monde de l'imprimé, il est impossible de trouver de similarité parfaite. En outre, elle ne peut être recherchée sur des critères autres qu'aspectuels, car des configurations géométriques semblables peuvent communiquer un contenu intellectuel très différent compréhensible seulement en contexte et par un œil expert³¹.

D'autre part, les diagrammes comportent souvent des labels, et quand un diagramme franchit des frontières linguistiques, les labels sont traduits dans un autre alphabet. Donc veut-on que le modèle soit sensible aux différences de labellisation ? Veut-on que le modèle considère comme similaire les diagrammes recopiés ayant traversés les barrières linguistiques ? Si oui il va falloir annoter comme tel les diagrammes avec des labels divergents, faisant ainsi baisser l'impact de la labellisation sur le score de similarité. Dans le cas contraire, on risque d'éloigner des doublons au titre de leur contenu textuel divergent.

30. Voir le chapitre 5 sur l'entraînement des modèles

31. Il convient d'annoter uniquement ce que l'algorithme est capable de comprendre, et il n'a accès qu'à l'image. Il est impossible de lui apprendre que deux diagrammes, bien que visuellement identiques, sont utilisés pour appuyer des arguments différents.

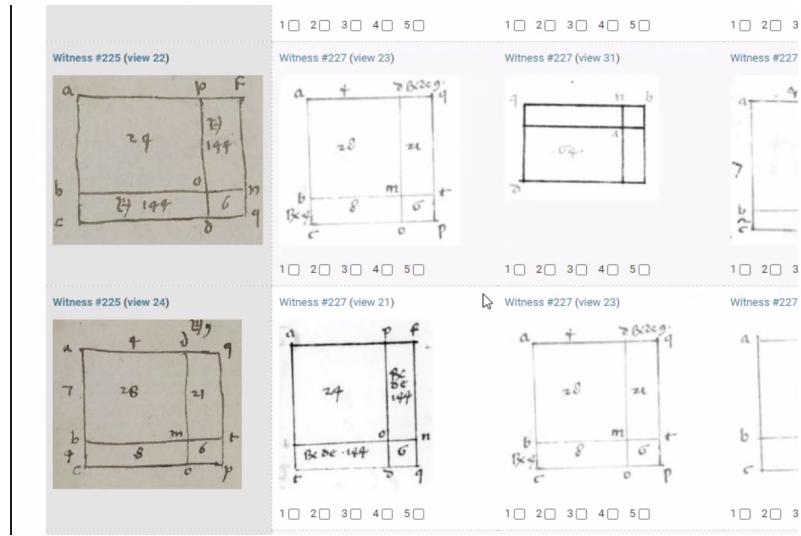


FIGURE 1.1 – Impact des labels sur la recherche de similarité.

Devant ces difficultés il a été décidé d'adopter une annotation binaire et basée uniquement sur le contenu graphique, partant du principe que l'assignation d'un score de similarité par le modèle ne vaut pas pour conclusion historiographique. Il sera toujours plus facile de ne pas prendre en compte un résultat, et la visée de cet entraînement est d'aboutir à un modèle assez extensif et généraliste. Il subsiste cependant des dissensions au sein de l'équipe entre une définition très stricte de la similarité, quitte à accepter des faux négatifs, et une définition plus souple (typiquement, sensible aux labels), quitte à créer des faux positifs. Pour concilier les besoins de tous les chercheur·ses, il sera recommandé d'annoter selon la définition la plus rigoureuse d'un point de vue scientifique, cette définition servant de métadonnée aux chercheur·ses. Les annotations les plus souples seront utilisées pour l'entraînement du modèle. Cette approche permet de garantir une précision dans la classification des données tout en offrant la souplesse nécessaire pour optimiser les performances des algorithmes d'apprentissage automatique.

Ce cas d'étude est assez caractéristique des enjeux tenant à la coordination au sein et entre des équipes de recherche dans le cadre du développement d'un outil commun. D'où le développement des outils d'annotation dans la plateforme commune assez généralistes pour que leur usage s'adapte à des besoins différents et complexes³².

1.3.3 Corpus

Les sources primaires d'EIDA couvrent un spectre géographique et temporel important. Orienté par les objectifs scientifiques d'EIDA – arriver à une représentation globale des continuités et divergences qui se tracent au cœur des pratiques astronomiques à travers

32. En réalité, les équipes de VHS sont aussi passées à un classement binaire pour prévenir des problèmes de coordination au sein de leurs équipes, gardant l'usage des trois autres catégories pour une prochaine étape de fine-tuning.

l’histoire, esquisser le voyage des sources à travers le temps et l’espace – le refus d’une vision eurocentrée justifie et explique une représentativité large, servie par la constitution de cinq grands corpus issus de sphères géographiques et temporelles diverses. Les manuscrits arabo-persans produits entre le VIII^e et le XIII^e siècle, des manuscrits latins médiévaux produits majoritairement entre le XIII^e et le XVI^e siècle, les manuscrits byzantins, produits entre les IX^e et XV^e siècles, les manuscrits sanskrits, à partir du XI^e, et enfin les sources chinoises datant du milieu du XVI^e siècle, après l’arrivée des premiers jésuites. Le support de ces dernières n’est pas nécessairement le manuscrit, elles prennent souvent la forme d’imprimés par blocs xylographiques dont les matrices sont réemployées dans plusieurs témoins. Elles présentent donc une forme hybride, sorte de pré-imprimé, qui posa des questions complexes pour la modélisation conceptuelle des données.

Plusieurs centaines de manuscrits sont numérisés pour chaque tradition. Sur ces numérisations, mises à disposition par les institutions patrimoniales qui conservent ces témoins, seront appliquées les traitements en prévision de l’analyse par les chercheur.ses, leur permettant de souligner les motifs qui sous-tendent la diffusion afro-euroasiens du modèle ptoléméen.

La gémellité avec VHS impose un effort de modularité pour pouvoir partager le modèle de données. Le travail de recherche de VHS est mené sur quatre corpus relevant, à l’instar de celui d’EIDA d’une grande diversité chronologique et géographique. Ces corpus se distinguent cependant par une grande diversité thématique : ils se composent de manuscrits et d’imprimés concernant les sciences des mathématiques et les sciences naturelles. Le premier corpus est le *Physiologus*, rédigé vers le II^e siècle à Alexandrie. Ce texte, l’un des plus populaires du Moyen Âge, a contribué à l’émergence de la zoologie chrétienne médiévale. Il a pour témoins 100 manuscrits grecs, dont treize sont illustrés et réalisés entre le XI^e et le XVI^e siècle, contenant environ 680 images d’animaux, de plantes et de minéraux. Le deuxième corpus est le *De materia medica* de Dioscoride, composé vers l’an 77 de notre ère. Ce traité pharmacologique destiné aux praticiens a été largement diffusé et copié. Il est conservé dans 65 manuscrits grecs, dont 17, réalisés entre le VI^e et le XVI^e siècle, sont illustrés et contiennent environ 8340 images de plantes, d’animaux et de minéraux. Les troisième et quatrième corpus se composent des planches de l’Encyclopédie de Diderot et d’Alembert (1751-1772), les témoins de ce travail couvrant une longue série de traités, dictionnaires et encyclopédies au fil desquels les illustrations ont été copiées et retravaillées. Leur étude se focalise aussi sur leur inclusion ultérieure dans des encyclopédies telles que l’Encyclopédie méthodique. Le troisième corpus se concentre sur l’Histoire naturelle (Zoologie), tandis que le quatrième porte sur les sciences mathématiques³³.

Le deux corpus jumeaux diffèrent aussi en taille : celui de VHS présente plus de 2000 témoins devant 300 pour EIDA.

33. VHS Project : Presentation, en-US, URL : <https://vhs.hypotheses.org/presentation> (visité le 12/05/2024).

Malgré ces divergences importantes, les données sont issus de spectres assez larges pour opérer des croisements dont l'exploitation peut s'avérer fructueuse : les corpus VHS présentent quelques diagrammes astronomiques, et les manuscrits EIDA présentent plusieurs images de plantes. La mise en commun des résultats permet d'établir des liens entre le Moyen Âge (focus d'EIDA) et la période moderne (focus de VHS) pour les domaines respectifs étudiés. Notamment, des chercheur.ses d'EIDA s'intéressent à la transition du manuscrit à l'imprimé. La possible jonction des corpus reste néanmoins problématique, et les modèles d'IA doivent être spécialisés sur les objets d'étude respectifs des deux projets.

Ces remarques sont révélatrices de l'entre-deux qui nous intéresse dans ce mémoire : bien que les projets aient un objectif et des objets d'étude précis, ceux-ci se trouvent élargis par le réseau institutionnel et les dynamiques collaboratives qui accompagnent la création des outils numériques. Le niveau de modularité de l'outil créé devra prendre en compte cette généralisation possible sans perdre de vue les objectifs initiaux des deux projets.

1.3.4 Modèle de données

L'application commune à VHS/EIDA est développée avec le *framework* Django et adossée sur une base de données gérée avec PostgreSQL. Le modèle de données conçu pour l'application a fait l'objet de réflexions et réadaptations pour répondre aux besoins de description des sources de VHS et EIDA, tout en restant suffisamment flexible pour être utilisé par d'autres projets souhaitant reproduire les méthodes employées par les deux projets. Le défi consiste à conceptualiser la donnée de manière suffisamment spécifique et assez généraliste.

Modèle initial

Le modèle de données initialement construit³⁴ pour l'application VHS prévoit l'existence des entités 'manuscrit' et 'imprimé', qui correspondent aux supports représentés dans les corpus d'EIDA et de VHS. Ces deux *états matériels* du texte sont reliés à une entité plus abstraite : le *Work*. S'appuyant sur la définition de 'work' que donne le CIDOC-CRM³⁵, un *Work* est indépendant de sa forme matérielle, c'est une production intellectuelle qui peut se manifester dans différentes sources, pouvant elles-mêmes présenter des variations.

34. Voir l'annexe A pour une illustration de l'évolution du modèle de données.

35. Le modèle conceptuel définit, dans un domaine donné, comment représenter la réalité. À ce titre il est indépendant de la manière dont on stocke les données informatiquement. Né dans le domaine des musées, CIDOC-CRM est un modèle conceptuel standardisé pour la modélisation des informations dans le domaine du patrimoine. Il vise à décrire et à rendre interopérables des objets du monde culturel en général. La notion d'événement se trouve au cœur du modèle, traduisant une approche monotonique des données : les objets décrits peuvent évoluer, on peut toujours décrire de nouveaux événements liés à cet objet. Ainsi, on n'aura potentiellement jamais terminé de décrire un objet en CIDOC-CRM.

Mais ce premier modèle va rapidement montrer des limites, notamment liées à la description d'une même œuvre divisée en plusieurs ouvrages, ou d'un ouvrage contenant plusieurs œuvres. De plus, la pertinence de la distinction absolue des manuscrits et imprimés ne permet pas une description pertinente de tous les objets : par exemple les sources chinoises, dont le support n'est pas nécessairement le manuscrit, prennent souvent la forme d'imprimés par blocs xylographiques dont les matrices sont réemployées. Ces constats ont mené à la refonte de ce modèle de données.

Première refonte

Le nouveau modèle est non seulement plus pertinent mais aussi déjà plus flexible. Il est centré autour de l'entité 'témoin' ou *Witness*, et est complété par d'autres tables pour décrire la diversité des sources et leurs différents modes d'existence. Ainsi, les séries héritent uniquement des imprimés et les témoins héritent à la fois des manuscrits et des imprimés. Une *Serie* est une édition en plusieurs volumes d'une œuvre imprimée. Le témoin ramène à chaque volume en tant qu'entité matérielle de référence, ou à un manuscrit. Il peut contenir un ou plusieurs 'travaux' (*Work*) ; un même *Work* peut correspondre à des séries et des témoins différents (table de relation *Content*). La table *Work* inclut donc les références au lieu et à l'auteur (clés étrangères), mais aussi l'information relative à la plage de temps durant laquelle cette œuvre a existé physiquement (durant laquelle on atteste l'existence de témoins). La ou les numérisations (*Digitizations*) sont reliées au *Witness* et contiennent les détails sur la source de numérisation et l'état des traitements via IIIF³⁶. Une entité Tag est utilisée comme moyen de différencier manuscrit, imprimé ou gravure, et la table de relation *Tag_Exemplar* fait le lien entre le témoin et les tags. La table *Role* établit des relations entre des contenus (*Content*), des séries (*Series*) et des personnes (*Persons*), attribuant à ces dernières les rôles d'auteur ou d'éditeur.

Les évolutions en cours pendant mon stage, détaillées en partie III³⁷, ont nécessité de faire des choix complexes, répondant à une double exigence *a priori* paradoxale : une plus grande flexibilité pour intégrer de nouvelles sources de données et une précision de description suffisante. De nouvelles questions ont été soulevées liées à la granularité et à la cohérence sémantique.

Ce paradoxe apparent met en lumière les défis et la complexité liés à la modélisation de la donnée, et les transformations successives reflètent des questionnements : comment rendre l'application la plus universelle possible, l'ouvrir à une large diversité de sources, sans renoncer à la qualité de la description des sources des projets de EIDA et VHS ? Comment, d'ailleurs, conceptualiser un modèle qui convienne aux deux projets et leurs objectifs ? Les solutions ci-dessus décrites sont satisfaisantes mais amenées à évoluer, montrant que la modularité se construit sur le temps long.

36. Voir le Chapitre 2 sur IIIF

37. Voir le Chapitre 7

En résumé, EIDA repose sur la constitution de jeux de données de diagrammes astronomiques extraits de sources relevant de multiples domaines linguistiques et culturels, sources connectées à l'échelle afro-eurasienne. Le projet développe des outils numériques permettant une approche critique des diagrammes astronomiques à large échelle, fournissant à ce titre une base solide pour une étude approfondie des schémas de circulation des connaissances astronomiques entre le Moyen Âge et l'époque moderne.

En *computer vision*, EIDA s'appuie sur les récents progrès en *deep learning*, et teste l'application d'une nouvelle génération de méthodes vectorisation d'images sur des sources historiques : manuscrits, *early prints* et xylogravures.

Au centre de ce projet réside une approche intrinsèquement transversale et interdisciplinaire, qui s'appuie sur des projets antérieurs et bénéficie de la collaboration avec des équipes dotées de compétences en ingénierie, ainsi qu'avec le projet ANR VHS ; bien que les objectifs de ce dernier divergent de ceux d'EIDA. Ainsi, au sein même de l'écosystème d'un même projet de recherche, les livrables doivent s'adapter aux usages et besoins des chercheur.ses, à la diversité de leurs questions de recherche et enfin à deux corpus différents dans toute leur hétérogénéité. De fait, la problématique de modularisation est au cœur du projet VHS/EIDA. Cette modularisation impose de créer des composants logiciels indépendants et réutilisables, facilitant ainsi l'adaptation du projet aux deux contextes, comme nous le développeront plus avant. En outre, pour garantir une approche conforme aux principes de l'open-source et de la science ouverte, une couche de standardisation doit être ajoutée. Cette standardisation vise à rendre les composants du projet interopérables, permettant ainsi une collaboration transparente et la réutilisation des résultats de recherche au sein de la communauté scientifique. En intégrant ces principes dans la conception et le développement du projet, il devient possible de promouvoir la transparence, la reproductibilité et l'accessibilité des outils techniques (modèles de vision) et des corpus enrichis.

Chapitre 2

Open-data et enjeux d'interopérabilité : la standardisation technique à grande échelle

Les acteurs impliqués dans la réalisation d'un outil tel que la plateforme commune AIKON ne se limitent pas aux participants du projets évoqués dans la partie précédente. Les potentiels utilisateur.rices futurs et les institutions détentrices des données sont aussi à prendre en compte, ils orientent les développements et les fonctionnalités des outils conçus.

Le traitement des sources iconographiques revêt des enjeux spécifiques du point de vue technique, juridique et de la disponibilité. La remise en circulation des résultats des traitements, à l'autre bout de la chaîne, présente d'autres défis et soulève des questions de visibilité, d'accès, et de médiation qui sont autant d'exigence à prendre en considération pour garantir l'utilité des résultats pour l'ensemble des utilisateur.rices et partenaires impliqués.

Enfin, la mutualisation du code induit des enjeux d'organisation devant aussi être pris en considération.

Ces enjeux contraignent le projet à rentrer dans un cadre technique rigoureux où plusieurs aspects clés doivent être respectés : l'utilisation de standards ouverts pour garantir la compatibilité et l'interopérabilité entre différents systèmes et outils, l'adoption de formats de données standardisés pour faciliter le partage et l'échange d'informations, la conformité avec les réglementations juridiques en matière de droits d'auteur et de protection des données, et la prise en compte de l'interaction avec des utilisateur.rices divers.

“une compatibilité technique et sémantique, l’interopérabilité des données aide au décloisonnement entre domaines d’expertise et permet à des projets de recherche aux objectifs divergents de s’appuyer sur un socle commun.”¹

1. S. Albouy, *Médiation des données de la recherche...*, p. xvii.

2.1 Ouverture des données : accès aux sources et partage des résultats de la recherche

La dernière décennie observe un changement de paradigme dans la recherche par la multiplication des actions concertées entre les institutions culturelles et patrimoniales, les établissements de l'enseignement supérieur, les chercheur.ses, les personnels et les étudiants du monde. La production scientifique sans frontières qui en résulte repose sur une nouvelle stratégie de coopération internationale dans le cadre de partenariats qui facilitent la circulation des connaissances. Or les outils numériques sont vecteurs d'affirmation de ces modes d'irrigation du savoir.

2.1.1 Enjeux de la récupération des sources numérisées

Dans le cadre d'un projet tel qu'EIDA, qui repose sur des traitements automatiques des sources par des algorithmes de vision par ordinateur, la disponibilité des sources numérisées constitue un principe fondamental de la démarche de recherche.

La numérisation, qui transforme un objet physique en fichier image (matrice de pixels), opérant un passage de la matière en données, est un premier pas pour accès et rendre visible les documents patrimoniaux, notamment les sources de la recherche destinées à être enrichies et étudiées. La numérisation n'a en effet pas vocation à interpréter mais à citer l'objet, elle offre une image la plus proche possible de l'objet en prenant en compte l'intégralité du support.

L'examen exhaustif des nombreuses sources inexploitées disponibles dans les archives et les bibliothèques à l'aide des méthodes classiques n'est pas viable. Cette limitation a conduit, ces dernières années, à l'émergence de nombreuses initiatives visant à numériser ces collections de documents historiques. Parmi ces initiatives, on peut citer Google Books Search (GBS) et l'Open Content Alliance (OCA). Objet d'une attention particulière par les bibliothèques et les musées depuis les années 90, la numérisation s'inscrit généralement pour les institutions dans un vaste programme de mise à disposition des documents qui peuvent alors circuler via les canaux d'internet, en facilitant la consultation, spécialement à destination du monde académique.

Si la numérisation vise avant tout, outre les questions de conservation, à donner accès, l'enjeux de la valorisation des documents, fonds et collections des bibliothèques et des musées, est également présent. Quand on numérisé pour faciliter la consultation, on vise un public déjà familier du contenu. L'affranchissement de l'hétérogénéité des supports permis par le passage au format numérique ouvre en outre de nouvelles voies d'exploitations techniques.

Certains grands opérateurs assurent alors des programmes ambitieux de numérisation et de valorisation des contenus culturels. La Bnf (*Bibliothèque Nationale de France*)

en est l'exemple le plus probant, numérisant plus d'un million de pages par mois à partir de ses collections patrimoniales². Depuis 2013, cette production est complétée par des documents numériques produits dans le cadre des accords conclus par BnF-Partenariats.

De vastes projets, à l'image de Gallica ou Européna, larges campagnes de numérisation accompagnées de plateformes de mise à disposition et de valorisation, voient alors le jour. Les collections se diffusent à l'international ; Internet devient un espace permettant à des utilisateur.rices aux profils variés d'exploiter ce patrimoine culturel numérique au-delà des limites physiques de consultation des documents.

Malgré le soutien des acteurs publics, les disparités nationales et internationales subsistent en matière de numérisation et mise en ligne des collections, disparités liées à la capacité des institutions à porter ces projets souvent coûteux. De nombreux facteurs rentrent en jeu : financements, qualité des données, volume de documents à traiter. Ces inégalités constituent un biais important à prendre en compte, notamment pour les projets de recherche s'appuyant – à l'instar d'EIDA – sur des sources internationales. On gardera donc à l'esprit que les corpus numérisés rassemblés auprès des institutions chinoises, indiennes et européennes ne sont qu'un fragment de la réalité des sources existentes, tout en étant suffisamment représentatif pour permettre de tirer des conclusions pertinentes.

Un autre défi tenant à la valorisation des collections et des projets de recherche concerne la question juridique : le respect du droit d'auteur. Toute diffusion d'une numérisation doit se faire dans le respect de l'ensemble des dispositions du code de la propriété intellectuelle, en France il se décline en deux grandes catégories. Le droit moral est perpétuel, inaliénable et imprescriptible, il comprend plusieurs prérogatives : le droit de paternité, de divulgation, la protection de l'intégrité de l'œuvre. Le droit moral ne peut être cédé ou vendu, et il subsiste même après la mort de l'auteur, étant alors transmissible à ses héritiers. Le droit patrimonial concerne les aspects économiques de l'œuvre et permet à l'auteur de tirer des revenus de son exploitation. Il inclut principalement les droits de reproduction et de représentation. Limité dans le temps, il dure toute la vie de l'auteur puis 70 ans après sa mort. Après cette période, l'œuvre tombe dans le domaine public et peut être librement utilisée par tous, tout en respectant le droit moral de l'auteur. Les principes fondamentaux des droits d'auteur sont largement reconnus à l'international, notamment grâce à la Convention de Berne et aux accords ADPIC. Cependant, les spécificités, notamment en matière de droit moral et de durée de protection, peuvent varier selon les législations nationales.

Dans le cas de corpus sur lesquels ne s'appliquent pas de droits de propriété intellectuelle, c'est-à-dire les corpus qui font partie du domaine public, il renvient à la Bibliothèque d'établir quelle licence va conditionner leur distribution et leur réutilisation. Elles n'imposent généralement pas de nouvelles restrictions, privilégiant la mise à

2. *La numérisation à la BnF*, fr, URL : <https://www.bnf.fr/fr/la-numerisation-la-bnf> (visité le 25/05/2024).

disposition. Il existe cependant des exceptions : même si les manuscrits et les textes originaux peuvent ne plus être sous droits d'auteur en raison de leur ancienneté, les copies numériques réalisées par la Bibliothèque du Vatican sont protégées par leurs politiques spécifiques de droit d'auteur. Toute reproduction ou diffusion sans autorisation est interdite, une considération importante dans l'optique de la diffusion des corpus de recherche à un large public.

2.1.2 Projet d'une plateforme web

La science ouverte renvoie au principes qui veulent rendre la recherche – ses méthodes et ses résultats – accessible à tous. Les chercheur.ses qui publient conformément à ses principes visent à plus de transparence, de collaboration et à une recherche plus étendue et plus efficace. Les principes de la recherche (ou science) ouverte veulent rendre les méthodes de recherche et les données résultantes librement disponibles, souvent via Internet, afin de soutenir la reproductibilité et, potentiellement, la collaboration entre des projets largement diffusés. À cet égard, elle est liée à la fois aux logiciels open-source et à la science participative : une approche de la recherche scientifique dans laquelle des membres du grand public contribuent au processus, ceci à divers degrés, de la collecte des données à l'encouragement à l'éducation scientifique informelle.

Conformément aux principes de la science ouverte, EIDA vise au long terme le développement d'une interface web incluant à la fois le front-end³ et le *back-office* pour collecter, étudier, éditer et visualiser des diagrammes, modelée sur la plateforme DISHAS existante pour les tables astronomiques. La nouvelle plateforme permettra à la communauté de recherche d'accéder aux outils utilisés pour produire les résultats publiés par EIDA, soit les outils de traitement ou la future plateforme d'édition, aidant ainsi la communauté scientifique à mener et publier des travaux portant sur les illustrations *a minima* à caractère historique, voire moins spécifiques⁴. L'application dédiée aux chercheur.ses de EIDA et VHS est développé en gardant à l'esprit son utilisation future par un public plus vaste que les seuls chercheur.ses affiliés à ces projets. Cette approche vise à rendre les outils accessibles et utilisables par une audience diversifiée, tout en restant centrée sur le monde académique.

À l'été 2024, il existe uniquement l'interface administrateur – ou *back-office* – pour la gestion de la base de données et la mise à disposition des modèles de vision. L'accès à cette plateforme est conditionnée par les droits admin accordés au chercheur.ses du projet, et à tous ceux qui le demandent. À terme, cette plateforme est amenée à être dotée d'une interface publique déployée en ligne, donnant un accès à la base de données

3. Le développement web frontal, ou *front-end*, désigne les productions d'une application avec lesquelles l'utilisateur.rice peut interagir directement.

4. Le niveau de spécialisation de ce système dépend justement des efforts de modularisation et d'interopérabilité dont la mise en œuvre sera discutée en partie III.

et aux résultats des traitements⁵. Les enjeux tenant à la construction de cette plateforme seront détaillés en partie III.

En suivant les principes FAIR (*Findable Accessible Interoperable Reusable*), la plateforme garantira l'ouverture des outils et des résultats de recherche, facilitant ainsi une collaboration et une diffusion plus larges des connaissances.

2.2 IIIF

IIIF (International Image Interoperability Framework) a révolutionné les approches pour la diffusion et la manipulation des images numériques.

Née au tournant des années 2010 sous l'impulsion de l'Université de Stanford, l'initiative émerge, dans un contexte de numérisation massive, de questionnements et de besoins spécifiques aux manuscrits médiévaux. Pour mettre côte à côte deux manuscrits issus de deux bibliothèques différentes, pour les annoter collectivement de transcriptions et commentaires, ou encore pour reconstruire virtuellement un témoin à partir de fragments dispersés, il fallait pouvoir s'affranchir des verrous techniques, ouvrir les silos de données et les systèmes cloisonnés, monolithiques, ayant prévalu jusque-là dans la conception des bibliothèques numériques. Un cadre global interopérable basé sur les standards du web, dans lequel des applications diverses peuvent dialoguer de manière normalisée avec des fournisseurs d'images, est alors créé.

Un consortium international est aujourd'hui chargé de maintenir et faire évoluer les spécifications techniques, et de les implémenter dans les logiciels et plateformes diverses (serveurs d'images, visualiseurs, outils d'annotation, systèmes de gestion de contenus, etc.), garantissant ainsi leur interopérabilité.

2.2.1 Les APIs de IIIF

Le cadre défini par le consortium IIIF repose sur plusieurs APIs (interface de programmation applicative), dont les deux principales sont l'API image et l'API présentation.

L'API Image est un “contrat” entre un serveur d'images et un client de visualisation, un mécanisme simple de fourniture d'images numériques en ligne, y compris en haute résolution, et permettant le zoom profond. Elle vise à rendre plus cohérente la diffusion en ligne des numérisations et faciliter leur exploitation dans différents environnements : l'image peut être affichée dans un visualiseur, ou dans un navigateur. L'API image se constitue ainsi comme la base de toutes les APIs IIIF, et la première à avoir vu le jour, elle peut être implémentée sans l'API Présentation.

5. Une version *démo* de la plateforme est actuellement en développement pour tester les modèles de vision sur des images personnelles et sans stockage. Elle sera déployée sur les serveurs de l'École des Ponts, en utilisant une instance de l'API développée dans le cadre du projet EIDA

Elle apporte une syntaxe d'URL (*Uniform Resource Locator*) standard et facilite la mise en cache pour accéder à une image et la manipuler à distance (avec la possibilité de paramétrier l'URL). Elle transmet ainsi des informations techniques sur l'image pour qu'un visualiseur puisse l'afficher dans différents contextes : en effet cette syntaxe d'URL concerne deux types de requêtes sur une même image : la requête de l'image elle-même (pixels) et celle d'information sur l'image (JSON (*JavaScript Object Notation*)).

Le modèle d'URL⁶ permet à un client d'appeler et manipuler une image selon des paramètres de région (region), de taille (size), de rotation (rotation), de qualité (quality) et de format (format). Grâce à cette URL paramétrable, à partir d'un même document mis en ligne, il est possible de créer plusieurs dérivés d'une même image. Par exemple il est possible de demander sa version pleine taille, ou simplement une portion de celle-ci (zone d'intérêt au sein de l'image), de la redimensionner, de la faire pivoter ou la retourner horizontalement ou verticalement, de la passer en noir et blanc, etc.

L'API Image permet donc de délivrer et partager des images en spécifiant un mécanisme qui ramènent au client les pixels d'une image et lui permet de manipuler l'image à distance à travers une syntaxe d'URL normée, mais elle ne traite qu'une seule image à la fois.

Intervient alors l'API Présentation, qui est à la fois un format d'échange et un modèle décrivant la représentation numérique d'un objet, sa structure interne, ses métadonnées, ses liens avec d'autres ressources. Toutes ces informations sont contenues dans un fichier appelé *Manifest* et structurés à l'aide du format JSON. C'est une sorte d'enveloppe virtuelle formant l'unité de distribution élémentaire dans l'univers IIIF. Il fournit des métadonnées essentielles et organise les images associées de manière cohérente pour permettre leur récupération et leur manipulation. C'est en général ce que manipulent les logiciels pour interagir avec une ressource, la visualiser, ou la transférer vers un autre outil.

Un *Manifest* IIIF représente le plus souvent le fac-similé numérique d'un objet physique (dans les cas qui nous occupent, il s'agit d'un manuscrit ou d'un imprimé). Mais un *Manifest* peut aussi bien représenter un objet virtuel composite en référençant de nouvelles composantes, notamment les annotations. C'est à ce titre que les technologies IIIF sont de plus en plus mobilisées dans des projets de reconnaissance et d'indexation automatique de formes ou de texte, basés sur l'intelligence artificielle.

2.2.2 Anatomie d'un *Manifest*

Le *Manifest* IIIF constitue la structure fondamentale pour la représentation des ressources visuelles. Il est composé d'une séquence ordonnée de "canevas", auxquels sont associés un ou plusieurs "contenus". Les modèles des différentes versions de l'API Présentation présentent quelques différences, mais en substance les types de *Manifest* restent

6. {scheme} ://{server}{/prefix}/{identifier}/{region}/{size}/{rotation}/{quality}.{format}

identiques. Leur tronc commun peut être représenté comme sur la figure 2.1.

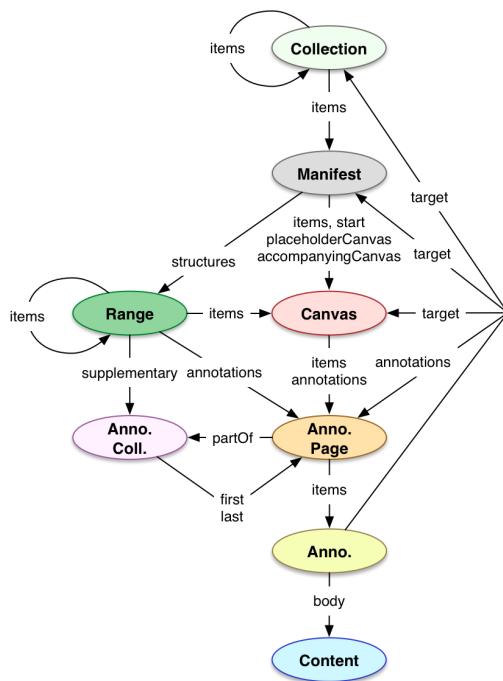


FIGURE 2.1 – Structure d'un *Manifest* IIIF.⁷

Un canevas peut être considéré comme un espace abstrait sur lequel un contenu peut être “peint”. Chaque canevas possède une largeur et une hauteur qui définissent un espace de coordonnées en deux dimensions, permettant ainsi de positionner précisément un ou plusieurs contenus.

Les contenus, servent à remplir un canevas, il s’agit typiquement du contenu visuel identifié grâce aux URLs définis par l’API image. L’association entre un Contenu et un canevas se réalise à travers une Annotation, qui peut soit occuper la totalité de l’espace du Canevas, soit cibler une zone spécifique de celui-ci. Un canevas peut donc aussi aussi être la cible d’autres types d’annotations (transcription, commentaires, tags etc.) ciblant des zones particulières au sein de cet espace de coordonnées.

Cette structure de données permet donc l’alignement des annotations et des images, y compris si l’utilisateur.rice producteur des annotations est différent du créateur du *Manifest*; ce qui facilite la collaboration de plusieurs acteurs autour d’un *Manifest* produit par une institutions.

Les annotations deviennent à ce titre un élément important et opérationnel dans le processus de localisation et d’indexation des coordonnées des diagrammes extraits par le modèle de vision artificielle. Elles permettent de lier à chaque page de manuscrit et d’afficher dans un visualiseur la localisation et l’étendue spatiale des objets détectés ; et ce en récupérant uniquement leurs coordonnées spatiales.

Pour l’entraînement, elles permettent de marquer et de délimiter les zones spécifiques

de l'image où se trouvent les objets que le modèle de vision artificielle doit identifier ; pour la correction des prédictions, elles permettent de modifier la taille et la place d'une annotation prédite ou d'en créer une nouvelle. La collaboration, enfin, est essentielle dans les projets impliquant l'annotation à grande échelle de vérités de terrain.

2.2.3 Avantages et limites du cadre IIIF

Interopérabilité et ouverture de la données

IIIF garantit l'accessibilité, favorisant la collaboration et l'échange entre institutions culturelles, universitaires et de recherche, participant à la visibilité des collections numériques des institutions. En termes d'interopérabilité, IIIF élimine les barrières techniques entre différentes plates-formes et systèmes, permettant une intégration transparente des collections d'images. Le cadre et ses protocoles associés facilitent la recherche et l'analyse d'images en permettant aux chercheurs d'agrégier des données provenant de multiples sources et de les examiner dans un environnement unifié, notamment avec des outils permettant l'annotation collaborative des images, tels que Mirador⁸ ou Universal Viewer⁹, ou SAS (*Simple Annotation Server*)¹⁰.

Légèreté

“[IIIF] permet de réunir des corpus de sources très diverses, sans devoir les restocker ; quel que soit le serveur d'où est publié le document en question, celui-ci peut être visualisé et comparé avec un document servi par n'importe quelle autre institution publiant des numérisations IIIF ; ce standard permet aussi, à partir d'une simple URI (*Uniform Resource Identifier*), de récupérer toutes les pages d'un volume particulier dans le bon ordre, et les métadonnées qui le décrivent.”¹¹

La solution apportée par l'API Image répond notamment à la problématique de l'optimisation du débit de transmission de l'information lors de la manipulation des éléments graphiques, celle-ci pouvant être de très haute résolution et de grande taille. En accédant directement à des extraits spécifiques de l'image d'origine via de simples transformations d'URL, l'API Image permet de la récupérer uniquement à la taille exacte requise pour une tâche particulière. Cette flexibilité réduit significativement la quantité de données transférées sur le réseau, améliorant la fluidité des échanges, essentielle pour les interfaces interactives et les processus automatisés. L'optimisation du temps de traitement constitue en effet un enjeux spécifique lié au traitement de l'image, plus lourde que le texte et plus gourmande en puissance de calcul.

8. <https://github.com/ProjectMirador/mirador>

9. <https://github.com/UniversalViewer/universalviewer>

10. SAS est un serveur d'annotations IIIF. Il est couplé à la version 2 de Mirador pour la partie cliente dédiée à la visualisation des images et à la création des annotations. La partie serveur sert à stocker les annotations de façon persistente. (<https://github.com/glenrobson/SimpleAnnotationServer>)

11. Robin Champenois et Béatrice Joyeux-Prunel, « Visual Contagions : extraire et tracer la circulation d'images dans des imprimés illustrés », dans *Humanistica 2023*, Genève, Switzerland, 2023 (Circulations), URL : <https://hal.science/hal-04108205> (visité le 15/05/2024), p.6.

Ainsi, en donnant accès aux images haute résolution avec des métadonnées structurées, IIIF facilite l'analyse visuelle et algorithmique des images. La mise en œuvre de ce protocole standardisé participe donc au renouvellement du dialogue entre le milieu de la recherche et les bibliothèques détentrices des données.

Cependant IIIF présente des limites dans son universalité. L'API Présentation est le modèle sous-jacent à l'écriture de tout Manifest. Elle se limite donc à spécifier la sémantique de présentation, et non la sémantique de description de l'objet lui-même. Il ne s'agit pas d'un schéma de métadonnées comme Dublin Core ou XML EAD¹², mais d'un format standard pour embarquer les métadonnées descriptives de l'objet et garantir l'interopérabilité des *Manifests* entre eux, pour autant un schéma ne porte pas une sémantique particulière. En pratique, toutes les institutions n'utilisent pas le modèle de la même manière, ce qui nécessite une adaptation des développements liés à IIIF pour prendre en compte ces variations particulières.

“L’application du standard varie ainsi d’une institution à l’autre, sans altérer le fonctionnement des outils de base de IIIF, mais en provoquant cependant une déperdition de son universalité pour des projets qui souhaiteraient développer des outils autour de ces manifestes.”¹³

À titre d'exemple, alors que le modèle de référence préconise l'utilisation de la clé *Canvas* pour regrouper les images (Fig. 2.2), certaines institutions, telles que la bibliothèque universitaire de Yale (Fig. 2.3), utilisent *Items* à la place, modifiant alors l'arborescence du *Manifest*¹⁴.

```

  "canvases": [
    {
      "@id": "https://iiif.slub-dresden.de/iiif/2/323548814/canvas/PHYS_0001",
      "@type": "sc:Canvas",
      "height": 1345,
      "width": 1000,
      "label": "Seite - ",
      "thumbnail": {
        "@id": "https://images.iiif.slub-dresden.de/iiif/2/data%2Fkitodo%2Ftracvaasa_323548814%2Ftracvaasa_323548814_tif%2Fjpeg%2F00000001.tif.medium.jpg/full/,150/0/default.jpg",
        "@type": "dctypes:Image"
      },
      "images": [ ... ] // 1 item
    },
    {
      "@id": "https://iiif.slub-dresden.de/iiif/2/323548814/canvas/PHYS_0002",
      "@type": "sc:Canvas",
      "height": 1424,
      "width": 1000,
      "label": "Seite - ",
      "thumbnail": {
        "@id": "https://images.iiif.slub-dresden.de/iiif/2/data%2Fkitodo%2Ftracvaasa_323548814%2Ftracvaasa_323548814_tif%2Fjpeg%2F00000002.tif.medium.jpg/full/,150/0/default.jpg",
        "@type": "dctypes:Image"
      },
      "images": [ ... ] // 1 item
    }
  ]
}

```

FIGURE 2.2 – Usage de 'Canvas' dans un *Manifest* IIIF par Bibliothèque Universitaire de Dresde.

12. Un schéma de métadonnées est un ensemble de règles qui définissent les éléments sémantiques à inclure dans une description structurée de données.

13. Jade Norindr, *Le traitement des sources historiques par la vision artificielle : l'exemple des manuscrits d'astronomie de tradition ptoléméenne*, fr, mémoire de master, École Nationale des Chartes, 2023, URL : <https://dumas.ccsd.cnrs.fr/dumas-04255677> (visité le 02/02/2024), p.25.

14. *Ibid.*

```

  "items": [
    {
      "type": "Canvas",
      "items": [
        {
          "id": "https://collections.library.yale.edu/manifests/oid/16960303/canvas/16969241/page/1",
          "type": "AnnotationPage",
          "items": [
            {
              "id": "https://collections.library.yale.edu/manifests/oid/16960303/canvas/16969241/image/1",
              "type": "Annotation",
              "motivation": "painting",
              "target": "https://collections.library.yale.edu/manifests/oid/16960303/canvas/16969241",
              "body": [
                {
                  "id": "https://collections.library.yale.edu/iiif/2/16969241/full/full/0/default.jpg",
                  "type": "Image",
                  "format": "image/jpeg",
                  "height": 3000,
                  "width": 2026,
                  "service": []
                }
              ]
            }
          ]
        },
        {
          "label": "...", // 1 item
          "thumbnail": [...], // 1 item
          "rendering": [...], // 2 items
          "height": 3000,
          "width": 2026,
          "metadata": [...] // 2 items
        }
      ]
    }
  ]
}

```

FIGURE 2.3 – Usage de 'Items' dans un *Manifest* IIIF par Bibliothèque Universitaire de Yale.

Le fichier reste lisible dans un visualiseur, mais la structure du *Manifest* change. Un projet d'exploitation des objets patrimoniaux reposant sur IIIF, à l'image d'EIDA, se doit de connaître les spécificités propres aux *Manifests* de chaque institution et d'en adapter le traitement. La récupération et le téléchargement des images sont gérés par des scripts dédiés qui parsent les *Manifests* déposés dans l'application. Leur écriture a nécessité une liste des institutions pourvoyeuses des sources et une connaissance des normes adoptées pour leur *Manifests*¹⁵. Ces scripts constituent un module indépendant de l'application, réutilisable et adaptable pour une réutilisation indépendante.

De plus, malgré les avantages évidents qu'elle offre en termes d'interopérabilité et d'accès aux ressources visuelles l'implémentation de IIIF n'est pas encore un réflexe systématique et global. Notamment le monde des musées, hors Amérique du Nord, exploite peu les API IIIF¹⁶. Ce contexte fragmente le paysage numérique des ressources culturelles et complexifie le développement de plateformes telles qu'AIKON. Il est alors essentiel de prendre en compte la diversité des formats de données avec lesquels elle pourrait interagir. Ainsi, en intégrant IIIF dans la plateforme, il est nécessaire d'envisager des mécanismes de compatibilité avec d'autres formats, en particulier ceux utilisés couramment dans le domaine des manuscrits numérisés, tels que les fichiers PDF et les images JPEG (*Joint Photographic Experts Group*). Cette approche permettra ainsi d'assurer une grande flexibilité et une interopérabilité optimale, garantissant la possibilité de traiter une large gamme de sources quel que soit le format dont dispose l'utilisateur.rice à l'entrée de la chaîne de traitement.

Ainsi IIIF s'est imposé dans la dernière décennie comme le standard et une brique technologique essentielle pour décloisonner les collections numériques des institutions, ou-

15. Script for automatic image retrieval from IIIF manifests by Segolène Albouy.

16. Gavin Mallory, *IIIF for museums, explained*, en, juill. 2019, URL : <https://blog.cogapp.com/iiif-for-museums-explained-49fd0560e1ba> (visité le 07/08/2024).

vrant de nouvelles perspectives pour l'exploration et le traitement du patrimoine visuel à l'échelle mondiale. Ce protocole contribue à répondre à un enjeu important : comment exposer sur le Web une collection patrimoniale de manière cohérente tout en l'adaptant aux formes diversifiées des pratiques ; proposant un environnement logiciel gratuit et accessible. Toutefois sa mise en œuvre généralisée et normalisée est encore en cours. Dans ce contexte, une chaîne de traitement des sources historiques doit être conçue de manière à prendre en compte cette réalité, en offrant une compatibilité avec d'autres formats de données et en s'adaptant aux spécificités locales des pratiques institutionnelles dans la mise en œuvre du standard.

2.2.4 Particularité de l'utilisation de IIIF au sein de la plate-forme

Chaque record de *Digitization* dispose d'un *Manifest* IIIF spécifique à EIDA, afin de remédier à la variabilité de structure des *Manifests* : à la soumission du formulaire d'enregistrement d'un *Witness* ou d'une *Serie*, si une *Digitization* a été ajoutée, les fichiers (déposés au format PDF ou JPEG, ou bien récupérés à partir du *Manifest* d'une institution) sont téléchargés, post-traités et renommés. Une vue génère le *Manifest* de la *Digitization* à la demande et il servira de base à l'indexation des annotations.

L'extraction manuelle ou automatiques de régions d'image entraîne la création d'une entité *Regions*, qui dispose de deux *Manifests* pour faire cohabiter différentes extractions (automatiques et manuelles). Les annotations issues de l'extraction automatique sont indexées sur un *Manifest* "auto". Elles sont inscrites "en dur" dans le *Manifest*, ne sont pas indexées dans SAS et ne peuvent pas être modifiées depuis Mirador. Un *Manifest* "v2" est utilisé pour les annotations réalisées avec SAS, donc effectuées manuellement, ou issues de la correction des annotations automatiques.

Ces stratégies de développement permettent de palier aux limitations imposées par le standard IIIF, tout en tirant partie de ses avantages et de son système d'annotation. Elles permettent aussi plus de flexibilité dans la méthode d'extraction des objets d'intérêt dans les images, en autorisant deux types d'annotation, manuelle et automatique.

Donc si les stratégies d'accès à la donnée se voudraient universelles, il existe des limitations techniques ou juridiques demandant une flexibilité dans les modes d'accès aux sources. Il faut donc laisser ouvert des modes d'accès plus "artisanaux" (fichiers stockés en local, métadonnées rentrées manuellement) que le standard IIIF, mais aussi garantir une gestion des droits d'accès aux sources et leur exploitation, afin de respecter les exigences des institutions détentrices.

Chapitre 3

État de l'art : IA et traitement du volume

La BnF constate le besoin d'automatisation pour traiter les volumétries croissantes des collections numérisées, en outre caractérisées par une variété considérable, et elle relève les tensions se jouant dans le traitement en masse de données hétérogènes.

"De plus en plus confrontées à des niveaux de volumétrie et de vitesse typiques des mégadonnées (big data), les collections numériques de la BnF, qui occupent aujourd'hui environ six pétaoctets, sont caractérisées par une variété considérable. Documents numérisés, tels que par exemple les livres et manuscrits consultables dans Gallica — la bibliothèque numérique de la BnF— ; documents nativement numériques comme les œuvres d'art vidéo, les logiciels, les bases de données, les archives de l'Internet ; métadonnées bibliographiques et données d'autorité décrivant les personnes, lieux, organisations, concepts... autant d'ensembles de données diverses en termes de structures, formats, qualité, contextes de production, fonctions et contenus. Ces ensembles ont des histoires différentes, issues des changements des supports et des multiples strates de pratiques documentaires accumulées au fil du temps. Leur hétérogénéité exige des traitements spécifiques et par conséquent des compétences et des méthodes particulières, aussi bien pour les conserver ou les communiquer que pour les analyser (cf Moiraghi, 2017). Cette hétérogénéité des données, qui découle de l'amplitude chronologique et de la vocation à l'encyclopédisme caractéristiques des bibliothèques nationales, s'ajoute à l'accroissement de la quantité des données en entrée et à l'accélération conséquente des temps de traitement. La tendance traditionnelle des bibliothèques à la systématisation des procédures doit dès lors trouver son équilibre face à la spécificité des données mais aussi des questions scientifiques propres aux projets de recherche qui les exploitent".¹

L'état de l'art montre une convergence vers l'utilisation de l'intelligence artificielle pour traiter efficacement des corpus de données de plus en plus vastes et complexes. Cette systématisation doit cependant être équilibrée par une compréhension fine des contextes

1. Emmanuelle Bermès et Eleonora Moiraghi, « Le patrimoine numérique national à l'heure de l'intelligence artificielle », *Revue Ouverte d'Intelligence Artificielle*, 1-1 (juill. 2020), p. 89-109, DOI : 10.5802/roia.5, p.6.

locaux et spécifiques des données, afin de garantir la pertinence et l'efficacité des outils développés.

Les explorations autour du traitement du patrimoine numérique menées à la BnF², en rapport étroit avec des projets de recherche en HN (*Humanités Numériques*), illustrent bien cette tension. Ces projets ont porté sur des ensembles de données balisés et spécifiques (tel que les sources documentaires numérisées autour de la guerre 14-18, les publicités de 1910 à 1920, illustrations du magazine de mode Vogue de 1920 à 1940 ou encore illustrations de papier peint³). L'absence de passage à l'échelle est significatif des difficultés d'un traitement et d'un enrichissement systématique et standardisé de très grands volumes de données très hétérogènes. Ainsi, chacun des projets reposait sur des méthodes et des techniques différentes en raison de la nature des données explorées et des finalités scientifiques propres à chaque projet.

Cependant des leçons ont été tirées : les résultats découlent d'un travail collectif et interdisciplinaire, et se sont appuyés sur des méthodes standardisées, par exemple pour l'extraction des données et métadonnées via des APIs⁴. Ces premières applications de l'IA ont démontré la nécessite de garantir la reproductibilité des méthodes et d'adopter des normes pouvant mettre en œuvre un cadre technique interopérable afin de faciliter la collaboration à plusieurs échelles : entre les projets de recherche d'un part, et d'autre part entre les disciplines et les corps de métier (notamment entre les chercheur.ses et les professionnels des bibliothèques).

Le projet *Gallic(orpor)a*⁵ – bénéficiant de l'appui du DataLab de la BnF – se détache néanmoins dans le paysage des projets estampillés BnF, car il aspire à s'éloigner le plus possible de son corpus, voulant concilier reproductibilité des résultats sur des données diverses et applicabilité à un large corpus issu des collections numérisées de la Bibliothèque Nationale. Il illustre en outre les problématiques susmentionnées : l'inscription dans un dialogue interdisciplinaire et dans des pratiques normalisées. L'ambition du projet porte sur le développement d'une chaîne de traitement automatisée pour la transcription et l'annotation des documents textuels historiques, en diachronie longue, en partant de leurs numérisations disponibles sur le portail Gallica, créant ainsi des corpus enrichis, et facilitant leur exploitation et leur valorisation. En effet les besoins des institution se déplacent

2. Ibid., Valérie Beaudouin et Zeynep Pehlivan, *Cartographie de la Grande Guerre sur le Web*, Research Report, Bibliothèque nationale de France ; Bibliothèque de documentation internationale contemporaine ; Télécom ParisTech, 2017, URL : <https://hal.science/hal-01425600> (visité le 29/07/2024), Guillaume Michez, *GallicaPix, un nouvel outil d'exploration iconographique*, juin 2021, URL : <https://gallica.bnf.fr/blog/21062021/gallicapix-un-nouvel-outil-dexploration-iconographique?mode=desktop> (visité le 29/07/2024), Ariane Bouchard, *Présentation du projet CORPUS à la BnF*, fr-FR, Billet, ISSN : 2608-1393, mai 2017, DOI : 10.58079/va8e

3. V. Beaudouin et Z. Pehlivan, *Cartographie de la Grande Guerre sur le Web...*, G. Michez, *GallicaPix, un nouvel outil d'exploration iconographique...*

4. E. Bermès et E. Moiraghi, « Le patrimoine numérique national à l'heure de l'intelligence artificielle »..., p.7.

5. <https://gallicorpora.github.io/>

de la transcription des textes à leur encodage sémantique automatique en XML-TEI, afin d'offrir aux utilisateurs de bibliothèques numériques de nouvelles options pour la fouille de données. Le projet *Gallic(orpor)a* s'inscrit dans cette dynamique, en exploitant le riche corpus de la BnF, qui met à disposition 193 265 manuscrits, dont 52 188 précédant 1800, et 1 182 471 livres imprimés, dont 160 335 précédant 1800⁶.

“Le nouveau défi à relever aujourd’hui est de transformer ces numérisations en des ressources enrichies, qui augmentent le texte extrait et repérable avec de la métadonnée et de l’analyse. Le texte brut et non annoté ne suffit plus pour la recherche en informatique appliquée aux documents historiques. De là vient l’impulsion pour le projet *Gallic(orpor)a*. Le projet envisage la mise en place d’un pipeline qui saisit un document numérisé depuis le portail Gallica et renvoie une ressource numérique très enrichie. En plus d’une description du texte repérable, la ressource présentera les données structurelles portant sur la mise en page, ainsi qu’une analyse linguistique du texte extrait et des métadonnées portant sur le document physique et le fac-similé numérique”.⁷.

En outre, le but sous-jacent du projet est de produire un prototype qui pourrait servir d'exemple de chaîne d'acquisition numérique pour les institutions patrimoniales. Par conséquent, ce projet se propose aussi d'être une preuve de concept d'un *modus operandi* pour l'extraction et l'annotation de textes très divers, créant une sorte de pipeline ultime. Mais avant tout, la chaîne de traitement est destinée à être applicable à un large corpus de documents mis à disposition par la BnF. Les documents du corpus visé proviennent de différentes époques (du XV^e au XVIII^e siècle) et présentent une grande diversité de mises en page, de langues (ancien français, moyen français, français classique), et de supports (manuscrits, imprimés). L'hétérogénéité des sources est censée faire preuve de faisabilité, et vérifier le potentiel de la méthode élaborée. Elle vise à prouver que le concept d'une chaîne de traitement généraliste peut être concrètement appliquée.

Gallic(orpor)a expose ainsi la plupart des questionnements et défis tenant au montage d'une chaîne de traitement unifiée applicable à une grande diversité de données. Les choix technologiques, tels que l'adoption de formats ouverts et interopérables, la prise en compte de la diversité des modes d'acquisition des données et la spécialisation des modèles d'IA, sont au cœur de ces enjeux. Par ailleurs, la question de l'ouverture des corpus annotés, essentielle pour l'apprentissage machine, est également un axe d'analyse à considérer.

6. Benoît Sagot, Laurent Romary, Rachel Bawden, Pedro Javier Ortiz Suárez, Kelly Christensen, Simon Gabay, Ariane Pinche et Jean-Baptiste Camps, « Gallic(orpor)a : Extraction, annotation et diffusion de l'information textuelle et visuelle en diachronie longue », dans *DataLab de la BnF : Restitution des travaux 2022*, Paris, France, 2022, URL : <https://hal.science/hal-03930542> (visité le 06/06/2024).

7. K. Christensen, A. Pinche et S. Gabay, « Gallic(orpor)a : Traitement des sources textuelles en diachronie longue de Gallica », dans *DataLab de la BnF*, Paris, France, 2022, URL : <https://hal.science/hal-03716534> (visité le 11/05/2024).

3.1 Formats standards

Une telle chaîne de traitement automatique va s'appuyer sur des protocoles, des cadres et des formats standards adaptés au contexte de la recherche.

La chaîne de traitement de *Gallic(orpor)a* prend en entrée les pages numérisées du document source, et fournit en sortie une ressource numérique composite exposant les métadonnées, la transcription du document par des modèles HTR et le texte dit *pré-éditorialisé*, obtenu en parsant les données récupérées des fichiers ALTO (*Analysed Layout and Text Object*)⁸ (la première transcription du texte), et en ignorant la mise en page. Ce dernier texte sert à la génération de l'objet suivant : le texte analysé par les outils TAL (*Traitemment Automatique du Langage*). Chaque élément de la sortie de la pipeline ouvre des possibilités d'usages différents des données textuelles. À l'issue de la chaîne de traitement, l'utilisateur obtient un fichier TEI destiné principalement à une exploitation philologique utilisant des approches de type “distant-reading”.

Pour rassembler les sorties des modèles et des traitements, XML-TEI semble particulièrement adapté.

Les principaux avantages du langage XML sont son interopérabilité et l'encodage sémantique qu'il permet. XML est un langage libre et documenté, respectant les recommandations du w3C (*World Wide Web Consortium*), et facilitant l'échange de données et la migration vers d'autres plateformes, logiciels ou formats. Il a l'avantage d'être lisible à la fois par les machines ou par l'œil humain.

La TEI est un langage XML. Elle a été d'abord développée comme un projet de recherche. L'idée originelle était de proposer un ensemble de recommandations sur la façon dont les chercheurs devraient créer des ressources textuelles *computable* (lisibles par ordinateur), qui soient adaptées aux besoins de la recherche – dans la mesure où un consensus existait sur le sujet –, mais qui soient également extensibles, puisque ces besoins changent et évoluent. Ces principes visent à établir un cadre uniifié et interopérable⁹ pour l'échange et le traitement de données textuelles dans la recherche en SHS, en prenant en compte les besoins spécifiques de différents domaines et projets.

L'atout majeur de la TEI est donc sa flexibilité et son aspect modulaire. Ses *guidelines*

8. L'OCR et l'HTR produisent un format basé sur XML et géré par un schéma, le format ALTO. ALTO est un des formats les plus couramment utilisés pour la conversion des textes contenus dans les documents textuels numérisés, il embarque de nombreuses informations, notamment il va représenter la structure physique de la page. Il va repérer le layout de la page, dans le layout les blocs de texte, dans les blocs de texte les paragraphe, dans les paragraphes les lignes, et dans les lignes les chaînes de caractères. On a aussi des informations sur le style (grosses ou petites caractères). Le format ALTO contient aussi des informations techniques sur la confiance de la reconnaissance. Il conserve toutes les coordonnées géométriques des contenus (textes, illustrations, graphiques) dans l'image et permet ainsi l'alignement de l'image et du texte, la superposition de la mise en page et de la prédiction des modèles (par exemple dans un fichier PDF multicouche). C'est de cette manière qu'on peut par exemple obtenir la surbrillance des mots recherchés sur l'image lors d'une requête dans le texte.

9. basée sur le langage XML, la TEI est indépendante de tout environnement logiciel, et donc idéale pour l'échange et la collaboration.

définissent plusieurs centaines de concepts différents, avec les instructions détaillées sur les éléments qui peuvent être utilisés pour les représenter informatiquement, mais les projets d'encodage n'en utilisent donc qu'un sous-ensemble restreint.¹⁰ C'est en partie à ce titre qu'elle est utilisée dans le cadre du projet *Gallic(orpor)a*, permettant un encodage léger mais adapté à l'exploitation des corpus enrichis. Elle est en outre un format pivot.

Les différents niveaux de description proposés constituent un autre point fort de la TEI. Elle propose des éléments pour l'encodage des aspects physiques de la page, purement visuel comme la mise en page, ou pour des considérations paléographiques (abréviations et forme développée, etc.). Mais aussi elle permet un balisage sémantique, analytique et des métadonnées. Le format XML-TEI embarque donc le contenu intellectuel (le texte annoté issu des post-traitements) comme la représentation matérielle du document source (notamment la mise en page encodée dans les sorties des modèles HTR au format ALTO), ainsi qu'un teiHeader riche pour les métadonnées¹¹.

Enfin, XML-TEI est le standard pour l'édition et la publication dans le milieu de la recherche¹².

EIDA trouve dans le format SVG (*Scalable Vector Graphics*) un équivalent à XML-TEI dans le domaine graphique, de par sa capacité à agréger les résultats de traitements variés, à assurer l'interopérabilité avec d'autres environnements logiciels et applicatifs, mais aussi son balisage sémantiquement riche. Sa structure basée sur des éléments XML le rend particulièrement adapté à des traitements informatiques ultérieurs.

Les moyens d'accès à la donnée reposent eux aussi sur des protocoles standards : via des API comme le service SRU (*Search/Retrieve via URL*) de la BnF, il est possible d'automatiser à la fois l'encodage des métadonnées et celui du texte issu du document facsimilé numérisé à partir d'une liste d'identifiants ARK (*Archival Resource Key*). Cependant, cette approche systématique se trouve confrontée à la diversité des sources de données. Si le passage de ALTO (issu de la transcription) à TEI est assez facilement généralisable, la génération automatique du teiHeader est hautement dépendante de l'origine des documents transcrits. Les moyens d'accès et la structure des notices bibliographiques diffèrent d'une institution à l'autre, et un script de récupération des métadonnées se voulant hautement modulaire devrait prendre en compte ces spécificités¹³. Ces remarques montrent les limites de l'idéal de l'interopérabilité universelle : malgré les tentatives de généralisation et les nombreux consortium, les bibliothèques et institutions ne présentent

10. La définition d'un modèle TEI (via une ODD (*One Document Does it all*)) restreint certains éléments et permet la personnalisation d'autres.

11. S. Gabay, A. Pinche et K. Christensen, *Gallic(orpor)a : Processing Gallica's historical sources*, Publisher : Université de Genève Published : UNIGE Data Science Day, sept. 2022, URL : <https://hal.science/hal-03819326> (visité le 08/05/2024).

12. *Ibid.*

13. Kelly Kristensen, *D'ALTO à TEI : Modélisation de transcriptions automatiques pour une pré-éditorialisant des textes*, fr, mémoire de master, École Nationale des Chartes, 2022, URL : <https://github.com/kat-kel/TNAH-Memoire>, p.117.

pas les métadonnées des documents selon les mêmes standards. Comme expliqué précédemment, cette problématique a aussi été soulevée dans le cadre d'EIDA, qui montre ainsi une souplesse accrue dans les modes d'accès aux données.

Gallic(orpor)a s'appuie non seulement sur des cadres et des formats standards, mais aussi sur les développements d'autres projets, que ce soit des outils de traitement ou des environnements applicatifs complets. L'interface de post-correction Pyrrha, par exemple, a été utilisée pour préparer les données de lemmatisation¹⁴, tandis que e-scriptorium a servi pour la création des corpus annotés pour l'entraînement des modèles d'HTR. Cette démarche illustre bien l'interdépendance des projets de recherche, notamment dans le domaine du *machine learning* : le partage des modèles comme des données destinées à leur entraînement accélère et facilite grandement les développements futurs.

3.2 La spécialisation des modèles

Si *Gallic(orpor)a* distingue clairement l'élaboration de la pipeline de traitement et d'enrichissement des données de la construction des modèles eux-mêmes (la chaîne de traitement étant destinée à tester et appliquer les modèles sur les données) le projet se donne également pour ambition de développer des modèles HTR et TAL personnalisés, capables de traiter des corpus de données larges et hétérogènes.

Un modèle HTR se spécialise sur une écriture particulière grâce à son entraînement, et cela suppose d'autoriser dans la pipeline une bascule entre différents modèles. De fait, elle est conçue pour implémenter tout modèle fourni par l'utilisateur lors de l'installation (à condition qu'il soit basé sur le moteur Kraken). Alors, la chaîne de traitement peut faire office de “coquille vide”, de prototype modulable contenant les briques fonctionnelles conditionnées par les données, leur spécificité, leur origine.

L'HTR désigne le processus d'extraction automatique du contenu textuel utilisant le *machine learning*, souvent faisant appel à des réseaux de neurones, pour identifier les caractères ou les mots à partir des *features* extraites. L'entraînement d'un modèle d'HTR nécessite de pouvoir itérer sur plusieurs étapes après soumission d'un *dataset* d'entraînement. Les étapes consistent en des traitements préliminaires sur l'image d'entrée, la segmentation de la page, la prédiction du texte par les modèles, sa correction par l'humain, puis le renvoi des données dans le *workflow* pour spécialiser le modèle (pour qu'il “apprenne”)¹⁵.

Les données pour l'HTR (des corpus annotés) ont été réalisés grâce à e-Scriptorium. e-Scriptorium¹⁶, porteuse du concept de modularité, est une plateforme logicielle qui offre un cadre de travail complet pour la transcription numérique de documents, qu'ils soient

14. B. Sagot, L. Romary, R. Bawden, *et al.*, « Gallic(orpor)a... ».

15. Le chapitre 4 de ce mémoire porte spécifiquement sur le fonctionnement des réseaux de neurones.

16. *eScriptorium - Homepage*, URL : <https://test2.fondue.unige.ch/> (visité le 15/06/2024).

imprimés ou manuscrits. Elle modélise l'ensemble du processus de transcription, de la préparation des images à la correction finale. En effet, e-Scriptorium propose une pipeline intégrant plusieurs étapes : l'importation d'images, la segmentation des lignes et blocs de texte, la transcription en elle-même (grâce aux modèle d'OCR et HTR), sa correction par l'œil humain, et l'exportation des transcriptions. Ainsi, il offre un environnement complet pour construire, tester et utiliser des modèles de transcription adaptés à différents types de documents.

e-Scriptorium dispose de plusieurs moyens d'être utilisé. L'application doit être déployée sur un serveur Web installé sur un ordinateur personnel ou sur une machine dédiée. Les capacités de calcul du matériel employé faisant ensuite la différence au moment de faire tourner Kraken (le moteur HTR), en particulier lors des entraînements. Elle dispose d'une interface graphique facilitant la prise en main des processus. Afin d'effectuer la transcription des labels présents dans les figures, EIDA prévoit d'insérer une instance de e-Scriptorium à sa propre pipeline. Cette solution technique démontre la grande modularité de e-Scriptorium, composant autonome ou bloc fonctionnel au sein d'un système plus complexe.

En développant la plateforme web AIKON, EIDA et VHS visent à apporter à l'image ce que e-Scriptorium permet pour le texte : la plateforme vise à devenir un outil modulable qui sous-tend une approche globale du traitement de l'image basée sur la vision artificielle.

3.3 La normalisation des données d'annotation

Gallic(orpor)a mène une réflexion méthodologique sur la portabilité des modèles, leur diffusion et le partage de grands ensembles de données annotées selon des normes communes. Les modèles d'IA, notamment ceux dédiés à la reconnaissance du texte manuscrit (HTR) et au traitement automatique des langues (TAL), requièrent des données d'entraînement spécifiques.

Mais si chaque projet de recherche annotait ses corpus selon ses propres exigences, ils engendreraient fatalement des silos de données non réutilisables. Pour garantir la réutilisation des données, il est impératif d'établir des normes et des standards.

Le projet porte alors le développement d'une syntaxe d'annotation générique pour harmoniser la segmentation des pages des vérités de terrain, afin de constituer des corpus d'entraînement réutilisables. *Gallic(orpor)a* propose une approche très inclusive en identifiant des éléments textuels communs à une large variété de documents, manuscrits comme imprimés. Cette démarche donne lieu à la définition d'un vocabulaire contrôlé permettant ainsi de construire des corpus annotés compatibles avec différents contextes¹⁷ :

17. “Using a common vocabulary to annotate zones called SegmOnto (that is still evolving), we have developed a generic workflow to analyse the layout, OCRise the text, and convert the ALTO output into minimally encoded TEI files (...).” Juliette Janes, A. Pinche, Claire Jahan et S. Gabay, « Towards automatic TEI encoding via layout analysis », dans *Fantastic future 21, 3rd International Conference on*

SegmOnto¹⁸

Les étapes de lemmatisation et d'étiquetage morphologique (POS-tagging) effectuées par les modèles de TAL sur le texte extrait des pages numérisées visent à normaliser le langage en réduisant les mots à leur forme canonique (lemme) et en identifiant leur catégorie grammaticale. Cette normalisation est essentielle pour faciliter des analyses ultérieures telles que la collation et la stylométrie. Elle permet une analyse comparative des textes malgré la grande variabilité inhérente aux langues historiques, pour lesquelles l'absence de normes orthographiques entraîne une grande diversité de graphies. La préparation des données, là aussi, a donné lieu à une réflexion méthodologique sur les défis liés à la standardisation des annotations linguistiques dans les corpus diachroniques.

Selon Gabay, *et al.* :

“With the development of big corpora of various periods, it becomes crucial to standardise linguistic annotation (e.g. lemmas, POS tags, morphological annotation) to increase the interoperability of the data produced, despite diachronic variations.”¹⁹

Gabay, *et al.*²⁰ relèvent la difficulté de mettre en œuvre un cadre technique qui prenne en compte les pratiques d'annotation déjà établies et propres à des états de la langue. Pourtant garantir une interopérabilité minimale avec les corpus existants est essentielle pour maximiser la valeur ajoutée des nouvelles données.

“Such a task cannot be done without taking into account longstanding annotation practices, in order to allow (minimal) interoperability with already existing datasets. Such a statement is sadly easier said than done, because EMF is an intermediary stage between medieval (12th-15th c.) and late modern and contemporary (from c. 1750) French, two states of language that tend to have different needs regarding annotation : EMF is then caught in between two (potentially incompatible) practices, one for each extreme of the continuum.”²¹

Il est particulièrement complexe de trouver un équilibre entre une description linguistique trop fine, qui pourrait limiter la réutilisabilité des corpus, et une description trop générale, qui pourrait manquer de précision. De plus, les besoins en annotation varient considérablement entre le français médiéval et le français moderne, et les systèmes d'annotation de ses deux états de la langue sont difficilement réconciliables : or une large partie des sources de *Gallic(orpor)a* se situent dans l'entre-deux. Trouver un compromis qui satisfait les exigences spécifiques de chaque période est complexe.

Artificial Intelligence for Librairies, Archives and Museums, Paris, France, 2021, URL : <https://hal.science/hal-03527287> (visité le 06/06/2024), p.2.

18. <https://github.com/SegmOnto>

19. S. Gabay, Thibault Clérice, J.B. Camps, Jean-Baptiste Tanguy et Matthias Gille-Levenson, « Standardizing linguistic data : method and tools for annotating (pre-orthographic) French », dans *Proceedings of the 2nd International Digital Tools & Uses Congress (DTUC '20)*, Hammamet, Tunisie, 2020, DOI : 10.1145/3423603.3423996, p.2.

20. *Ibid.*

21. *Ibid.*, p.2.

L’harmonisation des annotations vise à favoriser la diffusion des corpus de données sur HTR-United, une plateforme collaborative dédiée au catalogage de vérités de terrain pour l’HTR et l’OCR, principalement en français²². Cette base de données, hébergée sur GitHub, centralise des images et leurs transcriptions produites par différents projets de recherche, offrant ainsi une diversité de jeux de données diachroniques et géographiques pour l’entraînement de modèles HTR.

Gallic(orpor)a vise aussi à la diffusion des modèles en eux-même, qui peuvent alors être réutilisés et spécialisés. Par ailleurs, le projet s’appuie sur des outils existants, par exemple Deucalion, une boîte à outils de traitement automatique des langues (TAL) conçue pour être interopérable avec d’autres systèmes²³.

Les apports et les réflexions du projet *Gallic(orpor)a* reflètent la problématique globale de ce mémoire : comment concevoir des cadres pour l’environnement de la recherche en Humanités Numériques, tout en répondant aux exigences scientifiques de chaque projet. On voit se filer avec ce cas d’étude un écosystème complexe dans lequel s’agencent des systèmes et protocoles généraux et extensibles, adaptables aux besoins locaux, comme la TEI, ou SegmOnto, ou encore des API pour la récupération des données. De nombreux défis restent cependant à relever, montrant que l’interopérabilité universelle est un idéal difficilement atteignable.

L’utilisation de l’IA porte la question du partage des ressources à un niveau d’importance supérieur, puisque les modèles peuvent être spécialisés, et les corpus enrichis produits peuvent constituer des données d’entraînement. L’intervention du chercheur.se pour corriger les résultats des traitements automatiques est donc un aspect important à prendre en compte pour assurer la fiabilité scientifique de ses corpus annotés.

Gallic(orpor)a visait à l’élaboration d’un outil d’acquisition et d’enrichissement des données capable de se détacher d’un corpus spécifique et sur ce point se voulait preuve de concept, mais le projet n’a pas pleinement atteint ses objectifs, car la chaîne de traitement est restée très orientée vers les corpus de la BNF. Il a ainsi mis en évidence la complexité de concilier les exigences d’une infrastructure générique et les besoins spécifiques des données, montrant que la modularité s’inscrit avant tout dans le temps.

Comme le montre le cas de e-Scriptorium, une chaîne de traitement, si elle est assez généraliste pour tolérer une grande diversité de données, participe à la cohérence des pratiques et à la mutualisation des outils de la recherche. Les modalités d’accès (notebooks, plateforme, scripts etc.) impactent la prise en main et l’adoption des outils.

22. Alix Chagué, T. Clérice et L. Romary, « HTR-United : Mutualisons la vérité de terrain ! », dans *DHNord2021 - Publier, partager, réutiliser les données de la recherche : les data papers et leurs enjeux*, Lille, France, 2021, URL : <https://hal.science/hal-03398740> (visité le 09/08/2024).

23. <https://github.com/chartes/deucalion-model-af>

Plus généralement, les recherches menées sur l'enrichissement et l'exploration de larges corpus grâce aux outils d'IA ouvrent de nouvelles perspectives en terme de trouvabilité et de fouille de données visuelles ou textuelles. Cette valeur ajoutée est liée au passage du format JPEG à des formats balisés et sémantiquement riche qui permettent des exploitations ou l'indexation.

Conclusion partielle

Comme le dit Jacquot²⁴, face aux problématiques de traitement du volume, “une réponse peut résider dans la généralisation des outils et des méthodes ainsi que dans le partage des récits de projets de recherche antérieurs, forme d’apprentissage et d’échange importante.”

Un trait saillant dans la progression vers l’automatisation est la confrontation de deux besoins : la collaboration et la spécialisation. D’un côté, la communauté scientifique voudrait mutualiser les outils et les méthodes, afin d’encourager le partage d’une expertise interdisciplinaire, des moyens financiers, et d’améliorer la maintenabilité, contribuant ainsi à la robustesse des développements. Un tel outil, à l’image de la plateforme e-Scriptorium, qui porte et guide les processus de transcription de texte, et dont la portabilité facilite l’intégration dans divers écosystèmes de recherche, contribue aussi à la cohérence des méthodes scientifiques au sein d’une équipe de recherche, ou entre plusieurs projets. De l’autre côté, il faut prendre en compte les spécificités locales des données, les besoins des utilisateur.rices et les enjeux propres aux projets de recherche qui portent la création des outils de traitement. Les solutions doivent être adaptées aux particularités des corpus et aux questions de recherche spécifiques, ce qui peut parfois entrer en conflit avec la nécessité de généralisation.

On aura voulu montrer dans cette première partie que l’ambition de créer un outil générique – visant à l’enrichissement et la sémantification de la donnée – se heurte à des dynamiques collectives et à la multiplicité des acteurs, autant qu’elle en profite. La philosophie de l’ouverture, à la fois des données et du code, est un atout (libre accès aux sources, appui sur des scripts et des projets de recherche existants, rencontres avec d’autres projets, création des synergies avec d’autres équipes ou institutions et mutualisation des moyens financiers et humains), et une double contrainte. La première est la nécessité d’assurer une interopérabilité technique pour ouvrir en retour les corpus et garantir la reproductibilité des résultats. La deuxième porte sur la prise en compte de l’hétérogénéité des données, ce qui demande de trouver le juste équilibre entre précision et flexibilité en terme de description.

Nous sommes restés volontairement vagues sur la question des modèles de vision artificielle et les réseaux de neurones : ils sont pourtant la clé de voûte des processus d’en-

24. O. Jacquot, *Décrire, transcrire et diffuser un corpus documentaire hétérogène...*

richissement et d'exploration des données. La construction et l'entraînement des réseaux de neurones s'inscrit aussi dans une balance délicate entre généralisation et spécialisation. Les modèles doivent être suffisamment généralistes pour reconnaître et interpréter une large variété de motifs provenant de contextes divers, ce qui leur permet d'être applicables à la diversité des scénarios rencontrés et de gérer l'hétérogénéité des grands corpus : par exemple la grande variété orthographique, morphologique et de mise en page dans le cadre du projet *Gallic(orpor)a*. Mais l'efficacité des modèles de vision exige aussi leur spécialisation sur des jeux de données étroitement ciblés. Par conséquent, les chercheur.ses et ingénieur.es doivent constamment équilibrer ces deux exigences pour développer des modèles robustes, polyvalents, mais aussi suffisamment précis pour répondre aux besoins spécifiques des applications en recherche.

Deuxième partie

Mise en œuvre et exploitation de la
Computer Vision

Introduction partielle

Les diagrammes astronomiques offrent une perspective unique sur la compréhension et la diffusion du savoir à travers les siècles dans l'histoire afro-eurasienne. Des centaines de manuscrits numérisés et des milliers de diagrammes sont disponibles. Nous avons constaté dans la première partie de ce travail les dimensions et l'hétérogénéité importantes caractérisant ce corpus de recherche, ainsi que l'impossibilité de réaliser des annotations à grande échelle pour chaque tâche spécifique. Pour assister les chercheur.es dans le travail de fouille et d'analyse des documents historiques, diverses méthodes d'apprentissage profond (*deep learning*) ont été développées et implémentées sur la plateforme.

Il ne s'agit pas seulement d'implémenter des outils d'IA efficaces, mais aussi d'en exploiter les résultats. Pour EIDA, la finalité est de pouvoir produire des éditions scientifiquement satisfaisantes des diagrammes.

Qu'est-ce que la vision artificielle ? Comment rendre les modèles efficaces sur les sources historiques ? Notre propos s'articulera autour de trois axes principaux : le fonctionnement des modèles de vision, les enjeux tenant aux jeux de données pour leur entraînement, et la conceptualisation de l'outil d'édition qui repose sur les résultats des algorithmes.

Chapitre 4

Les modèles de *Computer Vision*

Dans des termes très simples, pour faire exécuter une tâche à la machine, deux solutions existent. La première consiste à écrire un programme, dont l'expert métier a explicité les règles. Le programme est entièrement rédigé par un.e développeur.se. Il effectue une tâche précise, chaque conjecture spécifique doit être prévue et son traitement clairement formulé. Si le code produit des erreurs, il doit être modifié par le.a développeur.se. La deuxième approche consiste à donner au programme la capacité de se modifier lui-même, sans que cette modification soit explicitement rédigée ; le programme programme. L'expert métier doit annoter des exemples et un seul algorithme (d'apprentissage) suffit pour traiter de multiples cas du réel. En cas d'erreur, il faut agir non plus sur le programme mais sur les exemples d'apprentissage. L'objectif est d'apprendre à généraliser pour prédire sur des exemples non vus pendant l'apprentissage¹.

“On peut ainsi opposer un programme *classique*, qui utilise une procédure et les données qu'il reçoit en entrée pour produire en sortie des réponses, à un programme *d'apprentissage automatique*, qui utilise les données et les réponses afin de produire la procédure qui permet d'obtenir les secondes à partir des premières.”²

L'émergence de l'apprentissage machine a alors ouvert de nouvelles perspectives en permettant de modéliser des interactions complexes entre les données. Les modèles de *deep learning* sont également plus résilients face aux variations et au bruit présents dans les données³. Ainsi, l'application de l'apprentissage profond se révèle particulièrement pertinente dans le cadre d'EIDA, car elle permet de relever le défi de la sémantification des images, y compris lorsqu'il s'agit de sources historiques complexes.

1. François Chollet et Jacqueline Isabelle Forien, *L'apprentissage profond avec Python*, Saint-Cyr-sur-Loire, 2020 (Les essentiels de l'IA), p.7.

2. Chloé-Agathe Azencott, *Introduction au machine learning*, 2e édition, Malakoff, 2022 (Info Sup), p.1-2.

3. Sahil Juneja, *Deep Learning vs Machine Learning — The Difference Explained!*, en-US, mars 2023, URL : <https://www.dataquest.io/blog/deep-learning-vs-machine-learning-the-difference-explained/> (visité le 05/07/2024).

4.1 Les réseaux de neurones, ou la généralisation prise au sens mathématique

Les images, qu'elles soient numériques ou non, résultent de la relation entre un phénomène visuel et le spectateur qui lui donne sens.

Or, “the computer knows no image, only numbers. It does not “see”, it only calculates”⁴.

Dans l'environnement numérique, une image matricielle ou *raster* est une grille de pixel, chaque pixel portant une valeur. 0 peut être interprété comme du blanc et 255 comme du noir, les valeurs intermédiaires représentant les nuances de gris. Ces nombres sont appelés valeurs de pixels, un pixel étant la plus petite portion de l'image que l'ordinateur sache distinguer. La perception de cette image résulte, d'un point de vue médiatique, de l'interaction de trois strates : l'information, le calcul par le medium et l'interprétation comme phénomène visuel.

L'information correspond aux binaires, se référant au système de représentation fondamental de la donnée par l'ordinateur. Un bit est l'unité de base de l'information en informatique, pouvant avoir une valeur de 0 ou 1. Pour une image en noir et blanc, chaque pixel porte un bit : un pixel peut être soit noir (0) soit blanc (1). Pour une image en niveaux de gris, chaque pixel porte 8 bits, et un pixel peut donc représenter 256 – de 0 à 255 – niveaux de gris (2^8). Pour décrire la couleurs il existe plusieurs codages, le plus élémentaire étant le code RVB (RGB en anglais) pour 'rouge, vert, bleu'. On décrit la quantité de rouge, de vert et de bleu qui composent la couleur, sur le même principe que les niveaux de gris. En ramenant chaque pourcentage sur une échelle de 0 à 255, on obtient un code RVB pour chaque pixel. Le code RVB d'une couleur, c'est donc trois nombres entiers de 0 à 255, représentés sur trois couches superposées. Puisqu'il faut 1 octet (8 bits) pour coder de tels entiers, chaque pixel demandera donc trois octets.

“A string of pixels on a display is nothing more than a sequence of brightness values on the visible light spectrum (red, green and blue) at first. However, we are able to calculate and use these color values as the criterion to sort images in a process involving these “low level features” of images. This process, however, is limited in its epistemic potential. In contrast, “high level features” are what those pixels represent ; the content of the images.”⁵

Le principe des algorithmes de vision par ordinateur consiste à passer de la détection des “caractéristiques de bas niveau” aux “caractéristiques de haut niveau”, comblant ainsi le “fossé sémantique” entre les pixels bruts et le contenu compréhensible⁶.

De l'IA au deep learning

4. H. Klinke, « Big Image Data within the Big Picture of Art History »..., p.21.

5. *Ibid.*, p.17.

6. *Ibid.*

Le *deep learning* est un sous-ensemble du *machine learning*, lui-même sous-ensemble de l’IA. On parle d’IA pour désigner tout ce qui imite l’intelligence humaine.

Le développement du *machine learning* correspond à un changement de paradigme au sein des computer sciences : la machine va *apprendre* à effectuer des tâches complexes. La capacité d’apprentissage justifie l’estampillage IA. En substance, au lieu d’avoir un ensemble explicite d’instructions, ce qui est le cas dans la programmation conventionnelle, on a un ensemble flexible de paramètres et on utilise une quantité énorme de données pour mettre à jour ces paramètres. Donc le *machine learning* ou apprentissage automatique consiste à “aborder une tâche en cherchant comment permettre à un ordinateur d’apprendre en lui fournissant des exemples accompagnés de la réponse correcte”⁷. La discipline repose d’une part sur les mathématiques, notamment la statistique (pour ce qui est de la construction des modèles et de leur inférence à partir de données) et d’autre part sur l’informatique et la *data science*, pour ce qui est de la représentation de la donnée et de l’implémentation efficace d’algorithmes d’optimisation (c’est ici que nous nous positionnons en tant que métiers de soutien à la recherche).

Le *deep learning* se situe toujours dans le paradigme de l’apprentissage à partir de données en masse, mais en utilisant un type de modèle particulier : les réseaux de neurones profonds. Un réseau de neurones profond, se présentant comme un ensemble de neurones interconnectées, est censé imiter le fonctionnement du cerveau humain : l’activation d’un neurone envoie un signal à un autre neurone du réseau, ce qui l’active, et ainsi de suite. La vision par ordinateur notamment a été révolutionnée par les techniques de *deep learning*, car dans l’image, l’information disponible au départ - l’intensité lumineuse - est représentée par des nombres réels, ce que manipulent justement les réseaux de neurones utilisés en apprentissage profond.

Le domaine de l’apprentissage profond repose sur les perceptrons multicouche profonds, soit un perceptron multicouche contenant suffisamment de couche – la définition de “suffisamment” variant dans la littérature.

Qu'est-ce qu'un réseau de neurones ?

Les réseaux de neurones artificiels sont des modèles paramétriques, potentiellement complexes, inspirés de la manière dont les réseaux de neurones biologiques du cerveau humain traitent l’information, et particulièrement efficaces pour modéliser des relations complexes et non linéaires entre des objets du monde ; c’est ce qui fait leur succès aujourd’hui. Ils connaissent divers domaines d’application : de la reconnaissance de discours à la vision artificielle, en passant par le traitement du texte.

Tirant la métaphore biologique, l’unité de calcul de base dans un réseau de neurone est appelé neurone mais aussi nœud ou simplement unité. Il reçoit un ou plusieurs input d’une autre unité (ou d’une source extérieure pour la première couche) et calcule une sortie. Chaque input a un poids associé, qui lui est assigné sur la base de son importance

7. Eugene Charniak, *Introduction au Deep Learning*, 2021, p.3.

56 relativement à d'autres inputs. Le noeud a pour fonction de rassembler et pondérer les inputs, avant de leur appliquer une *activation function* (la *fonction d'activation*) permettant d'extraire une représentation sémantique des données (appelées *feature*), qui peut constituer l'entrée d'un nouveau neurone.

Lorsque ces unités de calcul s'associent en réseau (on parle de *Neural Networks* ou réseaux de neurones) elles deviennent capables d'abstraire suffisamment l'information et de prendre des décisions correspondant aux *predictions* en sortie. Le réseau est organisé en couches, chacune constituée d'un groupe d'unités de calcul travaillant en parallèle et transmettant des valeurs à une autre couche. Autrement dit, chaque couche alimente la suivante. Les couches intermédiaires (entre l'entrée et la sortie) extraient donc des représentations profondes des données d'entrée, et au fil de l'avancée dans les couches de neurones, cette représentation se complexifie ; elle est utilisée pour obtenir la sortie finale. L'empilement d'un très grand nombre de couches donne la *profondeur* au réseau et justifie le *deep* de *deep learning*.

“Vous pouvez considérer un réseau profond comme une opération de distillation de l'informations en plusieurs étapes, où l'information passe par des filtres successifs et ressort de plus en plus épurée (c'est-à-dire de plus en plus utile pour une tâche donnée).”⁸

*Comment apprend-il ?*⁹

Le modèle de régression linéaire est un des modèles les plus simples permettant de comprendre les fondements de l'apprentissage automatique. Il permet de résoudre des problèmes unidimensionnels impliquant un input x et un output y . Prenons un exemple très simple : déterminer le prix d'une maison en fonction de sa superficie. Pour résoudre ce problème on a besoin d'une fonction linéaire, sa représentation graphique est une simple droite, initialisée aléatoirement et définie par deux paramètres : la pente et l'intersection avec l'axe des ordonnées. En donnant suffisamment de données au modèle – autrement dit en lui fournissant des exemples assortis de leur réponse, appelées également labels ou étiquettes – celui-ci peut apprendre comment ajuster ses paramètres pour que le modèle colle aux données disponibles. Tout l'apprentissage automatique repose sur cette capacité à mettre à jour les paramètres du modèle en itérant sur des exemples. Pour ce faire, le modèle a besoin d'une mesure quantifiable de sa qualité à un instant t , un moyen de mesurer son degré d'erreur. C'est le rôle de la *loss function*. Pour un modèle linéaire, la *loss function* donnerait la somme des distances de chaque point de données de la courbe au point de donnée réel. Plus généralement, la fonction de perte associe à chaque résultat une valeur indiquant à quel point ce résultat est “mauvais” ou du moins ne correspond pas aux attentes de la prédiction. Dans le processus d'apprentissage des paramètres et

8. F. Chollet et J. I. Forien, *L'apprentissage profond avec Python...*, p.12.

9. Ces explications sont inspirées de la présentation donnée par Syrine Kallel à l'occasion de la conférence EIDA 2024 (*EIDA*, URL : <https://imagine.enpc.fr/~kallelis/icdar2024/> (visité le 23/05/2024))

des poids du modèle, l'objectif est de minimiser cette valeur (Fig. 4.1). L'information de la *loss function* est donc utilisée pour mettre à jour les poids.

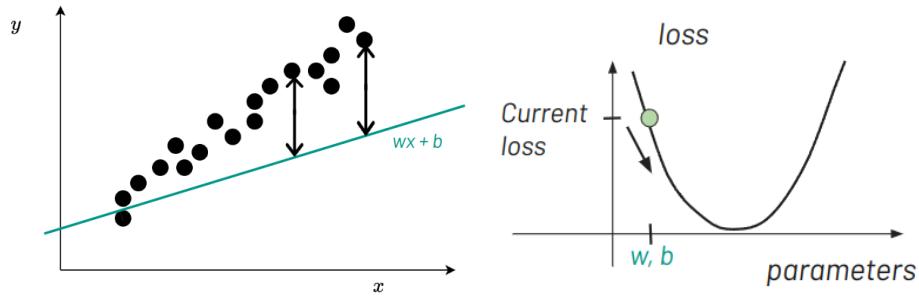


FIGURE 4.1 – Modèle linéaire et modélisation de la *loss*.

L'actualisation des poids passe par la modélisation de la perte en fonction des paramètres (Fig. 4.2).

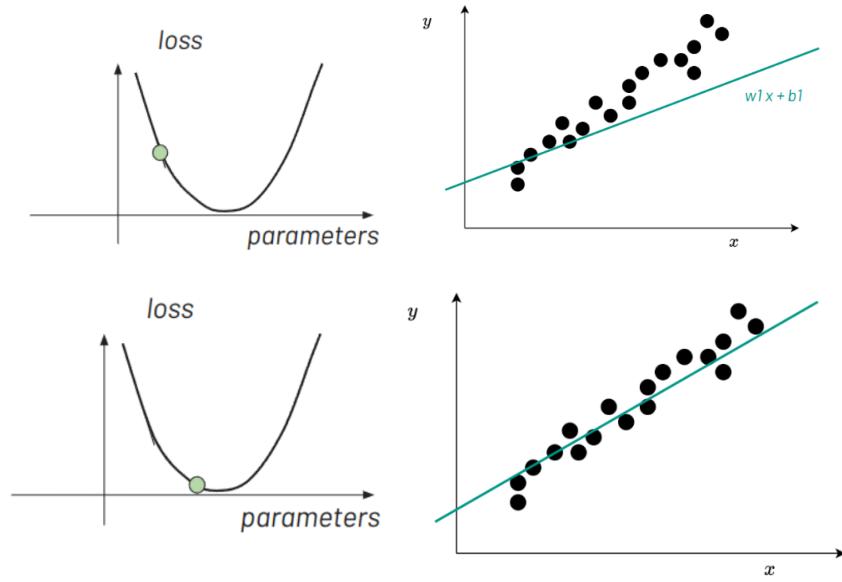


FIGURE 4.2 – Réduction de la *loss* et ajustement des poids en fonction.

L'application de cette nouvelle représentation permet de moduler les paramètres de manière à minimiser la fonction de perte. Le modèle, ainsi ajusté, génère des prédictions naturellement plus proches des données d'entraînement. En répétant ce processus de manière itérative, et en actualisant les paramètres à chaque itération pour réduire la fonction de perte, on obtient un ensemble de paramètres optimisés qui minimise efficacement la fonction de perte sur les données d'entraînement. On dit de ce modèle qu'il est *appris*, et il peut être évalué sur de nouvelles données (des datapoints inconnus constituant les données de test).

Ce processus peut être généralisé à un modèle intégrant plusieurs variables d'entrée : un modèle linéaire multidimensionnel. En effet, il serait simpliste de supposer que le prix d'une maison dépend uniquement de sa superficie. De nombreux facteurs influencent le prix, tels que l'année de construction et la localisation. Cependant, chaque paramètre n'exerce pas une influence équivalente sur les prédictions de sortie. Le modèle doit ainsi considérer la somme pondérée des n variables d'entrée, l'apprentissage consistant à apprendre ces poids, grâce au même processus itératif.

De l'apprentissage à l'apprentissage profond

Jusqu'à présent, nous avons abordé des structures mathématiques simples, représentées par des équations linéaires, qui permettent de modéliser des relations basiques entre les entrées et la sortie. Pour ces exemples, il n'est pas nécessaire d'utiliser des architectures neuronales complexes avec des connexions explicites : de tels réseaux (appelée *Multiple Layer Perceptron* ou *Perceptron multi-couche*) se composent d'une multiplication et d'association entre ces modèles linéaires organisés en couches. L'insertion de couches intermédiaires ou couches cachées (en anglais, *hidden layers*) entre l'entrée et la sortie de cette architecture complexe permet de construire des modèles de plus en plus sophistiqués.

Chaque neurone calcule une somme pondérée des entrées reçues, à laquelle il applique une fonction d'activation non linéaire, telle que la fonction sigmoïde, la tangente hyperbolique ou la fonction logistique. Cette fonction d'activation, appelée activation function, est conçue pour introduire de la non-linéarité dans le modèle : elle s'active pour des sommes pondérées élevées et modère ou annule les activations pour des sommes plus faibles. Ainsi, est créé un modèle paramétrique hautement non-linéaire qui permet de capturer des relations complexes et d'augmenter la capacité du modèle à généraliser à des données plus variées.

Le principe fondamental de l'apprentissage d'un perceptron multicouche, connu sous le nom de rétropropagation ou *backpropagation*, correspond au phénomène d'actualisation des poids. Initialement, tous les poids sont assignés de manière aléatoire. Pour chaque input dans le jeu de données d'entraînement, le réseau de neurones artificiel est activé et une sortie est observée. Cette sortie (dite *prédite*) est comparée à la sortie attendue, telle qu'annotée dans les données d'entraînement. S'il y a erreur, celle-ci est "propagée" à la couche précédente, et les poids sont ajustés en fonction. Le processus se répète jusqu'à ce que le taux d'erreur en sortie passe sous un *threshold* prédéterminé.

L'entraînement

Il faut distinguer l'algorithme d'apprentissage du modèle appris : le premier utilise les données pour produire le second, qui peut ensuite être appliqué comme un programme classique. Un algorithme d'apprentissage permet donc de modéliser un phénomène à partir d'exemples. Un modèle est un objet qui existe à un instant t , il est le résultat de l'entraînement d'un réseau de neurones et ne sera plus le même après ré-entraînement.

Le terme *entraînement* fait référence à un processus itératif composé de plusieurs

passages à travers le jeu de données, appelés époques (ou *epochs* en anglais). Au cours de chaque époque, les poids du modèle sont ajustés en fonction des erreurs observées. Le modèle module ainsi automatiquement ses poids au fil des itérations. La conclusion de l’entraînement est atteinte lorsque le modèle atteint un niveau de performance jugé satisfaisant.

Un modèle est dit “capable de généraliser” s’il est capable de modéliser et capturer des relations complexes entre les données d’entrée et les données de sortie.

“On appelle *généralisation* la capacité d’un modèle à faire des prédictions correctes sur de nouvelles données, qui n’ont pas été utilisées pour le construire”.¹⁰

Si le réseau de neurones “profond” est un modèle paramétrique dont les paramètres sont les poids de connection, ce modèle a donc d’autant plus de poids de connexion (de paramètres) qu’il a de couches intermédiaires et de neurones dans les couches. Pour éviter qu’ils ne “colle” trop aux données (et soit donc incapable de généraliser¹¹), l’apprentissage de “bons” modèles requiert souvent des quantités massives de data.

Après la phase initiale d’entraînement, il est possible de procéder à un ré-entraînement ou *fine-tuning*. Cette étape permet de rendre le modèle plus adapté à des données spécifiques, mais elle n’améliore pas nécessairement ses performances globales sur la tâche initiale. En d’autres termes, le ré-entraînement ou l’ajustement fin augmente la spécialisation du modèle pour des ensembles de données particuliers, ce qui peut influencer la capacité du modèle à généraliser à de nouveaux contextes.

Pour conclure, un modèle est avant tout une collection de poids qui s’adaptent grâce à un algorithme d’apprentissage pour réduire la *loss*. La phase d’entraînement consiste à tester des paramètres et des liens, et à affecter une pondération à ces liens. Toutes les pondérations, pour chaque couche, pour chaque neurone, constituent le modèle. Plus le phénomène à analyser est complexe plus le modèle est lourd (plus il a de poids).

4.2 Des traitements et des architectures diverses

L’implémentation de la vision artificielle change profondément les pratiques des historien.nes en leur offrant de nouveaux outils pour analyser rapidement et à grande échelle des corpus d’images. La machine devient capable d’identifier des motifs spécifiques et de les relier à d’autres sources, mettant à jour des réseaux et des narratifs complexes.

Le choix d’une architecture de réseau de neurones repose sur plusieurs considérations : la nature spécifique des données à traiter, la tâche particulière à accomplir, et l’efficacité computationnelle visée. Le défi consiste à trouver l’équilibre optimal entre l’efficacité de l’implémentation et la performance sur les données visées, car chaque type de

10. C.A. Azencott, *Introduction au machine learning...*, p.27.

11. le phénomène d’*overfitting* sera évoqué dans le chapitre 5

réseau possède des caractéristiques et des avantages propres qui le rendent plus ou moins adapté à des contextes particuliers.

Prenons pour exemple les trois tâches implémentées dans la plateforme EIDA et les questionnements qui ont motivé le choix d'une architecture.

4.2.1 YOLOv5 pour la détection : performance et légèreté

La tâche de détection d'objet consiste à identifier et localiser des objets d'intérêt dans une image. Cette tâche comprend deux sous-tâches principales : déterminer la classe de chaque objet détecté (ici, le diagramme) et localiser l'objet. Contrairement à la segmentation, la détection ne classe pas les pixels individuellement mais identifie des zones contenant chaque instance d'objet. En d'autres termes, elle ne délimite pas précisément l'objet au niveau des pixels, mais dessine une boîte (*bouding-box*) autour de celui-ci.

Pour la reconnaissance et la classification d'objets dans des images, les réseaux de neurones traditionnellement utilisés sont les CNN (*Convolutional Neural Network* ou réseau de neurone convolutionnel). Ces architectures sont composés de deux types de neurones agencés en plusieurs couches : les neurones de traitement et les neurones de mise en commun des sorties, dits de pooling. Ils procèdent de fait en plusieurs étapes : l'extraction de *features* des données d'entrée (phase de traitement), l'agrégation de ces *features* (phase de pooling), puis la résolution du problème (dans ce cas précis, la détection des objets).¹²

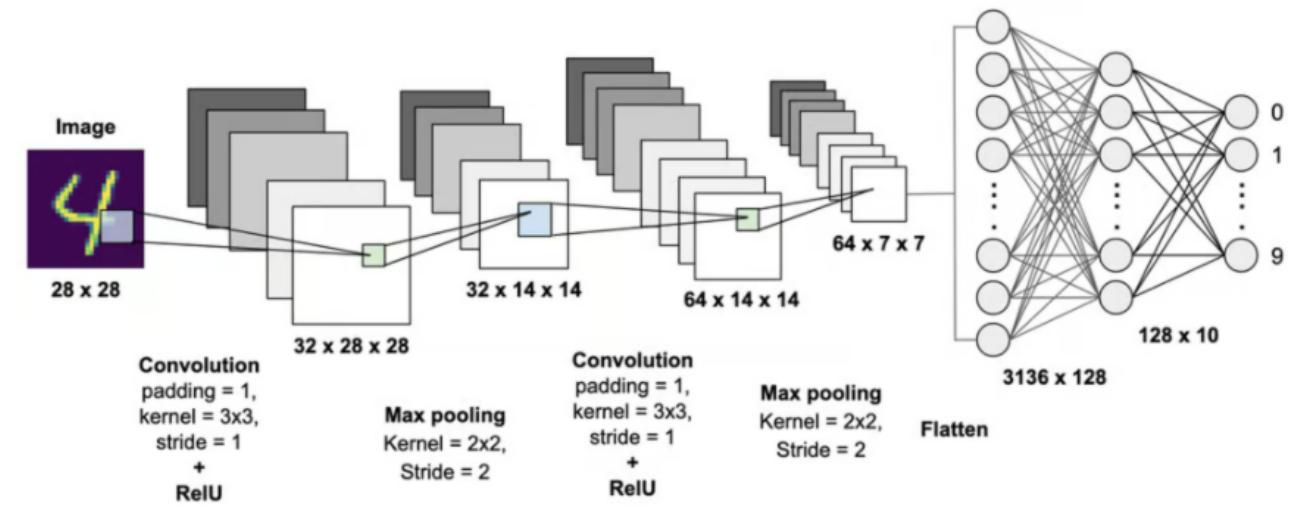


FIGURE 4.3 – Fonctionnement d'un CNN (*Convolutional Neural Network*).¹³

Les neurones de traitement des CNNS sont organisés en couches successives. Les

12. Sakshi Indolia, Anil Kumar Goswami, S. P. Mishra et Pooja Asopa, « Conceptual Understanding of Convolutional Neural Network- A Deep Learning Approach », *Procedia Computer Science*, International Conference on Computational Intelligence and Data Science 132 (janv. 2018), p. 679-688, DOI : 10.1016/j.procs.2018.05.069, p.680-681.

premières couches apprennent à détecter des caractéristiques simples (bords et textures par exemple), tandis que les couches plus profondes combinent ces caractéristiques pour reconnaître des formes plus complexes (visages et objets par exemple).

De plus, le CNN utilise un filtre convolutif, soit une matrice de valeurs qui sert à l'extraction des *features*. Il faut imaginer ce filtre comme une petite fenêtre qui glisse sur l'image. À chaque position, il calcule une valeur en multipliant les valeurs des pixels sous la fenêtre par les nombres correspondants du filtre et en additionnant les résultats. Les filtres convolutifs sont appliqués à toute l'image, ce qui permet de détecter une même caractéristique quelle que soit sa position dans l'image. Cette propriété est appelée *invariance par translation*¹⁴.

Un filtre partage les mêmes poids sur toute l'image et pour une couche donnée, de cette manière les CNNs réduisent considérablement le nombre de paramètres à apprendre par rapport à un réseau de neurones entièrement connecté. Cela les rend plus efficaces en termes de calcul et moins sujets au sur-apprentissage.

Ces propriétés (hiérarchie d'extraction, invariance par translation et partage des poids) améliorent la performance du réseau sur les images.¹⁵ C'est à ce titre qu'un CNN est employé pour la tâche de détection. YOLOv5, publié en 2020, est la cinquième version du modèle de détection d'objet et de segmentation d'images YOLO (*You Only Look Once*), développé à l'Université de Washington Joseph Redmon et Ali Farhadi et lancé en 2015. YOLO se distingue par son unique étape de détection. Cette subtilité change la donne, le rendant particulièrement rapide et précis¹⁶. Le modèle fonctionne en temps réel et peut être utilisé sur CPU (*Computing Processing Unit*) classique.

Current detection systems repurpose classifiers to perform detection. To detect an object, these systems take a classifier for that object and evaluate it at various locations and scales in a test image. Systems like deformable parts models (DPM) use a sliding window approach where the classifier is run at evenly spaced locations over the entire image. More recent approaches like R-CNN use region proposal methods to first generate potential bounding boxes in an image and then run a classifier on these proposed boxes. After classification, post-processing is used to refine the bounding boxes, eliminate duplicate detections, and rescore the boxes based on other objects in the scene. These complex pipelines are slow and hard to optimize because each individual component must be trained separately. We reframe object detection as a single regression problem, straight from image pixels to bounding box coordinates and class probabilities. Using our system, you only look once (YOLO) at an image to predict what objects are present and where they are. YOLO is refreshingly simple : [...] A single convolutional network simultaneously predicts multiple bounding boxes and class probabilities for those boxes.¹⁷

14. E. Charniak, *Introduction au Deep Learning...*, p.48.

15. S. Indolia, A. K. Goswami, S. P. Mishra, *et al.*, « Conceptual Understanding of Convolutional Neural Network- A Deep Learning Approach »..., p.680.

16. J. Büttner, J. Martinetz, H. El-Hajj, *et al.*, « CorDeep and the Sacrobosco Dataset... », p.7.

17. Joseph Redmon, Santosh Divvala, Ross Girshick et Ali Farhadi, *You Only Look Once : Unified, Real-Time Object Detection*, arXiv :1506.02640 [cs], mai 2016, DOI : 10.48550/arXiv.1506.02640, p.1.

En outre, son implémentation se veut aisée : YOLOv5, contrairement à ses prédecesseurs, est directement implémenté dans PyTorch¹⁸, permettant une intégration facile à un environnement de développement, puisqu'il nécessite moins d'adaptation que les versions précédentes fonctionnant avec le *framework* Darknet, basé sur le langage C¹⁹.

Pour conclure, YOLO est un choix qui allie performance et légèreté. Les CNNs sont traditionnellement choisis pour les tâches de détections de motifs dans des images, en raison de leur performance dans ce domaine. En outre YOLO est un modèle dit *off-the-shelf* : généraliste, réutilisable et simple d'implémentation, ce choix réduit les coûts de conception, de fabrication et de maintenance.

L'extraction permet ainsi d'isoler les éléments graphiques pertinents (diagrammes ou illustrations) au sein de grands corpus de documents. Leur traitement automatisé en batch offre une scalabilité élevée. En outre, l'extraction des unités sémantiques de base offre une granularité plus fine pour appliquer par la suite de nouveaux traitements (recherche de similarité ou vectorisation).

4.2.2 Similarité

Le choix d'un réseau de neurones pour la recherche de similarité entre les diagrammes extraits dépend de ce que l'on entend par similarité : elle peut être sémantique ou purement graphique, concerner un motif particulier ou un style global²⁰, et peut être mesurée de différentes manières²¹.

L'évaluation des similarités visuelles au sein d'un très grand corpus d'images implique en premier lieu une transformation des images en représentations numériques appropriées. Une approche classique consiste à transformer les images en représentations vectorielles basées sur les valeurs de pixels. Ces vecteurs servent de base au calcul de métriques de similarité.²².

Plusieurs approches permettent cette comparaison (par exemple le produit scalaire ou la distance euclidienne)²³. Cependant, la méthode la plus fréquemment utilisée en

18. PyTorch est une bibliothèque Python open-source. Permettant la représentation des données sous forme de tenseurs (équivalents multidimensionnels des matrices NumPy), elle permet l'entraînement des réseaux de neurones.

19. J. Norindr, *Le traitement des sources historiques par la vision artificielle...*, p.44.

20. “Quelle similarité chercher ? De multiples échelles sont possibles : celle du document (reproductions, duplicitas) ; celle de l'objet (reconnaissance d'un élément récurrent) ; celle du « style » ; ou même celle du motif (orientation d'un bras, position d'un corps, type de mise en page,...)” R. Champenois et B. Joyeux-Prunel, « Visual Contagions... », p.3.

21. Isabella Di Leonardo, Benoit Seguin et Frédéric Kaplan, « Visual Patterns Discovery in Large Databases of Paintings. » Dans *Digital Humanities 2016 : Conference Abstracts*, Kraków, 2016, p. 169-172, URL : <https://dh2016.adho.org/abstracts/348> (visité le 10/03/2024).

22. Les explications qui suivent s'inspirent de la présentation donnée par Sérgolène Albouy lors de la conférence EIDA 2024 EIDA...

23. Mathias Grønne, *Introduction to Embedding, Clustering, and Similarity*, en, oct. 2022, URL : <https://towardsdatascience.com/introduction-to-embedding-clustering-and-similarity-11dd80b00061> (visité le 10/03/2024).

vision artificielle est la *cosine distance*. Cette approche consiste à calculer le cosinus de l'angle entre les deux vecteurs à comparer. La méthode est particulièrement efficace car elle évalue l'orientation des vecteurs plutôt que leur magnitude, offrant ainsi une mesure de similarité relative, plus robuste. Cette métrique est particulièrement adaptée aux espaces vectoriels de haute dimension (plus de deux) et est largement utilisée en vision par ordinateur.

Elle ne donne toutefois pas de résultats optimaux quand elle est calculée sur une représentation basée sur les valeurs de pixels. Une représentation intermédiaire d'une image via un modèle par extraction des *features* constitue une représentation plus optimale de l'image. Les réseaux de neurones, conçus initialement pour des tâches de classification ou de détection d'objets, peuvent être réutilisés pour l'extraction de caractéristiques. En effet, les couches intermédiaires contiennent des représentations de plus en plus abstraites de l'image. Ces vecteurs de caractéristiques (ou *feature vectors*), de dimensionnalité réduite par rapport à la représentation pixellique, condensent l'information visuelle pertinente pour comparaison.

Ce que capturent les *features* au sein de l'image est assez insaisissable, néanmoins l'utilisation de la *cosine distance* sur ces vecteurs permet de mesurer l'alignement sémantique entre deux images, au-delà des simples similarités visuelles de bas niveau²⁴.

En choisissant le bon extracteur pour les *features* (la *backbone*²⁵ ou *feature net*)²⁶ et la bonne couche²⁷, la pertinence des vecteurs descripteurs se trouve optimisée.

Cette approche, cependant, n'est toujours pas suffisante pour capturer des éléments visuels répétés, indépendamment du style et de la forme exacte. Pour pallier à cette limite, il est possible d'utiliser la méthode Segswap²⁸. Initialement développée pour la recherche en histoire de l'art, cette méthode a été éprouvée sur l'analyse des œuvres issues des ateliers de Brueghel : ses assistants étaient connus pour répéter des petits motifs d'un tableau à l'autre. Le modèle prend deux images en entrée et renvoie un masque : un vecteur de correspondance qui *map* les deux parties de l'image qui se répètent.

24. Farley, *Multimodal embeddings concepts - Image Analysis 4.0 - Azure AI services*, en-us, févr. 2024, URL : <https://learn.microsoft.com/en-us/azure/ai-services/computer-vision/concept-image-retrieval> (visité le 10/03/2024).

25. La *backbone* est chargé d'extraire les *features* de l'image. Elle sert de base sur laquelle sont construites les couches supérieures du réseau, qui sont souvent spécifiques à une tâche particulière (classification, détection d'objets, segmentation). Les backbones sont souvent pré-entraînés sur d'énormes ensembles de données, ce qui leur permet de capturer les caractéristiques générales sur le monde visuel. Elles peuvent ensuite être réutilisées comme point de départ pour de nombreuses autres tâches, accélérant ainsi le processus d'apprentissage et le développement d'architectures spécifiques.

26. EIDA utilise MoCo (Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie et R. Girshick, *Momentum Contrast for Unsupervised Visual Representation Learning*, arXiv :1911.05722 [cs], mars 2020, DOI : 10.48550/arXiv.1911.05722), une architecture convolutionnelle construite pour un apprentissage non-supervisé.

27. Dans le cas présent la quatrième couche convolutionnelle.

28. X. Shen, Alexei A. Efros, Armand Joulin et M. Aubry, *Learning Co-segmentation by Segment Swapping for Retrieval and Discovery*, arXiv :2110.15904 [cs] version : 2, mars 2022, DOI : 10.48550/arXiv.2110.15904.

Cependant, l'application de l'algorithme Segswap à l'ensemble des paires d'images nécessite trop de ressources computationnelles. Pour optimiser le processus, il a été mis en place une approche hybride. Dans un premier temps, le score de *cosine distance* entre les *feature vectors* est calculé pour toutes les paires d'images. Cette étape permet de filtrer les paires d'images les moins susceptibles de contenir des motifs récurrents, réduisant ainsi considérablement le nombre de paires à soumettre à l'algorithme Segswap. Dans un second temps, l'algorithme Segswap est appliqué uniquement aux paires d'images ayant obtenu les scores de similarité cosinus les plus élevés.

Cette approche permet retrouver les variantes d'un même diagramme au fil des copies et des éditions, alignant ainsi plusieurs témoins en vue d'une édition critique. Les chercheur.ses pourront aussi s'appuyer sur Segswap pour retrouver les variations d'un même élément graphique, permettant de comparer les conventions visuelles en diachronie. L'utilisation de la vision artificielle rend la tâche réalisable sur une très grande base de donnée, ouvrant la voie à des études comparatives à grande échelle.

Cependant, ces algorithmes ne ramèneront pas les similarités dites "topologiques"²⁹. La vectorisation ouvre des perspectives pour une analyse plus fine des composants des diagrammes, afin de détecter les relations structurelles sous-jacentes.

4.2.3 La vectorisation : les transformers

L'implémentation de la vectorisation permet d'extraire automatiquement l'information géométrique contenue dans les diagrammes dans un format léger et sémantiquement riche. Les représentations vectorielles sont en effet significativement plus compactes que leurs homologues matricielles. Elles sont stockées dans un fichier appelé SVG pour *Scalable Vector Graphic*. Ce format utilise le langage XML pour encoder la représentation vectorielle du diagramme. Les SVGS offrent en outre une représentation du contenu de l'image indépendante de sa résolution. Cette invariabilité est rendue possible par l'utilisation de primitives géométriques (formes de base) et de leurs paramètres, plutôt que de pixels, pour encoder l'image³⁰. Dans le cas présent, trois types de primitives sont utilisés : les lignes, les cercles et les arcs, qui constituent les formes fondamentales des diagrammes du corpus EIDA.

Chaque *primitive* (forme géométrique élémentaire) est définie par un ensemble de paramètres pour pouvoir être tracée. Un cercle est encodé à l'aide de son rayon et des coordonnées de son centre ; la ligne est définie par les coordonnées des points de départ et d'arrivée ; l'arc par les coordonnées des points de départ, d'arrivée et du *midpoint*.

29. La transformation topologique d'un diagramme n'affecte pas son contenu mathématique. À une démonstration textuelle peuvent correspondre plusieurs représentations géométriques. Deux diagrammes peuvent donc être aspectuellement différents mais substantiellement équivalents.

30. *Tutoriel SVG - SVG (Scalable Vector Graphics)* / MDN, fr, juill. 2024, URL : <https://developer.mozilla.org/fr/docs/Web/SVG/Tutorial> (visité le 03/08/2024).

Cet encodage autorise un accès immédiat au contenu géométrique des diagrammes, et permet une manipulation fine des éléments graphiques : la machine peut exploiter ce format en s'appuyant sur les valeurs numériques qui définissent les primitives. L'implémentation de la vectorisation dans la plateforme vise alors à fournir un outil efficace pour étudier et analyser sémantiquement le contenu géométrique des diagrammes, ouvrant la voie aux divers post-traitements sous-tendant l'établissement d'une édition enrichie³¹.

Quels défis présentent les sources ? Et quelles méthodes sont adoptées en conséquence ?

La conversion d'images *raster* (matricielles, composées de pixels) de dessins en images vectorielles est un problème déjà exploré dans le domaine de la recherche en vision artificielle³². Les premières approches séquentielles, impliquant un prétraitement en deux étapes (filtrage, simplification et *Edge Detection*³³ suivi d'une détection de primitives³⁴, se sont révélées efficaces sur des images de haute qualité. Cependant, ces méthodes peinent à généraliser à des images plus complexes ou bruitées³⁵.

Or le corpus d'EIDA, composé de diagrammes astronomiques extraits de manuscrits historiques, présente des défis spécifiques qui compliquent leur traitement automatisé par les méthodes traditionnelles. Les documents sont bien souvent détériorés (taches ou atténuation des tracés) ; les traits manuels sont généralement irréguliers et présentent des variations dans l'épaisseur et la clarté des lignes ; les diagrammes peuvent être très simples comme présenter de nombreux traits enchevêtrés. La principale difficulté réside dans la coexistence du texte et des éléments graphiques. Le texte couvre fréquemment une partie du contenu géométrique ; parfois même, les diagrammes sont superposés avec des blocs de texte par manque de place³⁶.

La complexité de la donnée historique rend la tâche presque impossible aux méthodes de traitement séquentielles trop fortement dépendantes d'une détection des contours (première étape : *edge detection*) qualitative. Or les structures les plus fines sont souvent perdues lors de cette étape, ou bien les résultats souffrent de bruit excessif. La séquentialité accroît ainsi le risque d'erreurs cumulatives. D'autre part, lors de la dernière étape, les

31. Voir le chapitre 6

32. Vage Egiazarian, Oleg Voynov, Alexey Artemov, Denis Volkhonskiy, Aleksandr Safin, Maria Taktasheva, Denis Zorin et Evgeny Burnaev, « Deep Vectorization of Technical Drawings », 12358 (2020), arXiv :2003.05471 [cs], p. 582-598, DOI : 10.1007/978-3-030-58601-0_35.

33. John Canny, « A Computational Approach to Edge Detection », *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-8-6 (nov. 1986), Conference Name : IEEE Transactions on Pattern Analysis and Machine Intelligence, p. 679-698, DOI : 10.1109/TPAMI.1986.4767851.

34. **ref-noauthor_ransac_nodate**.

35. X. Hilaire et K. Tombre, « Robust and accurate vectorization of line drawings », *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28-6 (juin 2006), Conference Name : IEEE Transactions on Pattern Analysis and Machine Intelligence, p. 890-904, DOI : 10.1109/TPAMI.2006.127.

36. Syrine Kalleli, Scott Trigg, S. Albouy, M. Husson et M. Aubry, *Historical Astronomical Diagrams Decomposition in Geometric Primitives*, arXiv :2403.08721 [cs], mars 2024, DOI : 10.48550/arXiv.2403.08721.

motifs les plus petits³⁷ ont de grandes chances d'être manqués. Il n'est pas non plus possible d'appliquer un filtre sémantique aux détections, et les bordures des document sont souvent détectées comme des lignes. La dernière étape ralentit aussi considérablement le processus³⁸. En outre, la méthode séquentielle requiert une intervention humaine importante pour ajuster les paramètres en fonction des caractéristiques spécifiques de chaque jeu de données. Cette dépendance à l'expertise de l'utilisateur.rice restreint sa capacité de généralisation et rend difficile son application à des corpus de diagrammes hétérogènes.

Syrine Kalleli, doctorante à l'École des Ponts, propose une solution basée sur un modèle de type *transformers* pour remédier aux problèmes posés par les méthodes classiques³⁹.

Les transformers se basent sur une une méthode d'apprentissage de type *Seq2Seq* et sur une bipartition structurelle (encodeur-décodeur). L'apprentissage *séquence à séquence* est une technique d'apprentissage profond permettant d'associer une séquence de symboles à une autre séquence de symboles sans que cette mise en correspondance soit basée sur les symboles en eux-mêmes. Son application par excellence est la traduction automatique, mais elle connaît ces dernières années, de plus en plus d'application dans le traitement de l'image⁴⁰. Elle consiste à donner à la machine un *corpus aligné*, à savoir de nombreux exemples de paires, et d'entraîner un modèle à retrouver l'association. Le modèle opère en deux passes. La première, dont le but est d'extraire les *features* de l'image d'entrée, est appelée passe d'encodage. La passe de décodage vise l'élément cible : partant de ces *features*, elle prédit les primitives d'intérêt⁴¹.

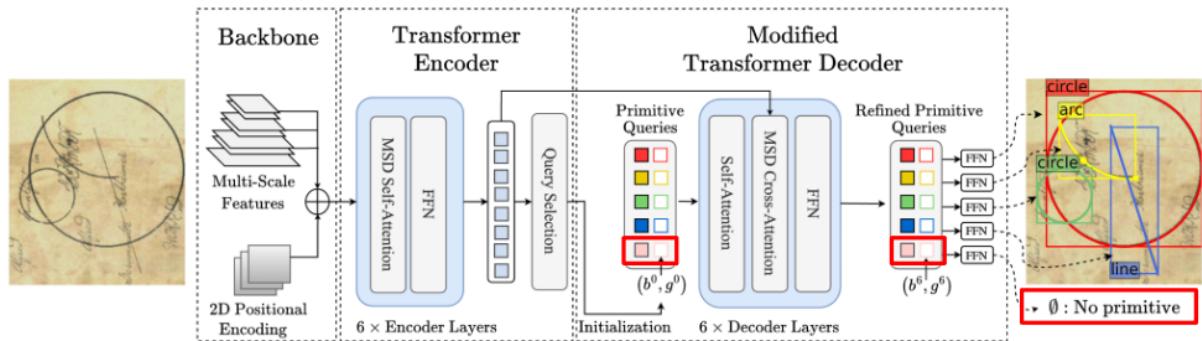


FIGURE 4.4 – Fonctionnement de la méthode de vectorisation, basée sur les transformateurs.⁴³

37. Il y a moins de pixels et RANSAC (*RANdom SAmple Consensus*) se base sur le nombre de pixel pour la détection des formes.

38. la dernière étape de détection (cf. (*RANSAC*, en, URL : <https://www.mathworks.com/discovery/ransac.html> (visité le 14/08/2024)), détecte les primitives une par une.

39. Id., *Historical Astronomical Diagrams Decomposition in Geometric Primitives...*

40. Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, *et al.*, *An Image is Worth 16x16 Words : Transformers for Image Recognition at Scale*, arXiv :2010.11929 [cs], juin 2021, DOI : 10.48550/arXiv.2010.11929.

41. E. Charniak, *Introduction au Deep Learning...*, p.90.

```

lines
[[134.81066895 457.17818451 779.38134766 248.8579483 ]
 [450.41806698 701.88641357 445.5779047 29.55957031]
 [ 82.28451538 235.88547516 751.17868042 460.67421722]
 [846.97953796 687.73287964 839.02302551 46.39743042]]

line_scores
[0.9833361 0.97366863 0.91139853 0.6446364]

circles
[[444.50769043 358.36816406 325.14129639]
 [122.31943512 357.18722534 99.33002472]
 [771.32922363 357.23934937 100.2789917]]

circle_scores
[0.9783766 0.955871 0.8885103]

arcs
[[720.67673302 494.20816517 745.99196815 474.5509901 740.33850098
 491.46079016]]

arc_scores
[0.31675324]

```

FIGURE 4.5 – Exemple de contenu d'un fichier NPZ

En outre, contrairement aux CNNs dont les performances reposent sur une convolution séquentielle de l'image, les modèles *seq2seq* doivent leur efficacité à la notion d'*attention*. Le mécanisme d'attention confère au modèle la capacité de hiérarchiser l'information contenue dans une image. En attribuant des poids variables aux différentes régions de l'image, le modèle est en mesure de focaliser son traitement sur les zones les plus pertinentes pour la tâche à accomplir, en utilisant un ensemble de *requêtes* (ou *queries*)⁴⁴. Chaque requête correspond à une primitive parmi le triplet initialement défini : segment, arc et cercle⁴⁵.

En somme, la méthode consiste à encoder l'image en représentations sémantiques ou *features*, et à effectuer un décodage des *features* en concentrant l'attention du modèle sur un ensemble de motifs à détecter, en l'occurrence trois types de primitives. Le modèle ramène alors les coordonnées de ces primitives : dans un premier temps, elles sont stockées dans un format Numpy : une structure de données permettant la manipulation d'arrays multidimensionnels (Fig. 4.5). Post-inférence, un script est capable de reconstituer les fichiers SVG au format XML.

```

<?xml version='1.0' encoding='us-ascii'?>
<svg xmlns="http://www.w3.org/2000/svg" xmlns:xlink="http://www.w3.org/1999/
  xlink" baseProfile="tiny" height="2597" version="1.2" width="2460" xmlns:
  inkscape="http://www.inkscape.org/namespaces/inkscape" xmlns:sodipodi="http
  ://sodipodi.sourceforge.net/DTD/sodipodi-0.dtd" inkscape:version="1.3 (0

```

44. *Ibid.*, p.95.

45. *EIDA...*

```

e150ed, 2023-07-21)">
<defs />
<image height="2597" width="2460" x="0" xlink:href="ms154_0139_1798
,1606,2460,2597.jpg" y="0" />
<circle cx="1265.9037" cy="1293.51" fill="none" r="506.6848" stroke="orange"
stroke-width="3" />
<circle cx="1260.7471" cy="1288.5127" fill="none" r="774.76514" stroke=
orange" stroke-width="3" />
<circle cx="1248.9653" cy="1293.2999" fill="none" r="1173.3875" stroke=
orange" stroke-width="3" />
<path d="M 947.62726 562.3929 L 2092.92 1338.452" fill="none" stroke="green"
stroke-width="3" />
<path d="M 91.71399 1291.111 L 2418.0642 1297.6548" fill="none" stroke="green
" stroke-width="3" />
<path d="M 1261.4149 2524.8396 L 1283.8896 109.190315" fill="none" stroke=
green" stroke-width="3" />
<path d="M 1819.7242 950.7695 L 1928.3906 1129.3221" fill="none" stroke=
green" stroke-width="3" />
<path d="M 1279.6703 109.75813 L 1633.5283 665.58594" fill="none" stroke=
green" stroke-width="3" />
<path d="M 792.355 186.13255 L 2428.4106 1307.5989" fill="none" stroke="green
" stroke-width="3" />
</svg>

```

Listing 4.1 – exemple de contenu d'un fichier SVG

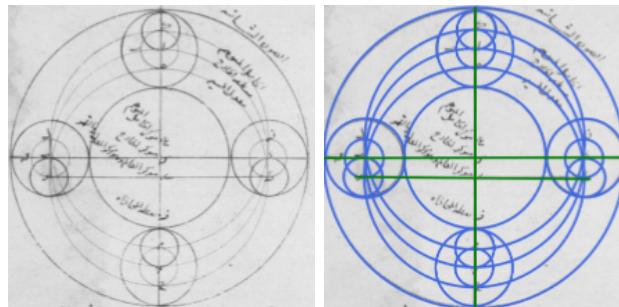


FIGURE 4.6 – Résultat de la vectorisation.

En résumé, la diversité des architectures de réseaux de neurones permet d'adapter les modèles aux spécificités des données et des tâches à traiter, tout en optimisant les performances et l'efficacité computationnelle. Cette flexibilité est essentielle pour répondre

aux multiples défis posés par le traitement automatique des documents historiques. Deux approches se distinguent : l'utilisation de modèles *off-the-shelf* ou d'architectures spécifiques. Traditionnellement, les ingénieur.es concevaient des architectures de réseau de neurones sur mesure, adaptées spécifiquement aux problèmes à résoudre. Cette approche, bien que permettant une optimisation fine des performances, est souvent coûteuse en temps et en ressources. Ces dernières années, l'émergence de modèles pré-entraînés sur des jeux de données massifs a ouvert de nouvelles perspectives. Ces modèles off-the-shelf (ou pre-trained models), déjà qualitatifs, demandent cependant à être spécialisés sur des corpus. Le *fine-tuning* consiste alors à adapter ces modèles à une tâche spécifique en les entraînant sur un jeu de données plus petit et plus pertinent.

L'approche de corpus de dimensions importantes, ne permettant pas une analyse manuelle, trop chronophage, est permise par la vision artificielle. L'utilisation des modèles de vision présente cependant des limites : notamment, la complexité de leur fonctionnement reste assez opaque aux non spécialistes, ce qui peut être un inconvénient là où la transparence et l'explicabilité sont importantes, typiquement pour des projets transversaux caractéristiques des humanités numériques. D'autant que la collaboration entre les chercheur.ses spécialistes des sources et les ingénieur.es *data-scientists* est essentielle : les deux pôles doivent collaborer pour entraîner les modèles et les rendre performants sur les données spécifiques. C'est pourquoi l'optimisation des modèles de vision ne s'obtient pas sans un effort du côté des équipes disciplinaires, qui vont réunir et annoter des jeux de données importants pour entraîner efficacement les modèles.

Chapitre 5

La donnée : générique ou caractéristique, réelle ou synthétique

Le *machine learning* repose sur deux piliers : l'algorithme d'une part, qui est la procédure que l'on fait tourner sur les données pour produire un modèle, et d'autre part les données, qui sont les exemples à partir desquels l'algorithme ajuste les poids du modèle (il *apprend*). L'entraînement consiste à faire tourner un algorithme d'apprentissage sur un jeu de données (très) important et accompagné de ses étiquettes ou annotations (correspondant aux résultats espérés du modèle). La disponibilité de corpus annotés assez divers et importants est un défi à relever, particulièrement dans le domaine de l'histoire des sciences. L'un des principaux challenges de la reconnaissance sémantique des éléments visuels dans les documents historiques est leur grande variabilité, ainsi que, et surtout, la rareté générale de jeux de données historiques cohérents axés sur leur détection. De plus, dans la majorité des cas où des éléments graphiques ont été rassemblés, ils n'ont pas été annotés selon des classes sémantiques issus d'ontologies normalisées. En outre, les *datasets* disponibles sont souvent trop spécifiques, concernant une période ou un support précis, tels que des manuscrits écrits à la main, des livres imprimés ou des journaux. On pourra par exemple citer le *Newspaper Navigator Dataset*¹ qui contient des éléments visuels issus de 16 millions de journaux des Etats-Unis publiés en 1789 et 1963 ou encore le *HORAE dataset*² qui se concentre sur les livres de prière de la fin du Moyen Âge. En histoire des sciences astronomiques ou mathématiques, aucun projet n'a encore entrepris de collecte d'envergure, hormis peut-être le projet Sphaera³ (travail sur les premiers manuscrits) ou le projet VHS, auquel EIDA se greffe.

1. Benjamin Charles Germain Lee, Jaime Mears, Eileen Jakeway, Meghan Ferriter, Chris Adams, Nathan Yarasavage, Deborah Thomas, Kate Zwaard et Daniel S. Weld, « The Newspaper Navigator Dataset : Extracting Headlines and Visual Content from 16 Million Historic Newspaper Pages in Chronicling America », dans *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, Virtual Event Ireland, 2020, p. 3055-3062, DOI : 10.1145/3340531.3412767.

2. HORAE - Hours : Recognition, Analysis, Editions, URL : https://heurist.huma-num.fr/heurist/?db=stutzmann_horae&website&id=378891 (visité le 14/08/2024).

3. The Sphere, URL : <https://sphaera.mpiwg-berlin.mpg.de/> (visité le 14/08/2024).

Cette rareté s'explique cependant : l'annotation des jeux de données représente une tâche chronophage qui nécessite des moyens humains importants et l'implication d'experts capables de reconnaître et de classer avec précision les éléments pertinents dans les images.

Par conséquent, il est nécessaire de trouver des solutions pour remédier à la rareté des jeux de données d'images de documents historiques annotés, telles que l'utilisation de techniques de génération de données synthétiques ou l'utilisation de modèle généraux tirant partie du *crowdsourcing* pour leur entraînement. On verra que ces stratégies sont néanmoins insuffisantes, et que la constitution de jeux de données réelles et contextuelles est nécessaire pour ne pas sacrifier la pertinence des modèles à l'efficience de leur fabrication.

5.1 Partir d'un modèle généraliste

Partir d'un modèle de vision généraliste et le *fine-tuner* est une pratique courante pour adapter des architectures de réseaux de neurones pré-entraînés à des tâches spécifiques. Cette approche permet de bénéficier des capacités de transfert d'apprentissage du *machine learning*. Le *fine-tuning* optimise le modèle pour qu'il réponde aux exigences particulières d'une application donnée, tout en réduisant le temps et les ressources nécessaires (humaines comme en données).

5.1.1 Utilisation de modèles off-the-shelf

Les modèles de vision *off-the-shelf* comme ResNet, VGG, Inception ou encore YOLOv5, ont déjà été entraînés sur un ensemble de données volumineux et générique. Leur utilisation, facilitée par leur architecture pré-formée, ne demande pas d'expertise trop poussée en *machine learning*. Ils permettent d'économiser du temps et des ressources : il n'est pas nécessaire d'entraîner le modèle à partir de zéro.

Pour l'entraînement initial, YOLOv5 utilise MS COCO⁴, vaste jeu d'images couvrant environ 1.5 million d'instances d'objets annotées selon 80 catégories relevant du quotidien (comme *personne*, *chaise* ou *voiture*) et 91 catégories dites "stuff", qui se réfèrent à des éléments de fond amorphes tels que le ciel ou l'herbe. MS COCO est largement utilisé dans diverses recherches et compétitions dans les domaines de base de la vision artificiel, tels que la détection d'objets, la segmentation d'instances, la génération de légendes... Cet ensemble de données se distingue par ses annotations détaillées et la complexité des scènes représentées, offrant ainsi une référence robuste pour entraîner et évaluer les modèles de *deep learning*.

Ultralytics utilise en outre ImageNet⁵ pour préformer ses modèles. ImageNet est

4. *COCO - Common Objects in Context*, URL : <https://cocodataset.org/#home> (visité le 20/06/2024).

5. *ImageNet*, URL : <https://www.image-net.org/> (visité le 20/06/2024).

une vaste base de données en accès libre d’images annotées contenant plus de 14 millions d’images regroupées en plus de 20 000 catégories, allant d’objets courants comme des fruits et des animaux à des objets plus spécifiques comme des instruments de musique ou des types de véhicules. Cette initiative est née du besoin de données d’entraînement et de validation pour le *machine learning*, et particulièrement pour la classification, tâche de base de la vision par ordinateur. Chaque image du jeu de données est annotée manuellement pour indiquer les objets présents dans l’image : le projet est participatif, ce qui assure la continuité de son enrichissement. ImageNet ne possède pas les droits sur les images, les chercheur.ses ou ingénieur.es qui souhaitent les utiliser à des fins non-commerciales (de recherche ou éducatives) ont accès par l’intermédiaire du site à des compilations de listes d’images disponibles sur le web, chaque liste correspondant à des concepts définis par le projet WordNet⁶.

Dans le domaine de la vision artificielle, il est très courant d’utiliser ces modèle pré-entraînés sur un grand jeu de données peu spécifique (qui en plus bénéficient de solutions d’implantation simplifiées), puis de l’affiner à partir d’un jeu de données plus petit et très spécifique au projet. Le modèle entraîné sur le jeu de données de base s’avère en effet trop généraliste : il apprend des caractéristiques qui peuvent être appliquées à l’ensemble du monde visuel. YOLOv5, employé pour l’extraction des diagrammes, sert ainsi de modèle générique mais demande à être affiné : l’étape de *fine-tuning* assurant la portabilité des caractéristiques générales vers les attentes du projet.

L’existence de tels modèles génériques est néanmoins indispensable pour des projets comme EIDA et VHS, pour lesquels les données et les sources disponibles sont en nombre important du point de vue des possibilités d’exploitation humaine (c’est pourquoi l’IA est employée) et de la faisabilité d’une annotation manuelle extensive, mais limitées vis-à-vis des besoins de la vision artificielle.

“Cette volonté de rendre accessible des outils adaptables, construits comme un socle solide pour des objectifs variés, permet l’intégration de la vision artificielle à des projets divers, sans demander de ces derniers de créer ou d’entraîner de zéro des réseaux de neurones ou des modèles de détections. La détection étant, en effet, une tâche de base de la vision par ordinateur,[...] il est peu pertinent d’allouer des ressources pour reproduire des techniques déjà appliquées par de nombreux projets, et les outils off-the-shelf offrent donc la liberté d’envisager la vision artificielle comme élément d’une chaîne de traitement des sources en allégeant les besoins et ressources – notamment humains et temporels que demandent le développement et l’application de ces techniques.”⁷

6. WordNet, en, URL : <https://wordnet.princeton.edu/homepage> (visité le 14/08/2024).

7. J. Norindr, *Le traitement des sources historiques par la vision artificielle...*, p.49.

5.1.2 fine-tuner le modèle

Les modèles très généralistes entraînés sur de vastes jeux de données extrêmement diversifiés (à l'image de YOLOv5), ne satisfont pas les exigences des projets de recherche EIDA et VHS : ils nécessitent un *fine-tuning* ou ré-entraînement. Le but du fine-tunning est de spécialiser le modèle sur les données caractéristiques. Un modèle qui a déjà connu un entraînement initial est performant sur une tâche donnée, il s'agit alors d'exploiter les représentations apprises précédemment et de spécialiser le modèle sur des données plus spécifiques et un contexte précis. Cette démarche permet de réduire la quantité de données nécessaire pour l'obtention d'un modèle à la fois pertinent et performant. Les modèles off-the-shelf comme YOLOv5 prévoient dans leur implémentation un *workflow* pour l'entraînement. La spécialisation des dernières couches est ainsi simplifiée par la mise à disposition d'outils permettant d'effectuer cette étape en écrivant peu de code. Il prévoit donc d'adapter la tâche de détection à des objets spécifiques dans un contexte particulier, typiquement les diagrammes dans des pages de manuscrit.

“Les projets EIDA et VHS entraînent, pour la détection d'illustrations dans les numérisations d'ouvrages, des modèles ayant pour base YOLOv5. Sans entraînement spécifique, en s'appuyant exclusivement sur le pré-entraînement fait avant la mise en ligne du modèle, les performances sont peu satisfaisantes : les jeux de données utilisés pour ce pré-entraînement, qu'il s'agisse d'ImageNet ou de MS COCO, sont en effet des jeux de données d'images réelles, faits pour l'apprentissage de la classification d'objets du quotidien, et ne sont donc initialement pas adaptés à la segmentation de pages de manuscrits ou à la détection d'illustrations. Il faut ainsi compter sur les propriétés de portabilité de l'apprentissage, qui assurent que ce pré-entraînement sur des images réelles permet d'accélérer le processus de développement du modèle, puisqu'elles permettent de s'intéresser immédiatement à l'entraînement à partir de données spécifiques, et réduisent le volume nécessaire.”⁸

À chaque étape de ré-entraînement, le modèle évolue, il devient un nouvel objet informatique, manipulable et implantable grâce à des scripts annexes. Dans le cadre de la collaboration EIDA/VHS, deux modèles d'extraction distincts sont ainsi obtenus, l'un est moins spécifique, adapté aux données VHS, et l'autre est spécialisé pour les diagrammes d'EIDA. L'objectif est d'affiner la détection et la reconnaissance d'objets en fonction des caractéristiques spécifiques présentes dans ces données et des attentes des deux projets jumeaux, d'où deux apprentissages différents sur deux jeux de données distincts.

5.1.3 Construction d'une vérité de terrain

Les expériences d'apprentissage supposent de séparer le jeu de données d'images annotées⁹ en trois groupes représentatifs, aléatoires et disjoints. Les proportions peuvent varier de plus ou moins 10%, mais une répartition typique serait la suivante. 80% de

8. *Ibid.*, p.44.

9. C'est-à-dire accompagnées des étiquettes correspondant à la prédiction attendue.

l'ensemble de départ constitue le *training set* ou *dataset* d'entraînement, utilisé pour ajuster les paramètres du modèle. Un échantillon d'environ 10% est dédié à la validation de l'entraînement en comparant des prédictions sur des données inconnues avec les résultats initiaux : il s'agit de vérifier que le modèle généralise bien et de comparer aux résultats des autres expériences d'apprentissage. Enfin, environ 10% constitue le *test set* utilisé pour évaluer les performances du modèle final. Une fois qu'un modèle est établi et qu'il fournit des résultats (estimés) satisfaisant, il est testé sur ce jeu de test pour en tirer des métriques. Il est important de distinguer la sélection d'un modèle de son évaluation quantitative, d'où la distinction entre validation et test.

5.1.4 Choisir et évaluer le modèle

Le *data-scientist* peut activer deux leviers lors d'un entraînement pour améliorer le modèle qui en sortira : la donnée en elle-même et le nombre d'itération sur les données.

Un modèle est dit *sur-appris* (ou *overfitted*) lorsqu'il capture le bruit présent dans les données d'entraînement, en plus des *patterns* pertinents pour l'application.

“un modèle qui sur-apprend est généralement un modèle trop complexe, qui “colle” trop aux données et capture donc aussi leur bruit”.¹⁰

Cela se produit lorsque le modèle est entraîné sur trop peu de données, ou qu'elles ne sont pas assez représentatives (trop spécifiques). Le modèle apprend alors à se concentrer sur les détails caractéristiques des données d'entraînement, mais il ne parvient pas à généraliser à de nouvelles données. En conséquence, il obtient de très bons résultats sur les données d'entraînement, mais ses performances se dégradent considérablement lorsqu'il est confronté à de nouvelles données.

L'inverse existe aussi : un modèle est en situation de *sous-apprentissage* (ou *underfitting*) lorsqu'il est trop simple pour capturer les *patterns* pertinents dans les données d'entraînement. Cela peut se produire pour plusieurs raisons, notamment un manque de données (si le modèle ne dispose pas de suffisamment de données pour apprendre les *patterns* complexes présents dans les données d'entraînement), trop de bruit ou de valeur erronées dans les données (le modèle ne sera donc pas capable de dégager de *patterns*, il ne capturera pas la relation entre les exemples d'entraînement et les valeurs cibles) ou le choix d'un modèle trop simple (pas assez profond) qui n'est pas adapté à la complexité des données (encore une fois il ne sera pas en mesure de capturer les relations subtiles entre les données et leurs étiquettes). Dans le cas du sous-apprentissage, le modèle obtient de mauvais résultats sur les données d'entraînement et sur les nouvelles données.

Choisir le modèle

La capacité de généralisation et le sur-apprentissage sont des enjeux majeurs en *machine learning* : il faut s'assurer que les modèles entraînés pour minimiser les erreurs de

10. C.A. Azencott, *Introduction au machine learning...*, p.29.

prédition sur un jeu d’entraînement défini seront capables de généraliser à l’ensemble des données d’intérêt. Pour ce faire, il est important de mettre en place un cadre expérimental rigoureux qui permette d’évaluer les modèles en évitant, tout particulièrement, le biais du sur-apprentissage. Sans anticiper l’overfitting, les développeur.ses peuvent entraîner et déployer un modèle hautement précis à première vue qui offrira en réalité des performances amoindries en production sur de nouvelles données¹¹.

Le jeu de validation est au cœur de ce protocole rigoureux : le modèle est entraîné sur le jeu d’entraînement, puis ses performances sont évaluées sur le jeu de validation. L’entraînement et l’évaluation peuvent être répétés plusieurs fois, à chaque itération les paramètres du modèle s’ajustent. L’*epoch* désigne le nombre d’itération sur le jeu de donnée : il s’agit de l’optimiser et de sélectionner le meilleur modèle parmi plusieurs candidats issus des N itérations. Le modèle final sélectionné est celui qui a obtenu les meilleures performances sur le jeu de validation : arrêter l’entraînement quand la mise à jour des paramètres n’améliore plus les performances sur le jeu de validation vise à prévenir l’overfitting. Souvent nommé *best.pt*, le meilleur modèle n’est donc pas nécessairement celui obtenu à l’issue de la dernière époque (*last.pt*).

Évaluer le modèle

Pour estimer la performance en généralisation d’un modèle, il est essentiel de l’évaluer sur des données qui n’ont été utilisées ni pour l’entraînement, ni pour la sélection de ce modèle, d'où la prévision du jeu de test (ou “test set”). Il est utilisé pour évaluer les performances du modèle sélectionné sur des données inconnues, autrement dit pour évaluer ses capacités de généralisation grâce à des métriques diverses. Parmi ces dernières, le *recall* désigne la proportion d’instances (diagrammes) détectées. La *précision* correspond au ratio de prédiction positive par rapport au nombre de prédiction total. Ces métriques demandent à être interprétées. Dans le cadre de la détection d’objets dans EIDA (extraction des diagrammes dans les pages de manuscrit) le recall est plus important que la précision. L’enjeu réside dans l’ajustement entre le retour de tous les diagrammes, en acceptant un certain nombre de faux positifs, et l’absence de faux positifs, mais avec le risque de manquer certains diagrammes. La première configuration présente le moins de risque : il est en effet plus aisé de supprimer un faux positif que de rechercher manuellement les diagrammes manquants dans les sources.

En résumé

Le jeu de validation est utilisé pour développer le modèle et sélectionner le meilleur candidat. Le jeu de test est utilisé pour évaluer la performance finale du modèle sélectionné sur des données non vues.

L’optimisation des performances passe aussi par une action sur la donnée. Une fois un modèle choisi et évalué à l’issu d’un entraînement, il faut examiner les résultats et réfléchir à la manière dont il est possible de l’améliorer en fournissant un retour d’information au

11. *Ibid.*, p.32.

système sous la forme d'un jeu de données annoté.

Après le dernier entraînement du modèle d'extraction sur les données d'EIDA uniquement (afin de le spécialiser sur les diagrammes uniquement), les résultats présentent encore des cas problématiques. Notamment, le modèle détecte comme diagramme et extrait les marginalia et les caractères chinois. Il faudra donc agir sur la *groundtruth* et placer d'avantage d'exemples négatifs de ces cas précis.



FIGURE 5.1 – Détection erronée des marginalia.

En outre, le modèle ne parvient pas à détecter les diagrammes occupant une page entière : YOLO est en effet conçu pour segmenter une page, ce qui devient problématique lorsqu'il rencontre des diagrammes l'occupant entièrement. Le problème vient en outre de la quantité de cas représentatifs : les sources contiennent peu de diagrammes en pleine page. Pour remédier à ce problème, une solution consiste à créer de fausses illustrations en pleine page en prenant les diagrammes déjà extraits et en étendant leurs marges, générant ainsi une *groundtruth* artificielle.

Ces données synthétiques peuvent en effet s'avérer cruciales pour l'entraînement des modèles, afin d'éviter le *sur-apprentissage*, car de manière générale, prévenir l'*overfitting* implique très souvent d'augmenter le volume de donnée ; volume qui fait défaut dans le cas des données historiques. C'est ce que nous détaillons ci-après.

5.2 Quelles données pour le spécialiser ?

Spécialiser un modèle d'intelligence artificielle implique de lui fournir des données pertinentes, diversifiées, et en quantité suffisante. Cependant, pour certains domaines, dont l'histoire fait partie, le volume de données disponible est insuffisant. Ce constat est d'autant plus vrai dans le cas des diagrammes issus de traités astronomiques : les corpus de documents scientifiques historiques contiennent généralement du texte en majeure

partie, des tables et des images, négligeant souvent les diagrammes.¹² De plus, ils sont dénués d'annotations précises sur les éléments constitutifs des pages ; c'est sans parler de l'inexistence d'un corpus de diagrammes dont les primitives sont annotées. Or l'annotation est une tâche chronophage et fastidieuse. Le recours aux données synthétique répond, mais en partie seulement, à ces problématiques.

5.2.1 *datasets* synthétiques

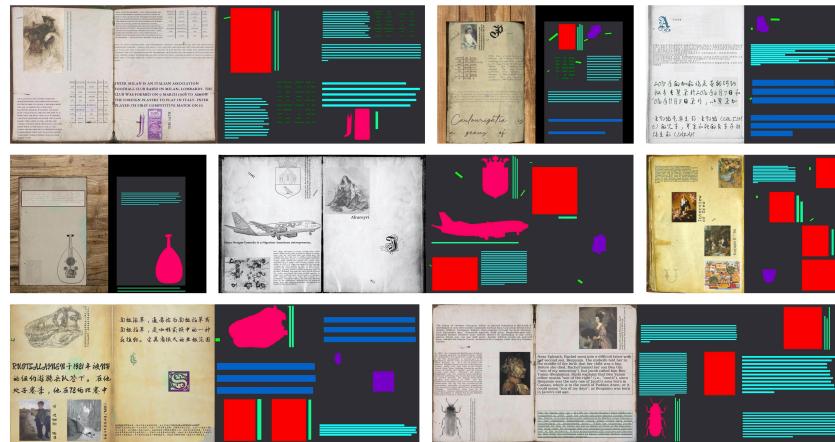
Les *datasets* synthétiques sont générés par des algorithmes ou des méthodes de simulation pour imiter des données réelles, sans être directement extraites de sources existantes. De tels jeux de données sont utilisés lorsque les données réelles sont limitées ou difficiles à obtenir, mais qu'il est cependant nécessaire de contrôler spécifiquement les caractéristiques des données d'entraînement¹³. La génération d'images a pour but de fabriquer des ensembles de données plus vastes, plus diversifiés, très variables et assez complexes, répondant aux caractéristiques des objets d'intérêt du projet, et surtout étiquetés automatiquement, sans recourir à l'annotation manuelle.

Ces données synthétiques sont assez ressemblantes et complexes pour être exploitées. Par exemple, docExtractor est un modèle off-the-shell (au même titre que YOLO) envisagé dans le cadre de la tâche d'extraction des diagrammes, et qui se veut spécifique aux données historiques, car il est entraîné sur des données produites par un générateur de documents historiques synthétiques : SynDoc¹⁴. SynDoc génère des images de manière aléatoire en combinant des éléments graphiques (fonds, images, texte et bruit) provenant d'un jeu d'image défini (constitué de 177 images de pages, 15 contextes, plus de 8000 œuvres d'art provenant de WikiArt, des lettrines générées à partir d'une lettre aléatoire avec 91 fonts possibles, et des dessins, schémas et textes tirés d'articles aléatoires sur Wikipedia, avec plus de 400 fonts). Les différents éléments s'agencent, intégrant sur le fond images, texte et bruit, offrant des combinaisons et des mises en pages assez complexes. Chaque élément de contenu est pré-annoté, éliminant ainsi le besoin d'annotations manuelles pour ces pages.

12. Exception faite du corpus S-VED (J. Büttner, J. Martinetz, H. El-Hajj, *et al.*, « CorDeep and the Sacrobosco Dataset... »), collection d'illustration très diverses contenant entre autre des diagrammes historiques. Mais les primitives ne sont pas annotées.

13. *Ibid.*

14. Tom Monnier et M. Aubry, « docExtractor : An off-the-shelf historical document element extraction », dans *2020 17th International Conference on Frontiers in Handwriting Recognition (ICFHR)*, arXiv :2012.08191 [cs], 2020, p. 91-96, DOI : 10.1109/ICFHR2020.2020.00027.

FIGURE 5.2 – Données synthétiques générées par SynDoc.¹⁵

Pour entraîner le modèle de vectorisation, il a de même été nécessaire d'utiliser des données synthétiques. Parce qu'annoter les primitives géométriques dans des images de diagrammes complexes est très chronophage, le modèle de vectorisation a été pré-formé sur des corpus artificiels générés dynamiquement. Le script de génération des données d'entraînement choisit aléatoirement un arrière-plan, y ajoute des mots, des nombres et des glyphes puis crée artificiellement un diagramme en insérant des segments, des cercles et des arcs. Le script est conçu pour que ces diagrammes aient une forte probabilité de présenter des formes très caractéristiques comme les cercles concentriques et tangents, les lignes parallèles et les arcs connectés, afin de simuler les structures typiques. Les primitives sont dessinées avec des variations aléatoires d'opacité, de largeur et de couleur. Les cercles peuvent être remplis ou vides. Enfin, du bruit est ajouté en appliquant un flou gaussien, et en supprimant de petites régions du diagramme pour imiter la dégradation des documents historiques. Les données d'entraînement ainsi générées présentent des configurations assez complexes.

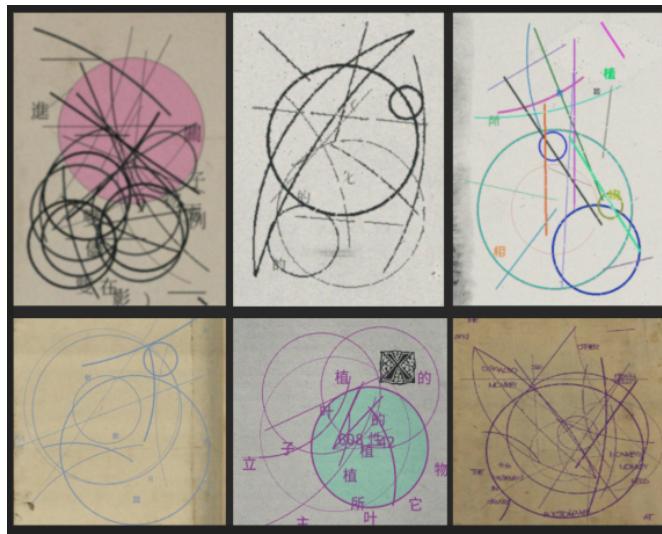


FIGURE 5.3 – Données synthétiques générées pour l’entraînement du modèle de vectorisation.¹⁷

Enfin, le modèle de similarité présente un troisième exemple, puisque SegSwap est pré-entraîné sur de la donnée synthétique. Le script de génération prend des parties aléatoires d’une images et les copie-colle au-dessus d’une autre image. Les trois images (source, cible et superposition) sont placées dans le même dataset d’entraînement, ainsi le modèle apprend à retrouver ce qui, dans la superposition, vient de la source, et ce qui vient de la cible.



FIGURE 5.4 – Données d’entraînement du modèle Segswap.

5.2.2 Les données réelles

S’appuyer sur les modèles *off-the-shelf*, sur de larges *datasets* généralistes, ou sur des données synthétiques permet une implémentation facilitée de la vision dans des projets et constitue une base solide. Toutefois, les sources tenant aux deux projets (VHS et EIDA) sont trop spécifiques pour se contenter de modèles généralistes ou formés sur des données artificielles. Même si ces derniers peuvent offrir des performances de base, ils risquent de manquer de précision et de sensibilité aux particularités des documents historiques. Les corpus artificiels présentent des configurations délibérément complexes pour s’approcher le plus possible des difficultés que le modèle pourrait rencontrer sur les données réelle.

Elles sont cependant irréalistes et insuffisantes pour permettre aux modèles de généraliser sur des diagrammes réels.

En atteste la comparaison des performances de docExtractor et YOLOv5 sur les données d'EIDA. docExtractor¹⁸, entraîné sur des données synthétiques mimant les documents historiques serait en théorie plus adapté au traitement d'images de pages de manuscrits, avec du texte et des illustration côté à côté, d'autant qu'il intègre des outils de traitement du texte (notamment pour la segmentation des lignes)¹⁹. Pourtant, sans fine-tuning sur des données réelles, il présente des performances équivalentes à celles de YOLOv5²⁰. Cela souligne que même les modèles off-the-shelf entraînés sur un corpus assez spécifique et complexe, mais synthétique, ne dispense pas d'un entraînement sur des données réelles, au même titre que les modèles très généralistes comme YOLOv5.

Alors, le modèle de base YOLOv5 tel que mis à disposition par Ultralytics est entraîné sur de grands ensembles de données réelles, ce qui constitue une base solide pour la classification des objets du monde. L'utilisation de SynDoc permet ensuite de compléter l'apprentissage initial en exposant le modèle à des exemples variés et spécifiques aux documents historiques, augmentant ainsi sa capacité de généralisation. Ces similis de manuscrits anciens offrent l'avantage de pouvoir être produits en grandes quantités et de couvrir un large éventail de scénarii et de configurations difficiles à obtenir dans des ensembles de données réelles. Puis le modèle est entraîné sur les données de VHS, qui sont de réelles pages de documents historiques contenant une large diversité d'illustrations. Ces données apporteront une dimension supplémentaire de pertinence au modèle, en l'exposant à des particularités des documents historiques réalistes. Enfin, YOLOv5 est entraîné sur les données d'EIDA, qui sont orientées spécifiquement vers les diagrammes, afin qu'il détecte uniquement ces derniers.

Quant au modèle de vectorisation développé par Syrine Kalleli²¹, il est formé sur des données synthétiques générées à la volée par un script. Mais le corpus de diagrammes d'EIDA est particulièrement caractéristique et le modèle n'aurait pu être optimal sans avoir appris sur des images de diagrammes issus de manuscrits réels. Un corpus d'entraînement de 303 diagrammes extraits de manuscrits et de gravures a donc été constitué et annoté par les historien.nes. Ces diagrammes sont issus de sources latines, arabes, grecques, hébreuses ou chinoises, datant du XII^e au XVIII^e siècle, et ils présentent en guise d'étiquettes plus de 3000 lignes, cercles et arcs. Le ré-entraînement a permis le transfert des connaissances acquises sur la tâche de détection des primitives sur les données réalistes.

Il sera également possible d'obtenir des meilleurs résultats sur la similarité grâce à

18. *Ibid.*

19. EIDA envisage l'implémentation d'un outil d'extraction et transcription des labels et des textes qui entourent les diagrammes

20. Id., *Le traitement des sources historiques par la vision artificielle...*, p.45.

21. S. Kalleli, S. Trigg, S. Albouy, *et al.*, *Historical Astronomical Diagrams Decomposition in Geometric Primitives...*

une évaluation des scores (qui constitue un jeu de données annotées) et le ré-entraînement du modèle, pour donner des résultats plus adaptés à la spécificité des données historiques.

D'ailleurs, cette étape d'annotation (le choix des exemples et des étiquettes) revêt des enjeux importants. L'apprentissage spécifique se fait à partir de données sélectionnées par les chercheur.ses : les exemples sur lequel l'algorithme d'apprentissage va itérer définissent le modèle. Il est nécessaire de constituer un échantillon de données aléatoire et représentatif, et de l'annoter en fonction de ce que l'on souhaite obtenir en prédiction.

L'annotation des jeux de données est non seulement une étape clé, mais aussi un bel exemple de collaboration chercheur.ses-ingénieur.es. Elle nécessite la définition de normes pertinentes et rigoureuses. Travail minutieux et chronophage, l'étiquetage des données peut engendrer des erreurs et du bruit dans les données, car elle implique la subjectivité des chercheur.ses et le regard parfois trop précis sur les sources desquels les annotateurs sont experts.

Voici un exemple rencontré lors de la préparation des données pour entraîner un modèle de segmentation du contenu textuel. Les sources arabes et chinoises sont particulièrement verbeuses et les diagrammes sont très souvent entourés des blocs de commentaires se mêlant alors aux légendes et aux labels. Doit-on considérer ces commentaires comme faisant partie des éléments que l'on souhaite identifier ou bien les ignorer ? Cette décision est importante car si on les ignore, le modèle risque de passer à côté d'éléments textuels pertinents. En revanche, si on les inclut, il ramènera des commentaires sans rapport direct avec le diagramme observé. On voit ici comment la binarité des modèles, qui se reflète dans les normes d'annotation, est problématique et constitue une limite au *machine learning*. Un compromis doit être trouvé entre l'automatisation, qui requiert une normalisation, des définitions claires et binaires, et la nuance dans l'interprétation des sources²².

La normalisation peut bénéficier à l'écosystème de recherche dans le domaine de l'HTR et de l'OCR. À ce titre, il est pertinent d'envisager l'utilisation du vocabulaire contrôlé SegmOnto pour l'annotation du contenu textuel entourant les diagrammes. Cela permettrait de créer des jeux de données réutilisables, à partager avec des projets poursuivant des objectifs similaires.²³ Encore une fois, un compromis doit être trouvé entre les besoins de description des chercheur.ses et les possibilités offertes par les vocabulaires contrôlés.

Un autre exemple concerne le dernier entraînement du modèle d'extraction : les résultats montrent que des diagrammes sont encore détectés en transparence. La question s'est alors posée de chercher à corriger ce défaut en donnant au modèle, à l'occasion d'un nouvel entraînement, d'avantage d'exemples négatifs (diagrammes visibles par transpa-

22. Dans le cadre du projet, il a toujours été plus intéressant d'opter pour une définition extensive des objets à détecter, car prévision d'une correction des traitements. Et il est plus facile de supprimer un élément pas pertinent que d'aller en rechercher un, surtout compte tenu de la taille des corpus des chercheur.ses. Vaut aussi pour la préparation des données pour l'entraînement du modèle d'extraction.

23. <https://segmonto.github.io/>

rence mais non annotés). Or il est préférable de se contenter de la correction ou suppression manuelle de ces prévisions erronées, garantissant que le modèle parvienne à détecter les diagrammes presque effacés.

Pour assurer la rigueur et la cohérence des annotations, les décisions prises entre les chercheur.ses et les ingénieur.es peuvent être l'objet d'une documentation ou d'ateliers d'annotation.

5.2.3 L'oeil de la machine : avantages et limites

Bien qu'il soit possible d'optimiser les performances d'un modèle d'apprentissage automatique en l'entraînant sur un ensemble de données spécifique, son interprétation des données reste limitée car fondamentalement binaire, ce qui le rend parfois déficient pour la recherche en histoire. Ainsi, il gèrera difficilement les cas limites et ambigüs. La décision d'inclure ou d'exclure ces cas particuliers de l'ensemble d'entraînement implique un arbitrage délicat. D'un côté, une inclusion trop restrictive peut compromettre les capacités de généralisation du modèle, c'est-à-dire sa capacité à s'adapter à de nouvelles données. À l'inverse, une inclusion trop permissive risque de dégrader la précision du modèle sur les cas plus typiques. Les chercheur.ses espérant obtenir un modèle maximaliste, quitte à accepter un certain degré d'erreur et de devoir supprimer les faux positifs, de nombreux cas limites ont été inclus. Le cas des diagrammes visibles en transparence (expliqué précédemment) en est un exemple éloquent.

Une autre difficulté réside dans la définition même du “diagramme astronomique”. Les limites de ce concept ne sont pas si claires et définitives pour les chercheur.ses, et pourtant le modèle a besoin d'une définition rigoureuse et cohérente. Il paraît en effet difficile de considérer les diagrammes astronomiques en dehors du contexte des pratiques d'autres sciences et disciplines connexes. Par exemple, Le *Flores Almagesti* – réécriture de l'Almageste datant du xv^e par l'astronome Giovanni Bianchini – présente une partie algébrique à l'ouverture mathématique, induisant la présence de nouveaux types de diagrammes d'inspiration euclidienne. Pour retracer la source de ces derniers, il est nécessaire de considérer les traités d'Euclide ou autres travaux d'algèbre. Ceux-ci ne sont pas des traités *astronomiques*, bien qu'il ne soit pas certain que ces distinctions contemporaines aient été aussi rigide à l'époque et aient eu un quelconque sens pour les acteurs historiques. Les sources byzantines confirment cette complexité : les diagrammes y sont nommés *katagraphai*, indépendamment du domaine scientifique auquel ils appartiennent. Également, de nombreux travaux astronomiques sont groupés dans des témoins qui contiennent des œuvres issus de domaines divers. C'est le cas avec les sources chinoises, comme le *Chongzhen lishu*, qui se présente généralement annexé d'une série de traités mathématiques. Par conséquent, les diagrammes euclidiens ont été gardés lors de la préparation des données, et l'algorithme de détection les classe comme “diagramme”, même s'ils ne constituent pas

l'objet principal des chercheur.ses.

En ce qui concerne les autres types de diagrammes non strictement astronomiques (géométriques, harmoniques, logiques, illustrations de constellations), une approche plus sélective a été adopté afin d'éviter un modèle trop maximalistes. Ces éléments, bien que potentiellement intéressants, n'ont pas été inclus dans la phase de détection automatique.

Ainsi l'œil de la *computer vision* constraint à des choix méthodologique potentiellement inconfortables, mais en même temps il peut aider à mesurer les impulsions des chercheur.ses, à mieux définir les objectifs de recherche et à prioriser les éléments les plus pertinents. Ainsi, la vision par ordinateur oblige les chercheur.ses à s'adapter à une logique algorithmique qui, tout en limitant certaines interprétations subjectives, offre l'opportunité de développer des modèles conceptuels et des méthodologies très rigoureuses.

L'optimisation d'un modèle revient à trouver un équilibre entre la disponibilité des données et leur réalisme. Trop peu de données, c'est un modèle qui sur-apprend – qui colle trop aux données – ou qui sous-apprend – qui est incapable de modéliser la relation entre l'entrée et la prédiction attendue. On ne peut donc pas se contenter des données réelles dont le volume fait défaut dans le cas des documents historiques. Néanmoins se limiter aux données trop générales ou synthétiques ne suffit pas non plus, au risque de construire un modèle trop généraliste. Le modèle doit se spécialiser sur un corpus et l'utilisation de données réelles est essentielle pour en capturer la complexité et la diversité.

L'annotation des jeux de données se trouve au cœur de cette tension : comment concilier la nécessité de standardiser les données pour assurer la cohérence et la réutilisabilité, tout en préservant la richesse et la complexité des données telles qu'elles sont perçues par les experts ? En effet, si la standardisation est indispensable pour former des modèles performant et combler le manque de données réelles, elle peut parfois aplatisir les nuances et les subtilités que seuls les chercheur.ses sont en mesure de saisir. Encore une fois, un équilibre doit être trouvé.

Chapitre 6

Un outils d'édition des diagrammes pour la cohésion des pratiques

L'astronome Martin Rees décrit le but de la science de la manière suivante :

“The aim of science is to unify disparate ideas, so we don't need to remember them all. I mean we don't need to recognize the fall of every apple, because Newton told us they all fall the same way.”¹

Cette remarque saisit une différence essentielle entre les sciences dures et les sciences humaines. Les chercheur.ses en sciences humaines s'intéressent aux pommes en raison de leur diversité. Ils veulent avoir une idée de ce qu'est une pomme générique, tout en observant les formes et nuances de chaque pomme. De même, les chercheur.ses d'EIDA veulent pouvoir produire une version unifiée des diagrammes présents dans les sources primaires, tout en gardant un accès aux diverses formes qu'ils ont pris à travers le temps et l'espace, comme on réussit à le faire pour le texte. Cependant le constat de l'absence de dispositif adéquat ou de recommandation pour l'édition des figures s'impose. Actuellement, chaque chercheur.se, projet ou éditeur utilise ses propres méthodes.

Ainsi, une des finalités du projet EIDA est de réfléchir aux normes d'une édition native numérique d'un corpus de diagrammes, qui autoriserait l'étude de leurs dimensions matérielles et historiographiques, ainsi qu'un outil numérique qui définirait des pratiques partagées dans la communauté scientifique.

6.1 Perspectives ouvertes par EIDA

Avant de mener une réflexion sur la définition des normes à adopter pour des éditions numériques des diagrammes, il est nécessaire de comprendre dans quel environnement technique, grâce à quels langages et quels formats elle pourra être implémentée. En effet,

1. Martin Rees, *Cosmic Origami and What We Don't Know*, en-US, nov. 2013, URL : <https://onbeing.org/programs/martin-rees-cosmic-origami-and-what-we-dont-know/> (visité le 14/07/2024).

dans la chaîne de traitement d’EIDA, si la vectorisation est une fin en soi, elle peut s’intégrer dans un projet plus ambitieux. La pipeline implémentant la vision artificielle extrait des formats *computable* (compréhensibles par les machines) constituant une base pour l’édition scientifique.

6.1.1 Transcription automatique des diagrammes

À l’heure actuelle, la transformation des diagrammes en objets informatiques repose souvent sur l’adaptation d’outils et de pratiques initialement conçus pour les textes.

Par exemple, l’Académie des Sciences de Berlin-Brandenburg mène depuis 1919 un projet monumental d’édition critique de l’œuvre de Gottfried Wilhelm Leibniz. Son pendant numérique, Leibniz-Online², utilise LATEX comme outil d’édition, tant pour les textes que pour les diagrammes. La méthode présente certaines limites. Premièrement, LATEX est un outil dédié au texte ; le tracé des diagrammes aura nécessité l’adaptation des packages dédiés aux schémas et la création de nouvelles commandes adaptées aux besoins du projet mais non standardisées, ce qui limite l’interopérabilité et interroge la pérennité de ses pratiques de transcription.

Ken Saito (Osaka Prefecture University), spécialiste en histoire des mathématiques grecques, met au point, au début des années 2010, un outil informatique nommé DRaFT³ (Fig. 6.1) pour l’édition des diagrammes présents dans les manuscrits mathématiques. Cet outil, permettant de poser des points, de les relier par des segments ou des courbes et d’ajouter des labels, génère des fichiers EPS (format vectoriel) intégrables dans des documents LATEX.

Reynaud relève, dans son travail de thèse⁴, les qualités de ce logiciel qui s’adapte relativement bien pour éditer les diagrammes imprimés ou incisés dans des tablettes mathématiques cunéiformes, et permet l’enregistrement de décisions éditoriales⁵.

Néanmoins, les formats TEX et EPS, bien adaptés à l’impression, présentent des limitations en termes de manipulation : les fichiers TEX sont figés une fois compilés, tandis que les fichiers EPS nécessitent des logiciels propriétaires pour être modifiés. De plus, la transcription manuelle des diagrammes demeure une étape chronophage. Enfin, les deux méthodes présentent des limitations pour rendre compte de toutes les variations infimes de la structure et de l’épaisseur des lignes présentes dans les diagrammes dessinés

2. *Leibniz Online - Leibniz-Edition Berlin*, URL : <https://leibniz-berlin.bbaw.de/de/leibniz-online> (visité le 10/07/2024).

3. *DRaFT download*, URL : https://www.greekmath.org/draft/draft_index.html (visité le 14/08/2024).

4. Adeline Reynaud, *Les diagrammes mathématiques paléo-babyloniens : catalogue, propriétés, matérielles, rôles dans les raisonnements*, These en préparation, Université Paris Cité, 2017, URL : <https://theses.fr/s197780> (visité le 26/08/2024), p.292.

5. “Pour chacune des lignes ainsi définies, il est ensuite possible de choisir selon son statut (présente sur le diagramme, reconstituée, bordure du document, etc) un style différent (plein, tirets, pointillés, etc).” *Ibid.*

à la main⁶.

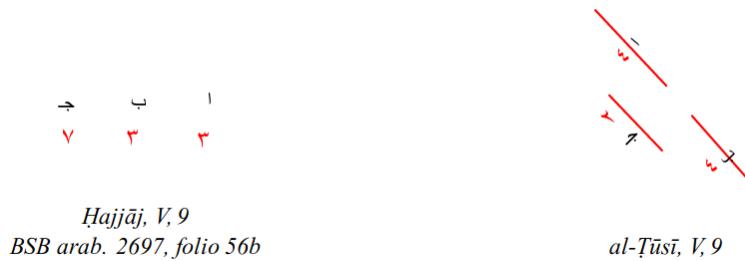


FIGURE 6.1 – Édition et comparaison de diagrammes des *Éléments* d’Euclide par Greg De Young⁸, utilisant DRaFT.



FIGURE 6.2 – Préparation de l’édition d’un diagramme mathématiques paléo-babylonien d’une tablette cunéiforme avec le logiciel DRaFT¹⁰.

De manière générale, l’absence de normes pour l’édition numérique des diagrammes laisse aux projets le choix des outils et méthodes de transcription, créant un paysage disparate sans pratiques uniformes. Actuellement, les projets emploient plutôt des outils complets mais complexes comme GeoGebra ou Inkscape – logiciels dont la maîtrise représente un défi pour certains chercheurs – et procèdent manuellement.

Pour pallier l’absence de solutions uniformes, l’IA se présente comme une alternative. Telle qu’implémentée dans EIDA, elle permet d’automatiser génération d’images SVG à partir d’un large corpus d’images, avec des corrections éventuelles qui restent beaucoup

6. “When a line has a very noticeable thickness, for example, should one choose points on the outer edge of the line, the inner edge, or somewhere in the middle? When the scribe produces a distinctly rounded vertex to a quadrilateral, should one choose the point of the vertex to be somewhere in the rounded portion of the line or should the editor choose a point outside the line where the two sides would intersect if continued rectilinearly?” Gregg De Young, « Diagrams in ancient Egyptian geometry : Survey and assessment », *Historia Mathematica*, 36–4 (nov. 2009), p. 321–373, DOI : 10.1016/j.hm.2009.02.004.

moins chronophages qu'un tracé manuel de chaque numérisation. Le SVG est en outre un format léger, interopérable et manipulable.

6.1.2 Idiomes et layers : des formats computables et des couches sémantiques

“One way to understand the structure of a diagram is to break it down into its component shapes.”¹¹

Dans le fichier SVG (obtenu en sortie du modèle de vectorisation) chaque objet géométrique est défini avec des éléments `<path/>` et `<circle/>`. La forme encodée dans la balise correspond à la forme dessinée dans le témoin et la décrit à l'aide de paramètres mathématiques. Avec ces deux formes fondamentales il est possible d'encoder et tracer une grande diversité de complexes géométriques. Par ailleurs le SVG, basé sur le langage XML, autorise un balisage des composants individuels en ensembles significatifs (grâce aux balises `<g/>` ou `<symbol/>` notamment)¹², lorsqu'il est nécessaire de traiter une série d'unités.

Tout comme XML est utilisé pour encoder de manière structurée des données textuelles, ce langage peut être mis au service de l'enrichissement sémantique de l'élément graphique. Les fichiers SVG peuvent ainsi embarquer des métadonnées. Les balises XML, en plus de la description du contenu géométrique, peuvent porter des attributs tels que des identifiants uniques (`id`), des propriétés visuelles (couleur : `color`, épaisseur de trait : `stroke-width`)¹³ – qui peuvent être extraits automatiquement à partir des images grâce à la *computer vision* – et d'autres attributs sémantiques. Ces données enrichissent considérablement la représentation numérique de l'objet et ouvrent la voie à des analyses dynamiques et à des visualisations interactives. En effet, comme le dit Vitali-Rosati, l'édition numérique ne vise plus seulement la production d'un objet ; elle doit assurer un balisage fin des informations qui permette ensuite une exploitation automatique.

“éditer” ne se réduit plus à mettre en forme pour l'impression et la diffusion sur papier : il s'agit aussi de structurer et de mettre en forme pour le format numérique qui a ses propres particularités. Il ne suffit pas d'essayer de mimer le papier en produisant des versions homothétiques, mais il est nécessaire d'essayer de comprendre comment l'information peut être agencée, structurée et diffusée dans les environnements numériques.”¹⁴

11. Christine Roughan, « Digital editions and diplomatic diagrams », dans *Proceedings of the First International Conference on Digital Access to Textual Cultural Heritage*, event-place : Madrid, Spain, New York, NY, USA, 2014 (DATECH '14), p. 77-82, DOI : 10.1145/2595188.2595189, p.79.

12. *Tutoriel SVG - SVG (Scalable Vector Graphics) / MDN...*

13. *Référence des attributs SVG - SVG (Scalable Vector Graphics) / MDN*, fr, juill. 2024, URL : <https://developer.mozilla.org/fr/docs/Web/SVG/Attribute> (visité le 03/08/2024).

14. Benoît Epron et Marcello Vitali-Rosati, *L'édition à l'ère numérique*, Paris, 2018 (Repères), URL : <https://papyrus.bib.umontreal.ca/xmlui/handle/1866/20642>, p.104.

Sur un mode *close reading*, il devient possible de développer des solutions pour isoler, manipuler et transformer les composants individuels du diagramme. Cela permet de réorganiser, coloriser ou animer les éléments. En tant qu'objets largement interopérables, les SVGs peuvent être intégrés dans des pages web, des documents numériques, et des applications, assurant une accessibilité universelle.

L'encodage en XML du format SVG permet une interprétation directe par la machine. Cette caractéristique ouvre la voie à des analyses de grande échelle, dites de *distant reading*, qui exploitent les capacités de traitement de corpus de données massifs. En effet, la structure arborescente et les attributs sémantiques des fichiers SVG servent d'appui à des algorithmes de fouille de données et de reconnaissance de similarités. Ces algorithmes peuvent identifier des motifs récurrents, des structures similaires ; de fait il devient possible de mettre à jour des relations entre plusieurs diagrammes.

“Digital formats, when encoded meaningfully, can be well-suited for quantitative analyses.”¹⁵

Le contenu géométrique n'est toutefois pas le seul idiome à représenter dans l'optique d'une édition, et actuellement des discussions sont en cours concernant l'extraction automatique des labels. Une des propositions est d'utiliser un format JSON, dans un premier temps, contenant les coordonnées des textes et la direction de lecture. Si les labels sont transcrits, le contenu sera certainement¹⁶ encodé en XML-ALTO, format qui condense la disposition et la transcription du contenu textuel. Pour enrichir automatiquement les SVGs avec du texte, des scripts sont capables d'extraire les coordonnées et le contenu textuel dans des éléments SVG.

Enfin, l'implémentation de bi-classifications automatiques pourraient ajouter de l'information globale sur la figure : le caractère poly/monochromatique et la présence/absence de graduations.

Sur l'extraction de ces idiomes sémantique repose la perspective d'une recherche de similarité fondée sur des critères explicites : un contenu géométrique (type de primitives, relations spatiales, position), une couleur (en se basant sur la valeur des attributs), ou encore le contenu textuel (transcription et position). La pipeline d'EIDA vise à décomposer l'image en une série de représentations qui constituent autant de calque exploitables et dans lesquels des scripts peuvent rechercher des similarités. En particulier grâce au SVG, l'information contenue dans l'image devient éditabile, manipulable, enrichie. Grâce à ce système de calques, les images acquièrent une richesse sémantique, facilitant des utilisations dans des contextes de fouille, et surtout d'édition scientifique (Fig.6.3).

15. C. Roughan, « Digital editions and diplomatic diagrams »..., p.82.

16. La tâche de transcription du texte entourant les diagrammes est un défis important en raison de la grande diversité des alphabets, des langues et des symboles utilisés, ainsi que de la segmentation spécifique des blocs de texte. À l'été 2024, les chercheurs sont encore à la phase de préparation des données pour entraîner l'algorithme de segmentation. Le format de sortie reste une supposition.

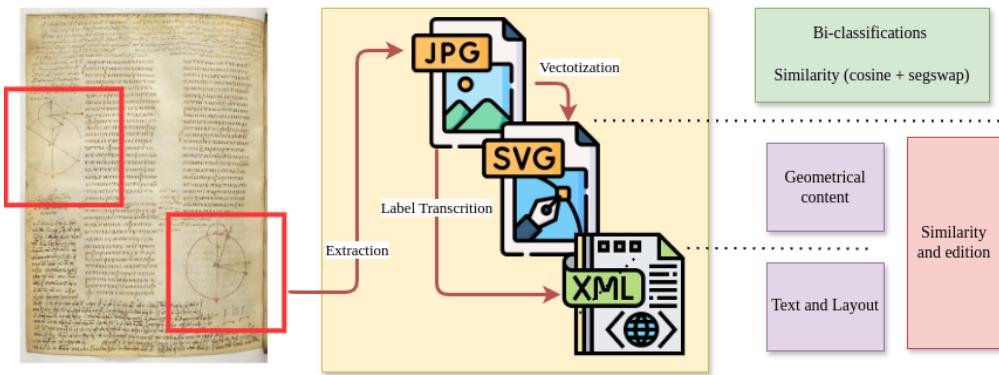


FIGURE 6.3 – Extraction de représentations sémantiquement riches et exploitation.

6.1.3 Une plateforme qui porte une méthode

L'intégration de l'IA dans AIKON ne se limite pas à l'automatisation des tâches : elle doit s'inscrire dans un environnement de recherche numérique complet, dédié à l'analyse approfondie de diagrammes astronomiques. Le constat de l'absence d'outil d'édition dédié à ces objets justifie l'intégration d'un processus d'édition numérique au sein d'AIKON. La plateforme et son interface utilisateur ont en effet l'ambition de favoriser la cohérence des méthodes et leur diffusion au sein de la communauté scientifique, contribuant ainsi à l'établissement de pratiques communes pour l'étude des diagrammes astronomiques.

Le modèle de données d'AIKON pourra évoluer pour intégrer les concepts de *Transcription* et *EditedDiagram*, s'inspirant de la distinction Témoin/Édition. La *Transcription* représentera la version informatique d'un diagramme source, reliée à son document d'origine, tandis que l'*EditedDiagram* correspondra à une nouvelle édition construite à partir d'une ou plusieurs transcriptions, et reliée à un chercheur.

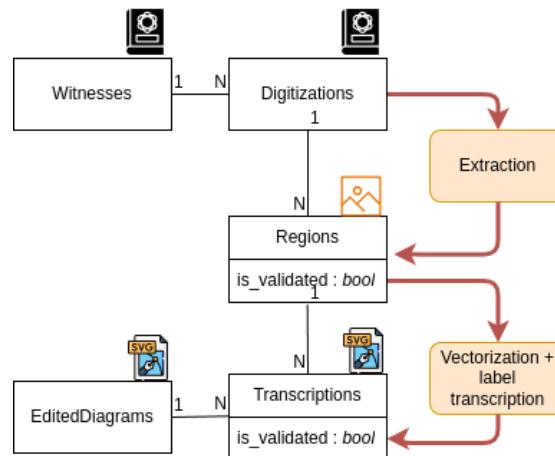


FIGURE 6.4 – Modèle de données hypothétique.

On voit sur la figure 6.4 qu'une édition se fond sur une transcription automatique

générée par les modèles de vision. Cette approche, tout en offrant des avantages en termes d'efficacité, nécessite une intervention humaine pour corriger les erreurs potentielles des algorithmes, engendrant ainsi l'existence de multiples versions d'une même transcription au sein de la plateforme.

La question est importante, car les pistes de correction envisageables divergent selon l'approche adoptée. Une étude fondée sur le *distant reading* d'un corpus volumineux et axée sur la recherche de similitudes ne requiert pas le même niveau d'intervention qu'une analyse fine d'un diagramme dans le cadre d'une édition critique. L'ontologie CIDOC-CRM permet, avec sa notion d'*événement*, de représenter de modéliser des changements, et de les relier aux objets, aux personnes et éventuellement à un espace-temps associé. En ce sens, il soutiendra certainement les chercheur.ses et les équipes d'ingénieur.es lorsque se posera la question de modéliser des versions des transcriptions.

En conclusion, la vectorisation occupe une place centrale dans le projet. Convertissant les images en représentations numériques hautement structurées, elle permet à la machine de saisir la sémantique de leur contenu. Le format SVG autorise en outre la manipulation : la correction ou l'enrichissement par l'ajout de métadonnées et d'attributs sémantiques dans les balises. Ces propriétés en font le format de base sur lequel faire reposer une interface d'édition.

Si la vectorisation offre un socle technique solide pour l'édition numérique, elle ne résout pas pour autant les questions fondamentales liées à l'interprétation des images, qui relèvent d'une expertise disciplinaire spécifique.

6.2 Une problématique d'unification des pratiques

L'industrie éditoriale destinée au milieu académique détermine depuis le XIX^e siècle les standards desquelles héritent l'édition contemporaine¹⁷. Or, dans les traités historiques d'astronomie, il n'existe pas de norme pour la reproduction des diagrammes. Face à ce constat, il est nécessaire de mener une réflexion sur les enjeux épistémologiques et historiographiques du partage des pratiques d'édition des diagrammes, et sur les besoins de la mise en œuvre d'une édition numérique des diagrammes.

6.2.1 Sortir des pratiques aléatoires

Adolphe Rome, dans l'introduction du premier volume de son édition des commentaires de Pappus et Théon sur l'Almageste de Ptolémée, a remarqué à propos des figures qu'"à leur manière, elles constituent également un texte à éditer"¹⁸. En tant qu'objet

17. B. Epron et M. Vitali-Rosati, *L'édition à l'ère numérique...*

18. Commentaires de Pappus et de Théon d'Alexandre sur l'Almageste, ed. by A. Rome (Rome, 1931), i, p. xxiv. Boris Jardine et Nicholas Jardine, « Critical Editing of Early-Modern Astronomical Diagrams », *Journal for the History of Astronomy*, 41-3 (août 2010), Publisher : SAGE Publications Ltd,

d'étude à part entière, porteur d'indices important, ils doivent être considérés au même titre que le texte comme des témoins des pratiques des sciences astronomiques et de la transmission des connaissances

Or il n'existe ni conventions, ni normes pour une édition des traités astronomiques qui retracent efficacement les métamorphoses des images comme du texte de manière équivalente ; ce qui est surprenant car après tout, les éditeurs sont confrontés, avec les diagrammes, aux même problématiques que celles concernant les textes, à savoir des erreurs et variations. B. Jardine et N. Jardine parlent même de “remarkable failure to take seriously the editing of diagrams”¹⁹, et ce même dans les éditions numériques.

“Text can be handled with TEI-compliant XML; a similar standard does not yet exist for diagrams.”²⁰

Les éditeurs font alors des choix dans la manière de représenter les diagrammes. Notamment, dans les manuscrits, il arrive régulièrement que les figures rentrent en contradiction avec le texte qui les accompagne. Jardine et Jardine remarquent alors que les éditions canoniques de ces grands traités d'astronomie corrigent les diagrammes afin de les rendre conformes au texte. Mais les modifications sont apportées sans commentaire, une pratique inenvisageable en ce qui concerne le texte.²¹

En outre, les éditeurs ont tendance à moderniser les conventions visuelles pour adapter les diagrammes à leurs lecteurs contemporains. En voici un exemple typique :

“In Heiberg’s edition [of Archimedis *Archimedis Opera Omnia*], medieval methods of representing diagrams are quietly abandoned in favor of more modern methods. Heiberg is not alone in silently modernizing the diagrams in his editions : many editors of mathematical texts diverge from the manuscripts in their presentation of diagrams. In some cases they “correct” shapes ; in other they add more generality to diagram figures which in the manuscripts were overspecified. (For instance a manuscript diagram might have a right triangle but the text state that any triangle would do – often modern editors choose to draw an irregular triangle instead so that their readers are not misled into thinking that the proof requires a right triangle.)”²²

La pratique de la correction ou adaptation dénote une perception très contemporaine du diagramme comme simple accompagnement des textes scientifiques auxquels ils sont soumis. Les illustrations sont considérées comme outil de compréhension du contenu textuel destiné au lecteur. Or les diagrammes médiévaux et du début de l'époque moderne

p. 393-414, DOI : 10.1177/002182861004100307, p.394.

19. *Ibid.*, p.393.

20. C. Roughan, « Digital editions and diplomatic diagrams »..., p.77.

21. B. Jardine et N. Jardine donnent en exemple des cas où la casse des labels est modifiée pour s'aligner avec le texte : *Kepleri opera omnia*, ed. by Ch. Frisch (Frankfurt and Erlangen, 1858 ; facs. Hildesheim, 1971), i, 282, Fig. 9 N. Jardine's, in *The birth of history and philosophy of science : Kepler's A defense of Tycho against Ursus with essays on its provenance and significance* (Cambridge, 1984 ; rev. edn, 1988). V. Bialas's, *Johannes Kepler : Gesammelte Werke*, xx/1 (Munich, 1988). (B. Jardine et N. Jardine, « Critical Editing of Early-Modern Astronomical Diagrams »..., p.411)

22. C. Roughan, « Digital editions and diplomatic diagrams »..., p.78.

ne se contentent pas d'illustrer le texte. Il est probable (c'est la position des chercheurs du projet EIDA) qu'une grande variété de pratiques et de manières d'utiliser les diagrammes aient coexisté dans l'astronomie ancienne, tissant des relations complexes aux textes et aux données tabulaires. Par conséquent, imposer un principe d'édition unique, qui arrête un emploi des diagrammes, revient à contraindre artificiellement les sources, en occultant la diversité des représentations. Il est nécessaire de trouver, grâce au numérique, un environnement éditoriale plus plastique qui permette de représenter différentes modalités d'utilisations des diagrammes et de leur rapport aux autres éléments des sources (textes et tables).

Dans l'édition papier, les tentatives d'édition pour les diagrammes reposent sur leur mise en texte. Les variations sont codifiées et expliquées²³. Les corrections, reconstructions, avec modernisation des conventions font également l'objet d'un accompagnement textuel²⁴. Selon B. Jardine et N. Jardine, bien que plus satisfaisantes que la modification sans signalement, ces pratiques, trop dépendantes des présupposés de l'éditeur et des objectifs éditoriaux (l'établissement du *stemma*, l'attribution de la production ou encore une étude de la réception et la transmission), méritent d'être questionnées.

Une première problématique concerne la mise en page et le placement du diagramme. L'adoption de la convention consistant à centrer l'image entre deux blocs de texte masque la manière dont l'emplacement peut s'avérer signifiant, les diagrammes centrés servant souvent de bases aux démonstrations textuelles, tandis que les diagrammes marginaux constituent généralement des gloses ou des commentaires sur le texte. La question de l'emplacement ne devrait pas être complètement éludée, mais comment faire cela avec l'imprimé sans sacrifier la lisibilité²⁵ ?

Un second problème concerne l'observation des conventions graphiques. Leur compréhension n'est évidente que dans un contexte et une culture donnée (si on identifie immédiatement un code barre ou un QR Code, un alien qui arrive sur Terre ou un scribe du Moyen Âge pourrait croire à une simple tâche). Sans une étude des normes utilisées par les acteurs historiques, le jugement des variations et des erreurs demeure fondamentalement ambiguë : il est difficile de distinguer ce qui est le produit d'une convention et donc un standard, et ce qui relève de l'erreur, ou de l'innovation ; il est impossible de trancher avec certitude. Or ces variations, selon l'objectif de l'édition, tendent à être éludées, les éditeurs les considérant comme des erreurs minimes qui ne nécessitent pas un signalement.²⁶

23. Greg De Young, « Editing a collection of diagrams ascribed to al-Hajjāj : an initial case study »..., Voir les édition de.

24. B. Jardine et N. Jardine, « Critical Editing of Early-Modern Astronomical Diagrams »..., p.395.

25. *Ibid.*, p.400.

26. “The basic task of editing diagrams centers around things that go wrong with the reproduction of diagram lines and points as diagrams are copied from one manuscript to another. Just as when editing a verbal text, the principles of editing include the recording as clearly as possible of all significant variations within the tradition and indicating transparently any editorial interference in the text.” Greg De Young,

Une étude exhaustive de l'ensemble du matériau visuel est indispensable au risque de rentrer dans un cercle vicieux herméneutique empêchant toute étude de cette riche grammaire visuelle, mais elle est complexe à mettre en œuvre, notamment dans un format *print*²⁷. C'est pourquoi B. Jardine et N. Jardine plaident pour une approche qui met en évidence les conventions historiques (la disposition des diagrammes par rapport au texte, les utilisations de la perspective, de la couleur etc.) des diagrammes plutôt que de les uniformiser.

Pour résumer, les méthodes d'édition des diagrammes contenus dans les traités éludent des informations importantes. Les éditeurs sont confrontés à un dilemme : concilier la production d'une version unifiée et lisible tout en préservant la richesse et la diversité des informations contenues dans les sources originales. Les supports imprimés présentant des limites face à cet objectif, le recours à l'édition numérique est une voie à explorer. Les chercheurs d'EIDA s'y penchent pour élaborer une méthode d'édition capable de rendre compte de la complexité des processus de création, de transmission et de réception des diagrammes.

6.2.2 L'édition numérique : une solution ?

“We predict that this brave new world of digital editing will eventually produce a much better understanding and appreciation of the many lively roles of diagrams in early-modern astronomy.”²⁸

Quelles sont les avantages et les apports d'une édition numérique ?

Un format enrichi

Les langages de balisage comme le XML (utilisé dans le format SVG) permettent d'expliciter le sens et la valeur de certaines parties des diagrammes, ce qui s'avère très utile d'un point de vue scientifique, car les données deviennent exploitables à un niveau de détail très fin.

Le format SVG permet de verrouiller ensemble des complexes géométriques et/ou des labels, d'effectuer des recherches de similarité dynamique, ou encore d'insérer des identifiants. Ces derniers rendent alors possible des formes de citation permettant de documenter et justifier des choix éditoriaux.

Le projet *Archimedes* (College of the Holy Cross), par exemple, présente une méthode d'édition numérique et diplomatique des diagrammes mathématiques de tradition

« Editing a collection of diagrams ascribed to al-Hajjāj : an initial case study »..., p.229. La notion de “significant variation” est problématique, car les chercheurs n'accordent pas tous la même importance à tous les types de variation pouvant affecter le diagramme.

27. B. Jardine et N. Jardine, « Critical Editing of Early-Modern Astronomical Diagrams »..., p.398.

28. *Ibid.*, p.411.

grecque présents dans les sources médiévales, en utilisant le *framework* CITE (*Collections, Indices, Texts, and Extensions*) comme outil de citation. Il y est utilisé pour lier une édition diplomatique d'un diagramme à la preuve de sa transcription, à la proposition mathématique associée, et pour faire tenir ensemble toute l'édition numérique du traité²⁹.

L'architecture CITE³⁰ a été initialement développée pour le projet *Homer Multitext*³¹. Elle fournit une méthode pour l'identification, la récupération, la manipulation et l'intégration de données d'une bibliothèque numérique via des URN (*Uniform Resource Name*s), notamment via des mécanismes pour relier les différents objets. CITE offre un cadre, une méthode pour la création et la gestion d'éditions numériques complexes, en particulier pour faire tenir ensemble des éléments hétérogènes (comme les images et leurs transcriptions), associant le tout dans un index qui organise la donnée en triplets.

Initialement pensé pour le texte, l'utilisation du standard de citation CITE dans le cadre du projet *Archimedes* montre la manière dont il peut être appliqué à l'édition numérique des diagrammes, combinant texte et images et permettant l'identification et le lien entre les éléments à un niveau de granularité très fin.

Le standard de citation CITE, traditionnellement utilisé dans l'analyse textuelle, a fait l'objet d'une adaptation pour répondre aux besoins de l'édition diplomatique numérique des diagrammes dans le cadre du projet *Archimedes*. Il permet d'atteindre un niveau de granularité très fin dans l'identification et l'interconnexion des éléments constitutifs des documents, qu'ils soient textuels ou visuels. À titre d'exemple, chaque objet géométrique dont il est question dans les propositions mathématiques du texte sont associés à deux ensembles de labels distincts : ceux utilisés dans le texte pour le désigner et ceux qui le marquent directement sur la figure, les deux pouvant différer. L'objet est au premier ensemble via son identifiant (attribut `id`) assigné à partir des labels qui lui réfèrent dans le texte (le cercle ABC du texte a un identifiant `circleABC`). Ils sont reliés au second ensemble de labels via des triplets rassemblés dans l'index et selon la syntaxe : `circleABC hasLabel A`, par exemple. Cependant, si le projet *Archimedes* a besoin de mettre du liant entre les entités numériques, c'est au titre d'une édition complète des traités, comprenant textes et illustrations, ce qui n'est pas le projet de la plateforme d'édition d'EIDA, relativisant la pertinence de l'utilisation du cadre CITE dans le cadre d'une édition uniquement des illustrations³².

29. C. Roughan, « Digital editions and diplomatic diagrams »...

30. Christopher W. Blackwell et Neel Smith, « The CITE Architecture : a Conceptual and Practical Overview », dans *The CITE Architecture : a Conceptual and Practical Overview*, 2019, p. 73-94, DOI : 10.1515/9783110599572-006.

31. *The Homer Multitext project*, URL : <https://www.homermultitext.org/> (visité le 09/07/2024).

32. C. Roughan, « Digital editions and diplomatic diagrams »..., p.81.

Interactivité

Une édition numérique autoriserait des formes d'interactivité impossibles à obtenir dans le *print* :

Déjà, des scripts de reconnaissance des formes et des labels, basés sur la lecture du format SVG, permettraient d'identifier automatiquement la présence ou absence de certains éléments graphiques dans les sources. La modulation des critères de recherche, en incluant ou excluant certains calques (par exemple, en considérant ou non la couleur), permet d'affiner les résultats, prenant en compte toute la complexité des figures. De telles recherches effectuées dans le corpus de sources pour retrouver des configurations similaires offriraient la possibilité d'opérer un tri dynamique dans le corpus de sources, avec pour ambition d'automatiser la génération d'un appareil critique.

La possibilité d'avoir un accès direct à l'image source – sans altération de la lisibilité de l'image éditée – renforce la rigueur scientifique de l'édition. Cette approche permet d'éviter les biais interprétatifs liés à une description textuelle et offre aux chercheur.ses une vision synoptique des variantes au sein d'un corpus. De plus, la capacité à contextualiser chaque diagramme en le reliant à sa page de manuscrit d'origine facilite l'analyse des facteurs matériels liés à sa réception et sa transmission.

Enfin, les visualisations d'une édition pourraient inclure un environnement de type *sandbox* pour des comparaisons et superpositions interactives. Les utilisateur.rices pourraient comparer les images originales (les témoins) et éditées, et accéder à la page source pour une vérification détaillée des décisions éditoriales et corrections. Cette multiplicité de perspectives faciliterait l'analyse approfondie des diagrammes, l'édition numérique pourrait ainsi résolument appuyer les analyses historiographiques et les processus d'interprétation des chercheur.ses, ce que souligne Roughan :

“When mathematical documents are digitized, diagrams should receive the same attention as text does : both are vehicles of the mathematical argument.”³³

6.2.3 Exigences et dissensus

À l'heure de la rédaction de ce mémoire, les modalités des divers types d'éditions suscitent encore de nombreuses interrogations parmi les chercheur.ses d'EIDA. En effet, si le texte hérite d'une tradition éditoriale de près de deux siècles et des apports de l'écriture et de la philologie, les normes restent à inventer pour le matériau graphique, ce qui soulève inévitablement des questions épistémologiques et heuristiques. Le système inventé pour le texte est-il transférable ? C'est toute la question : certains chercheurs veulent s'inscrire dans les traditions philologiques textuelles, d'autres entendent s'en écarter pour mieux s'adapter à l'objet image.

33. *Ibid.*, p.78.

DISHAS, dont les objets d'étude sont les tables astronomiques, arrête trois types d'édition : l'édition “diplomatique” d'une table originale (type A), l'édition “critique” basée sur plusieurs sources (type B), et l'édition “recalculée” qui s'appuie sur un modèle mathématique reconstruit par l'éditeur et pas nécessairement lié à une quelconque source. Déjà, DISHAS optait pour des dénominations délibérément neutres (types A, B, C) pour les types d'édition des tables, afin de pouvoir s'éloigner des catégories canoniques et héritées de la tradition textuelle. Un alignement avec les trois types d'édition présents dans DISHAS constitue un angle d'approche mais ce système n'est pas complètement transférable non plus.

Type A ou édition diplomatique ?

Dans DISHAS, les éditions de type A s'appuient sur un seul témoin et se veulent au plus proche du texte. Il n'est cependant pas possible de parler d'édition diplomatique dans la mesure où aucune donnée de mise en page n'est reproduite. Elles constituent ensuite la base des éditions “critiques” qui s'appuient sur plusieurs témoins.

“Plusieurs éditions peuvent être réalisées d'un même original item, et une édition peut elle-même s'appuyer sur plusieurs éditions préalables. Le travail de transcription d'une table étant sujet à interprétation, il est ainsi considéré dans DISHAS qu'une édition critique de plusieurs témoins constitue en réalité l'édition de plusieurs éditions. Une simple transcription constitue déjà en ce sens, une édition classique (éditions de type A [...]).”³⁴

Dans le cadre du projet EIDA, la catégorie correspondante s'appuierait sur une vecto-risation automatique corrigée. Cela soulève la question de l'intervention des chercheur.ses : quand passe-t-elle de la correction à l'édition ? Les chercheur.ses doivent définir ce qui constitue pour eux une “édition type A” et ce qu'elle apporte de plus par rapport à la simple correction d'une transcription par des modèles d'IA.

Le Projet *Archimedes*, dans sa vision de l'édition diplomatique des diagrammes, choisit de maintenir un parallèle avec le texte, comme l'explique Roughan :

“For a comparison : digitally, diplomatically transcribing the text of a manuscript accomplishes certain goals. Such a transcription makes the text in the manuscript more accessible to those unfamiliar with the manuscript's paleography and more accessible when it is faded or obscured. It also makes the text in the manuscript machine-actionnable and allows for annotations to mark certain features (personal names, unclear text, expanded abbreviations, etc.).”³⁵

“A digital, diplomatic transcription therefore should make the diagram in the manuscript more accessible, make the information contained in the diagram machine-actionable, and allow for editorial annotations.”³⁶

34. S. Albouy, *Médiation des données de la recherche...*, p.24.

35. C. Roughan, « Digital editions and diplomatic diagrams »..., p.78.

36. *Ibid.*

Les éditions diplomatiques du projet *Archimedes* apportent une valeur ajoutée par rapport à une simple image raster, en améliorant la lisibilité et en réduisant le bruit de la numérisation. Enrichies par rapport à une simple transcription vectorielle, elles intègrent des identificateurs uniques normalisés pour chaque élément géométrique. Ces identifiants combinent le terme contemporain désignant la forme avec les labels issus de la proposition mathématique associée.

Il y a donc un enjeu, dans le cadre d'EIDA, à penser des protocoles statuant sur des états (au sens presque chimique du terme) d'une transcription, et à articuler ces états avec les protocoles de versionning. Les chercheur.ses devront alors décider sur quelle version baser leurs éditions critiques.

Sur quels objets baser une édition corrigée ?

Dans DISHAS, pour réaliser une édition de type C, le.la chercheur.se récupère les paramètres d'une table témoin et recalcule l'ensemble de ses valeurs à l'aide de modèles mathématiques contemporains. Ce type d'édition permet, entre autres, de comparer les techniques arithmétiques anciennes et modernes, en observant par exemple les variations de certaines constantes mathématiques. Dans le modèle de données, les tables corrigées renvoient aux scénarios de calcul utilisés.

Une édition corrigée d'un diagramme repose sur la proposition mathématique qu'il illustre et correspond à ce qui est textuellement décrit. Une telle édition permet d'observer les déviations par rapport à la démonstration (témoignant des modalités de réception) ou bien les phénomènes d'*overspecification*, c'est-à-dire la tendance à représenter des formes très spécifiques comme des triangles rectangles ou équilatéraux, alors que la preuve textuelle ne les mentionne pas (suggérant plutôt des représentations quelconques). Cependant, dans EIDA, ces éditions ne s'appuient sur aucun objet informatique, car les textes des manuscrits ne sont pas intégrés à la plateforme, ce qui est scientifiquement insatisfaisant et fait douter de la possibilité d'intégrer ce type d'édition dans la plateforme.

Peut-on vraiment parler d'édition critique ?

Une édition critique s'appuie sur un ensemble de témoins dont les vectorisations ont été corrigées et figure donc un appareil critique. Le défi consiste à imaginer un outil d'édition scientifique des diagrammes pensé pour donner un accès aux variantes pour chaque complexe de primitives et/ou label verrouillé sous l'œil expert du.de la chercheur.se. Les débats portent sur les modalités d'accès aux variations.

Certains chercheur.ses du projet défendent une codification des types de variantes directement sur le diagramme³⁷. Ce système, conçu initialement pour le texte et l'imprimé,

37. “Electronic tools such as DRaFT allow us to indicate editorial actions by altering the weight of solid lines or by using various forms of dashed lines.” Greg De Young, « Editing a collection of diagrams ascribed to al-Hājjāj : an initial case study »..., p.229.

repose sur une séquentialité et se transfère mal à l’immédiateté des images. En effet, l’image présente une densité d’information qui se manifeste en une seule fois, contrairement au texte qui se déploie progressivement. Par ailleurs, de nombreux symboles graphiques et expressions visuelles sont déjà utilisées dans les sources par les acteurs historiques. Ces deux problématiques (la densité d’information et la difficile dissociation des annotations éditoriales des éléments du diagramme) rendent le résultat rapidement chargé et illisible. De plus, l’absence d’un langage symbolique partagé au sein de la communauté scientifique limite les possibilités de collaboration autour de cette méthode, pouvant susciter des problèmes d’interprétation³⁸. Par ailleurs, est-il même possible de répertorier et symboliser de manière exhaustive tous les types de variation possibles ? Cette méthode a néanmoins l’avantage d’être transposable facilement du numérique à l’imprimé.

La proposition défendue par d’autres chercheur.ses est celle d’une représentation résolument numérique (quitte à ce qu’elle le soit uniquement). La visualisation de l’édition devrait alors fournir un appareil critique interactif, uniquement visuelle, généré automatiquement sur la base de l’analyse systématique des fichiers SVG. Ainsi l’édition se détacheraît de toute description textuelle comme du langage visuel des diagrammes. L’idée initiale consisterait à générer un appareil critique qui ne montrerait pas les variantes mais les invariantes, via une recherche dynamique de complexes similaires dans le corpus de sources, avec système de calques (couleur, épaisseur, etc.) à prendre en compte ou ignorer.

Cependant, dans la tradition philologique textuelle, un certain nombre de notions ont été définies et fondent la légitimité d’une édition critique, spécifiquement les concepts d’appareil critique et de collation.

“The critical apparatus of readings may be positive, that is mention all the witnesses in each note, or negative, that is mention only the witnesses disagreeing with the chosen reading.”³⁹

Soit la phrase “Tom sort pour acheter une pizza”. Les manuscrits A, B, C, D ont “pizza”, tandis que les manuscrits E, F, G ont “piazza”. “Pizza” est la bonne lecture. Un appareil positif notera : “pizza A B C D : piazza E F G” tandis qu’un appareil négatif notera : “pizza] piazza E F G”. La différence tient essentiellement dans la manière de séparer le “bon mot” de “la variante”. Finalement, tant l’appareil positif que négatif nécessitent de tabler sur ce qui est juste et ce qui est faux.

La collation désigne l’acte d’enregistrer les différences entre les textes préservés dans les manuscrits. Ces différences peuvent concerner : des fautes d’orthographe, des inversions de mots ou de syntagmes, des suppressions, des mots illisibles, des corrections faites par le scribe du texte principal, des corrections faites par d’autres, etc. Le mot collation vient du latin *collatio* qui signifie ‘rassemblement, assemblage’. En ce sens, le terme peut donc être

38. “Since there is as yet no strong consensus on how to indicate various forms of editor interference within diagrams, there is the potential for confusing or contradictory indicators to develop.” *Ibid.*

39. Sébastien Moureau, « The Apparatus Criticus » (, 2015), URL : <https://dial.uclouvain.be/pr/boreal/object/boreal:155652> (visité le 14/08/2024), p.348.

employé pour désigner le rassemblement de tous les diagrammes témoins qui présentent une configuration similaire, grâce à une recherche de similarité dans le corpus de fichiers SVGs sur lequel l'édition est basée.

Mais pour qu'une information figure dans un appareat critique au sens philologique du terme, c'est-à-dire pour qu'elle soit pertinente pour l'intérêt d'une édition, elle doit présenter une variation par rapport à un cadre invariant. Dans des termes simples, si tous les manuscrits avaient "pizza", l'appareat critique serait vide car il n'y aurait rien de remarquable à ce sujet. Si les manuscrits présentaient deux autres graphies, et une suppression, les trois variantes seraient répertoriées, distinguées et associées ensemble aux témoins qui les présentent. C'est pourquoi "appareat critique" ne serait pas un bon terme dans le langage des éditeurs pour désigner les résultats d'une simple recherche de similarité dans les fichiers SVGs⁴⁰.

La mise en œuvre d'une collation automatique visant à produire un appareat critique positif d'un diagramme nécessite une double segmentation : une segmentation en classes de similarité (invariantes vs. variantes) et une classification fine des variantes selon le type de transformation visuelle. Cette classification pose le défi d'identifier, parmi la multitude de variations possibles, celles qui sont à la fois détectables par les algorithmes et pertinentes pour l'analyse.

Cela démontre que l'élaboration d'une typologie des éditions de diagramme ne peut se faire par simple analogie avec les textes. Les classifications existantes, notamment celle proposée par DISHAS, fournissent un cadre de référence utile. Cependant, la spécificité des opérations éditoriales sur les images nécessite une adaptation et une extension de ces classifications. Ces enjeux terminologiques soulignent le besoin d'une réflexion collective pour parvenir à une standardisation des termes et des concepts.

En conclusion, afin de sortir des pratiques aléatoires et de garantir un cadre cohérent et partagé des pratiques éditoriales, une unification des méthodes d'édition de diagrammes astronomiques s'impose. En s'affranchissant des contraintes de la page et de l'imprimé, une édition numérique ouvre la voie à la définition de pratiques normalisées – servant la fonction de légitimation de l'édition scientifique – et de nouvelles formes d'interactivité – appuyant la fonction de médiation de l'édition⁴¹. L'illustration acquiert ainsi un statut équivalent à celui du texte, en tant que vecteur d'arguments scientifiques et témoin des pratiques des acteurs historiques. Néanmoins, de nombreuses questions restent en suspens quand à la mise en œuvre pratique de ces nouvelles méthodes.

40. Paolo Trovato, *Everything You Always Wanted to Know about Lachmann's Method*, Google-Books-ID : VZqlBAAQBAJ, 2014, PAGE.

41. Sur les trois fonctions de l'édition, voir B. Epron et M. Vitali-Rosati, *L'édition à l'ère numérique...*

Conclusion partielle

L’application de techniques de *computer vision* aux documents historiques se heurte à divers obstacles. Les modèles pré-entraînés sur des données génériques peinent à capturer les spécificités visuelles et les complexités propres aux données réelles. En outre, l’entraînement de modèles spécifiques est contraint par la rareté de jeux de données historiques annotées. Les annotations, aussi, doivent être assez précises pour satisfaire les chercheur.ses et assez extensives pour permettre au modèle de généraliser.

Toutefois l’émergence de nouvelles architectures de réseaux de neurones et de nouvelles méthodes d’apprentissage profond permet d’améliorer les performances des modèles sur des documents spécifiques. Par ailleurs, la génération de données synthétiques constitue une approche prometteuse pour pallier le manque de données réelles annotées. En simulant des documents historiques, en imitant les dégradations et le trait manuel qui les caractérise, il est possible d’enrichir les jeux de données d’entraînement. Enfin, la collaboration entre les ingénieur.es de la donnée et les spécialistes permet de définir des protocoles d’annotation adaptés. En trouvant un compromis entre la finesse d’analyse requise par les chercheur.ses et les contraintes de représentativité des modèles, il est possible de créer des jeux de données réelles suffisamment qualitatifs. Pour affiner les résultats, il est nécessaire d’intégrer un processus de correction des sorties des modèles dans la boucle d’apprentissage, afin d’améliorer progressivement leur précision en récupérant des jeux de données annotés à partir des prédictions. Ces processus de vérification sont aussi essentiels pour garantir la pertinence des résultats et compenser la binarité de l’œil de la machine.

Malgré cette étape de correction, l’automatisation des processus grâce aux modèles de vision constitue un gain de temps pour les chercheur.ses et permet de traiter des corpus dont la taille aurait rendu le traitement manuel inenvisageable : ils permettent d’extraire les illustrations des documents, puis de chercher, pour chaque entité, leurs homologues les plus similaires, et enfin de générer des représentations numériques sémantiquement riches et manipulables à partir des images (SVGs). Ces dernières pourrait ainsi servir de base à la création d’un outil qui systématiserait les modalités d’édition des diagrammes, proposant une méthodologie rigoureuse.

En effet, un outil définit un cadre et des pratiques, il porte une vision des sources et une méthode. C’est ce sur quoi portera la partie III. On voudra montrer quelles réflexions, défis et solutions tiennent au développement d’une plateforme qui, tout en fournissant un

cadre méthodologique solide, offre une grande souplesse d'utilisation, permettant aux chercheur.ses de personnaliser leurs *workflows* en fonction de leurs besoins spécifiques. En somme, comment concevoir un véritable Système d'Information extensif capable de soutenir l'ensemble du processus de traitement des illustrations présentes dans les documents historiques, de l'acquisition des données à leur exploitation scientifique.

Troisième partie

**Une plateforme pour une méthode
reproductible et transposable**

Introduction partielle

La modularité des projet EIDA/VHS repose en grande partie sur la construction de la plateforme web AIKON : outil complet pour la recherche sur des documents historiques, permettant leur import, leur stockage, leur analyse (par différentes méthodes) puis leur visualisation. Le but est que tout projet tenant des HN puisse utiliser la plateforme. Ainsi l'application vise à être réutilisable et le code sera mis à disposition en accès libre sur GitHub.

Cette plateforme, pensée pour être agnostique quant au domaine d'application spécifiques, doit être capable d'accueillir des corpus de données divers pour tout chercheur.se ou projet en étude visuelle souhaitant adopter la méthodologie qu'elle porte. Elle doit également prendre en compte la diversité des ressources humaines et matérielles disponibles. Elle doit autoriser l'intégration modulable d'outils et algorithmes, notamment ceux issus du domaine de la *computer vision*, pour traiter des motifs visuels et des objets présents dans les sources. En adoptant cette approche, elle se veut environnement de recherche collaboratif où les chercheur.ses peuvent partager leurs données, méthodes et résultats.

Le défi consiste donc à créer une application aux fonctionnalités suffisamment spécifiques pour répondre aux besoins des deux projets qui portent son développement (VHS et EIDA), tout en étant suffisamment généralistes pour que la plateforme puisse être réemployée à l'avenir par d'autres projets. L'objectif de l'application AIKON est alors de porter une méthode de traitement et d'assurer sa possible transposition dans des environnements techniques et disciplinaires différents. La stratégie de développement consistera alors en l'assemblage de briques techniques, indépendantes le plus possible, et réutilisables seules ou ensemble.

Le développement d'architectures flexibles et évolutives comporte des défis liés à la conception de systèmes adaptables à des besoins changeants et à des environnements hétérogènes. Il implique la modularisation des modèles de données, l'interopérabilité des formats et protocoles, l'architecture du code, les contraintes liées aux ressources matérielles diverses (accès au *hard-ware* et à la puissance de calcul) et les enjeux liés à l'expérience utilisateur.rice (*UX (User eXperience)/UI (User Interface)*). Dans cette partie nous verrons comment la conception de l'application AIKON entend répondre à ses enjeux.

Quelles sont les complexités liées à la conception d'un système d'information capable de s'adapter à des domaines d'application variés et d'évoluer dans le temps ? Nous ex-

plorent deux piliers de cette problématique : l'optimisation des processus internes et la conception d'interfaces utilisateur.rice intuitives, capable de guider l'utilisateur.rice dans ces processus.

Préambule : importance d'une plateforme modulaire

L'ouverture des codes, outils et données dans le contexte de la recherche en HN, et spécifiquement impliquant l'IA, est importante pour plusieurs raisons.

Traçabilité et Reproductibilité

L'ouverture du code source de la plateforme garantit la traçabilité des pipelines d'analyse, favorisant ainsi la transparence des méthodes de recherche. La plateforme offre un cadre de travail reproductible et permet de modéliser et standardiser les processus d'analyse, assurant ainsi une cohérence méthodologique au sein des équipes de recherche et entre différents projets. Cette approche contribue à renforcer la confiance dans les résultats scientifiques et à faciliter leur validation par la communauté. Elle favorise aussi le partage des méthodes au sein de la communauté des chercheur.ses.

Modularité, Solidité et Pérennité

L'objectif est de développer un outil qui pourra être réutilisé, maintenu, et pourra évoluer dans le temps, contrairement à l'usage qui voit souvent des outils de recherche devenir obsolètes et difficiles à maintenir une fois les projets de recherche terminés.

Cette problématique est particulièrement prégnante dans les HN, où les outils développés sont fréquemment abandonnés à la fin des financements, faute de stratégie de pérennisation. Par exemple, la difficulté de maintenir la plateforme DISHAS est exprimée par l'équipe DH, ne serait-ce qu'en raison du coût humain lié au temps passé à son maintien. Qui plus est, l'accord des financements repose bien souvent sur la justification par des livrables techniques innovants. Et donc les outils développés précédemment ne sont pas maintenus et abandonnés.

Un outil qui fonctionne aujourd'hui peut rapidement devenir obsolète ou inadapté si sa conception n'intègre pas une capacité d'évolution. En construisant un SI (*Systèmes d'Information*) modulaire, il devient possible de prolonger la vie de l'outil. Dans une optique de sobriété et dans la philosophie de l'open-source, cette démarche a aussi pour objectif d'éviter à de multiples projets le développement d'outils aux fonctionnalités similaires.

Démocratisation des outils de deep learning

L'ouverture des données historiques ou patrimoniales, mais aussi des technologies qui permettent de les manipuler (outils low-tech, publication du code, mais aussi et no-

tamment outils d'IA) produisent de riche dynamiques de collaboration. Si l'utilisation de la vision artificielle tend à se généraliser dans le traitement de ces données, les projets de recherche ne disposent pas toujours des compétences en interne pour intégrer le *machine learning*, ni les ressources temporelles, financières et humaines nécessaires. Ils peuvent alors profiter du soutien d'une communauté et de l'existence d'outils génériques spécialisables, réduisant ainsi les coûts de développement. En effet, les algorithmes généralistes ne sont généralement pas adaptés aux données historiques et sont donc difficiles à utiliser off-the-shelf sans ajustement. AIKON inclue donc des moyens d'export dans des formats divers, pour créer des vérités de terrain et envisager l'entraînement des modèles, suivant l'exemple de e-Scriptorium.

Il y a donc un enjeu à penser des outils standards, en réfléchissant techniquement à des solutions maintenables sur le long termes grâce à des formats interopérables et des infrastructures flexibles, pour permettre non seulement la reproductibilité des résultats mais aussi une réutilisation des données produites dans des contextes divers.

Chapitre 7

Processus et Fonctionnalités

La modularisation est une stratégie de développement applicatif visant à décomposer un système complexe en unités de code autonomes, appelées modules. Cette décomposition a pour objectif de minimiser les couplages entre ces modules, facilitant ainsi leur développement, leur maintenance et leur réutilisation.

“La programmation modulaire est une solution favorisée pour la création d’une application réutilisable, puisqu’elle permet le développement de modules indépendants qui répondent à des besoins spécifiques, et qui peuvent être réemployés par d’autres projets sans être dépendants du reste de l’application. Ainsi, les différents traitements appliqués aux numérisations déposées par les utilisateur.rices font appel à des outils divers, indépendants, dédiés chacun à une tâche spécifique du *workflow* : cette disposition permet l’amélioration et la modification de chacun des modules sans impacter la structure globale de l’application, et permet également la récupération d’éléments spécifiques par des programmes futurs. Ainsi, le cœur de l’application EIDA/VHS (récemment baptisée AIKON) permet le dépôt, le stockage et l’affichage des numérisations d’ouvrage, ainsi que la correction des annotations. La détection d’objet est gérée par une API[...].”¹

Comment penser des processus adaptables et un *workflow* flexible ? Et quel est l’impact sur les (infra)structures de la plateforme ? En somme, d’un point de vue technique, comment se manifeste l’extensivité de l’outil AIKON ?

7.1 Module de base

Le module de base est un package pour la gestion documentaire, duquel l’application ne peut se détacher. Celui-ci inclut tout d’abord des formulaires pour l’intégration des documents dans la base de données. Le modèle de données permet de décrire différentes entités qui, bien que liées dans leurs métadonnées, peuvent être intégrées indépendamment. Le module de base permet également la création de *Manifests IIIF* pour chaque

1. J. Norindr, *Le traitement des sources historiques par la vision artificielle...*, p.52.

numérisation, permettant ensuite la visualisation des documents grâce aux outils open-source dédiés. De ce fait, l’indexation de zones d’image peut être réalisée manuellement via l’interface Mirador intégrée à SAS. Ce noyau fonctionnel inclut en outre la sélection de lots de documents (le “panier”), sur lesquels pourront être effectués des traitements groupés paramétrables.

Les briques fondamentales offrent donc les fonctionnalités essentielles de gestion documentaire (intégration, modèle de données, IIIF). Les traitements, quant à eux, sont gérés par des modules séparés, et c’est sur cette structure que repose la modularité et l’évolutivité de l’application.

Ci-après nous donnons une description détaillée de certaines de ces fonctionnalités de base.

7.1.1 Description des données

Le module de base contient un modèle de données suffisamment extensif pour décrire efficacement une diversité de données, allant de documents textuels historiques à des tableaux en histoire de l’art. La tripartition entre témoin (*Witness*), série (qui contient un ensemble de témoins), et contenu permet un alignement avec des corpus très diversifiés et des données potentiellement hétéroclites, telles que des manuscrits, des documents épistolaire, des inventaires de galeries d’art, et même pourquoi pas des cartes...

Pour ouvrir à cette large diversité de données, la liste des types de pagination témoin doit être étendue *a minima* d’un nouveau type “other”, émancipant l’enregistrement des mentions de pagination. Les développements futurs prévoient aussi la création d’un système pour ajouter facilement un nouveau type² de *Witness* (tel que peinture, catalogue, etc.).

Au fil des développements, des débats ont émergé autour de l’ajout dans le modèle de données d’un niveau de granularité supplémentaire pour décrire des images ou zones d’images unitaires (*Graphical Elements*), créant ainsi une entité détachée du fait qu’elle provienne d’une extraction dans un document. Cette solution aurait permis une description plus détaillée et plus fine des images, importante pour des projets axés sur des images uniques, et aurait favorisé un élargissement du spectre des types de sources pris en charge. L’utilisateur.rice aurait pu soit importer une image unique (et de manière optionnelle, la lier à un *Witness*) via un formulaire, soit sélectionner une région d’image d’intérêt au sein des extractions (annotations SAS), laquelle serait enregistrée comme *Graphical Element*, puis l’enrichir de métadonnées. Dans les deux cas l’enregistrement d’un *Graphical Element* aurait donné lieu à la création d’une *Digitization* au format JPEG.

2. Le type de témoin est une métadonnée rentrée par l’utilisateur.rice lors de l’enregistrement du *Witness* dans la base de donnée. Il choisit le type dans une liste, originellement manuscrit, imprimé ou gravure sur bois.

Sans l'unité de description *Graphical Element*, les régions d'images sont créées uniquement via les annotations SAS.

L'intégration de cette entité au sein du modèle aurait offert plusieurs avantages en termes de cohérence et de flexibilité. En s'alignant sur les structures existantes (*Witnesses* et *Series*), elle aurait permis une manipulation plus intuitive des images, facilitant ainsi les opérations de recherche et la création de *Sets* personnalisés. De plus, elle aurait rationalisé la gestion des annotations SAS, permettant de sélectionner les plus pertinentes dans la multitude existante.

Cependant, cette approche présente des limites, et on peut trouver des alternatives. Tout d'abord, la coexistence de *Graphical Elements* avec les annotations SAS, générées par des processus distincts, aurait pu créer une certaine confusion quant à leur nature et à leur méthode de création. De plus, la multiplication potentielle de milliers d'enregistrements aurait pu impacter les performances de la base de données et complexifier les requêtes. Enfin, le lien sémantique ambigu et sujet à interprétation subjective entre *Graphical Element* et *Witness* aurait compliqué les possibilités de corrélation.

Compte tenu de ces limites, il a semblé préférable de maintenir les annotations SAS pour identifier les instances de base du modèle, sans créer de nouvelle unité de description. La solution actuelle reste donc basée sur la création manuelle ou automatique de zones dans les images via IIIF et SAS, évitant les problèmes de redondance et de confusion. Bien que l'entité *Graphical Element* n'ait pas été implémentée, les fonctionnalités d'annotation et de sélection d'images sont assurées par d'autres mécanismes. L'outil Mirador permet d'associer des tags aux zones d'image, offrant ainsi une première couche d'enrichissement sémantique. La sélection dans un *set* personnalisé sera possible en gardant en mémoire une référence contenant des coordonnées du *crop*. De plus, l'importation d'images individuelles est réalisable en les considérant comme des *Witness partielles*, ce qui permet de les intégrer dans le *workflow* existant. Toutefois l'enrichissement sémantique à un niveau de granularité fin restera limité ; et la dépendance à l'outil SAS constitue une potentielle dette technique, susceptible de restreindre les évolutions futures du système.

Afin de mieux répondre aux exigences de modularité, l'évolution du modèle de données s'oriente non pas vers une description individuelle des documents, mais vers la gestion des traitements. Cette évolution implique la création d'une entité *Treatment* liée à des ensembles de données (*Document Set* et *Region Set*) potentiellement hétérogènes.

7.1.2 Principe du Traitement

Le but fondamental de la plateforme est de pouvoir effectuer plusieurs actions sur les objets de la base. Afin d'assurer une meilleure traçabilité et plus de flexibilité, la plateforme abandonne les lancements automatiques des processus³ au profit d'un système

3. C'était initialement le cas de l'extraction des entités, dont le lancement était lié à une méthode de classe liée à la *Digitization* après soumission d'un formulaire d'ajout d'un *Witness* ou d'une *Serie*. L'action

basé sur l'entité *Treatment*. Chaque traitement est associé à un ensemble d'objets traités ensemble (*Document Set* ou *Region Set*), à un jeu de paramètres et à un résultat. Ces informations sont stockées dans une table dédiée. Cette approche facilite la gestion et le trackage des processus (notamment, les utilisateur.rices sont notifiés par e-mail à la fin du *processing*), permet aux utilisateur.rices de consulter un historique de leurs actions et offre la possibilité de créer des *workflows* personnalisés en passant par un formulaire de lancement unique mais extensif.

En permettant de regrouper des documents de types différents (*Works*, *Series*, *Witnesses*) dans des *Document Sets*, on offre à l'utilisateur.rice la flexibilité de lancer des actions sur des ensembles d'entités hétérogènes et granulaires. Le traitement est ensuite réparti sur les entités de niveau inférieur (les témoins). Les *Witnesses* ainsi sélectionnés peuvent être soumis à une large gamme de traitements : des fonctions déjà implémentées comme l'exportation (avec choix du format), l'extraction, la vectorisation, la recherche de similarité ; ou de nouveaux traitements personnalisés, tels que la visualisation sur une frise chronologique ou une carte. La modularité de la plateforme est assurée par un formulaire de lancement configurable, permettant de l'adapter à différents scénarios d'utilisation, et à l'ajout de modules personnalisés.

Le *Region Set* fonctionne similairement au *Document Set*, à un niveau de granularité inférieur (à l'échelle de la zone d'image)⁴.

7.1.3 Extraction manuelle des zones d'image

Le choix de la méthode d'extraction des régions d'intérêt dans les documents constitue un élément clé de la modularité de la plateforme. Les utilisateur.rices peuvent opter pour une extraction manuelle ou une extraction automatique basée sur des algorithmes de vision par ordinateur, adaptée aux traitements à plus grande échelle.

Après importation d'un enregistrement, le flux de travail procède à la création de *Manifests IIIF* pour chaque numérisation (*Digitization*) afin de permettre une visualisation grâce à la plateforme Mirador. Le module de base autorise par la suite l'extraction manuelle de zones d'intérêt au sein des images. Cette fonctionnalité est particulièrement utile pour les projets ne souhaitant pas recourir à des méthodes entièrement automatisées de vision par ordinateur. L'outil SAS permet de créer des annotations, c'est-à-dire de définir des régions d'intérêt spécifiques dans les numérisations, et de les indexer directement dans les *Manifests IIIF* correspondants, enrichissant ainsi les ressources numériques. De plus, les développements futurs prévoient la possibilité d'importer des fichiers d'annotation pré-existants en format .TXT afin de pouvoir les indexer manuellement. Par conséquent, le

se lançait immédiatement après enregistrement des images d'une numérisation dans la plateforme.

4. À l'été 2024, l'entité n'existe pas encore dans la base de données, mais le processus d'envoi du traitement et les modes de communication entre l'application et l'API prévoient la possibilité de lancer l'inférence des modèles sur un ensemble de régions extraites.

workflow de base ne comporte aucun traitement automatique basé sur la vision (et de fait éventuellement trop gourmand en puissance de calcul).

L'extraction, qu'elle soit manuelle ou automatique, constitue le fondement du reste des processus. Une interface est disponible pour sélectionner un ensemble de documents et effectuer des actions spécifiques sur les témoins annotés, via le formulaire de traitement qui s'étend selon un choix de module configuré. Ainsi l'utilisateur.rice n'est pas limité par un contexte initial, à l'origine deux étapes indissociables et incontournables (importation et extraction), pour pouvoir effectuer d'autres actions. Cette modularité permet de s'affranchir d'un *workflow* linéaire et prédéfini, offrant ainsi une plus grande adaptabilité aux besoins spécifiques et aux ressources matérielles des projets.

Pour conclure, l'existence de ce module de base répond à des besoins élémentaires des projets de recherche en études visuelles. Il fournit un outil qui permet d'agréger toutes les sources primaires qui concernent le sujet, de décrire les sources et de les mettre en relation. Il offre en outre la possibilité d'extraire et visualiser des contenus d'intérêt (les "crops" d'images), ciblant ainsi les instances de base qui intéressent les chercheur.ses.

7.2 Des tâches complémentaires : parallélisation du *workflow*

Bien que les étapes de traitement automatiques implémentées au dessus du module de base aient une cohérence logique, dans les faits le seul principe obligatoire est la présence de l'entité de base (la région d'image) sur laquelle effectuer l'action. Ensuite, les traitements ne suivent pas obligatoirement une séquence chronologique stricte. La première étape de sélection de zones d'images d'intérêt (automatique ou manuelle) peut être seule considérée comme nécessaire et suffisante pour démarrer une analyse approfondie des sources. La refonte des processus prend donc en compte la possibilité de procéder dans un ordre aléatoire.

7.2.1 Modules optionnels

Le *workflow* commence par l'enregistrement d'un document (*Witness* ou *Serie*) avec ses métadonnées, et l'import d'une numérisation associée. Il est alors visualisable.

L'application contient des modules optionnels à activer ou désactiver selon les besoins. Les modules d'ores et déjà proposés utilisent la *computer vision*, ils sont aujourd'hui au nombre de trois : l'extraction, la recherche de similarité et la vectorisation. Les outils de *deep learning* sont mis à disposition via une API accueillant les modèles pour l'inférence et séparée de l'application permettant la gestion des données, le lancement des traitements, la visualisation et la correction des résultats.

L'extraction automatique détecte les zones d'intérêt dans les numérisations et les indexe dans le système SAS. La tache se déclenchaient initialement systématiquement lors de l'enregistrement du témoin dans la base de données (la méthode de classe qui permettait l'enregistrement du *Witness* lançait un *thread* en même temps). Pour plus de souplesse dans les processus, l'annotation automatique – en tant que module optionnel – est aujourd'hui déclenchée par une action de l'utilisateur.rice, via le formulaire de lancement de traitement. Si le chercheur.se fait le choix d'appliquer une extraction automatique sur sa/ses sources, une requête API lance l'inférence avec les scripts de YOLOv5 sur le GPU Dishas-ia, et déclenche l'annotation automatique de diagrammes dans les scans numérisés. À l'issue de ce processus, l'API génère un fichier texte structuré, contenant pour chaque image son identifiant unique et les coordonnées précises des régions détectées comme étant des diagrammes. Ces informations sont ensuite renvoyées à l'application, le fichier est enregistré et les informations qu'il contient sont indexées sous forme d'annotations dans le *Manifest IIIF* correspondant à chaque numérisation.

La présence de ces régions découpées au sein des images numérisées constitue une condition préalable indispensable à l'exécution de traitements ultérieurs.

Le principe du traitement groupé normalise le processus de lancement des tâches et le formatage des données à envoyer à l'API. Un fichier JSON, vecteur de la communication entre les deux composants, structure une liste d'URLS IIIF, chacune pointant vers une image ou une région d'image d'intérêt, ainsi que des métadonnées descriptives.

Le JSON est formaté par une fonction appelée `prepare_request`, et qui existe pour tous les modules, ci-dessous, par exemple, pour l'extraction :

```
#vhs/app/regions/utils.py
def prepare_request(witnesses, treatment_id):
    manifests = {}

    try:
        for witness in witnesses:
            if witness.has_regions():
                log(
                    f"[regions_request] Witness {witness.get_ref()} already has
                    regions"
                )
                pass
            else:
                digits = witness.get_digits()
                for digit in digits:
                    manifests.update({witness.get_ref(): digit.gen_manifest_url()
                        })
    
```

```

if manifests:
    return {
        "experiment_id": f"{treatment_id}",
        "documents": manifests,
        "model": f"{EXTRACTOR_MODEL}", # Use only if specific model is
            desire
        "callback": f"{APP_URL}/{APP_NAME}/get-regions", # URL to which
            the regions file must be sent back
        "tracking_url": f"{APP_URL}/{APP_NAME}/api-progress",
    }
else:
    return {
        "message": f"Regions were already extracted for all the selected
            {WIT}es"
        if APP_LANG == "en"
        else f"Les régions ont déjà été extraites pour tous les {WIT}s
            sélectionnés"
    }

except Exception as e:
    log(
        f"[prepare_request] Failed to prepare data for regions request",
        e,
    )
    raise Exception(f"[prepare_request] Failed to prepare data for regions
        request")

```

Listing 7.1 – Fonction `prepare_request` du module d'extraction

La fonction reçoit une liste de *Witnesses* générée par une tâche appelée par la méthode `treatment_post_save` de l'entité *Treatment* :

```

# vhs/app/webapp/models/treatment.py
@receiver(post_save, sender=Treatment)
def treatment_post_save(sender, instance, created, **kwargs):
    if created:
        get_all_witnesses.delay(instance)

```

Listing 7.2 – Crédit à la création d'un *Treatment*.

```

# vhs/app/webapp/tasks.py
@celery_app.task
def get_all_witnesses(treatment):
    try:

```

```

    witnesses = treatment.get_witnesses()
    treatment.start_task(witnesses)

except Exception as e:
    treatment.on_task_error(
        {
            "error": f"Error when retrieving documents from set: {e}",
            "notify": treatment.notify_email,
        },
    )
)

```

Listing 7.3 – Tâche qui rassemble les *Witnesses* et lance l'action sur les documents.

Ce formatage normalisé permet d'adapter tout le processus de lancement à une variété de tâches. L'envoi d'une requête HTTP (*Hypertext Transfer Protocol*) à l'*endpoint* d'API appropriée déclenche l'exécution du traitement. Une fois l'inférence terminée, les résultats sont retournés à l'application et une notification est envoyée à l'utilisateur.rice. L'instance *Treatment* sert aussi à standardiser le suivi en temps réel de l'état d'avancement de chaque tâche (Fig. 7.1).

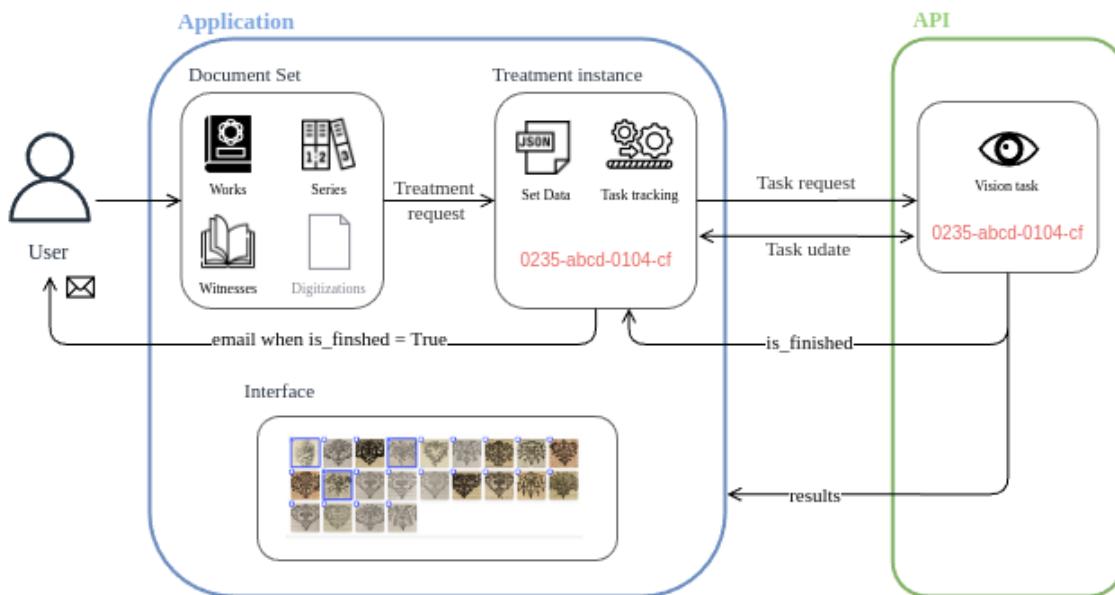


FIGURE 7.1 – Fonctionnement de l'envoi d'un traitement à l'API.*⁵

Le système est ainsi conçu pour être évolutif. Les utilisateur.rices peuvent envisager d'étendre ses capacités en y intégrant de nouveaux modules. Parmi les fonctionnalités potentielles, on peut citer : des outils d'édition des diagrammes (ce qui intéresse spécifiquement EIDA), des algorithmes de classification et de clustering, un modèle de recon-

5. Les schémas suivis d'un astérisque sont prélevés ou inspirés de la documentation technique interne créée par les développeur.ses du projet EIDA et de l'ISCD (*Institut des sciences du calcul et des données*).

naissance de filigranes, ou encore l'intégration d'une pipeline de transcription de textes comme e-Scriptorium.

La conception modulaire du système permet de configurer des pipelines de traitement personnalisés. Chaque module peut être activé ou désactivé indépendamment des autres. Les traitements peuvent être combinés selon les besoins spécifiques. Par exemple, le calcul de similarité et la vectorisation peuvent être réalisés parallèlement, et sans passer par une étape d'extraction automatique préalable, si les documents ont été annotés manuellement.

Module *Similarité* : workflow détaillé

Le module *similarity retrieval* inclut le calcul de scores de similarité selon différentes méthodes (*cosine distance*, Segswap, etc.)⁶ et la catégorisation de ces similarités (exact match, partial, personnel, etc.)⁷. La recherche de similarité ne fonctionne que sur les images de diagrammes déjà extraites, identifiées grâce à des URLs et accessibles grâce à l'API Image de IIIF. La tâche est lancée sur un ensemble de témoins. Pour chaque document sélectionné, le GPU reçoit une liste d'URLS IIIF pointant vers des régions d'intérêt. Les images correspondantes sont téléchargées et des descripteurs visuels (*features*) sont extraits grâce au réseau de neurones. L'attribution d'un indice de similarité se fait par la comparaison de plusieurs métriques des vecteurs de descripteurs. Les résultats sont stockés dans un fichier NumPy. Ce fichier, qui contient les scores de similarité pour toutes les paires d'images comparées, est retourné à l'application.

L'interface utilisateur.rice permet une exploration interactive de ces résultats. Pour une image donnée, il est possible de visualiser dynamiquement les images les plus similaires provenant d'autres documents. Cette fonctionnalité repose sur l'agrégation des scores de similarité au niveau de chaque image.

Module *Vectorisation* : workflow détaillé

La tâche de vectorisation consiste en la génération de fichiers SVGS à partir des images extraites. Tout comme la recherche de similarité, cette opération s'applique à des régions d'images préalablement indexées. La vectorisation peut être intégrée à différents points du pipeline de traitement ; elle peut être exécutée avant ou après la recherche de similarité, en fonction des besoins.

Comme pour la similarité, une liste d'URL structurées dans un fichier JSON est envoyée par une requête HTTP POST : sa réception lance la chaîne de traitement dans l'API, qui télécharge les images à partir des URLs spécifiées dans le fichier JSON. Les SVGS résultants de l'inférence avec le modèle sont compressés dans un fichier ZIP. Le fichier

6. Voir le chapitre IV, 2.2

7. Voir le chapitre suivant.

ZIP contenant les SVGs est renvoyé à l'application via une requête HTTP POST. Enfin, l'application enregistre le fichier ZIP sur le serveur, rendant ainsi les SVG accessibles à l'utilisateur.rice.

L'inférence du modèle de vision nécessite des ressources computationnelles importantes. L'implémentation dans l'API doit prendre en compte ces exigences de performance et garantir en conséquence une scalabilité adéquate. Pour éviter la surcharge et gérer les multiples requêtes sur le GPU, l'API utilise un système de threading. Ce système, nommé Dramatiq, gère les tâches de fond et traite les requêtes entrantes en imposant leur traitement séquentiel. Dramatiq fonctionne avec Redis, un magasin de valeurs-clés, servant de base de données intermédiaire qui assigne les tâches aux *workers*. Dramatiq gère la bonne tenue des tâches, faisant l'intermédiaire entre les *workers* et le système de *login*. De la même manière, pour la réception ou l'envoi de données volumineuses, l'application emploie Celery, un système de threading similaire.

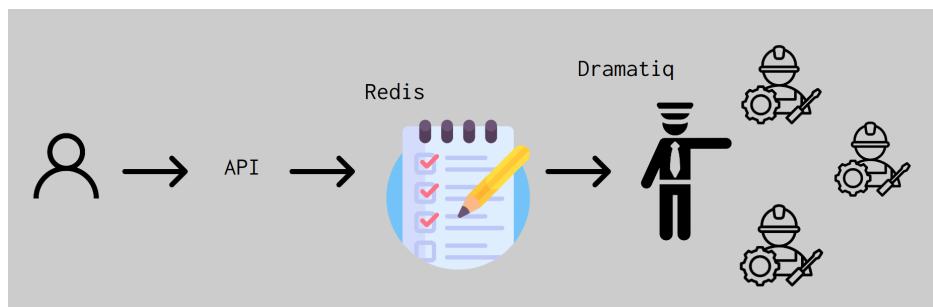


FIGURE 7.2 – Fonctionnement du système de *Threading* de l'API.

Le cœur de ce système modulaire réside dans le développement de trois interfaces principales qui guideront le chercheur.se dans les processus : la recherche d'entités, leur sélection dans un panier et le déclenchement des traitements.

7.2.2 Alternance d'étapes automatiques et d'étapes manuelles

Prenant en compte les biais potentiels des traitement d'IA, l'implémentation des tâches de vision doit être pensée de manière à autoriser la correction de l'inférence des modèles, d'une part pour avoir des résultats finaux plus pertinents, et d'autre part pour produire des données annotées dédiés au *fine-tuning* des modèles.

En effet :

“Data analysis for Art History means the introduction of quantitative methods into a field that has used exclusively qualitative methods throughout its existence. However, these quantitative methods are not a substitute for conventional methods, but an addition. Results of data analysis can help to answer to our research questions, but may also spark new hypotheses that can be followed using a bouquet of methods that Art History has. For some, Big Data seems to be the end of theory

but one could also argue that Data Analytics helps us to see the big picture and create new hypotheses for qualitative research.⁴⁸ This way, the computer remains a tool in the hands of the researcher, or as Steve Jobs defined the role of the computer to the human, "a bicycle for the mind".⁸

Les méthodes quantitatives sur la donnée de masse sont une étape préliminaire à la recherche qualitative. Les résultats nécessitent toujours une interprétation ; ils ne remplacent pas les hypothèses, mais en sont la prémissse.

"L'automatisation n'exclut pas un certain nombre d'actions manuelles sur les sources, et l'utilisation d'algorithmes de vision artificielle ne remplace pas le regard critique que peuvent porter les chercheur.ses sur les sources : il est donc nécessaire, dans les chaînes de traitement définies par les projets, de trouver un équilibre entre étapes automatiques et étapes manuelles de traitement des sources historiques, pour assurer, en finalité, des résultats les plus pertinents possibles."⁹

"Face à ce constat, il est nécessaire, pour les ingénieurs, de concevoir une chaîne de traitement qui va au-delà des étapes automatisées, et qui prend en compte la nécessité d'interventions manuelles entre chaque traitement pour un ajustement des résultats et une analyse primaire des données de sortie. Pour espérer produire des recherches aux résultats pertinents, il est ainsi nécessaire de savoir intégrer l'outil qu'est le *deep learning* à un *workflow*, et de trouver l'équilibre entre automatisation et tâches manuelles."¹⁰

L'IA est employée comme un outil de pré-traitement des sources, permettant de naviguer dans un corpus volumineux, de créer des rapprochements pour identifier les phénomènes de copie et d'emprunts, et/ou d'éditer les sources dans un format sémantiquement riche qu'est le SVG. Ces traitements permettent aux chercheur.ses de gagner un temps significatif comparé à une analyse et une transcription manuelle des documents. L'établissement de la chaîne de traitement se doit néanmoins de prendre en compte les limites de la *computer vision*. Le *workflow* doit comporter suffisamment d'étapes de correction manuelle, qui restent moins chronophages qu'une recherche et édition manuelle des illustrations dans chaque source étudiée, et n'en sont que d'autant plus nécessaires pour assurer la pertinence des analyses.

Pour optimiser les performances de l'algorithme de détection, il est nécessaire d'exclure ou d'inclure certains cas limites lors de l'entraînement du modèle¹¹. Le modèle de détection final peut alors manquer des diagrammes pertinents pour les chercheur.ses ou, à l'inverse, détecter des illustrations sans rapport avec leur objet d'étude.

De même, l'algorithme de similarité présente ses propres limites¹². Les chercheur.ses ont besoin de poser un regard expert sur leurs corpus, et d'effectuer manuellement les

8. H. Klinke, « Big Image Data within the Big Picture of Art History »..., p.28.

9. J. Norindr, *Le traitement des sources historiques par la vision artificielle...*, p.70.

10. *Ibid.*

11. Voir le chapitre 5, 2.3

12. Voir le chapitre 1, 3.2

rapprochements entre les images, ou bien de qualifier avec une plus grande finesse le type de similarité existant entre deux images

Une étape de vérification manuelle permet donc de corriger les défauts des modèle, notamment sa tendance à la binarité qui ne correspond pas à la complexité du réel. Il est alors nécessaire de travailler des interfaces de visualisation et de correction des résultats, et de modéliser ces processus de vérification dans le *workflow*. Les données récupérées sont donc inspectées par l’œil critique du chercheur.se. La réduction de l’automatisation au profit d’une intervention humaine contrôlée favorise une meilleure adéquation aux besoins spécifiques de chaque projet de recherche, bien que les chercheur.ses bénéficient la rapidité de traitement de l’intelligence artificielle pour le traitement de corpus volumineux. Cette approche hybride, combinant l’efficacité de l’apprentissage profond et la précision de l’expertise humaine, optimise ainsi la qualité globale des résultats.

La table *Regions* intègre un attribut booléen `is_validated` indiquant si les annotations automatiques associées à une numérisation ont été approuvées par un chercheur.se. Pareillement, afin d’évaluer la pertinence des scores attribués par l’algorithme de recherche de similarité, une catégorie est associée à chaque paire d’images, permettant ainsi d’évaluer la ressemblance de ces paires selon une nomenclature prédéfinie. Cette catégorisation, établie par les chercheur.ses, sert donc à évaluer la qualité des scores de similarité attribués par le modèle et constitue une métadonnée précieuse pour une utilisation personnalisée ou pour l’amélioration du modèle. Pour pallier aux manquements de l’algorithme, les utilisateur.rices peuvent aussi créer manuellement des paires d’images (attribut `is_manual` de la table *Regions_Pair*). De même, les SVGs ramenés en sortie du modèle de vectorisation devront faire l’objet d’une correction manuelle, dans une interface dédiée, par les chercheur.ses, avant d’envisager leur exploitation.

Ainsi, pour que les avantages de la *computer vision* soient pleinement exploités, il est important de disposer d’interfaces utilisateur.rice interactives pour la visualisation et la correction des résultats, facilitant leur compréhension et permettant de tester différentes hypothèses (Fig. 7.3).

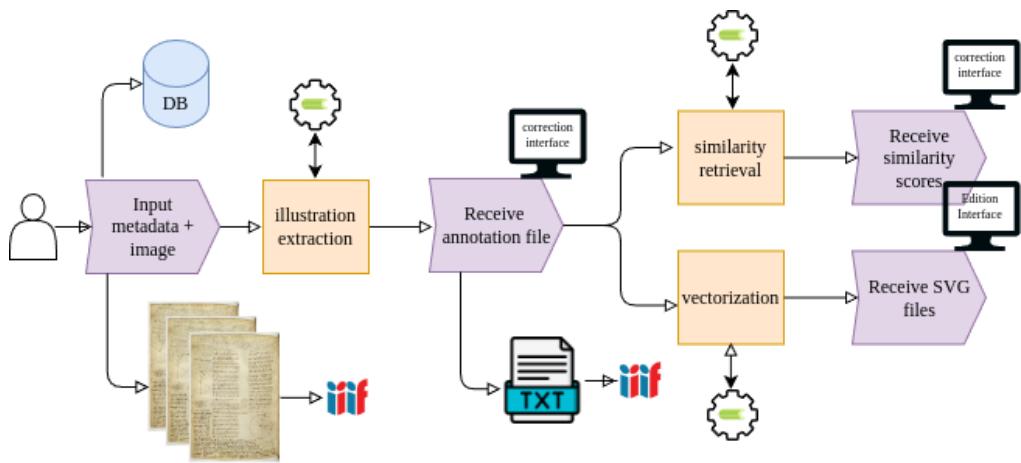


FIGURE 7.3 – Workflow schématique.*

Pour conclure, l’application, conçue selon une architecture modulaire, s’adresse aussi bien aux utilisateur.rices recherchant des fonctionnalités de base qu’aux équipes de recherche souhaitant personnaliser leur environnement de travail. En offrant un ensemble de modules génériques, le projet EIDA fournit un point de départ solide pour des projets plus ambitieux. Les équipes d’ingénieurs peuvent ainsi se concentrer sur le développement de modules spécifiques répondant à leurs besoins précis, sans avoir à reconstruire l’ensemble de la plateforme. Cette approche favorise l’innovation et permet de créer des solutions sur mesure en réduisant le coût de développement.

7.3 (Infra)structures

La construction d’une plateforme extensive et modulaire pour démocratiser l’accès à un outil de gestion et de traitement de la donnée visuelle implique une réflexion approfondie sur les architectures matérielles et logicielles. Pour toucher des publics diversifiés de la communauté de la recherche, il est essentiel de penser des infrastructures matérielles diverses, plus ou moins puissantes et abordables, intégrant ou non des composants comme les GPU (qui permettent d’accélérer les calculs intensifs nécessaires à l’IA). En effet, la gestion efficace des ressources, la scalabilité, et la performance sont des aspects à prendre en compte pour que ces outils puissent être utilisés de manière fiable. Parallèlement, l’architecture logicielle doit être flexible et évolutive.

7.3.1 Hardware : une API sur le GPU

La séparation physique de l’inférence des modèles tient un rôle important dans l’ouverture de la plateforme.

Le type d’infrastructure de calcul, notamment le CPU ou le GPU, implique des différences dans le traitement et l’analyse des données, chacun offrant des capacités dis-

tinctes adaptées à des besoins spécifiques. Un CPU (Central Processing Unit) est le processeur principal d'un ordinateur, conçu pour gérer une large gamme de tâches générales et basiques, et utilisé pour les besoins quotidiens. Un GPU (Graphics Processing Unit) est spécialisé dans le traitement des éléments graphiques. Il est conçu pour effectuer un grand nombre de calculs simples en parallèle grâce à ses nombreux coeurs, ce qui le rend extrêmement efficace pour des tâches nécessitant un traitement massif et simultané de données. L'utilisation d'un GPU est souvent nécessaire pour les tâches d'IA, notamment en vision artificielle, car ces tâches impliquent souvent des opérations de calcul intensives et parallélisables. Un GPU, avec sa capacité à gérer des milliers de *threads* en parallèle, permet d'accélérer l'entraînement et l'inférence des modèles de vision artificielle, rendant le traitement plus rapide et plus efficace que sur CPU.

Discover-Demo est une API développée comme un module de l'application, répondant au besoin de séparer les algorithmes de vision du reste de l'application. Cette séparation permet une plus grande flexibilité dans l'utilisation des ressources de calcul. Elle tourne sur le GPU Dishas-ia, dédié quasi exclusivement aux besoins de l'équipe d'histoire des sciences du SYRTE.

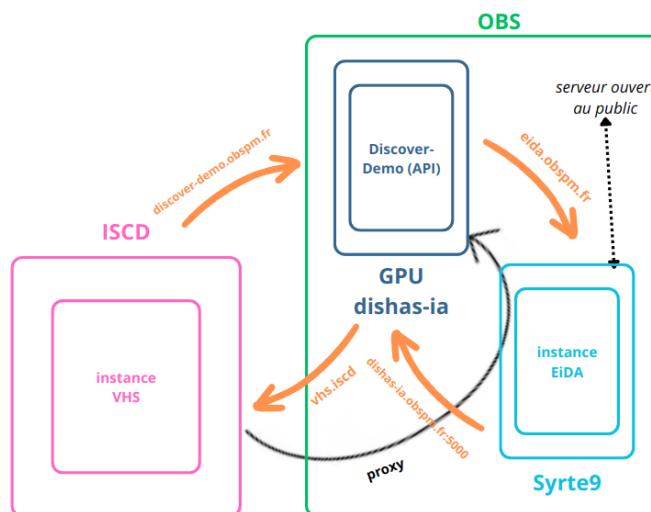


FIGURE 7.4 – Organisation et communication des infrastructures.

Malgré l'importance accordée à l'IA, l'interface web et l'API associée sont conçues pour une analyse complète des documents historiques, allant de leur importation et stockage à leurs traitements (divers) et visualisations. Dans une optique d'extensivité, elles ne doivent être rattachées à aucun processus d'analyse prédéterminée. Ainsi, toutes les étapes peuvent être effectuées manuellement ou à l'aide d'algorithmes automatisés. L'application de base n'intègre pas de traitement de *computer vision*, mais permet de gérer une base de données et des sources avec leurs numérisations, utilisant le standard IIIF. Elle permet l'indexation manuelle de zones d'images dans SAS via l'interface Mirador,

permettant la sélection de zones d'images d'intérêt. Pour cela, l'extraction automatique de zones d'images est séparée en un nouveau module, mais les fonctionnalités de base de l'application incluent toujours les outils nécessaires pour effectuer des annotations manuelles de régions. Cela comprend toutes les fonctions pour indexer un fichier texte dans SAS, visualiser les régions annotées, et exporter les résultats. Il devient alors envisageable d'importer des résultats de traitement (fichiers d'annotation de régions ou de paires de régions similaires) et de les indexer manuellement pour permettre leur visualisation et analyse ultérieure.

Chaque traitement peut donc être réalisé via l'inférence des modèles de vision sur GPU (comme c'est le cas pour EIDA grâce à l'API), par l'import d'un fichier de résultats, manuellement, ou potentiellement par des méthodes locales sur CPU (YOLOv5, par exemple, est assez léger pour tourner en local). La plateforme permet ainsi une adaptation à des environnements matériels divers, laissant la possibilité de réaliser les traitements soit automatiquement via l'IA, soit manuellement.

Séparer les modèles de vision c'est aussi permettre une bascule vers des modèles spécialisés. Les modèles développés dans le cadre du projet sont disponibles mais peuvent facilement être réentraînés pour correspondre spécifiquement aux données de l'utilisateur.rice, prenant en compte les besoins de sa recherche.

Pour conclure, grâce à cette séparation des composants *hard-ware*, la plateforme répond efficacement à une diversité de besoins et permet son intégration dans des projets aux ressources matérielles variées. Même sans ressource matérielle capable de faire tourner les modèles de vision, les utilisateur.rices peuvent toujours exploiter la plateforme web. Cette conception offre un accès aux outils et méthode à des utilisateur.rices divers, allant des projets sans ingénieur dédié pour le développement, aux équipes de recherche disposant de leurs propres ingénieurs, en passant par des doctorants indépendants ayant des compétences en programmation mais sans accès à un serveur. L'outil est pensé pour s'adapter à des environnements variés, des configurations légères fonctionnant en local, jusqu'à des projets disposant de ressources matérielles importantes comme un GPU.

7.3.2 Software : des modules séparés

Le modèle MVC (Model-View-Controller) est une architecture logicielle qui segmente une application en trois composantes interconnectées. Le Modèle est chargé de la gestion des données et de la logique métier de l'application, assurant la manipulation et l'administration des informations. La Vue est responsable de la présentation visuelle des données, les mettant en forme visuellement dans un *template*. Le 'Contrôleur', sert d'intermédiaire entre le 'Modèle' et la 'Vue' : il reçoit les entrées de l'utilisateur.rice via la 'Vue', traite ces entrées, puis interagit avec le 'Modèle' pour actualiser les données et, enfin, met à jour ces modifications dans la Vue. Cette séparation des préoccupations permet une or-

ganisation plus rigoureuse du code, facilitant ainsi la maintenance, la réutilisabilité et le développement parallèle de chaque composante.

Le cycle action → mise à jour → affichage induit par ce patron est bien adapté aux applications web, il est à ce titre utilisé par nombre d'entre elles, dont EIDA fait partie, et par de nombreux *frameworks*, Django y compris.

Bien que le modèle MVC offre déjà une structure prenant en compte la séparation des préoccupation, EIDA cherche à aller au-delà, proposant une architecture encore plus flexible. La plateforme est conçue pour permettre aux développeur.es d'ajouter ou de supprimer des fonctionnalités de manière indépendante. Cette approche permet de personnaliser l'application en fonction des besoins spécifiques de chaque projet, sans avoir à modifier le cœur du système. Les utilisateur.rices peuvent ainsi partir de la base de la plateforme et la compléter avec des modules sur mesure.

Voici une transcription de l'arborescence des fichiers de l'application :

```
app/
    config/
    logs/
    mediafiles/
    regions/
    similarity/
    vectorization/
        templates/
        __init__.py
        const.py
        tasks.py
        urls.py
        utils.py
        views.py
    webapp/
    webpack/
        __init__.py
        manage.py
        requirements-base.txt
        requirements-dev.txt
        requirements-prod.txt
    cantaloupe/
    celery/
    docs/
    gunicorn/
    sas/
```

```

scripts/
  .gitignore
  .pre-commit-config.yaml
README.md
run.sh

```

Il a été créées plusieurs unités fonctionnelles pouvant inclure leurs vues, *templates*, utilitaires, etc. Cette approche permet aux développeur.es de découper l'application tout en factorisant le code dédié à plusieurs tâches, ainsi les modules partagent des *statics*, un fichier de configuration global et des fonctions utilitaires.

Chaque sous-dossier dans `app/` représente un module fonctionnel. Le répertoire `webapp/` contient le module de base, tandis que `webpack/` est dédié aux interfaces¹³. Les modules additionnels, autonomes, peuvent s'interfacer les uns avec les autres, et être développés puis testés de indépendamment. Pour un exemple détaillé du contenu des fichiers, une description du module dédié à la vectorisation développée pendant se trouve en annexe B.

La variable `INSTALLED_APPS` du fichier de configuration global permet de personnaliser l'application en activant ou désactivant les modules souhaités.

Modularité

Cette architecture s'inscrit dans une stratégie de développement applicatif ouverte et évolutive. La division en unités fonctionnelles et indépendantes permet d'ajouter ou de supprimer les fonctionnalité complémentaires au module de base. Ce cadre de développement modulaire garantit que l'application reste adaptable aux exigences évolutives des chercheur.es et des institutions partenaires, la rendant plus robuste et tolérante aux usages extérieurs et autorisant alors le réemploi du code par des projets ayant besoin d'effectuer des traitements divers sur du matériel documentaire pictural numérisé. AIKON est utilisable par des projets divers, ce qui réduit le temps de développement et ouvre la voie à des partenariats, aidant alors à pérenniser les outils.

Maintenance

L'organisation modulaire du code facilite également la maintenance et les mises à jour. Avec une telle structure, il devient plus facile d'isoler les composants pour le développement et le débogage. Les développeur.es peuvent travailler sur un module spécifique sans interférer avec les autres parties du projet. De plus, l'implémentation d'un module indépendant provoquera moins de conflit lors du déploiement.

En conclusion, AIKON est composée de plusieurs modules applicatifs, elle est exten-

13. Voir le chapitre suivant

sible et facile à maintenir (l'organisation des processus se reflétant dans l'arborescence de fichiers), tout en facilitant la gestion des dépendances et l'optimisation des performances.

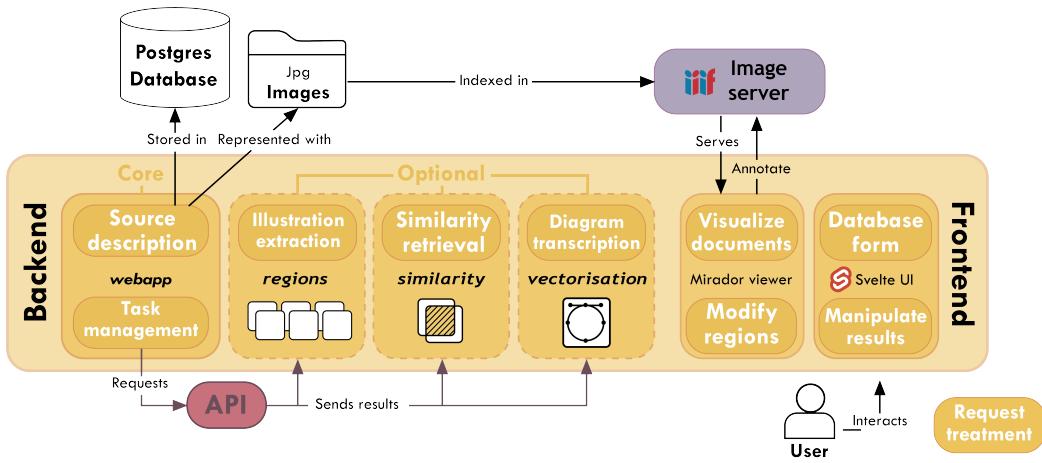


FIGURE 7.5 – Représentation schématique des fonctionnalités et processus de la plate-forme.*

Cette approche modulaire est particulièrement bénéfique pour les projets complexes nécessitant une collaboration entre plusieurs équipes de développement (c'est d'ailleurs le cas dans le cadre de la collaboration EIDA/VHS), assurant la qualité et la robustesse du code, et permet la modularité de la plateforme.

Chapitre 8

Interfaces

L’interface crée une zone d’échange et de contact entre l’application et l’utilisateur.rice, permettant d’échanger des informations grâce à l’adoption de règles communes. En cela, elle est bien plus qu’une simple couche superficielle. Elle constitue le point de rencontre entre deux mondes : celui de l’humain et celui de la machine. Concevoir une interface, c’est penser à la manière dont les utilisateur.rices vont interagir avec un système. C’est définir les éléments visuels, les commandes, les actions possibles et la logique qui régit ces interactions. L’interface est ainsi l’aboutissement d’un processus de design qui vise à rendre une expérience utilisateur.rice aussi intuitive et efficace que possible. Et, à ce titre, elle jouera un rôle décisif dans la mise à disposition et l’adoption des méthodes d’analyse basées sur l’IA.

L’intégration du *deep learning* aux pratiques des chercheur.ses en sciences historiques peut être favorisée par le développement de la plateforme AIKON, qui leur permet d’exploiter simplement ces outils pour traiter leurs sources. L’interface graphique sert donc d’intermédiaire entre les chercheur.ses et les algorithmes de vision artificielle. Elle facilite en outre le dépôt et la gestion structurée des données sources et des métadonnées associées et simplifie l’accès aux outils de traitement de l’images. Cette plateforme à interface graphique doit être adaptée à une grande diversité de documents à traiter, et pensée pour accueillir des utilisateur.rices divers, aux compétences numériques et aux questions de recherche variées.

“[L]’activité de conception des plateformes digitales, chères aux Humanités Numériques en tant que ces dernières se veulent productrices d’outils et d’instruments, doit être considérée par nature comme relevant d’un travail de design et, par conséquent, en intégrer la culture créative et la philosophie dès le commencement, dans l’esprit d’un « design des programmes » (Masure, 2014) qui doit permettre de développer et d’améliorer le design des plateformes. Car, on le sait, la cause principale de la réussite (et donc de l’adoption par une communauté) d’un service numérique réside dans la haute qualité d’expérience utilisateur.rice (User eXperience) qu’il est capable de délivrer aux usagers.”¹

1. Frédéric Clavert, *2DH — Design & Digital Humanities : le design comme méthode pour les*

L’UX *design*, approche centrée utilisateur.rice, vise à optimiser les interactions entre un utilisateur.rice et un système. Il s’appuie sur des méthodes issues des sciences cognitives pour concevoir des interfaces intuitives et efficaces. L’UI *design*, partie intégrante de l’UX, se concentre sur l’aspect visuel de l’interface, en cherchant à créer une expérience esthétique et cohérente. Le projet EIDA illustre parfaitement l’importance de l’UX/UI *design* dans le domaine de la recherche. En plaçant l’utilisateur.rice final (le chercheur.se) au centre de ses préoccupations, EIDA vise à développer une plateforme qui le guidera dans l’adoption des méthodes et la prise en main de la chaîne de traitement. Ainsi un outil de recherche, pour être efficace, doit disposer d’une interface utilisateur.rice pensée pour optimiser les *workflows* des chercheur.ses. Une bonne ergonomie, une navigation fluide et une organisation claire des informations permettent d’assurer la cohérence des méthodes et processus. Ainsi, l’outil pourra devenir un véritable levier de productivité, modélisant des processus unifiés au sein des équipes. Une interface intuitive et agréable facilite l’apprentissage et l’utilisation de l’outil, même par des utilisateur.rices ayant des niveaux d’expertise variés ; elle favorise ainsi l’engagement et l’adoption, par le biais de l’outil numérique, d’une méthodologie.

8.1 Faire le lien avec les chercheur.ses

Le *back-office* d’une application web conçue pour gérer les données de la recherche, à l’instar d’AIKON, est l’interface interne, non accessible au public. Cette interface offre un espace de travail centralisé aux chercheur.ses pour saisir, modifier, supprimer et organiser les données de manière structurée. Elle assure également la gestion des utilisateur.rices et des droits d’accès, garantissant ainsi la sécurité et la confidentialité des informations.

L’enjeu majeur du SI AIKON est de structurer et d’uniformiser les pratiques de recherche. Il entend se positionner comme outil efficace pour la standardisation et l’optimisation des *workflows*, offrant un cadre méthodologique rigoureux. En regroupant l’ensemble des données de recherche au sein d’une infrastructure commune, ce système rationalise les opérations de gestion de données, réduit les risques de perte et facilite les recherches. Grâce à l’implémentation de métadonnées descriptives, de formats de données standardisés et de règles de validation, il assure la cohérence sémantique et structurelle des données. Par ailleurs il facilite la collaboration, s’offrant comme véritable “laboratoire”. Enfin, en modélisant des protocoles d’analyse des données, il contribue à la transparence de la recherche et la reproductibilité des méthodes.

Ces buts peuvent être atteints en réfléchissant au design d’interface.

8.1.1 Avantages des interfaces graphique pour la recherche

Une interface peut s'adapter à des utilisateur.rices aux profils variés. Cela implique notamment de prendre en compte les questions de multilinguisme (assurer *a minima* le bilinguisme français et anglais) et de littératie numérique. Comparée à des scripts “artisanaux”, l’interface graphique présente l’avantage d’offrir une abstraction des complexités techniques, permettant aux chercheur.ses de se concentrer sur l’interprétation des résultats plutôt que sur la mise en œuvre des algorithmes. De plus, elle assure une cohérence méthodologique et facilite la collaboration en proposant un environnement de travail uniifié. Cette approche permet donc de gagner en efficacité et en reproductibilité, tout en démocratisant l’utilisation des outils de vision par ordinateur au sein de la communauté scientifique.

Ainsi le *back-office* doit se doter d’interfaces ergonomiques et faciles à prendre en main. En effet, le public de chercheur.ses présente des profils différents en terme de compétences numériques, c’est pourquoi il faut veiller à l’intuitivité des fonctionnalités.

“Programmer ou savoir utiliser les logiciels ? Au cours d’une expérience d’une base de données collaborative associée à un projet de d’analyse d’images, les interfaces et la base de données ont été créées par l’informaticien, avec les contraintes d’ergonomie et après conseil des historien.nes. Cela a bien fonctionné. La conclusion à en tirer est que les chercheur.ses doivent d’abord connaître les logiciels développés pour eux, pas nécessairement savoir programmer. Des logiciels puissantes existent, il n’est pas toujours nécessaire de réinventer la roue. L’ignorance informatique des historien.nes doit être évitée, mais il est nécessaire de disposer d’une bonne culture informatique, autorisant un bon apprentissage des logiciels.”²

Une documentation a été rédigée pour aider les chercheur.ses à prendre la plateforme en main : elle est présentée en annexe D.

8.1.2 Construction logique de l’interface utilisateur.rice

La plateforme propose deux grands axes logiques de navigation, correspondant à la gestion des documents d’un côté et à l’application des traitements de vision artificielle de l’autre. Elle s’adapte alors aux utilisateur.rices qui souhaitent utiliser l’IA comme à ceux qui ont essentiellement besoin d’un outil de gestion de données.

Cette séparation se reflète sur la page d’accueil présentant une hiérarchie visuelle claire (deux blocs séparés ‘Entités’ et ‘Mes Tâches’). Cette structure répétée dans la *navbar* du *header* présentant deux menus déroulant distincts : ‘Manage Documents’ et ‘Computer Vision Tasks’.

2. F. Clavert, Aurelien Berra et Franziska Heimburger, « L’historien programmeur ? », dans *THATCamp Paris 2012 : Non-actes de la non-conférence des humanités numériques*, Code : THATCamp Paris 2012 : Non-actes de la non-conférence des humanités numériques, Paris, 2012 (La Non-Collection), DOI : 10.4000/books.editionsms.305, p.35.

Donner accès à la base de données : saisir les sources et les consulter

La plateforme AIKON est conçue pour agréger des données moissonnées à partir de multiples SI. La structure de l'interface d'administration, pensée pour accompagner le chercheur.se dans la récupération des données et le peuplement de la base, est calquée sur le modèle de données sous-jacent.

Le processus d'ajout de nouvelles entités (*Witness, Serie, Work*) est effectué via des formulaires conçus de manière à guider l'utilisateur.rice dans la saisie des informations, permettant d'assurer la cohérence et la qualité des métadonnées. Ces formulaires, directement liés à la structure des tables de la base de données, proposent des champs prédéfinis et des listes déroulantes ou de l'autocomplétion pour normaliser les valeurs. Quand il rentre les métadonnées des entités principales, le chercheur.se a accès à de nouveaux formulaires pour remplir les tables de relation (lieux, langues, personnes, rôles...), constituant la liste des valeurs normalisées. Garantir la cohérence et la qualité de la donnée implique tout de même de mettre en place des règles de saisie strictes, éventuellement rassemblées dans une documentation utilisateur.rice pour favoriser leur suivi et leur respect.

Le processus de saisie des métadonnées constitue en effet une étape cruciale dans la construction et la valorisation d'une base de données de recherche. Ces données descriptives, associées aux objets numériques ou physiques, permettent d'en décrire précisément toutes les caractéristiques (type, format, date de création, langue, auteurs, lieu, institution de conservation, etc.) ; ce afin de faciliter ensuite la trouvabilité. Des métadonnées riches et structurées permettent par la suite de mettre en place des mécanismes de recherche plus performants pour les chercheur.ses.

Les formulaires permettent aussi de décrire les liens entre les entités. Le formulaire de saisie du *Witness* permet d'importer une ou plusieurs *Digitizations*, et de renseigner un ou plusieurs *Contents*. Le *Content* est associé à un *Work* préalablement décrit et enregistré, et à un ou plusieurs acteurs historiques. De cette manière, un *Work* peut être relié à plusieurs *Witnesses* et un *Witness* à plusieurs contenus donc plusieurs *Works*. L'entité *Serie* concerne plutôt le monde de l'imprimé, elle est reliée à plusieurs *Volume* d'une *Édition* et le formulaire permet d'importer une *Digitization* par volume.

Chacune des entités principales dispose aussi d'une page qui liste l'ensemble de ses enregistrements, permettant une circulation libre, sur le mode de la découvrabilité. L'affichage de la liste des *Witnesses*, *Series* et *Works* met en avant un ensemble de métadonnées liées à chaque élément saisi. Cette organisation par entité permet de conserver la cohérence avec la base de données relationnelle. Dans cette logique, les pages de listes agrégeant les données conduisent aux notices d'enregistrement unique, permettant de compléter ou modifier leurs métadonnées.

Depuis ces interfaces, des fonctionnalités de recherche ont été mises en place pour filtrer les résultats à l'aide de plusieurs critères. Les formulaires de recherche amènent un

autre mode de navigation, celui de l'accès direct à la ressource.

List of series

| Work | Volume | Publication place | Dates | Publisher | Conservation place |
|------------------------------|--|-------------------|-----------|-----------------|---|
| 徐光啓 崇禎曆書 Chong zhen li shu | ▶ 黃赤道距度表 Huang chi dao ju du biao Sinica 897 日躔曆指 Ri chan li zhi Sinica 897 交食表 Jiao shi biao Sinica 897 | Beijing | 1629-1637 | 徐光啓 (1562-1633) | Oxford Bodleian Libraries, University of Oxford |

FIGURE 8.1 – Capture d'écran du formulaire de recherche des *Series*.

Le *backend* administrateur constitue en ce sens une interface graphique pour la base de données, les administrateurs pouvant rentrer les données, accéder aux ressources qu'ils ont ajoutées, parcourir celles ajoutées par d'autres, et ajuster les métadonnées.

Traitements et Résultats

La plateforme permet la sélection des entités dans les listes, et ainsi la constitution de *Document Sets* dont le contenu est consultable et modifiable dans une interface dédiée et accessible depuis la barre de navigation. Ce sont sur ces ensembles de documents que peuvent être appliqués des traitements de nature variée (export, recherche de similarité, vectorisation etc.), déclenchés par l'utilisateur.rice. Le formulaire de lancement est accessible directement sur l'interface de consultation du *set*, ainsi que via barre de navigation du header. Le traitement d'un document unique passe aussi par la création d'un *Document Set*.

Une fois un traitement effectué, l'utilisateur.rice reçoit un e-mail avec le lien vers une page de résultats, encore en construction à l'été 2024. Elle permettra de comparer les résultats d'un même traitement (par exemple l'extraction) en fonction de différents paramètres (typiquement en fonction de différents modèles).

Les résultats des traitements sont aussi accessibles à l'échelle d'un *Witness*, via sa cartouche descriptive dans la page de liste. Un bouton mène à une interface qui traite la donnée à un niveau de granularité inférieur, le diagramme – ou l'illustration découpée. L'utilisateur.rice navigue entre l'affichage des régions extraites, des similarités et des vectorisations dans différents onglets.

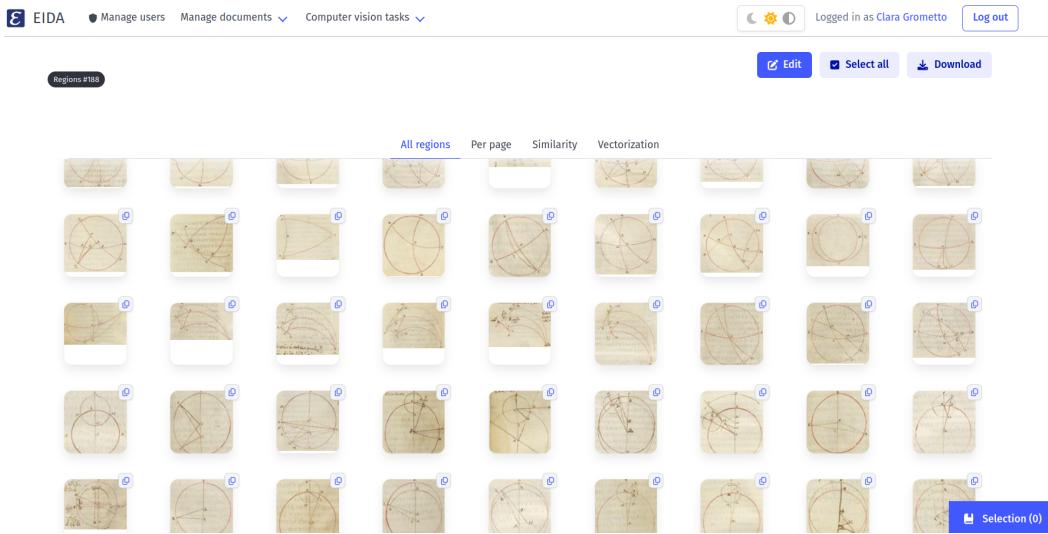


FIGURE 8.2 – Capture d’écran de la visualisation des extractions type *dump*, les onglets permettent de naviguer entre les interfaces de visualisation des résultats des traitements.

Pour l’extraction, deux visualisations sont proposées : un affichage synthétique de toutes les régions extraites (Fig. 8.2) et une représentation qui les remet en contexte, présentant en regard la page de manuscrit originale. C’est via ces visualisations que se fera la sélection des *Region Sets* (Fig. 8.3).

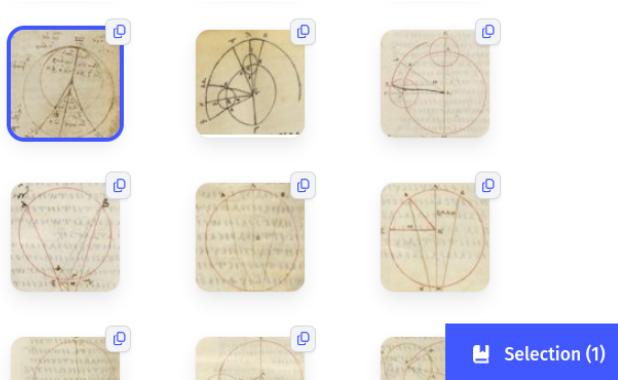


FIGURE 8.3 – Sélection des *Region Sets*.

Un onglet donne accès à une interface pour explorer les similarités calculées par l’algorithme. Elle liste les témoins comparés. Sans sélection, seules les zones extraites du témoin requête apparaissent. L’utilisateur.rice peut afficher les similarités calculées pour un ou plusieurs témoins, avec possibilité de filtrer selon les catégories assignées par les chercheur.ses. Les 10 meilleurs scores, recalculés dynamiquement, sont affichés par ordre décroissant. Un système de seuil pour masquer les images avec un score bas, pourra également être implémenté pour améliorer encore la lisibilité des résultats. Actuellement, les similarités avec un score inférieur à 25 sont déjà exclues pour réduire le temps de calcul.

Le nombre de résultats affichés pourrait aussi être ajusté. En résumé, cette interface est pensée pour optimiser l'exploration et l'analyse des similarités détectées par le modèle, grâce à des filtres superposables.

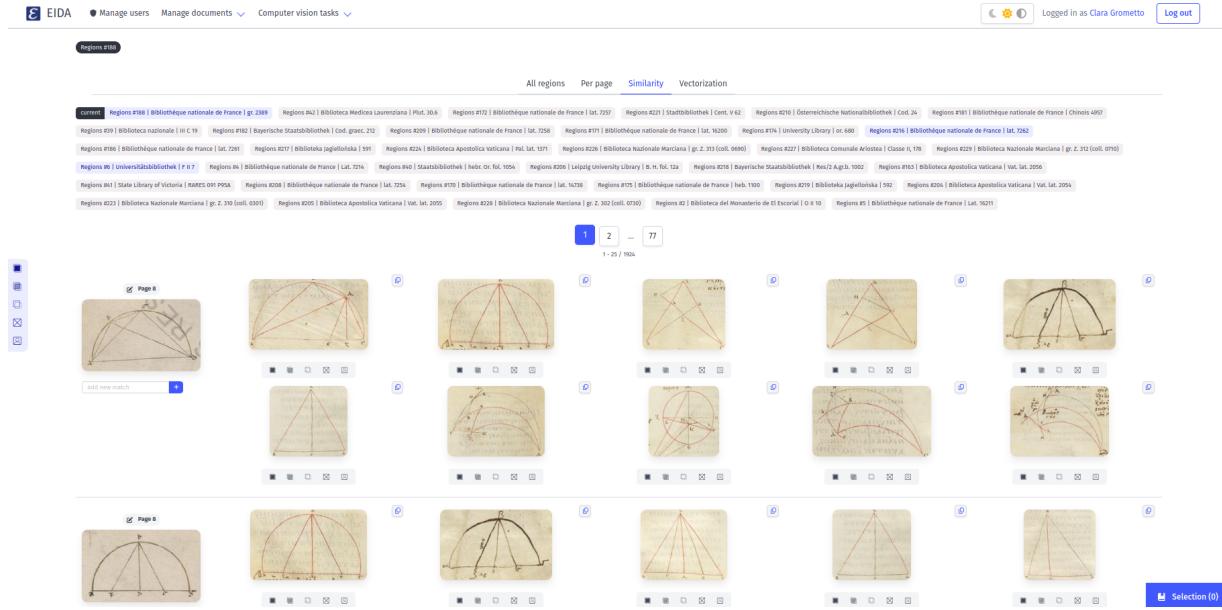


FIGURE 8.4 – Capture d'écran de l'interface de visualisation des similarités.

La visualisation actuelle des vectorisations (Fig. 8.5) permet une exploration préliminaire des données, mais reste sommaire (simple superposition du SVG à l'image JPEG). Des améliorations sont envisagées, notamment pour implémenter l'export et la sélection en vue de l'établissement des éditions.

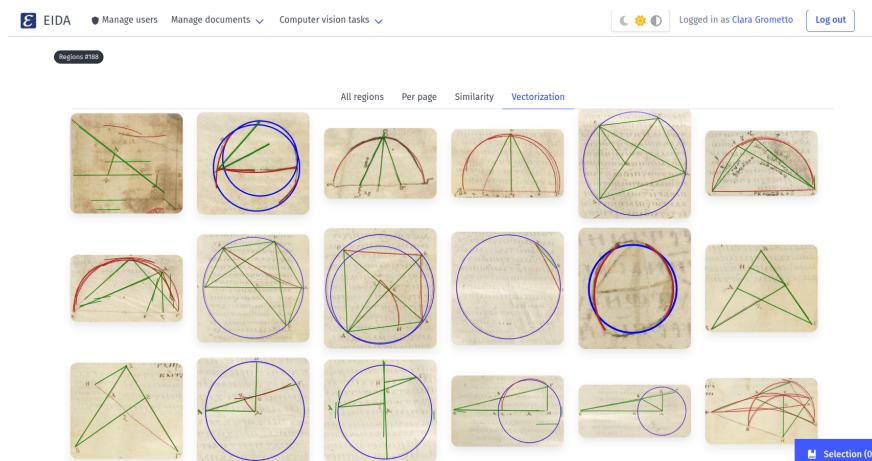


FIGURE 8.5 – Capture d'écran de l'interface d'affichage des vectorisations.

Anoter et exporter

S'inspirant des fonctionnalités proposées par la plateforme e-scriptorium pour le texte, AIKON vise à offrir un environnement de travail complet pour la sémantification et l'enrichissement des documents numérisés présentant des éléments visuels. Grâce à la plateforme, les utilisateur.rices peuvent donc non seulement effectuer des traitements automatiques sur leurs données et les corriger, mais aussi exporter les résultats dans différents formats interopérables, et donner accès au *Manifest IIIF* annoté. Les premières données de prédiction (*silver data*), soumises à une validation humaine, viennent alors enrichir des corpus d'entraînement (*gold data*), permettant ainsi d'affiner progressivement les modèles d'apprentissage profond et d'améliorer la précision des résultats, alimentant une boucle de rétroaction entre l'homme et la machine.

À ce titre, une attention particulière est portée à la vérification manuelle des résultats. Sur l'interface de visualisation des similarités, un champ est prévu pour ajouter manuellement celles qui auraient pu être omises par les algorithmes. L'identifiant de l'extraction, copiable depuis l'onglet de visualisation, est utilisé comme référence à verser dans le champ de saisie.

L'évaluation du calcul de similarité entre images repose sur cinq critères d'annotation prédéfinis (Fig. 8.6), permettant une qualification plus précise du concept. Ces catégories, associés à chaque image similaire, constituent des métadonnées importantes pour les chercheur.ses et servent de filtre pour la visualisation des résultats. Mais surtout, l'ensemble formé par les paires d'images annotées forme un corpus d'entraînement pour améliorer les résultats du modèle.

| | |
|--|----------------|
| | Exact Match |
| | Partial Match |
| | Semantic Match |
| | No Match |
| | Personal Match |

FIGURE 8.6 – Récapitulatif des catégories de similarité et leurs icônes associées.

Il est par ailleurs possible de supprimer les extractions non pertinentes depuis l'interface de visualisation de celles-ci (Fig. 8.7). L'utilisateur.rice peut aussi accéder au visualiseur Mirador en cliquant sur la page de manuscrit se trouvant en regard, afin d'y

ajouter des annotations : diagrammes ou autre objet d'intérêt que l'algorithme aurait manqué.

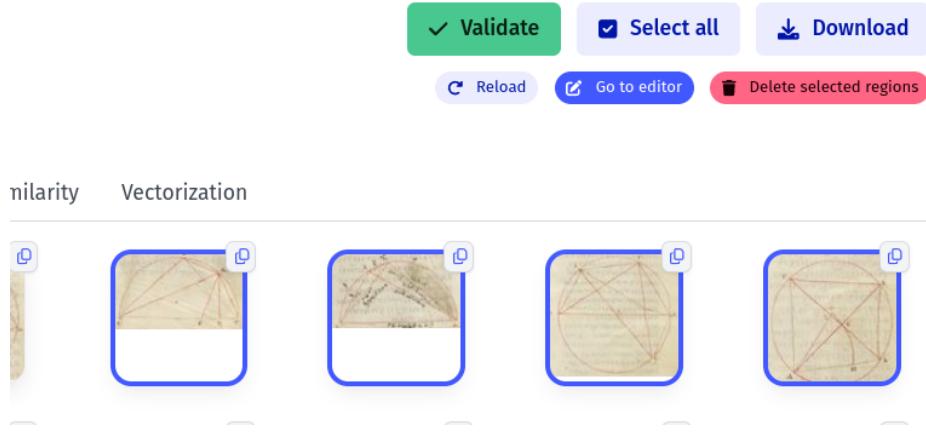


FIGURE 8.7 – L’interface permet la validation des extractions.

La correction des prédictions s’inscrit donc dans un processus itératif qui permet de construire des bases de données annotées et d’améliorer les performances des modèles de classification, de détection, ou les calculs de similarités. La prochaine étape consiste à implémenter un éditeur SVG, intégrant des fonctionnalités de création et de modification de ce format.

En conclusion, le *back-office* d’une application web pour la gestion des données est un outil qui peut améliorer l’efficacité, la qualité et la transparence de la recherche. Il offre aux chercheur·ses un environnement de travail collaboratif pour gérer les données et les intégrer dans une chaîne de traitement basée sur des outils de vision par ordinateur. Le parcours dans la plateforme est pensé pour guider les processus de saisie et de traitement des données, et pour l’exploitation des résultats.

La communauté scientifique accorde désormais une importance primordiale à la visibilité et la réutilisabilité des bases de données³, un enjeu majeur se situe donc dans l’ouverture et l’accessibilité de celles-ci. La conception de l’interface utilisateur·rice s’inscrit alors dans la perspective de son ouverture au public. Une fois la base de données constituée et enrichie, elle a vocation à être exposée. L’implémentation d’APIs assurera l’interopérabilité avec d’autres plateformes et facilitera les échanges. En outre, la perspective de l’exposition et la médiation des données impacte les choix techniques concernant l’interface.

3. « L’Open Data a-t-il tenu ses promesses ? », dans ADEMEC, Paris, France, 2023.

8.2 Choix techniques

“Qu’elle suive des normes ou pas, la réflexion sur l’interface, sur le rendu graphique d’un projet de recherche devrait être présente dès l’origine du processus. Cela demande, comme pour la gestion des métadonnées, une attitude en apparence paradoxale mais nécessaire pour mener un projet de recherche aujourd’hui : l’anticipation des problématiques de valorisation dès l’amont du projet.”⁴

Cette affirmation souligne l’importance d’une interface utilisateur.rice intuitive et performante, non seulement pour favoriser l’adoption d’un outil par la communauté scientifique, mais aussi pour prévoir la valorisation des résultats de recherche. Les décisions techniques liées à l’interface répondent ainsi à un double objectif : fournir aux chercheur.ses un environnement de travail, et faciliter la diffusion et la médiation des données de la recherche auprès d’un public de non-spécialistes, anticipant la création d’une plateforme publique.

8.2.1 Interface Admin Django vs. vues personnalisées

Une fonctionnalité puissante et intéressante de Django est son interface d’administration automatique, qui a servi d’interface à EIDA dans un premier temps. Cette interface lit les métadonnées du modèle pour fournir rapidement un outil de gestion centrée sur celui-ci. Via cette interface *backend* les utilisateur.rices administrateurs peuvent gérer les contenus du site. Cependant, son utilisation est limitée à des fins de gestion interne et ne suffit pas pour servir de base à la création d’une interface distribuable ou publique. Cette fonctionnalité est pertinente si une interface centrée sur les tables et les champs de base de données est nécessaire. Cependant, pour une interface qui fait abstraction des détails d’implémentation de la base de donnée, centrée davantage autour des processus, l’interface d’administration automatique de Django devient insuffisante et il faut construire des vues personnalisées.

La construction d’une vue dans Django passe par l’écriture d’un template HTML dans lequel le python injectera les données structurées par une fonction définie dans le fichier `view.py`. Le fichier `urls.py` est aussi un composant essentiel du *framework*. Il sert à définir les mappages entre les URLs des requêtes entrantes et les vues correspondantes. En d’autres termes, il gère le routage des requêtes HTTP en les dirigeant vers les fonctions qui traiteront ces requêtes, faisant elles-mêmes l’intermédiaire entre les données et leur affichage.

Écrire des vues dans Django sans utiliser de *framework front-end* génère donc directement le HTML à partir des *templates*. Les interactions utilisateur.rice sont gérées par des requêtes HTTP traditionnelles. Ce mode de développement, simple et rapide à mettre en place, est particulièrement adapté pour des applications avec des interactions limitées

4. O. Jacquot, *Décrire, transcrire et diffuser un corpus documentaire hétérogène...*

ou des besoins modérés en termes de dynamisme. Dans l'interface EIDA, il a permis la création de vue personnalisées et adaptées aux visualisation des sorties de l'inférence avec les modèles d'extraction, de similarité et de vectorisation. La description de certains de ces développements est présentée en annexe C.

Au mi-terme de mon stage, a été prise la décision de refondre les interfaces et d'adopter un *framework front-end*, offrant une expérience utilisateur.rice plus fluide grâce à une gestion différente des interactions entre le *back-end* et le *front-end*.

8.2.2 Utilisation d'un *framework front*

“La structure de l’application ainsi que l’architecture client/serveur conditionnent la manière dont est traitée l’information. Ainsi, pour assurer aux internautes une expérience de navigation agréable au sein du front office, les problématiques de vitesse de chargement doivent influencer directement la manière dont sont pensés les développements. S’il est préférable d’effectuer la manipulation de données dans la partie back-end de la plateforme, il est cependant nécessaire de savoir repenser les méthodes de requêtage et de gestion des données, afin d’offrir aux utilisateur.rices des interfaces optimales et ergonomiques.”⁵

Pour les applications web utilisant le protocole HTTP lors d'une action de l'utilisateur.rice, le contenu de la vue est recalculé puis envoyé au client. Or, pour la manipulation d'ensembles de données lourds dans une même interface, il est intéressant d'avoir recours à des méthodes de programmation asynchrone au sein du *front-end*, de façon à réduire les durées de chargement à l'ouverture de la page.

Les *frameworks front-end* sont des collections de ressources (fichiers, bibliothèques, structures de code, etc.) permettant de créer une interface utilisateur.rice plus dynamique et réactive, essentiellement pour les applications nécessitant une manipulation de contenus visuels, offrant ainsi une expérience utilisateur.rice plus fluide et agréable. Ils sont conçus pour optimiser les performances en gérant efficacement le DOM (*Document Object Model*) et en permettant des mises à jour en temps réel sans recharger la page entière. De plus, grâce à la virtualisation des données et à la manipulation asynchrone, ils permettent de charger et de rendre les données lourdes de manière progressive, réduisant ainsi les temps de latence et améliorant la réactivité globale. Par ailleurs, ils vont faciliter l'adoption de normes et standards pour le développement de la partie frontale, donnant un rendu plus uniifié.

Pour comprendre le fonctionnement d'un tel outil, une digression s'impose à propos de l'architecture client/serveur, sur laquelle repose les applications web.

5. S. Albouy, *Médiation des données de la recherche...*, p.97.

Paradigme client/serveur ; *back-end/front-end* ; SPA

L'architecture client/serveur définit le mode de communication entre deux programmes : le client, qui envoie des requêtes, et le serveur, qui y répond. Dans une plate-forme en ligne, le serveur envoie les pages affichées sur le navigateur du client, et le client (utilisateur.rice) envoie des requêtes au serveur en agissant sur l'interface.

Cette architecture se reflète dans la distinction entre le développement *back-end* et *front-end*. Le *back* concerne la partie du code invisible à l'utilisateur.rice (par exemple, en Python pour EIDA), tandis que le *front* concerne la partie visible, influençant directement ce qui est affiché à l'écran (en HTML, CSS et JavaScript). Le code Python est exécuté sur le serveur, tandis que le HTML, CSS et JavaScript sont interprétés par le navigateur du client.

Les méthodes de traitement des données varient : le serveur, souvent plus puissant, exécute plus rapidement les procédures lourdes en Python. C'est pourquoi les données sont formatées en *back-end* avant de les transmettre eu *front-end*. Toutefois, ce mode de traitement des ressources peut être optimisé pour certaines situations. Les méthodes dites asynchrone effectuent des traitement en *front-end*, afin de réduire les temps de chargement et de fluidifier l'expérience utilisateur.rice. JavaScript récupère des données du serveur sans entraver l'exécution du reste du code. Cette technique permet au navigateur de continuer à fonctionner de manière fluide et réactive, même lorsqu'il attend une réponse du serveur. Lorsqu'une requête asynchrone est envoyée, JavaScript utilise des APIs telles que Fetch ou XMLHttpRequest pour interagir avec le serveur en arrière-plan. Une fois la réponse du serveur reçue, une fonction de rappel (callback) ou une promesse (promise) traite les données retournées sans avoir interrompu les opérations en cours. Cette approche est largement utilisée par les applications web modernes, permettant une mise à jour dynamique du contenu et une meilleure expérience utilisateur.rice, en évitant les temps d'attente et les blocages qui pourraient survenir avec des requêtes synchrones traditionnelles : on les nomme SPA (*Single Page Application*), par opposition aux MPA (*Multiple Page Application*) classiques.

Une SPA est une application web qui charge une seule page HTML et met à jour dynamiquement le contenu au fur et à mesure que l'utilisateur.rice interagit avec l'application. Lors du chargement initial, le serveur envoie les fichiers HTML, CSS et JavaScript nécessaires. La page HTML chargée inclut des zones réservées (comme des éléments div) où le contenu sera remplacé dynamiquement. Les SPA utilisent généralement des *frameworks* JavaScript tels que React, Angular, Vue.js ou Svelte qui gèrent le rendu du contenu directement dans le navigateur, sans dépendre du côté serveur. À mesure que l'utilisateur.rice interagit avec l'application (par exemple, en cliquant sur des liens ou des boutons), le *framework* intercepte ces actions et charge de nouveaux contenus ou données de manière asynchrone. Ce processus se fait souvent en effectuant des appels API pour récupérer des

données d'un serveur ou d'une base de données. Lorsque de nouvelles données sont récupérées, le *framework* met à jour le DOM de manière dynamique sans recharger la page entière (Fig.8.8).

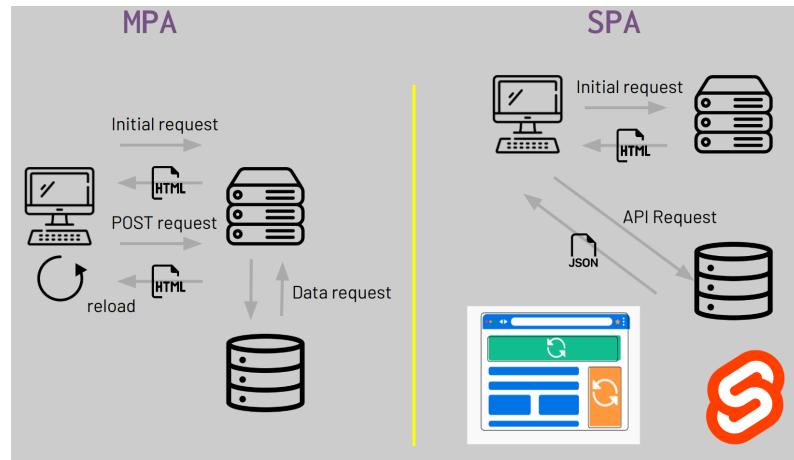


FIGURE 8.8 – Fonctionnement schématique d'une SPA.

Leur mode de fonctionnement donnent aux SPA rapidité et fluidité : après le chargement initial, elles ne récupèrent et ne rendent que le contenu qui doit être mis à jour, plutôt que de recharger toutes les ressources chaque fois que l'utilisateur.rice interagit avec l'application.

AIKON n'est pas une SPA. C'est une application classique qui donne à certaines pages le fonctionnement d'une SPA, pour permettre plus d'interaction en conservant fluidité et performance. Cela est possible grâce à l'utilisation de requêtes asynchrones, permises par l'adoption d'un *framework front-end* : Svelte, ajouté de manière incrémentielle à la base de code existante.

Svelte

Pour gérer efficacement un volume de contenu visuel et éviter le temps de latence au chargement de certaines pages, AIKON s'appuie donc sur des requêtes asynchrones et sur Svelte.js, choisi pour ses performances.

Une page construite avec Svelte est à considérer comme un ensemble de *Components* (composants), unités de code réutilisables et auto-suffisantes. Les composants sont écrits en utilisant la *Svelte syntax*, un langage propre à Svelte : il reprend les trois langages de base HTML, CSS et JavaScript, mais sans les diviser. On les retrouve tous les trois dans les mêmes fichiers (en gardant une base de HTML : c'est avant tout un langage de balisage). Le développement est alors moins verbeux, et le langage est plus simple et lisible, et donc plus maintenable.

Svelte est davantage un compilateur qu'une *Library*, c'est ce qui le distingue et optimise ses performances. Ce nouveau paradigme permet une répartition optimisée du

travail entre le serveur et le client. Contrairement à d'autres *frameworks front-end*, avant que le code Svelte ne soit chargé dans le navigateur, il passe par une étape de compilation en JavaScript pur côté serveur, lors de la phase dite de *build*. Le code Svelte, écrit avec sa syntaxe propre, est transformé en JavaScript *vanilla*, très performant⁶. Le code exécuté dans le navigateur est léger, rapide et optimisé, puisqu'il n'y a pas de *framework* lourd à exécuter ; et l'expérience utilisateur.rice en est d'autant plus fluide.

L'étape de compilation permet aussi d'intégrer la réactivité directement dans le langage. Chaque variable déclarée dans un composant Svelte est, par défaut, un objet réactif. Lorsqu'une nouvelle valeur est affectée à cette variable (via `=`, `+=`, `-=`, ou `$`⁷), Svelte ne se contente pas d'effectuer une simple assignation. Il surcharge les opérateurs d'un appel à une sorte de fonction, un effet de bord, généré par le compilateur, qui va détecter les dépendances (identifier tous les éléments du DOM ou autres expressions qui dépendent directement ou indirectement de cette variable), et mettre à jour le DOM (planifier une mise à jour de ces éléments afin d'afficher le nouvel état à l'écran).

“Updating some count value — and all the things that depend on it — should be as simple as this :

```
count += 1;
```

Since we're a compiler, we can do that by instrumenting assignments behind the scenes :

```
count += 1;
$$invalidate('count', count);
```

Importantly, we can do all this without the overhead and complexity of using proxies or accessors. It's just a variable.”⁸

Cette approche permet de se passer de DOM virtuel. Le *virtual DOM* est un objet JavaScript, une copie de l'arbre des composants (le vrai DOM, celui qui est interprété et rendu dans le navigateur), utilisé par le plupart des autres *framework* front comme React.js, Angular.js ou Vue.js. À chaque trame de rendu, le *framework* compare un à un les paramètres de chaque nœud de l'arbre des composants entre la copie de l'ancien (*virtual DOM*) et le nouveau DOM. S'il y a un paramètre à changer, le *framework* ne rafraîchit le rendu que des nœuds qui en dépendent. Mais cette comparaison des nœuds un à un est coûteuse. Svelte suit directement les dépendances entre valeurs uniquement grâce au langage. Lorsque une donnée change, Svelte sait exactement quels éléments de l'interface

6. Le JavaScript *vanilla*, ou JavaScript pur, désigne l'utilisation du langage JavaScript sans recourir à bibliothèques ou *frameworks* externes. C'est la forme de JavaScript la plus fondamentale, directement intégrée dans les navigateurs web, permettant d'écrire des scripts légers et performants pour interagir avec le DOM et réaliser diverses fonctionnalités côté navigateur.

7. L'opérateur `$` assigne une variable à une autre non déclarée, à cette dernière Svelte ajoutera automatiquement une déclaration `let`, et entre les deux se crée une dépendance, un lien durable. Où que la première variable change dans le code, le changement sera détecté et affectera la seconde. Svelte crée ainsi une sorte d'arbre des dépendances insensible à l'ordre des assignations.

8. Rich Harris, *Svelte 3 : Rethinking reactivity*, en, avr. 2019, URL : <https://svelte.dev/blog/svelte-3-rethinking-reactivity> (visité le 16/08/2024).

doivent être mis à jour, sans avoir à parcourir l'ensemble de l'arbre des composants. Ce paradigme permet plus d'efficience et améliore la réactivité.

Les composants Svelte sont conçus pour être autonomes et indépendants. Cependant, il est fréquent de devoir partager des données entre plusieurs composants. C'est là qu'interviennent les *stores*. Un *store*, en Svelte, est un objet indépendant de la hiérarchie des composants (il est typiquement déclaré en dehors de tout composant, dans un fichier séparé), qui contient les variables avec leur état, et permet leur partage entre différents *Components*. Il fonctionne comme un dépôt centralisé où la donnée est stockée et mise à jour. Les composants qui ont besoin d'accéder aux données d'un store s'y "abonnent". Lorsque la valeur d'une variable dans un store est modifiée, le changement est automatiquement propagé et tous les composants abonnés sont mis à jour, affichant alors le nouvel état des données à l'écran. L'état de l'application est géré de manière centralisée, ce qui améliore la lisibilité du code.

Dans le cas d'AIKON, où l'interface utilisateur.rice doit gérer de manière dynamique un grand nombre d'images et de filtres complexes, Svelte s'est avéré un choix judicieux. La capacité de Svelte à mettre à jour l'interface en temps réel, en réponse aux actions de l'utilisateur.rice, comme l'application de filtres sur les résultats de recherche de similarité, permet de conserver la fluidité dans l'interaction. Les multiples critères de filtrage – témoins à comparer, catégories de similarité – et leur interdépendance – les catégories à afficher reposent sur la catégorisation apposée à chaque image – rendent l'interface particulièrement dynamique, et Svelte permet la gestion de cette complexité.

Essentiellement, Svelte reprend l'idée d'une SPA pour développer toute ou une partie d'une interface. Il propose cependant une alternative intéressante aux *frameworks frontend* traditionnels. En compilant le code à la volée et en générant du JavaScript pur, Svelte offre un optimisation du temps de latence, notamment pour le développement d'applications web qui restent réactives malgré le volume de données.

8.2.3 Le design d'interface

La dernière étape de l'UX/UI *design* consiste à travailler sur le rendu visuel du produit afin que l'interface soit claire, lisible, attrayante et simple d'utilisation. Il s'agit par exemple de choisir la bonne couleur afin de rendre un bouton visible, choisir la bonne police pour plus de lisibilité, etc.

Un *framework CSS* est une bibliothèque pré-conçue de fichiers CSS qui permettent de simplifier et standardiser le développement de la mise en page et du design des sites web. Ces *frameworks* fournissent des outils, des styles, et des composants réutilisables pour créer des interfaces utilisateur.rice de manière plus efficace et cohérente. Notamment, ils simplifient l'obtention d'un résultat *responsive*, s'adaptant à différents mode d'affichage et plusieurs écrans. Pour le développement de la plateforme, Bulma a été utilisé.

En outre, l'utilisation des fonctionnalités avancées comme les fonctions ou les variables personnalisables, permettent aux développeur.es de créer des styles plus dynamiques, et surtout réutilisables, ce qui est particulièrement pertinent dans le cadre du développement collaboratif afin d'aligner la charte graphique.

Les variables personnalisées sont définies à l'aide du préfixe -- et peuvent être réutilisées dans tout le document CSS. Elles permettent de centraliser la gestion des couleurs et d'autres valeurs répétitives. Par exemple, il est possible de définir une couleur que prendront les éléments au survol grâce à son code hexadécimal ou les valeurs RGB, en lui assignant un nom (--on-hover). La variable sera alors réutilisable avec la même facilité que les variables de couleur intégrées dans Bulma (comme --bulma-danger-light).

Il est aussi possible d'assigner des variables par défaut :

```
:root {
    --bulma-body-family: "Fira Sans", sans-serif;
    --default-color: hsl(229, 68%, 46%);
    --accent-grey: hsl(0, 0%, 14%);
```

Listing 8.1 – Créations de variables CSS personnalisées.

Utiliser la pseudo-classe :root pour définir des variables CSS garantit qu'elles sont accessibles partout dans le document, en utilisant la fonction var(), comme ci-dessous :

```
.navbar-link:hover {
    background-color: inherit;
    color: var(--default-color);
}
```

Listing 8.2 – Utilisation de variables CSS personnalisées.

Pour aller plus loin dans la standardisation, il est possible de définir des styles thématiques.

```
[data-theme="light"] {
    --contrasted: #f3eff1;
    --accent: hsl(0deg 0% 0% / 5%);
    --selected: hsl(234deg 100% 96%);
    --selected-text: hsl(237deg 82% 32%);
    --deleted: var(--bulma-danger-light);
    --deleted-text: var(--bulma-danger-dark);
}

[data-theme="dark"] {
    --contrasted: #252424;
    --accent: hsla(0, 0%, 100%, 0.05);
    --selected: #000528;
```

```
--selected-text: #7281f9;
--deleted: var(--bulma-danger-dark);
--deleted-text: var(--bulma-danger-light);
```

Listing 8.3 – Définition d'un thème sombre et d'un thème claire grâce aux classes CSS.

Dans l'exemple ci-dessus, des variables personnalisées sont utilisées pour définir un mode clair et un mode sombre. Les styles sont appliqués en fonction de l'attribut `data-theme` sur un élément HTML (`<body data-theme="light">` ou `<body data-theme="dark">`). Cela permet de basculer facilement du mode sombre au mode clair, sans redéfinir les modalités du changement dans chaque élément affecté.

Les *media queries* sont des techniques CSS qui permettent d'appliquer des styles en fonction des caractéristiques du périphérique sur lequel le contenu est affiché, comme la largeur de l'écran, la résolution, ou dans le cas illustré ci-dessous, la préférence du schéma de couleurs de l'utilisateur.rice.

```
@media (prefers-color-scheme: dark) {}
@media (prefers-color-scheme: light) {}
```

Listing 8.4 – Utilisation des *media queries*.

En somme, le design d'interface rend l'expérience de l'utilisateur.rice plus agréable en prenant en compte ces préférences d'affichage (mode ou taille d'écran) et en offrant une identité visuelle cohérente. Il contribue à l'adoption des outils au sein des équipes de recherche. Par ailleurs, l'utilisation d'un *framework CSS* rationalise le développement *front-end*, simplifiant le travail de l'équipe de développement, qui peut alors se concentrer sur les fonctionnalités clés plutôt que sur des aspects purement esthétiques.

On aura voulu montrer dans cette section l'importance du travail des interfaces, qui font le lien entre le chercheur.se et la donnée d'une part, et entre le chercheur.se et des pratiques d'autre part. Plus que de simples ornements, des interfaces performants impactent l'analyse des données, ainsi que l'adhésion et l'efficacité des utilisateur.rices. Elles constituent le pont entre les chercheur.ses, souvent peu familiers des bases de données, et les données complexes qu'ils produisent. Une interface permet de faciliter la navigation entre les sources, même si elles sont hétérogène, et permet aux chercheur.ses de retrouver facilement les données dont ils ont besoin, grâce à des fonctionnalités de recherche pertinentes. L'utilisation de formulaires de saisie réduit les erreurs et permet une gestion normalisée des métadonnées, favorisant la trouvabilité et la réutilisation des données par d'autres chercheur.ses. Des outils de visualisation leur permettent une exploration de leurs données centrée sur l'élément graphique, unité de base du modèle.

L'interface permet aussi de guider le chercheur.se dans un protocole. Unifier l'accès aux données et aux modèles de vision par ordinateur implique d'élaborer des *workflows*, facilitant la mise en œuvre de méthodes standardisées. L'interface facilite leur prise en main, favorisant ainsi la cohérence des méthodes au sein des équipes, ou entre plusieurs équipes de recherche (à l'instar de l'annotation des prédictions, destinées à être réutilisés pour l'entraînement des modèles). Une interface bien conçue peut donc faciliter notamment le travail en équipe, améliorant le partage des pratiques.

Afin de répondre à ces exigences, le travail de front est clé. La plateforme AIKON a été pensée pour proposer des interfaces performantes pour les fonctionnalités de recherche et de visualisation diverses. Ce travail pour rendre la plateforme réactive et accueillante anticipe en outre la valorisation et la médiation des données de la recherche vers un public plus large (via la construction d'une plateforme publique). Les choix techniques sont faits dans ce sens, l'utilisation d'un *framework front-end* permettant d'améliorer les performances et de garantir une fluidité d'interaction, tandis que l'adoption d'un *framework CSS* assure une uniformité et une cohérence visuelle, dotant la plateforme d'une identité graphique forte.

Conclusion partielle

Le travail des chercheur.ses sur de grandes quantités de données nécessitent le développement d'outils adaptés. La performance des modèles d'IA ne suffit pas : il faut les rendre opérationnels au sein des environnements de recherche. Là se situe l'enjeu de la plateforme AIKON, dont l'objectif est de rendre compatibles les outils techniques et les pratiques des ss. Elle a pour ambition de façonner un véritable SI incluant les outils de *computer vision* pour systématiser les traitements, des fonctionnalités de gestion documentaire, le tout en préservant le rôle central de l'interprétation humaine.

L'objectif est aussi de concevoir une plateforme adaptable à une multitude de projets de recherche en études visuelles. En adoptant une architecture modulaire, divers projets peuvent l'exploiter comme une ‘coquille’, en personnalisant les protocoles selon leurs besoins spécifiques. Cette flexibilité contribue à la pérennité et l'évolution de l'outil.

À l'heure actuelle, l'application est à déployer sur des serveurs personnels. Cependant, l'objectif à long terme est de développer une plateforme en ligne, à l'exemple d'e-Scriptorium, dédiée à l'analyse automatique de documents multimodaux. En offrant des fonctionnalités de correction des traitement et d'exploration de résultats, elle permettra la création de corpus enrichis. Et en rationalisant les méthodes de travail des historiens, elle ouvrira de nouvelles perspectives en matière de collaboration et de partage des outils comme des données. Se conformant aux principes FAIR et aux normes internationales d'interopérabilité, elle pourra faciliter la migration des données vers d'autre systèmes (via des API), et permettre ainsi les comparaisons avec d'autres corpus. Elle favoriserait ainsi son intégration dans des écosystèmes de recherche plus larges.

Conclusion

“Les outils façonnent la pensée. Ce que nous pouvons penser et ce que nous pouvons dire résulte d'une dynamique dans laquelle les outils et les techniques jouent un rôle fondamental.”⁹

Epron et Vitali-Rosati avancent l'idée selon laquelle les outils ne sont pas de simples instruments passifs, mais des acteur dans la construction de la pensée. Ils ne se limitent pas à faciliter des tâches, mais façonnent une conceptualisation du monde et des idées. Les outils de la recherche, qu'ils soient numériques ou analogiques, structurent des gestes intellectuels. Les outils numériques ne sont pas de simples remédiations des outils traditionnels. Ils offrent de nouvelles possibilités, de nouvelles façons d'interagir avec l'information et de la manipuler. Ils peuvent catalyser les dynamiques collaboratives autour de ces nouvelles méthodes et ces nouvelles perceptives. Un cadre technique contient un cadre conceptuel, et il y a donc un enjeu à forger des outils de traitement de la donnée qui vont dans le sens du partage des pratiques.

Ce travail s'est attaché à exposer l'importance et les moyens de partager le développement et l'utilisation des outils numériques de la recherche, en prenant pour exemple l'application du *deep learning* à la sémantification et l'enrichissement des sources en histoire de l'astronomie. Les questionnements, impliquant la modélisation, l'accès à la donnée, et l'élaboration des modèles de *computer vision*, s'orientent vers la conception d'un SI dédié à des documents numérisés variés et hétérogènes, et au traitement de leurs éléments visuels. Les outils de *deep learning* sont mis à disposition via une plateforme web modulaire et réemployable dans différents contextes (AIKON).

La première partie de ce mémoire, qui portait sur le contexte disciplinaire des projets EIDA/VHS, a mis en évidence la complexité des données traitées, ainsi que la diversité des besoins des chercheur.ses. En conséquence, une modélisation de données flexible et interopérable, ne renonçant pas aux exigences de description des sources, a constitué un axe de réflexion. Par ailleurs, la variété des modes d'accès aux données est une piste d'analyse importante. Les données visuelles et les besoins spécifiques des historien.nes, notamment l'accès à leurs propres images stockées localement, nécessitent des solutions techniques adaptées. Bien que IIIF offre un cadre prometteur pour l'échange de données volumineuses, il ne constitue pas une solution universelle.

9. B. Epron et M. Vitali-Rosati, *L'édition à l'ère numérique...*, p.31.

Dans un second temps, nous avons présenté sur les enjeux et défis liés à l'utilisation du *deep learning* et de la *computer vision* dans le traitement des données historiques, notamment les moyens de gérer la faible disponibilité de corpus annotés. Nous nous sommes ensuite penchés sur une application des traitements d'IA, en dressant les exigences fonctionnelles d'un outil d'édition des SVGS (sorties de l'algorithme de vectorisation) intégré à la plateforme. Cet outil a vocation à fédérer les pratiques des chercheur.ses pour l'édition des diagrammes astronomiques.

La troisième partie est consacrée aux spécificités techniques de la conception d'une plateforme modulaire destinée à héberger les outils de gestion de données et les instruments de *computer vision*. Cette plateforme présente une architecture modulaire, reposant sur une séparation claire des fonctionnalités dans des modules et des composants applicatifs distincts. L'inférence des modèles, qui exige une puissance de calcul importante, est optimisée par l'utilisation d'un GPU sur lequel tourne une API prévue à cet effet. Par ailleurs, le choix d'une bibliothèque *front-end* permet de créer des interfaces utilisateur.rice performantes, facilitant ainsi l'interaction des chercheur.ses avec un processus de travail itératif, alternant phases de traitement par les modèles de vision et corrections afin de garantir la qualité des résultats. Ces derniers constituent des corpus annotés qualitatifs qui pourront être réutilisés dans le cadre de l'entraînement des modèles.

La modularité des outils : pourquoi ?

Cette étude de cas a montré l'importance de l'accessibilité et du partage au sein de la communauté de la recherche, non seulement des données, mais aussi des outils qui vont permettre de les traiter. Cette démarche favorise à la fois l'évolutivité et la pérennité des outils numériques, permet d'établir des pratiques communes au sein de la communauté scientifique et, enfin, démocratise l'accès à des outils innovants qui ouvrent de nouvelles perspectives d'analyse des sources.

Les outils taillés trop spécifiquement sur un projet ou sur une question de recherche sont fragiles. Leur cycle de vie est intimement lié aux financements, ce qui les expose à l'obsolescence dès lors que les ressources s'épuisent. De tels développements, souvent coûteux en temps et en moyens (financiers comme humains), ne garantissent pas la pérennité des outils. Pour assurer leur survie et favoriser leur évolution, il est alors intéressant de s'inscrire dans des projets collaboratifs, et en conséquence, d'adopter une approche modulaire. En fédérant les efforts et en mutualisant les ressources, il est possible non seulement de stimuler l'innovation mais aussi de garantir la maintenance à long terme de ces outils, les rendant ainsi plus robustes et pérennes.

L'ouverture des outils et l'extensivité des chaînes de traitement des données a aussi des implications scientifiques d'importance. Ces aspects inscrivent une démarche scientifique dans l'optique de la collaboration et du partage des pratiques, favorisant leur reproductibilité, et renforçant la transparence des méthodes. On l'aura évoqué dans le chapitre 6, la diversité des pratiques d'édition scientifique des diagrammes présents dans

les traités d'histoire des sciences témoigne de la richesse des points de vue sur les sources, mais peut aussi constituer un obstacle à la coopération au sein du champ de recherche. La multiplication des outils, des méthodologies et des normes entraîne une fragmentation des pratiques et remet en question l'utilité scientifique des contenus produits. Chaque chercheur.se ou chaque équipe développe ses propres outils, avec ses propres formats, et ses normes adaptées à leur angle d'approche épistémologique et heuristique. Devant ce constat, l'élaboration d'un outil d'édition numérique peut alors porter un cadre de travail partageable.

S'adapter à l'hétérogénéité des données et des contextes matériels ne favorise pas seulement la collaboration et la cohérence des pratiques au sein d'un champ de recherche, mais facilite également l'accès aux outils d'IA pour l'analyse à grande échelle des corpus, ouvrant ainsi de nouvelles perspectives sur les sources.

La plateforme AIKON et son interface charpentent alors un véritable environnement de recherche numérique, dédié à l'analyse des sources, orienté vers l'interprétation par les chercheur.ses. En centralisant les ressources et en offrant des fonctionnalités d'annotation et de visualisation, cet outil fait office de nouveau laboratoire, et change la manière dont les chercheur.ses interagissent avec les données historiques, permettant au domaine des études visuelles d'exploiter le paradigme du *big data*. Elle est conçue pour s'adapter au traitement de documents numérisés divers, pour exploiter les éléments graphiques qu'ils contiennent.

Un outil généraliste et adaptable à plusieurs projets de recherche : comment ?

Le cas AIKON a en outre permis d'identifier les principaux enjeux liés à la mise en œuvre de la modularité. Nous nous sommes interrogés sur les fondements d'un outil suffisamment polyvalent et offrant une base solide pour permettre sa spécialisation.

Premièrement, la modélisation de la donnée est un aspect à considérer. Se référer à des modèles conceptuels standards est alors essentiel pour disposer de concepts larges et adaptables aux données du patrimoine. Si le modèle de données adopté n'est pas taillé spécifiquement pour décrire les sources de EIDA, il est néanmoins suffisamment précis pour répondre aux besoins actuels des chercheur.ses, et bénéficie de sa généralité pour s'adapter aux sources de VHS et à une grande diversité de types de documents, permettant l'étude des éléments graphiques qu'ils contiennent.

Le domaine de recherche de la *computer vision*, dans son essence même, illustre en outre les dynamiques collaboratives dans la construction des instruments de traitement de la donnée. En partant de modèles génériques pré-entraînés, il est possible de construire des architectures complexes spécialisées sur des tâches et des données particulières. Pour assurer la cohérence et la réutilisabilité des données, il est important d'encourager l'utilisation de vocabulaires contrôlés et de normes pour l'annotation des données.

Une attention particulière a été portée aux interfaces de la plateforme AIKON, pour faciliter le travail des chercheur.ses et la diffusion de données. Un outil doté d'une interface

graphique accueille un vaste éventail d'utilisateur.rice et anticipe en outre la médiation des résultats des recherches vers un public plus large.

D'autre part, le recours à des protocoles standards et interopérables, tels que le IIIF, assure la compatibilité des systèmes. Les formats de sortie des traitements sont aussi des formats libres et manipulables : TXT, Numpy ou SVG.

La personnalisation s'exprime par ailleurs dans le fait qu'AIKON peut être déployée dans différents environnements matériels, la possibilité d'installation en locale démontrant sa légèreté et sa portabilité. Mais les problématiques liées à l'ouverture et la modularité du code concernent aussi la conception d'une architecture qui puisse s'adapter à une augmentation du volume de données et du nombre d'utilisateur.rices. Ainsi les besoins côté ingénierie vont parfois rentrer en conflit avec cette volonté de légèreté si importante pour inclure un maximum de contextes de recherche et d'utilisateur.rices différents. La gestion de volumes de données croissants ou des outils d'IA plus lourds requiert du matériel et des architectures applicatives plus robustes que celles développées avec les moyens des laboratoire de recherche en SHS. Pour l'instant, AIKON est taillé pour – et utilisé sur – des corpus qui, bien que larges, restent limités comparés aux collections des bibliothèques par exemple. Un passage à l'échelle reposera sur l'implication de consortiums ou d'institutions, qui disposent des technologies et des infrastructures capables de gérer de grandes quantités de données de manière efficace. Nous conclurons et ouvrirons sur cette idée : la recherche constitue le terreau de l'innovation ; néanmoins, son passage à l'échelle institutionnelle exige des compétences en ingénierie qui transcendent ceux des projets de recherche, même collaboratifs.

Annexes

Annexe A

Évolution du modèle de données

A.1 Modèle de données initial de l'application VHS

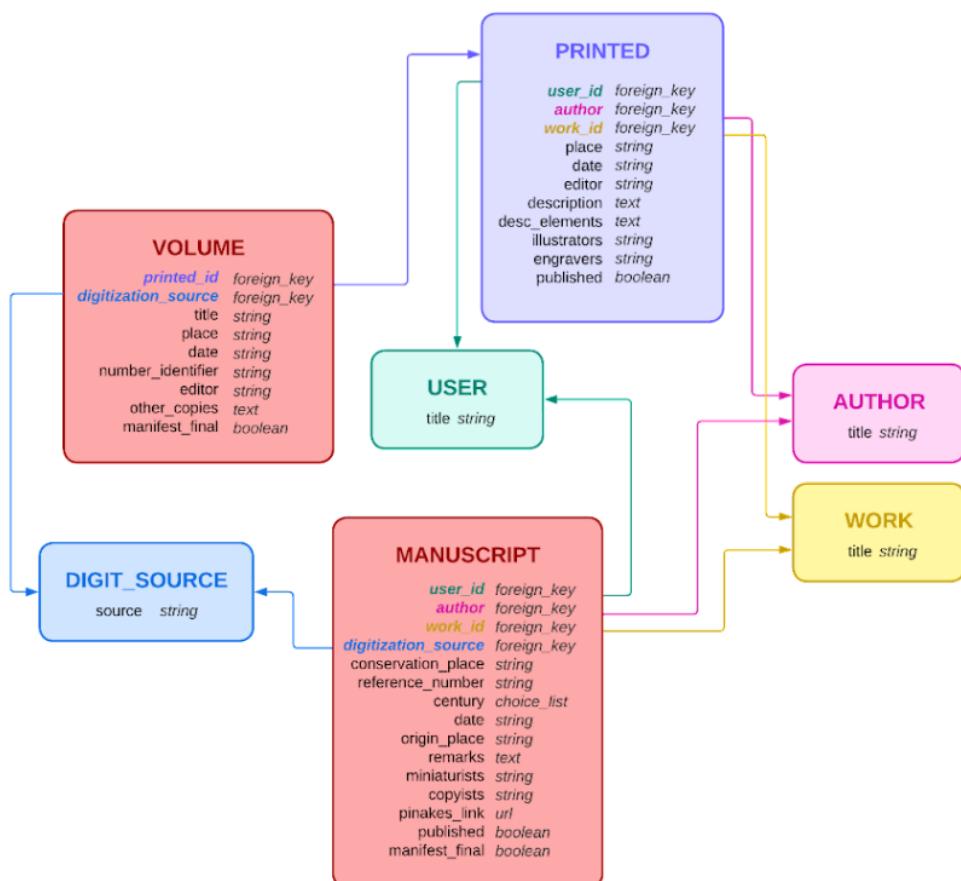


FIGURE A.1 – Modèle de données de l'application VHS avant refonte.*

A.2 Modèle de données de l'application VHS/EIDA

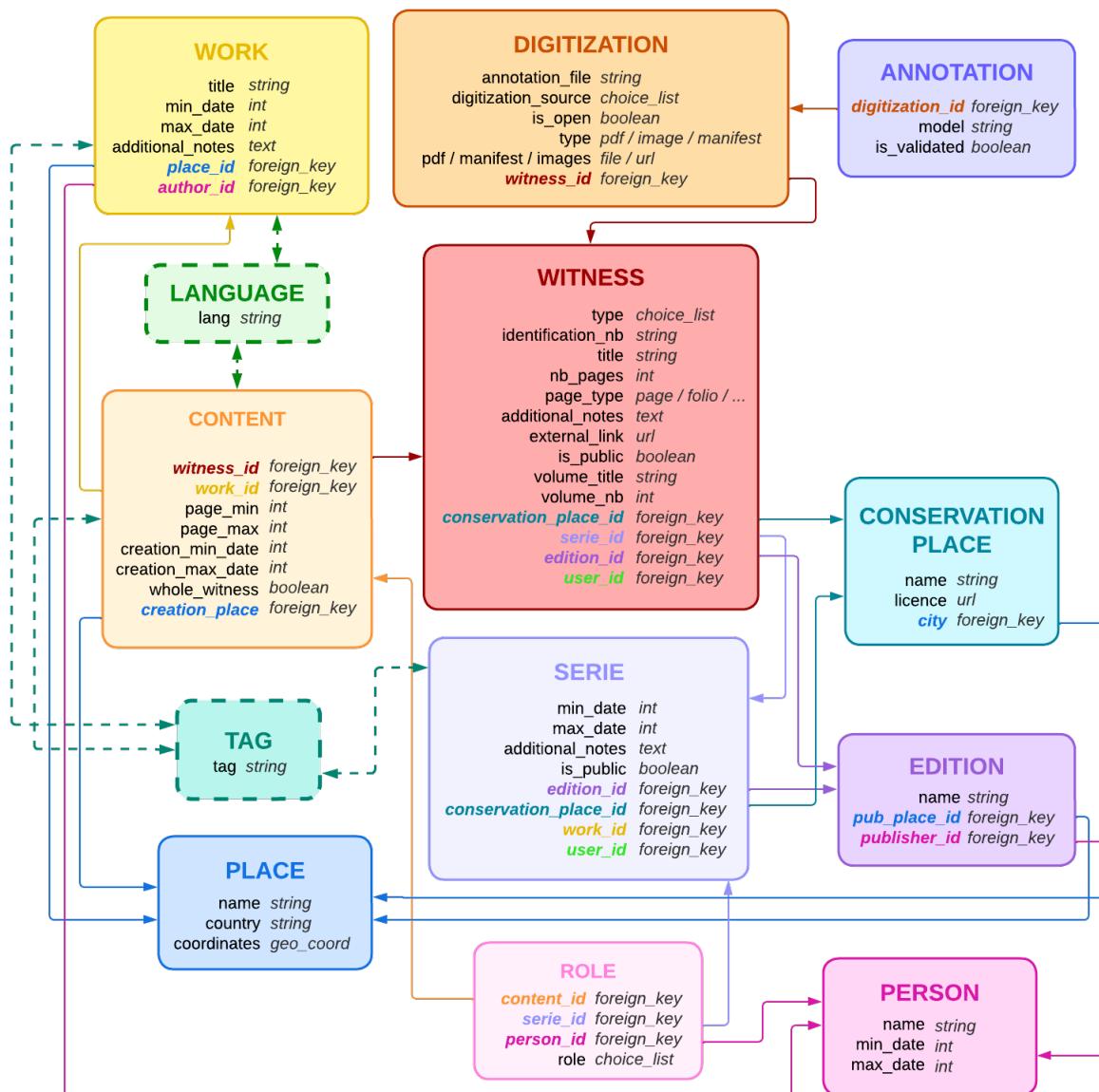


FIGURE A.2 – Nouveau modèle de données de EIDA appliqu   à VHS.*

A.3 Modèle de données de l'application AIKON

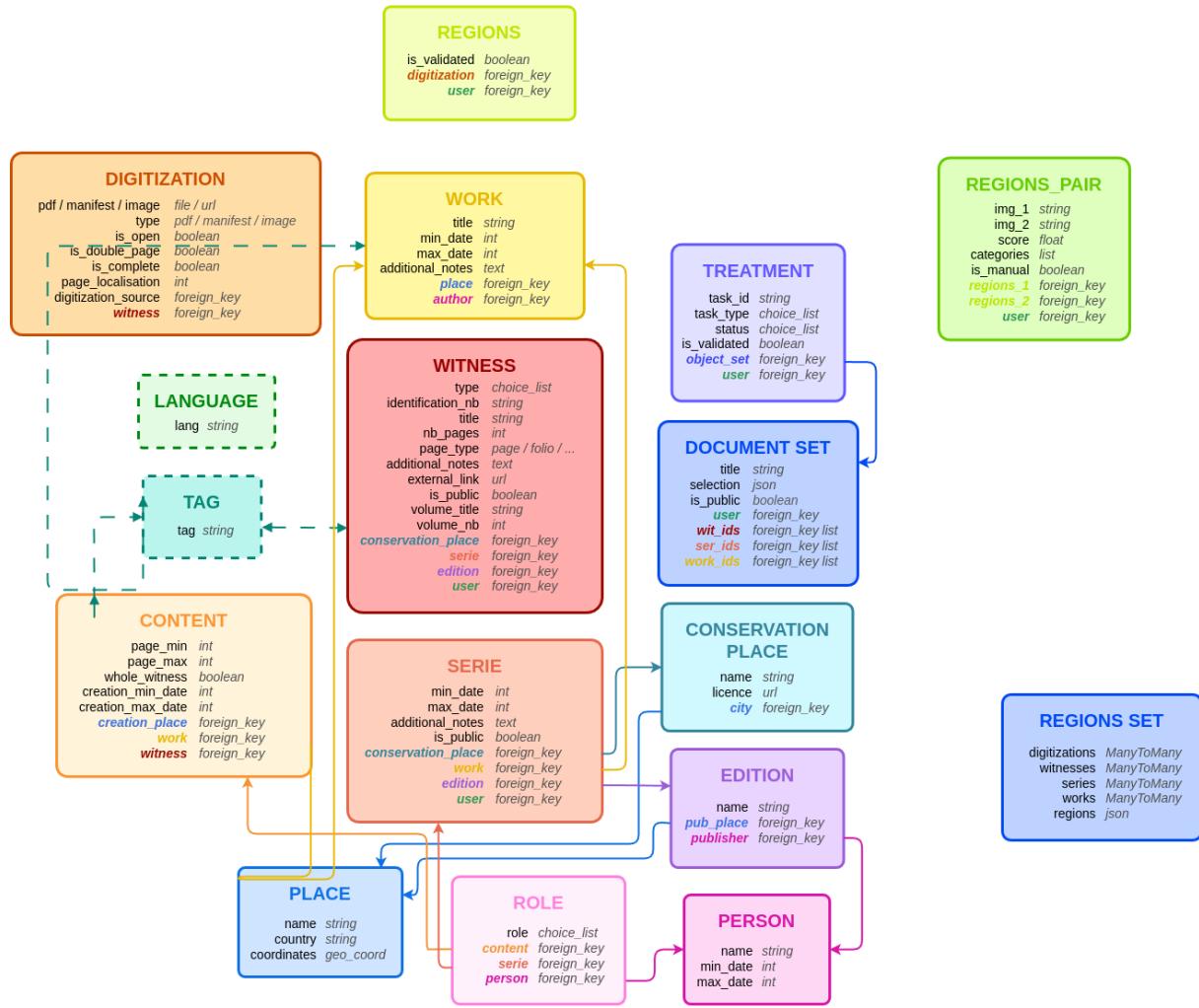


FIGURE A.3 – Nouveau modèle de données en vue de la refonte de l'application AIKON.*

Annexe B

Module vectorisation : description des développements

Les méthodes développées pendant mon stage ont posé les bases d'un *workflow*, qui a très rapidement évolué. Ces méthodes ont ensuite été itérativement raffinées pour orienter la plateforme vers plus de modularité, notamment grâce à l'implémentation de l'instance *Treatment*, sur laquelle repose désormais le lancement des actions. Cette annexe présente le développement du *workflow* de lancement de la vectorisation les *Witnesses*, ainsi que son évolution.

B.1 Workflow

- Requête utilisateur.rice dans l'application AIKON ;
- Lancement d'un *Treatment* :
 - Identification des *Witnesses* ;
 - Parser les *Manifests* et ramener les URLs des régions d'images dans une liste ;
 - Création d'un fichier JSON
 - Envoie du JSON via requête HTTP POST à l'API *endpoint* ;
- Inférence du modèle dans Discover-Demo :
 - Vérification du JSON ;
 - Lancement d'une tâche ;
 - Enregistrement des images ;
 - Inférence avec le modèle : écriture des fichiers SVGS ;
 - Envoi d'un ZIP contenant les fichiers à l'application via requête HTTP POST ;
- Récupération du ZIP au *endpoint* de l'application ;
 - Dézip, lecture et écriture des SVGS en *backend* dans l'application AIKON ;

- Vérification dans le répertoire de résultats (dossier dans `mediafiles`) que deux fichiers du même nom n'existent pas ;
- S'il existe déjà un fichier du même nom, il est écrasé par le nouveau ;
- Apparition des visualisations dans l'interface ;

B.2 Settings

B.2.1 AIKON

Le fichier de configuration (`.env`) de l'application doit contenir l'URL de l'API à laquelle elle se connecte. Afin d'activer les fonctionnalités de vectorisation, il est aussi nécessaire de spécifier le module correspondant dans les paramètres de configuration.

```
# Computer vision apps to install
ADDITIONAL_MODULES=regions,similarity,vectorization
```

Listing B.1 – Extrait du fichier de configuration de l'application.

B.2.2 Discover-Demo

Dans le fichier `.env`, la variable `INSTALLED_APPS` contient la liste des modules à charger. Pour activer les fonctionnalités, notamment la vectorisation, il est nécessaire d'ajouter les modules correspondants à cette liste.

```
# apps (folder names) to be imported to the API
INSTALLED_APPS=dticlustering,watermarks,similarity,region,vectorization
```

Listing B.2 – Extrait du fichier de configuration de l'API.

B.3 Envoi d'un traitement de vectorisation

B.3.1 Initialisation de la requête

Le processus de vectorisation était initialement déclenché via l'interface d'administration Django, en utilisant le mécanisme des `@admin.action` (des fonctions appelées avec une liste d'objets sélectionnés depuis la page de liste pour modification, elles agissent donc au niveau du témoin). Cette approche, bien que fonctionnelle, présentait des limites : notamment concernant le suivi des traitements et leur application sur différents objets de la base.

Une refonte des processus a été entreprise par Jade Norindr, Ségolène Albouy et Fouad Aouinti. Cette évolution a conduit à la création d'un formulaire dédié au lancement

de tous les traitements, standardisant ainsi les modes de communication avec l’API. Ce formulaire, accessible depuis l’interface utilisateur.rice, permet d’appliquer le traitement sur un ensemble d’objets précédemment sélectionnés.

Add new treatment

Task type
vectorization

Document set
Search...

Notify by email
Send an email when the task is finished

Submit

FIGURE B.1 – Capture d’écran du formulaire d’envoi d’un traitement dans AIKON.

Le lancement de la vectorisation se fait donc désormais au niveau de l’entité *Treatment* reliée à un ensemble de témoins. Le formulaire de lancement permet de créer une instance de cette entité, et par conséquent, tous les témoins associés seront soumis au processus de vectorisation.

B.3.2 Création du json

Initialement la fonction utilitaire `vectorization_request_for_one` était utilisée pour formater un fichier JSON à envoyer au *endpoint* de l’API. Afin de traiter par lots un ensemble de témoins, la fonction `vectorization_request` permettait d’itérer sur une liste de témoins, en appelant récursivement la fonction `vectorization_request_for_one` pour chacun d’entre eux.

```
try:
    response = requests.post(
        url=f"{CV_API_URL}/vectorization/start",
        json={
            "doc_id": regions.get_ref(),
            "model": f"{VECTO_MODEL_EPOCH}",
            "images": get_regions_urls(regions),
            "callback": f"{APP_URL}/{APP_NAME}/get-vectorization",
        },
    )
```

Listing B.3 – Extrait de la fonction gérant l’envoi de JSON à l’API dans `app/vectorization/utils.py`

Avec les récents développements, le JSON utilisé pour communiquer avec l'API est généré par la fonction `prepare_request`, présente dans le fichier `utils.py` de chaque module.

La boucle externe itère sur les *Witnesses* et vérifie qu'il n'y a pas déjà de vectorisation effectuée. La boucle interne itère sur les annotations des *Witnesses*. Le dictionnaire `regions_dic` est créé pour mapper les références des témoins annotés aux URLs des annotations. La fonction renvoie un dictionnaire contenant les données nécessaires pour la requête de vectorisation.

```
def prepare_request(witnesses, treatment_id):
    regions_list = []
    regions_dic = {}

    try:
        for witness in witnesses:
            if witness.has_vectorization():
                log(
                    f"[vectorization_request] Witness {witness.get_ref()} already
                     has vectorizations"
                )
                pass
            else:
                regions_list.extend(witness.get_regions())

        if regions_list:
            for regions in regions_list:
                regions_dic.update({regions.get_ref(): get_regions_urls(regions)
                                    })

    return {
        "experiment_id": f"{treatment_id}",
        "documents": regions_dic,
        "model": f"{VECTO_MODEL_EPOCH}",
        "callback": f"{APP_URL}/{APP_NAME}/get-vectorization", # URL to
                    which the SVG zip file must be sent back
        "tracking_url": f"{APP_URL}/{APP_NAME}/api-progress",
    }

else:
    return {
        "message": f"No regions to vectorize for all the selected {WIT}es
                   "
    }
```

```

        if APP_LANG == "en"
        else f"Pas de regions à vectoriser pour tous les {WIT}s
              sélectionnés"
    }

except Exception as e:
    log(
        f"[prepare_request] Failed to prepare data for vectorization request
         ",
        e,
    )
    raise Exception(
        f"[prepare_request] Failed to prepare data for vectorization request
         "
    )

```

Listing B.4 – Structurer les URLs de régions d’images.

La fonction reçoit une liste de *Witnesses*, générée par une tâche appelée par une méthode `post_save` de l’instance de *Treatment* :

```

# vhs/app/webapp/models/treatment.py
@receiver(post_save, sender=Treatment)
def treatment_post_save(sender, instance, created, **kwargs):
    if created:
        get_all_witnesses.delay(instance)

```

Listing B.5 – Méthode `post_save` du *Treatment*.

```

# vhs/app/webapp/tasks.py
@celery_app.task
def get_all_witnesses(treatment):
    try:
        witnesses = treatment.get_witnesses()
        treatment.start_task(witnesses)
    except Exception as e:
        treatment.on_task_error(
            {
                "error": f"Error when retrieving documents from set: {e}",
                "notify": treatment.notify_email,
            },
        )

```

Listing B.6 – Ramener tous les *Witnesses* à partir des entités reliées aux *Treatment* et lancer de la tâche.

B.4 Discover-Demo : module vectorization

B.4.1 Réception à l'*endpoint* de l'API :

L'*endpoint* (dans `api/app/vectorization/routes.py`) sert de passerelle pour lancer la tâche de vectorisation. Il valide la requête entrante, extrait les paramètres nécessaires et transfère la tâche au processus en arrière-plan pour son exécution.

```
@blueprint.route("start", methods=["POST"])
@shared_routes.get_client_id
@shared_routes.error_wrapper
def start_vectorization(client_id):
    """
    A list of images to download + relevant data
    """
    if not request.is_json:
        return "No JSON in request: Vectorization task aborted!"

    json_param = request.get_json()
    console(json_param, color="cyan")

    experiment_id = json_param.get("experiment_id")
    documents = json_param.get("documents", {})
    model = json_param.get("model", None)

    notify_url = json_param.get("callback", None)
    tracking_url = json_param.get("tracking_url")

    return shared_routes.start_task(
        compute_vectorization,
        experiment_id,
        {
            "documents": documents,
            "model": model,
            "notify_url": notify_url,
            "tracking_url": tracking_url,
        },
    )
```

Listing B.7 – *Endpoint start_vectorization*.

B.4.2 Configuration de la tâche de fond

Dans le fichier `api/app/vectorization/tasks.py`, la fonction `compute_vectorization` définit un acteur Dramatiq qui exécute la tâche de vectorisation en *background*. Elle initialise une nouvelle instance de la classe `LoggedComputeVectorization`, en lui passant le logger et les paramètres fournis. Elle appelle la méthode `run_task` sur l'objet créé, qui va lancer le processus.

```
@dramatiq.actor(time_limit=1000 * 60 * 60, max_retries=0, queue_name=VEC_QUEUE)
def compute_vectorization(
    experiment_id: str,
    documents: dict,
    model: str,
    notify_url: Optional[str] = None,
    tracking_url: Optional[str] = None,
    logger: TLogger = LoggerHelper,
):
    """
    Run vecto task on lists of URL
    """

    vectorization_task = LoggedComputeVectorization(
        logger,
        experiment_id=experiment_id,
        documents=documents,
        model=model,
        notify_url=notify_url,
        tracking_url=tracking_url
    )
    vectorization_task.run_task()
```

Listing B.8 – Tâche `dramatiq compute_vectorization`.

B.4.3 inférence avec le modèle

La classe `LoggedComputeVectorization` gère les processus de vectorisation en utilisant les paramètres fournis. La méthode `run_task` lance sur l'instance le téléchargement des données, l'inférence du modèle, le traitement des résultats (renvoi des SVGs sous forme de ZIP) et la notification, avec journalisation tout au long du processus.

Dans le fichier `api/app/vectorization/lib/vectorization.py` :

```
class ComputeVectorization:
    def __init__(self,
```

```

        experiment_id: str,
        documents: dict,
        model: Optional[str] = None,
        notify_url: Optional[str] = None,
        tracking_url: Optional[str] = None,
    ):
        self.experiment_id = experiment_id
        self.documents = documents
        self.model = model
        self.notify_url = notify_url
        self.tracking_url = tracking_url
        self.client_id = "default"
        self.imgs = []

    def run_task(self):
        pass

    def check_dataset(self):
        if len(list(self.documents.keys())) == 0:
            return False
        return True

    def task_update(self, event, message=None):
        if self.tracking_url:
            send_update(self.experiment_id, self.tracking_url, event, message)
            return True
        else:
            return False

class LoggedComputeVectorization(LoggingTaskMixin, ComputeVectorization):
    def run_task(self):
        if not self.check_dataset():
            self.print_and_log_warning(f"[task.vectorization] No documents to
                download")
            self.task_update("ERROR", f"[API ERROR] Failed to download documents
                for vectorization")
            return

        error_list = []

```

```

try:
    for doc_id, document in self.documents.items():
        self.print_and_log(
            f"[task.vectorization] Vectorization task triggered for {doc_id} !"
        )
        self.task_update("STARTED")

        self.download_dataset(doc_id, document)
        self.process_inference(doc_id)
        self.send_zip(doc_id)

        self.task_update("SUCCESS", error_list if error_list else None)

except Exception as e:
    self.print_and_log(f"Error when computing vectorizations", e=e)
    self.task_update("ERROR", "[API ERROR] Vectorization task failed")

def download_dataset(self, doc_id, document):
    self.print_and_log(
        f"[task.vectorization] Dowloading images...", color="blue"
    )
    for image_id, url in document.items():
        try:
            if not is_downloaded(doc_id, image_id):
                self.print_and_log(
                    f"[task.vectorization] Downloading image {image_id}"
                )
                download_img(url, doc_id, image_id)

        except Exception as e:
            self.print_and_log(
                f"[task.vectorization] Unable to download image {image_id}", e
            )

def process_inference(self, doc_id):
    model_folder = Path(MODEL_PATH)
    model_config_path = f"{model_folder}/config_cfg.py"
    epoch = DEFAULT_EPOCHS if self.model is None else self.model
    model_checkpoint_path = f"{model_folder}/checkpoint{epoch}.pth"
    args = SLConfig.fromfile(model_config_path)

```

```

args.device = 'cuda'
args.num_select = 200

corpus_folder = Path(IMG_PATH)
image_paths = glob.glob(str(corpus_folder / doc_id) + "/*.jpg")
output_dir = VEC_RESULTS_PATH / doc_id
os.makedirs(output_dir, exist_ok=True)

model, criterion, postprocessors = build_model_main(args)
checkpoint = torch.load(model_checkpoint_path, map_location='cpu')
model.load_state_dict(checkpoint['model'])
model.eval()

args.dataset_file = 'synthetic'
args.mode = "primitives"
args.relative = False
args.common_queries = True
args.eval = True
args.coco_path = "data/synthetic_processed"
args.fix_size = False
args.batch_size = 1
args.boxes_only = False
vslzr = COCOVisualizer()
id2name = {0: 'line', 1: 'circle', 2: 'arc'}
primitives_to_show = ['line', 'circle', 'arc']

torch.cuda.empty_cache()
transform = T.Compose([
    T.RandomResize([800], max_size=1333),
    T.ToTensor(),
    T.Normalize([0.485, 0.456, 0.406], [0.229, 0.224, 0.225])
])

with torch.no_grad():
    for image_path in image_paths:
        try:
            self.print_and_log(
                f"[task.vectorization] Processing {image_path}", color="blue"
            )
            # Load and process image

```

```

im_name = os.path.basename(image_path)[-4]
image = Image.open(image_path).convert("RGB")
im_shape = image.size
input_image, _ = transform(image, None)
size = torch.Tensor([input_image.shape[1], input_image.shape[2]]))

# Model inference
output = model.cuda()(input_image[None].cuda())
output = postprocessors['param'](output, torch.Tensor([
    im_shape[1], im_shape[0]]).cuda(), to_xyxy=False)[0]

threshold, arc_threshold = 0.3, 0.3
scores = output['scores']
labels = output['labels']
boxes = output['parameters']
select_mask = ((scores > threshold) & (labels != 2)) | (((scores > arc_threshold) & (labels == 2)))
labels = labels[select_mask]
boxes = boxes[select_mask]
scores = scores[select_mask]
pred_dict = {'parameters': boxes, 'labels': labels, 'scores': scores}
lines, line_scores, circles, circle_scores, arcs, arc_scores = get_outputs_per_class(pred_dict)

# Postprocess the outputs
lines, line_scores = remove_duplicate_lines(lines, im_shape, line_scores)
lines, line_scores = remove_small_lines(lines, im_shape, line_scores)
circles, circle_scores = remove_duplicate_circles(circles, im_shape, circle_scores)
arcs, arc_scores = remove_duplicate_arcs(arcs, im_shape, arc_scores)
arcs, arc_scores = remove_arcs_on_top_of_circles(arcs, circles, im_shape, arc_scores)
arcs, arc_scores = remove_arcs_on_top_of_lines(arcs, lines, im_shape, arc_scores)

# Generate and save SVG

```

```

    self.print_and_log(f"[task.vectorization] Drawing {image_path}
        ", color="blue")
    #shutil.copy2(image_path, output_dir)
    #décommenter cette ligne si on veut obtenir les images dans le
    #répertoire de sortie
    diagram_name = Path(image_path).stem
    image_name = os.path.basename(image_path)
    lines = lines.reshape(-1, 2, 2)
    arcs = arcs.reshape(-1, 3, 2)

    dwg = svgwrite.Drawing(str(output_dir / f"{diagram_name}.svg")
        , profile="tiny", size=im_shape)
    dwg.add(dwg.image(href=image_name, insert=(0, 0), size=
        im_shape))
    dwg = write_svg_dwg(dwg, lines, circles, arcs, show_image=
        False, image=None)
    dwg.save(pretty=True)

ET.register_namespace('', "http://www.w3.org/2000/svg")
ET.register_namespace('xlink', "http://www.w3.org/1999/xlink")
ET.register_namespace('sodipodi', "http://sodipodi.sourceforge
.net/DTD/sodipodi-0.dtd")
ET.register_namespace('inkscape', "http://www.inkscape.org/
namespaces/inkscape")

file_name = output_dir / f"{diagram_name}.svg"
tree = ET.parse(file_name)
root = tree.getroot()

root.set('xmlns:inkscape', 'http://www.inkscape.org/namespaces
/inkscape')
root.set('xmlns:sodipodi', 'http://sodipodi.sourceforge.net/
DTD/sodipodi-0.dtd')
root.set('inkscape:version', '1.3 (0e150ed, 2023-07-21)')

arc_regex = re.compile(r'[aA]')
for path in root.findall('{http://www.w3.org/2000/svg}path'):
    d = path.get('d', '')
    if arc_regex.search(d):
        path.set('sodipodi:type', 'arc')
        path.set('sodipodi:arc-type', 'arc')

```

```

        path_parsed = parse_path(d)
        for e in path_parsed:
            if isinstance(e, Line):
                continue
            elif isinstance(e, Arc):
                center, radius, start_angle, end_angle, p0, p1 =
                    get_arc_param([e])
                path.set('sodipodi:cx', f'{center[0]}')
                path.set('sodipodi:cy', f'{center[1]}')
                path.set('sodipodi:rx', f'{radius}')
                path.set('sodipodi:ry', f'{radius}')
                path.set('sodipodi:start', f'{start_angle}')
                path.set('sodipodi:end', f'{end_angle}')

        tree.write(file_name, xml_declaration=True)

    self.print_and_log(f"[task.vectorization] SVG for {image_path} drawn",
                      color="yellow")

except Exception as e:
    self.print_and_log(f"[task.vectorization] Failed to process {image_path}", e)

self.print_and_log(f"[task.vectorization] Task over", color="yellow")
)

def send_zip(self, doc_id):
    """
    Zip le répertoire et envoie ce répertoire via POST à l'URL spécifiée.
    """
    try:
        output_dir = VEC_RESULTS_PATH / doc_id
        zip_path = output_dir / f"{doc_id}.zip"
        self.print_and_log(f"[task.vectorization] Zipping directory {output_dir}",
                          color="blue")

        zip_directory(output_dir, zip_path)
        self.print_and_log(f"[task.vectorization] Sending zip {zip_path} to {self.notify_url}",
                          color="blue")

        with open(zip_path, 'rb') as zip_file:

```

```

        response = requests.post(
            url=self.notify_url,
            files={
                "file": zip_file,
            },
            data={
                "experiment_id": self.experiment_id,
                "model": self.model,
            },
        )

        if response.status_code == 200:
            self.print_and_log(f"[task.vectorization] Zip sent successfully
                                to {self.notify_url}", color="yellow")
        else:
            self.print_and_log(f"[task.vectorization] Failed to send zip to {
                                self.notify_url}. Status code: {response.status_code}", color
                                ="red")

    except Exception as e:
        self.print_and_log(f"[task.vectorization] Failed to zip and send
                           directory {output_dir}", e)

```

Listing B.9 – Classe ComputeVectorization.

B.5 Réception des résultats dans AIKON

Pour que les vectorisations soient retournées de l’API à AIKON après l’inférence du modèle, un *endpoint* est créé, dont le routage avec la bonne URL est définie dans le fichier `urls.py` du module.

```

path(
    f"{APP_NAME}/get-vectorization",
    receive_vectorization,
    name="get-vectorization",
),

```

Listing B.10 – Routage de l’*endpoint* pour la réception de vectorisations.

Cet *endpoint* appelle la fonction `receive_vectorization` du fichier `views.py` du module `vectorization`, qui reçoit un fichier .ZIP via une POST request de l’API. Grâce à une fonction utilitaire du fichier `utils.py` (`save_svg_files`), l’archive est extraite, les

fichiers sont lus et écrits dans le dossier `mediafiles`. Si les fichiers existent déjà, ils sont écrasés et réécrits.

```
@csrf_exempt
def receive_vectorization(request):
    """
    Endpoint to receive a ZIP file containing SVG files and save them to the
    media directory.
    """
    if "file" not in request.FILES:
        return JsonResponse({"error": "No file received"}, status=400)

    file = request.FILES["file"]
    # treatment_id = request.DATA["experiment_id"]

    if file.name == "":
        return JsonResponse({"error": "File name is empty"}, status=400)

    if file and file.name.endswith(".zip"):
        try:
            temp_zip_path = default_storage.save("temp.zip", file)
            temp_zip_file = default_storage.path(temp_zip_path)

            save_svg_files(temp_zip_file)
            default_storage.delete(temp_zip_path)

        return JsonResponse(
            {"message": "Files successfully uploaded and extracted"}, status
            =200
        )
        except Exception as e:
            return JsonResponse({"error": str(e)}, status=500)
    else:
        return JsonResponse({"error": "Unsupported file type"}, status=400)
```

Listing B.11 – `Endpoint receive_vectorization` pour le retour des vectorisations dans l’application.

Fonction utilitaire pour traiter le contenu du ZIP :

```
def save_svg_files(zip_file):
    """
    Extracts SVG files from a ZIP file and saves them to the SVG_PATH directory
    .

```

```

"""
# Vérifie si le répertoire SVG_PATH existe, sinon le crée
if not os.path.exists(SVG_PATH):
    os.makedirs(SVG_PATH)

try:
    with zipfile.ZipFile(zip_file, "r") as zip_ref:
        for file_info in zip_ref.infolist():
            # TODO do not save jpg file
            # Vérifie si le fichier est un fichier SVG
            if file_info.filename.endswith(".svg"):
                file_path = os.path.join(
                    SVG_PATH, os.path.basename(file_info.filename)
                )

            # Supprime le fichier existant s'il y en a un
            if os.path.exists(file_path):
                os.remove(file_path)

            # Extrait le fichier SVG et l'écrit dans le répertoire
            # spécifié
            with zip_ref.open(file_info) as svg_file:
                with open(file_path, "wb") as output_file:
                    output_file.write(svg_file.read())
except Exception as e:
    log(f"[save_svg_files] Error when extracting SVG files from ZIP file", e)
    return False
return True

```

Listing B.12 – Fonction utilitaire pour le traitement du contenu de l'archive.

Annexe C

Interfaces : développement des vues personnalisées

Le début de mon stage s'est focalisé sur le développement d'interfaces. Plus spécifiquement, j'ai travaillé sur l'affichage des 'régions d'image', autrefois nommées 'annotations' (correspondant aux extractions), ainsi que des vectorisations associées. J'ai pu ainsi approfondir ma compréhension du *framework* Django en me familiarisant avec le processus de création de vues personnalisées. Néanmoins, ces vues ont été amenées à évoluer au fil des développements effectués par l'équipe de développeur.ses du projet. En effet, le principe de la 'vue' est d'interfacer le rendu graphique et le modèle de données, et ce dernier a été amené à évoluer au cours de mon stage. Cette annexe présente les développements effectués et certaines de leurs évolutions.

C.1 views.py pour traiter la logique des requêtes

Ce fichier contient les fonctions ou classes appelées 'vues' qui traitent les requêtes HTTP envoyées par l'utilisateur.rice. Chaque vue correspond généralement à une fonctionnalité spécifique de l'application. Elle récupère des données, souvent formatées grâce à des méthodes de classes, elle peut les traiter, et retourne une réponse HTTP (comme une page HTML, un fichier JSON, etc.).

Lorsque l'utilisateur.rice visite une URL spécifique, Django appelle la vue correspondante pour générer la réponse.

Par exemple, l'URL qui affiche les résultats de l'extraction peut appeler la fonction suivante :

```
# /app/webapp/views.py

@login_required(login_url=f"/{APP_NAME}-admin/login/")
def show_all_annotations(request, anno_ref):
    passed, anno = check_ref(anno_ref, "Annotation")
```

```

if not passed:
    return JsonResponse(anno)

if not ENV("DEBUG"):
    credentials(f"{SAS_APP_URL}/", ENV("SAS_USERNAME"), ENV("SAS_PASSWORD"))

_, all_annos = formatted_annotations(anno)
all_crops = [
    (canvas_nb, coord, img_file)
    for canvas_nb, coord, img_file in all_annos
    if coord
]

paginator = Paginator(all_crops, 50)
try:
    page_annos = paginator.page(request.GET.get("page"))
except PageNotAnInteger:
    page_annos = paginator.page(1)
except EmptyPage:
    page_annos = paginator.page(paginator.num_pages)

return render(
    request,
    "show_crops.html",
    context={
        "anno": anno,
        "page_annos": page_annos,
        "all_crops": all_crops,
        "url_manifest": anno.gen_manifest_url(version=MANIFEST_V2),
        "anno_ref": anno_ref,
    },
)

```

Listing C.1 – Vue pour l'affichage des extraction en « dump ».

La fonction envoie au *template* des métadonnées, ainsi que les coordonnées des annotations, pour chaque page annotée. Les régions d'images seront appelées dynamiquement dans le *template* grâce à la reconstruction à la volée des URLs IIIF.

L'action utilisateur.rice pour exporter l'ensemble des *crops* de diagramme d'un témoin en JPEG et sous forme de ZIP appelle la vue suivante :

```

@login_required(login_url=f"/{APP_NAME}-admin/login/")
def export_all_crops(request, anno_ref):

```

```

passed, anno = check_ref(anno_ref, "Annotation")
if not passed:
    return JsonResponse(anno)

if not ENV("DEBUG"):
    credentials(f"[SAS_APP_URL]/", ENV("SAS_USERNAME"), ENV("SAS_PASSWORD"))

urls_list = []

_, all_annos = formatted_annotations(anno)
all_crops = [
    (canvas_nb, coord, img_file)
    for canvas_nb, coord, img_file in all_annos
    if coord
]

for canvas_nb, coord, img_file in all_crops:
    urls_list.extend(gen_iiif_url(img_file, 2, f"[c[0]}/full/0") for c in
                     coord)

return zip_img(urls_list)

```

Listing C.2 – Vue pour l'export de l'ensemble des diagrammes du témoin affiché.

Enfin, une vue a été créée pour exporter uniquement les diagrammes sélectionnés. La liste est créée côté navigateur grâce à du JavaScript.

```

@login_required(login_url=f"/{APP_NAME}-admin/login/")
def export_selected_crops(request):

    urls_list = json.loads(request.POST.get("listeURL"))

    return zip_img(urls_list)

```

Listing C.3 – Vue pour l'export d'une liste de diagrammes.

Les vues font appel à de nombreuses autres fonctions utilitaires ou méthodes de classe, par exemple `formatted_annotations`, situé dans un fichier du dossier utilitaire dédié aux annotations IIIF (`app/webapp/utils/iiif/annotation.py`). Elle génère une liste structurée d'annotations. Pour chaque image, dans les données fournies par l'entité *Regions*, la fonction récupère les annotations associées. Si des annotations existent pour une image donnée, elle extrait pour chacune ses coordonnées et son identifiant.

C.2 url.py pour le routage des urls

Ce fichier définit les correspondances entre les URLs entrées par l'utilisateur.rice et les vues dans `views.py`. Il lie les requêtes URL spécifiques aux fonctions appropriées. Lorsqu'une requête est faite par l'utilisateur.rice, Django utilise `urls.py` pour déterminer quelle vue doit être exécutée en fonction de l'URL.

Indexation des vues dans le fichier `urls.py` :

```
#/app/config/urls.py
from app.webapp.views import (
    ...
    show_all_annotations,
    export_all_crops,
    export_selected_crops,
    ...
)
```

Listing C.4 – Contenu du fichier `urls.py` : import des views.

```
urlpatterns = [
    ...
    path(
        f"{APP_NAME}/<str:anno_ref>/show-all-annotations",
        show_all_annotations,
        name="show-all-annotations",
    ),
    path(
        f"{APP_NAME}/export-crops/<str:anno_ref>",
        export_all_crops,
        name="export-crops",
    ),
    path(
        f"{APP_NAME}/export-selected-crops",
        export_selected_crops,
        name="export-selected-crops",
    ),
    path(
        f"{APP_NAME}/<str:anno_ref>/show-vectorization",
        show_vectorization,
        name="show-vectorization",
    ),
]
```

```

),
...
]
```

Listing C.5 – Contenu du fichier urls.py : routage des fonctions et des URLs.

C.3 Les *templates*

Les *templates* sont des fichiers HTML (avec éventuellement des balises Django) qui définissent la présentation de la réponse envoyée à l'utilisateur.rice. Ils permettent de séparer la logique de l'application (dans les vues) de la présentation (dans les *templates*).

```
{#/app/webapp/templates/show_crops.html#}

{% block header %}
    <div class="toolbar">
        <div class="title">
            <b>{{ anno|capfirst }}</b>
        </div>

        <div style="display: flex; flex-direction: row;">
            <a href="{% url 'export-crops' anno_ref=anno_ref %}">
                <button class="export-button" type="submit">
                    <i class="fa-regular fa-file-zipper"></i>&ampnbsp
                    {% if APP_LANG == "en" %}
                        Download all
                    {% else %}
                        Télécharger toutes les
                    {% endif %}
                    annotations
                </button>
            </a>

            <form id="export" action="{% url 'export-selected-crops' %}" method="post">
                {% csrf_token %}
                <input type="hidden" id="listeURL" name="listeURL" value="">
                <button class="export-button" type="submit">
                    <i class="fa-regular fa-file-zipper"></i>&ampnbsp
                    {% if APP_LANG == "en" %}
                        Download selected annotations
                    {% else %}
```

```

        Télécharger les annotations sélectionnées
    {% endif %}
    </button>
</form>

{% if anno.is_validated %}
    <a href="{% url 'show-annotations' anno_ref=anno_ref %}" target=""
       _blank">
        <button class="edit-button" type="submit">
            <i class="fa fa-pencil"></i>&ampnbsp
            {% if APP_LANG == "en" %}
                Edit
            {% else %}
                Éditer les
            {% endif %}
            annotations
        </button>
    </a>
    {% endif %}
</div>
</div>
{% endblock %}

{% block content %}

<div class="tabs-crops">
    <div class="row">
        <div class="tab-buttons">
            <button class="btn-change active-tab">{% if APP_LANG == "en" %}
                Page view
            {% else %}
                Vue page
            {% endif %}
            </button>
            <button class="btn-change">{% if APP_LANG == "en" %}
                Dump view
            {% else %}
                Vue générale
            {% endif %}
            </button>
        </div>
    </div>
</div>

```

```

</div>

<div class="tab-bodies">
<div class="row" style="display:block;">
<div class="column">
<table class="anno-table" style="margin-top: 0em;">
{%
  for canvas_nb, coord, img_file in page_annos %}
<tr class="anno-row">
<td class="anno-td page-col">
<a href="{{ img_file|img_to_iif }}" target="_blank">

</a>
<h3><a href="{{ img_file|img_to_iif }}" target="_blank">Page {{ canvas_nb }}</a></h3>
</td>
<td class="anno-td anno-col">
{%
  for coords, id in coord %}
<div id="ill_{{ id }}" class="anno-div">
{%
  with region_full=coords|add:'/full/0'
  region=coords|add:'/,200/0' %}
<a href="{{ img_file|img_to_iif:region_full }}" target="_blank">

</a>
<br>
<input id="bbox_{{ id }}" type="checkbox"
       name="crop_checkbox" value="{{ img_file|img_to_iif:region_full }}"
       onchange="updateSelectedImageURLs()">
<label for="bbox_{{ id }}">
{%
  if APP_LANG == "en" %}Select{%
  else %}Sélectionner{%
  endif %}
</label>
{%
  endwith %}
</div>

```

```

        {% endfor %}
    </td>
</tr>
{% endfor %}
</table>
{% include 'includes/pagination.html' with current_page=
page_annos %}

</div>
</div>

<div style="display:none;">
<div class="grid-container" style="margin-top: .5em;">
    {% for canvas_nb, coord, img_file in all_crops %}
        {% for coords, id in coord %}
            {% with region_full=coords|add:'/full/0' region=coords|
add: '/250/0' %}
                <div class="grid-item">
                    <a href="{{ img_file|img_to_iif:region_full }}" target="_blank">
                        
                    </a>
                    <h3><a href="{{ img_file|img_to_iif }}" target="_blank">Page {{ canvas_nb }}</a></h3>
                    <input id="bbox_{{ id }}" type="checkbox" name="crop_checkbox" value="{{ img_file|img_to_iif:region_full }}" onchange="updateSelectedImageURLs()">
                    <label for="bbox_{{ id }}">
                        {% if APP_LANG == "en" %}Select{% else %}
                            Sélectionner{% endif %}
                    </label>
                </div>
            {% endwith %}
        {% endfor %}
    {% endfor %}
</div>
</div>
</div>

```

```

<script>

    var checkboxes = document.querySelectorAll('input[name="crop_checkbox"]');
    var selectedImageURLs = [];

    // Fonction pour mettre à jour la liste des URL sélectionnées
    function updateSelectedImageURLs() {
        // Réinitialiser la liste des URL sélectionnées
        selectedImageURLs = [];

        // Parcourir toutes les cases à cocher
        checkboxes.forEach(function(checkbox) {
            // Vérifier si la case est cochée
            if (checkbox.checked) {
                // Récupérez l'URL de l'image à partir de la value de la case
                // cochée
                var imageURL = checkbox.value;
                // Ajouter l'URL de l'image à la liste des URL sélectionnées
                selectedImageURLs.push(imageURL);
            }
        });
    };

    // Afficher la liste des URLs sélectionnées
    console.log(selectedImageURLs);
    var jsonString = JSON.stringify(selectedImageURLs);

    // liste => hidden input
    document.getElementById("listeURL").value = jsonString;
    // Vérifier si la valeur du champ est correcte
    console.log(document.getElementById("listeURL").value);
};

Array.from(document.querySelectorAll('.tabs-crops')).forEach((tab_container
, TabID) => {
    const registers = tab_container.querySelector('.tab-buttons');
    const bodies = tab_container.querySelector('.tab-bodies');

    Array.from(registers.children).forEach((el, i) => {
        el.setAttribute('aria-controls', `${TabID}_${i}`);
        bodies.children[i]?.setAttribute('id', `${TabID}_${i}`);

        el.addEventListener('click', (ev) => {
            let activeRegister = registers.querySelector('.active-tab');

```

```

        activeRegister.classList.remove('active-tab')
        activeRegister = el;
        activeRegister.classList.add('active-tab')
        changeBody(registers, bodies, activeRegister)
    })
})

function changeBody(registers, bodies, activeRegister) {
    selectedImageURLs = [];

    checkboxes.forEach(function(checkbox) {
        checkbox.checked = false;
    });

    Array.from(registers.children).forEach((el, i) => {

        if (bodies.children[i]) {
            bodies.children[i].style.display = el == activeRegister? 'block'
                ~: 'none'
        }

        el.setAttribute('aria-expanded', el == activeRegister? 'true'~: 'false')
    })
}
</script>

{% endblock %}

```

Listing C.6 – Template HTML pour l'affichage des crops de diagrammes.

Ce qui donne deux affichages possibles pour les extractions :

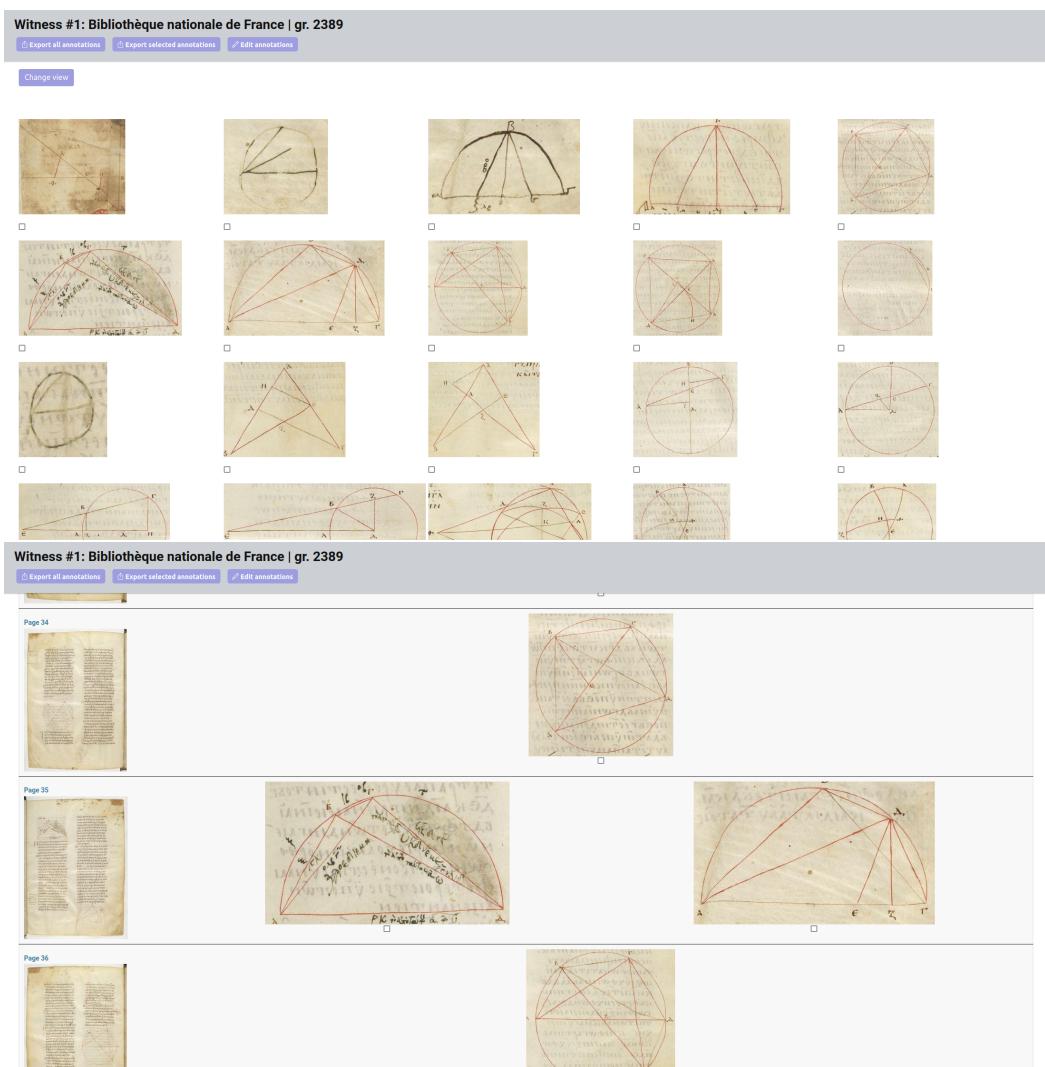


FIGURE C.1 – Affichage des extractions, vue *dump* et avec la page de manuscrit en regard.

Des vues similaires ont été créées pour afficher les résultats de la vectorisation.

C.4 Autres développements

Vues pour l'export des SVGs et des images JPEG :

```
@login_required(login_url=f"/{APP_NAME}-admin/login/")
def export_all_images_and_svgs(request, anno_ref):
    passed, anno = check_ref(anno_ref, "Annotation")
    if not passed:
        return JsonResponse(anno)

    if not ENV("DEBUG"):
        credentials(f"{SAS_APP_URL}/", ENV("SAS_USERNAME"), ENV("SAS_PASSWORD"))
```

```

urls_list = []
path_list = []

_, all_annos = formatted_annotations(anno)
all_crops = [
    (canvas_nb, coord, img_file)
    for canvas_nb, coord, img_file in all_annos
    if coord
]

for canvas_nb, coord, img_file in all_crops:
    urls_list.extend(gen_iiif_url(img_file, 2, f"{c[0]}/full/0") for c in coord)
    vecto_path = f"{img_file[:-4]}_{''.join(c[0] for c in coord)}.svg"
    # Vérifie si le chemin existe
    if os.path.exists(os.path.join(SVG_PATH, vecto_path)):
        path_list.append(vecto_path)

return zip_images_and_files(urls_list, path_list)

```

Listing C.7 – Vue pour l'export de tous les SVGs et des images associées.

```

@login_required(login_url=f"/{APP_NAME}-admin/login/")
def export_selected_imgs_and_svgs(request):
    images_list = json.loads(request.POST.get("liste_images"))
    urls_list = []
    paths_list = []
    for element in images_list:
        if is_url(element):
            urls_list.append(element)
        else:
            paths_list.append(element)
    return zip_images_and_files(urls_list, paths_list)

```

Listing C.8 – Vue pour l'export des SVGs et des images JPEG sélectionnés.

Des fonctions utilitaires sont utiles pour assurer l'export :

```

def zip_img(img_list, zip_name=f"{APP_NAME}_export"):
    buffer = io.BytesIO()
    with zipfile.ZipFile(buffer, "w") as z:
        for img_path in img_list:

```

```

    img_name = f"{url_to_name(img_path)}.jpg"
    if urlparse(img_path).scheme == "":
        z.write(f"{IMG_PATH}/{img_name}", img_name)
    else:
        response = requests.get(img_path)
        if response.status_code == 200:
            z.writestr(img_name, response.content)
        else:
            log(f"[zip_img] Fail to download img: {img_path}")
            pass

    response = HttpResponse(
        buffer.getvalue(), content_type="application/x-zip-compressed"
    )
    response["Content-Disposition"] = f"attachment; filename={zip_name}.zip"
    return response


def zip_files(filenames_contents, zip_name=f"{APP_NAME}_export"):
    # filenames_contents = [(filename1, content1), (filename2, content2), ...]
    buffer = io.BytesIO()
    with zipfile.ZipFile(buffer, "w") as z:
        for filename_content in filenames_contents:
            filename, content = filename_content
            z.writestr(filename, content)

    response = HttpResponse(
        buffer.getvalue(), content_type="application/x-zip-compressed"
    )
    response["Content-Disposition"] = f"attachment; filename={zip_name}.zip"
    return response


def zip_images_and_files(img_list, file_list, zip_name=f"{APP_NAME}_export"):
    buffer = io.BytesIO()
    with zipfile.ZipFile(buffer, "w") as z:
        # Ajouter des images à partir des URLs ou du répertoire local
        for img_path in img_list:
            img_name = f"{url_to_name(img_path)}.jpg"
            if urlparse(img_path).scheme == "":
                try:

```

```

        z.write(f"{SVG_PATH}/{img_name}", img_name)
    except FileNotFoundError:
        log(f"[zip_images_and_files] Local image not found: {img_path}")
    else:
        response = requests.get(img_path)
        if response.status_code == 200:
            z.writestr(img_name, response.content)
        else:
            log(f"[zip_images_and_files] Fail to download image: {img_path}")

# Ajouter des fichiers à partir du répertoire mediafiles
for file_path in file_list:
    try:
        with open(f"{SVG_PATH}/{file_path}", "rb") as f:
            z.writestr(file_path, f.read())
    except FileNotFoundError:
        log(f"[zip_images_and_files] Local file not found: {file_path}")

response = HttpResponse(
    buffer.getvalue(), content_type="application/x-zip-compressed"
)
response["Content-Disposition"] = f"attachment; filename={zip_name}.zip"
return response

```

Listing C.9 – Fonctions utilitaires pour compresser et télécharger les images.

Template pour afficher les *crops* de diagrammes avec les vectorisations associées. Le *template* et le rendu sont très similaire à ceux concernant l'affichage des diagrammes uniquement :

```

{% extends "admin/base_site.html" %}

{% load static %}

{% block title %}{{ anno }}{% endblock %}

{% block header %}
<div class="toolbar">
    <div class="title">
        <b>{{ anno|capfirst }}</b>

```

```

    </div>
<div style="display: flex; flex-direction: row;">
    <a href="{% url 'show-all-annotations' anno_ref=anno_ref %}" target="_blank">
        <button class="export-button" type="submit">
            <i class="fa-solid fa-eye"></i>
            {% if APP_LANG == "en" %}
                Visualize all
            {% else %}
                Voir toutes les
            {% endif %}
            annotations
        </button>
    </a>
    <a href="{% url 'export-all-imgs-and-svgs' anno_ref=anno_ref %}">
        <button class="export-button" type="submit">
            <i class="fa-regular fa-file-zipper"></i>&ampnbsp
            {% if APP_LANG == "en" %}
                Download all
            {% else %}
                Télécharger toutes les
            {% endif %}
            vectorisations
        </button>
    </a>

<form id="export" action="{% url 'export-img-and-svg' %}" method="post">
    {% csrf_token %}
    <input type="hidden" id="liste_images" name="liste_images" value="">
    <button class="export-button" type="submit">
        <i class="fa-regular fa-file-zipper"></i>&ampnbsp
        {% if APP_LANG == "en" %}
            Download selected annotations
        {% else %}
            Télécharger les vectorisations sélectionnées
        {% endif %}
    </button>
</form>
</div>

```

```

    </div>
{%
  endblock %}

{%
  block content %}
}

<div class="tabs-crops">
  <div class="row">
    <div class="tab-buttons">
      <button class="btn-change active-tab">{%
        if APP_LANG == "en" %}
          Page view
        {% else %}
          Vue page
        {% endif %}
      </button>
      <button class="btn-change">{%
        if APP_LANG == "en" %}
          Dump view
        {% else %}
          Vue générale
        {% endif %}
      </button>
    </div>
  </div>
</div>

<div class="tab-bodies">

  <div class="row" style="display:block;">
    <div class="column">
      <table class="anno-table" style="margin-top: 0em;">
        {% for canvas_nb, coord, img_file in page_annos %}
          <tr class="anno-row">
            <td class="anno-td page-col">
              <a href="{{ img_file|img_to_iif }}" target="_blank">
                
              </a>
              <h3><a href="{{ img_file|img_to_iif }}" target="_blank"
                ">Page {{ canvas_nb }}</a></h3>
            </td>
            <td class="anno-td anno-col">

```

```

    {% for coords, id in coord %}
        <div id="ill_{{ id }}" class="anno-div grid-item"
            style=" display: flex; flex-wrap: wrap;
            justify-content: center;">
            {% with region_full=coords|add:'/full/0' region=
                coords|add:'/,200/0' %}
                <a href="{{ img_file|img_to_iif:region_full
                }}" target="_blank">
                    
                </a>
                
                <a href="{% url 'img-and-svg' img_file=
                    img_file coords=coords %}" target="_blank
                    ">{% if APP_LANG == "en" %}Visualize{{%
                    else %}Visualiser{{% endif %}}</a>
                <br>
                <input type="checkbox" name="vecto_checkbox"
                    value='["{{ img_file|jpg_to_none }}_{{ coords }}.svg", "{{ img_file|img_to_iif:region }}"]' onchange="
                    updateSelectedImageURLs()">
                <label for="checkbox">
                    {% if APP_LANG == "en" %}Select{% else %}
                        Sélectionner{% endif %}
                </label>
                {% endwith %}
            </div>
            {% endfor %}
        </td>
    </tr>
    {% endfor %}
</table>
{% include 'includes/pagination.html' with current_page=page_annos
%}
</div>
</div>

```

```

<div style="display: none;">
<div style="margin-top: 5%;">
    <div class="grid-container">
        {% for canvas_nb, coord, img_file in all_crops %}
            {% for coords, id in coord %}
                <div class="grid-item">
                    <div style="display: flex; flex-direction: row;">
                        {% with small=coords|add:'/full/0' %}
                            <a href="{{ img_file|img_to_iif:small }}" target="_blank">
                                
                            </a>
                            
                            <a href="{ % url 'img-and-svg' img_file= img_file coords=coords %}" target="_blank">{% if APP_LANG == "en" %} Visualize {% else %} Visualiser {% endif %}</a>
                        </a>
                    </div>
                    <input type="checkbox" name="vecto_checkbox" value='["{{ img_file|jpg_to_none }}_{{ coords }}.svg", "{{ img_file|img_to_iif:small }}"]' onchange="updateSelectedImageURLs()">
                {% endwith %}
            </div>
            <h3><a href="{{ img_file|img_to_iif }}" target="_blank">Page {{ canvas_nb }}</a></h3>
        </div>
    {% endfor %}
    {% endfor %}
</div>
</div>
=</div>

</div>

```

```

</div>

<script>
    document.addEventListener('DOMContentLoaded', () => {
        const tabContainers = document.querySelectorAll('.tabs-crops');

        tabContainers.forEach((tabContainer, tabID) => {
            const registerButtons = tabContainer.querySelectorAll('.tab-buttons
                .btn-change');
            const tabBodies = tabContainer.querySelectorAll('.tab-bodies > div')
            ;

            registerButtons.forEach((button, index) => {
                button.setAttribute('aria-controls', `${tabID}_${index}`);
                if (tabBodies[index]) {
                    tabBodies[index].setAttribute('id', `${tabID}_${index}`);
                }
            }

            button.addEventListener('click', () => {
                setActiveTab(registerButtons, tabBodies, button);
            });
        });
    });

    function setActiveTab(registerButtons, tabBodies, activeButton) {
        registerButtons.forEach((button, index) => {
            if (button === activeButton) {
                button.classList.add('active-tab');
                if (tabBodies[index]) {
                    tabBodies[index].style.display = 'block';
                }
            } else {
                button.classList.remove('active-tab');
                if (tabBodies[index]) {
                    tabBodies[index].style.display = 'none';
                }
            }
            button.setAttribute('aria-expanded', button === activeButton? 'true': 'false');
        });
    }
}

```

```

        }

    });

var checkboxes = document.querySelectorAll('input[name="vecto_checkbox"]');
var selectedImages = [];

function updateSelectedImageURLs() {
// Clear the selectedImages array before processing checkboxes
selectedImages.length = 0; // More efficient way to clear the array

// Loop through all checkboxes
checkboxes.forEach(function(checkbox) {
    if (checkbox.checked) {
        try {
            // Parse the JSON string from checkbox value (handle potential
            // errors)
            var parsedImages = JSON.parse(checkbox.value);
            // Add parsed images to selectedImages array (assuming correct
            // format)
            selectedImages.push(...parsedImages);
        } catch (error) {
            console.error("Error parsing JSON from checkbox:", checkbox.value,
                error);
        }
    }
});
}

// Convert the selectedImages array to JSON string (handle empty array)
var jsonString = selectedImages.length > 0? JSON.stringify(selectedImages)
    : "";

// Update the hidden input "liste_images" with the JSON string
document.getElementById("liste_images").value = jsonString;

// Log the JSON string for verification (optional)
console.log(selectedImages);
console.log(jsonString);
}

</script>
{% endblock %}

```

Listing C.10 – Template HTML pour afficher les extractions et leurs vectorisations.

C.5 Manipuler le format SVG

Le stage m'a également permis de me familiariser avec le format SVG. J'ai développé une interface basique permettant de manipuler des fichiers SVGs, en implémentant des fonctionnalités de base (Fig. C.2) :

- supprimer une primitive (en cliquant dessus) ;
- passer les primitives en noir et blanc ;
- *fade* l'image en arrière plan ;
- télécharger au format JPEG l'image et le format vectoriel superposés, et en gardant les modifications effectuées.

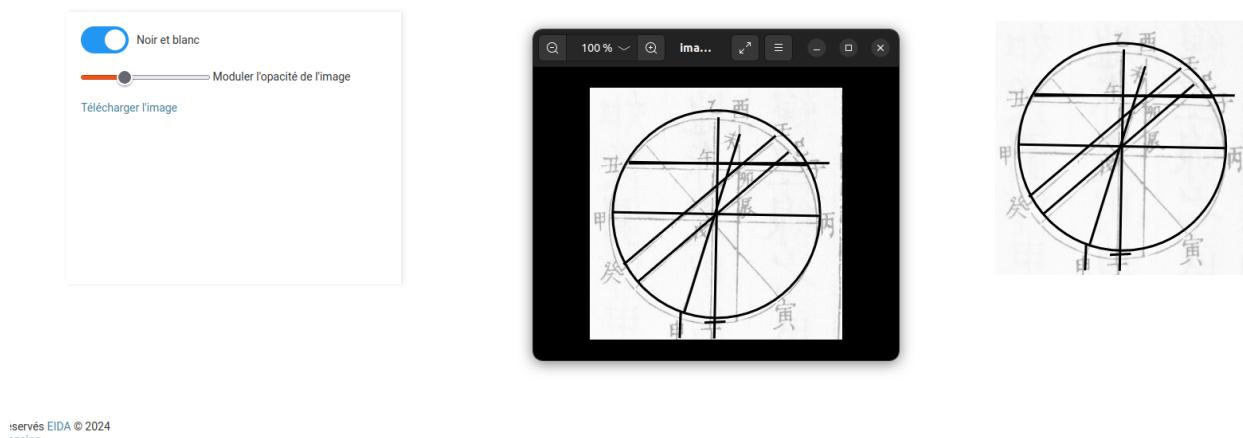


FIGURE C.2 – Interface de manipulation des fichiers SVGs, avec exemple d'image modifiée téléchargée.

J'ai pu ainsi approfondir mes connaissances en JavaScript et découvrir les spécificités du format SVG. La stratégie a consisté à écrire le contenu des fichiers dans le DOM et ajouter de l'interaction grâce à du JavaScript.

```
{% extends "admin/base_site.html" %}

{% load static %}

{% block title %}{{ regions }}{% endblock %}
```

```

{% block extrahead %}
    {{ block.super }}
    <link rel="stylesheet" href="{% static 'css/tools.css' %}">
    <link rel="stylesheet" href="{% static 'css/style.css' %}">
    <link href="https://cdn.jsdelivr.net/npm/bootstrap-icons@1.8.0/font/
        bootstrap-icons.css" rel="stylesheet">

{% endblock %}

{% block header %}
<div class="toolbar">
    <div class="title">
        <b>{{ regions|capfirst }} | page {{ canvas_nb }}</b>
    </div>
</div>
{% endblock %}

{% block content %}
<div class="container">
    <div class="sidebar">
        <div class="switch-container">
            <label class="switch">
                <input type="checkbox" id="change-stroke-color">
                <span class="slider round"></span>
            </label>
            <label for="change-stroke-color">{{ if APP_LANG == "en" }}All lines
                in black{{ else }}Noir et blanc{{ endif }}</label>
        </div>

        <div>
            <input type="range" min="0" max="1" step="0.01" value="1" id=""
                opacityRange">
            <label for="opacityRange">{{ if APP_LANG == "en" }}Toggle Image
                Opacity{{ else }}Moduler l'opacité de l'image{{ endif }}</label>
        </div>

        <div>
            <a id="downloadLink" href="#" download="image.jpg">Télécharger l'
                image</a>
        </div>
    </div>
</div>

```

```

</div>

<div class="image-container">
  {{ svg_content|safe }}
</div>
</div>

<script>
  document.addEventListener('DOMContentLoaded', function() {

    const svgElement = document.querySelector('svg'); // Sélectionner l'
      élément SVG
    const imageElement = document.querySelector('svg image');

    // Charger l'image de fond
    function fond() {
      {%- with small=coords|add:'/full/0' %}

      const backgroundImage = "{{ img_file|img_to_iif:small }}";
      {%- endwith %}

      imageElement.setAttribute('xlink:href', backgroundImage);
    }

    fond();

    // Changer la couleur des traits
    var toggleSwitch = document.getElementById('change-stroke-color');
    var elementsWithStroke = document.querySelectorAll('[stroke]');
    var originalStrokeColors = [];

    elementsWithStroke.forEach(function(element) {
      originalStrokeColors.push(element.getAttribute('stroke'));
    });

    function changeStrokeColorToBlack() {
      elementsWithStroke.forEach(function(element) {
        element.setAttribute('stroke', 'black');
      });
    }

    function restoreOriginalStrokeColors() {
      elementsWithStroke.forEach(function(element, index) {
        element.setAttribute('stroke', originalStrokeColors[index]);
      });
    }
  });
</script>

```

```

        element.setAttribute('stroke', originalStrokeColors[index]);
    });
}

toggleSwitch.addEventListener('change', function() {
    if (this.checked) {
        changeStrokeColorToBlack();
    } else {
        restoreOriginalStrokeColors();
    }
});

// Modifier l'opacité de l'image
const opacityRange = document.getElementById("opacityRange");
opacityRange.addEventListener("input", function() {
    const opacityValue = this.value;
    imageElement.style.opacity = opacityValue;
});

// Faire disparaître les éléments onclick
function removeElementOnClick(event) {
    event.target.remove();
}

const circles = document.querySelectorAll('svg circle');
const paths = document.querySelectorAll('svg path');

circles.forEach(circle => {
    circle.addEventListener('click', removeElementOnClick);
});

paths.forEach(path => {
    path.addEventListener('click', removeElementOnClick);
});

// Convertir le SVG modifié en image (PNG ou JPEG)
function convertModifiedSVGToImage(format, quality, callback) {
    // Sérialiser le SVG modifié
    const svgString = new XMLSerializer().serializeToString(svgElement);

    // Créer un canvas pour dessiner l'image
}

```

```

const canvas = document.createElement('canvas');
const context = canvas.getContext('2d');

// Charger l'image de fond dans le canvas
const image = new Image();
image.setAttribute('crossOrigin', 'anonymous');
image.onload = function() {
    // Le canvas a la même taille que l'image
    canvas.width = image.width;
    canvas.height = image.height;

    // Remplir le canvas avec un fond blanc
    context.fillStyle = 'white';
    context.fillRect(0, 0, canvas.width, canvas.height);

    // Appliquer l'opacité de l'image de fond
    const imageOpacity = parseFloat(window.getComputedStyle(imageElement).opacity);
    context.globalAlpha = isNaN(imageOpacity)~? 1.0~: imageOpacity;
    context.drawImage(image, 0, 0, canvas.width, canvas.height);

    // Réinitialiser l'opacité pour tracer le SVG
    context.globalAlpha = 1.0;

    // Dessiner le SVG modifié sur le canvas
    const svgBlob = new Blob([svgString], { type: 'image/svg+xml;charset=utf-8' });
    const DOMURL = window.URL || window.webkitURL || window;
    const svgUrl = DOMURL.createObjectURL(svgBlob);

    const svgImage = new Image();
    svgImage.onload = function() {
        // Dessiner le SVG à la bonne taille
        context.drawImage(svgImage, 0, 0, canvas.width, canvas.height);

        // Télécharger l'image convertie au format spécifié
        const dataUrl = canvas.toDataURL(`image/${format}`, quality);
        callback(dataUrl); // Utiliser le callback pour manipuler le dataUrl
    };
    svgImage.src = svgUrl;
}

```

```

    };
    image.src = imageElement.getAttribute('xlink:href');
}

// Fonction de téléchargement
function downloadImage(format, quality) {
    convertModifiedSVGToImage(format, quality, function(dataUrl) {
        const downloadLink = document.createElement('a');
        downloadLink.href = dataUrl;
        downloadLink.setAttribute('download', `image.${format}`);
        document.body.appendChild(downloadLink);
        downloadLink.click();
        document.body.removeChild(downloadLink);
    });
}

// Lien pour télécharger en JPEG
const downloadLink = document.getElementById('downloadLink');
downloadLink.addEventListener('click', function(event) {
    event.preventDefault();
    downloadImage('jpeg', 0.8);
});

});
</script>

{% endblock %}

```

Listing C.11 – *Template* HTML pour la manipulation des fichiers SVGS.

Cette interface est associée à une vue qui renvoie au *template* les informations nécessaires, notamment le contenu du fichier SVG.

```

@login_required
def show_crop_vectorization(request, img_file, coords, regions, canvas_nb):
    svg_filename = f"{jpg_to_none(img_file)}_{coords}.svg"
    svg_path = os.path.join(SVG_PATH, svg_filename)

    if not os.path.exists(svg_path):
        print(f"File {svg_path} not found")

    with open(svg_path, "r", encoding="utf-8") as file:

```

```

    svg_content = file.read()

    return render(
        request,
        "crop_vecto.html",
        context={
            "img_file": img_file,
            "coords": coords,
            "svg_content": svg_content,
            "regions": regions,
            "canvas_nb": canvas_nb,
        },
    )
)

```

Listing C.12 – Vue pour relier les données à leur affichage dans l’interface de manipulation.

C.6 Nouvelles interfaces

Avec la refonte des interfaces effectuée par les développeuses du projet, ces codes ne sont actuellement plus utilisés dans l’application AIKON, l’utilisation d’un *framework front-end* ayant totalement changé les logiques d’affichage. J’ai cependant eu l’opportunité de poser fondations de la page d’accueil de la plateforme (Fig. ??) :

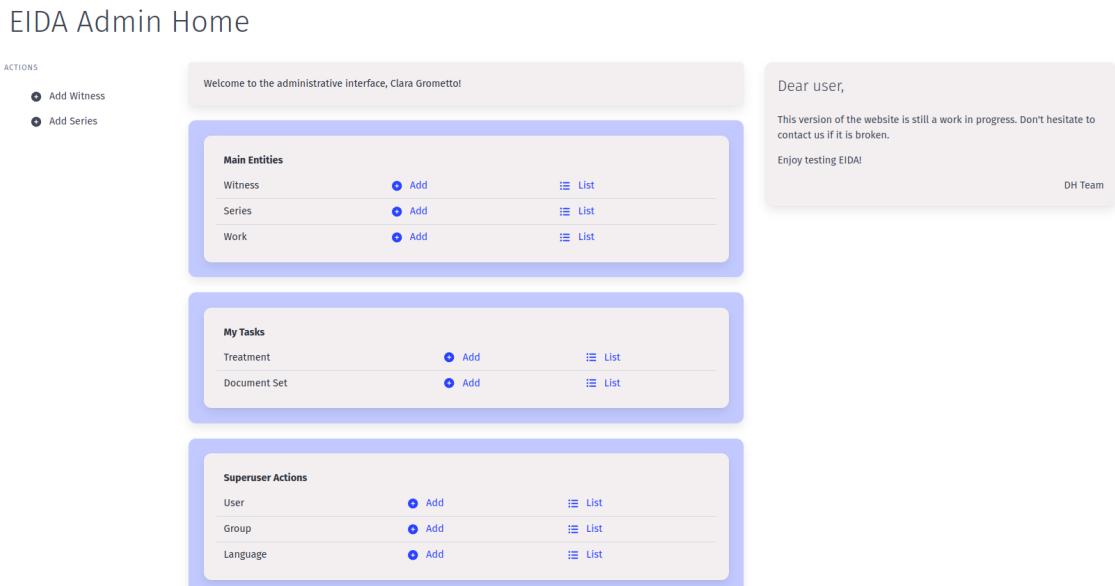


FIGURE C.3 – Page d’accueil de la plateforme AIKON

Annexe D

Nouvelles interfaces : documentation

Getting Started with the EiDA Interface

This guide should help you efficiently navigate and utilize the platform's features for accessing lists, viewing item information, launching treatments and visualizing the results. If you need further assistance, please refer to the DH Team 🧑.



Working with existing data

Items

A screenshot of the EIDA Admin Home dashboard. It shows a navigation bar with 'EIDA', 'Manage users', 'Manage documents', 'Computer vision tasks', a user icon, and 'Logged in as Jade Norindr' with a 'Log out' button. Below the bar is a main title 'EIDA Admin Home'. On the left, there's a sidebar with 'ACTIONS' and buttons for 'Add Witness' and 'Add Series'. The central area contains a 'Main Entities' section with three items: 'Witness', 'Series', and 'Work', each with an 'Add' button and a 'List' button. To the right, there's a message box with a welcome message, a note about the website being a work in progress, and a 'DH Team' signature.

Navigate to a list of elements:

- Once logged in, locate the different menus on the dashboard.
 - The Nav Bar on the top of the page
 - A central Block
- Click on the "List" option from one of them.
- You will be redirected to the list page, where you can see all available Witnesses, Series or Works.

The IIIF logo takes you to the Witness's manifest.

Click on the ref. to change metadata linked to the item.

Bibliothèque nationale de France | gr. 2389

Témoin # Bibliotheque nationale de France | gr. 2389

sziege 2024-02-13 16:16

Afficher les régions

Cote gr. 2389

Oeuvre Ptolemy | Almagest

Lieu de création -

Dates 801-900

Nombre de page 376 ff.

Langue Ancient greek

Ajouter à la sélection

Biblioteca del Monasterio de El Escorial | O II 10

Témoin # Biblioteca del Monasterio de El Escorial | O II 10

admin 2023-12-14 13:06

Afficher les régions

Cote O II 10

Oeuvre -

Lieu de création -

Dates -

Sélection (0)

The 'Regions' button allows you to view the predictions from the treatments applied to the Witness.

The 'Add to Set' button will add the witness to your personal set.

Retirer de sélection

Ajouter à la sélection

Retirer de sélection

Ajouter à la sélection

Retirer de sélection

Sélection (1)

Sélection (0)

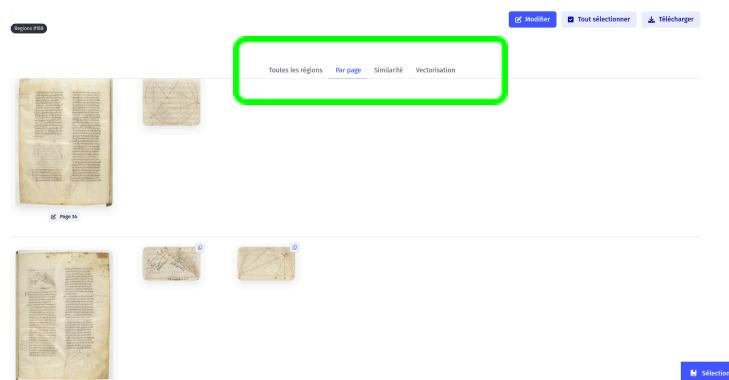
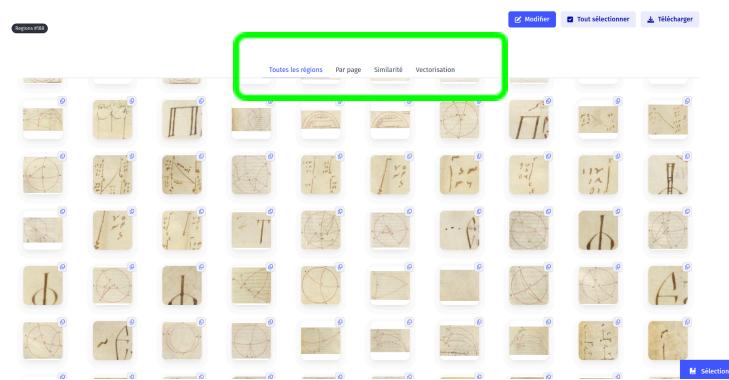
Manage your set directly from the list page.

OR Click on "Selection" to manage your Set ([more details here](#)).

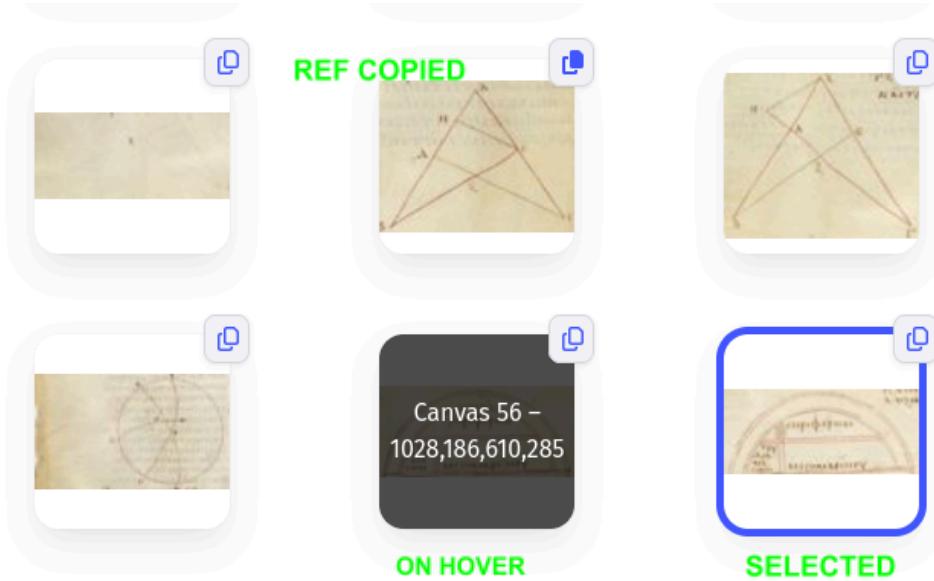
Visualize Predictions

Extraction

- Navigate to the “Regions” section 
- You will be redirected to the *Regions* page, where you can visualize the results of the diverse Treatments applied on the Witness of interest.
- Switch between the *dump* view and the *per page* view. Clicking on the page will lead you to the Mirador viewer.



- Hover the diagram crop for the canvas number and the coordinates
- Click to create a selection of diagrams on which you want to apply Treatments (e.g. Export)
- Copy the ref. thanks to the icon on the top right of the image



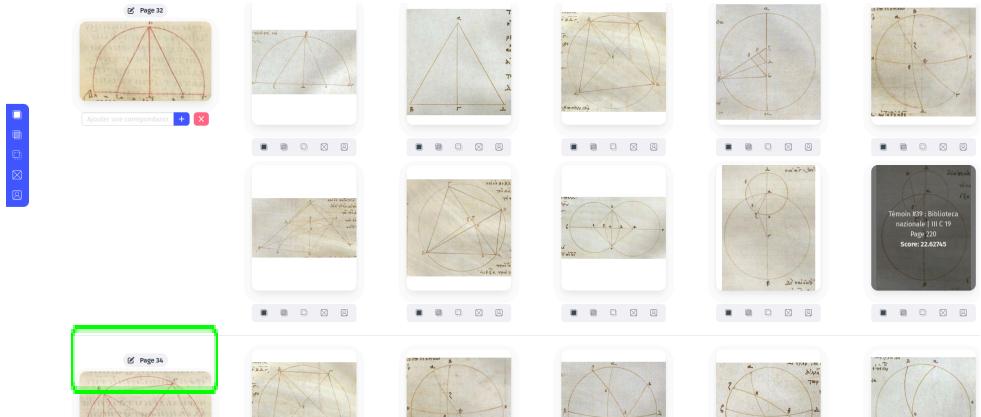
Export Results

- Select
- Click on the "Download" button.
- Follow the prompts to download or save the region's crops.

Similarity

- Navigate to the "Similarity" tab
- You have all the witnesses for whom similarities have been calculated with the current witness. Select at least one of them to visualize the 10 best similarity results.

- Hover on the image to get the ref. and metadata
- Clicking on the page number leads you to the Mirador Viewer



Vectorization

EIDA Manage users Manage documents Computer vision tasks Logged in as Jade Norindr Log out

“Bibliothèque nationale de France | gr. 2389” regions

Regions #188

All regions Per page Similarity Vectorization

1 2 ... 5

1 - 49 / 232

Selection (0)

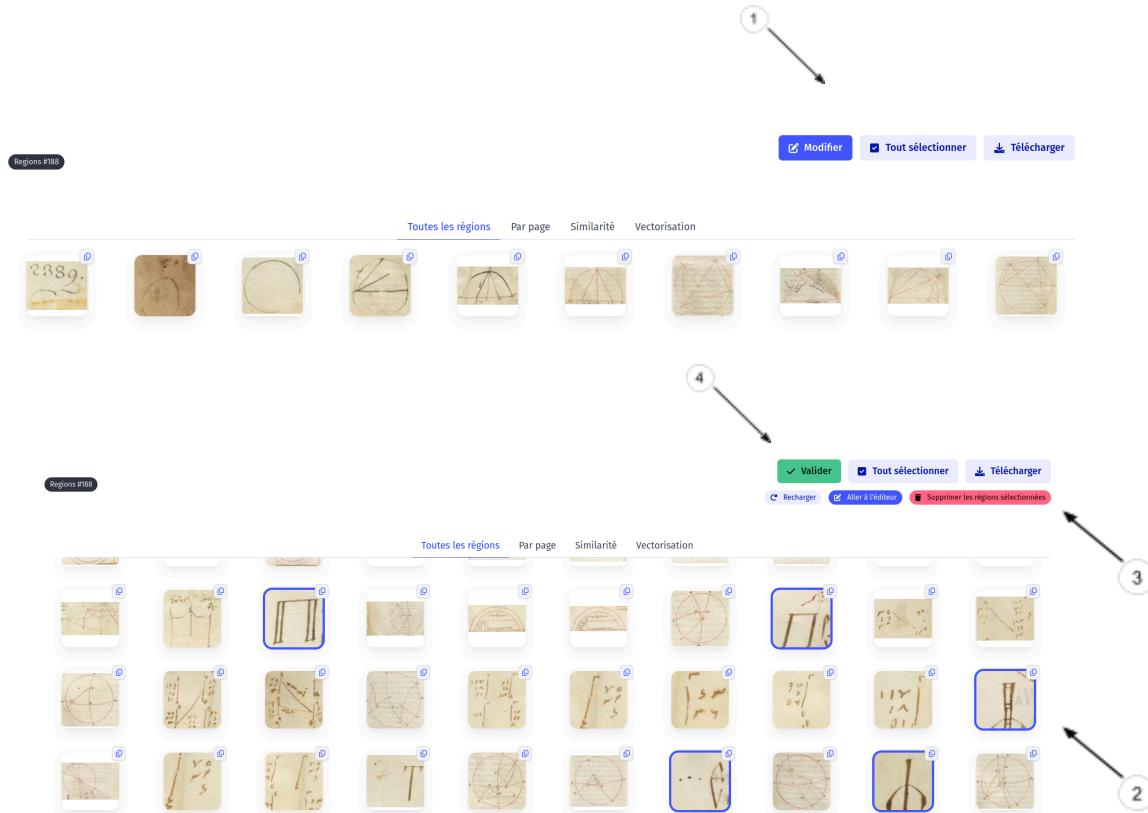
- Click on the Vectorization tab to view the vectorized diagrams
- For now, you can only visualize them as such, and hover the images to see the original JPEG image

Correct Predictions

Extraction

- (1) Click on “Edit” to open the set of commands
- (2) Select the annotations you need to delete
- (3) Click on “Delete”

- (4) Click on Validate



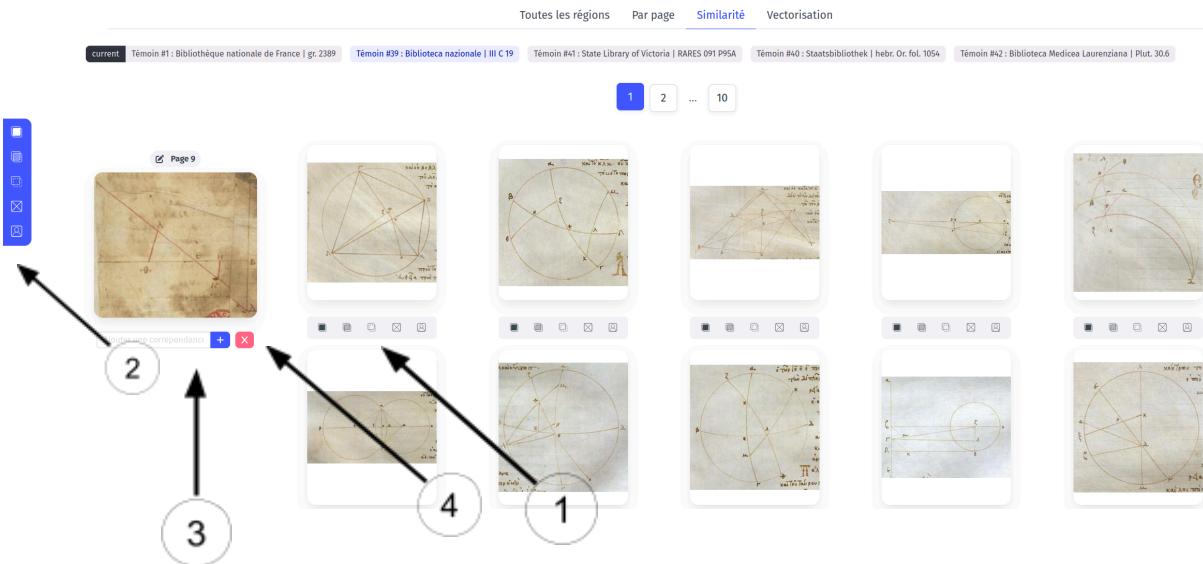
Similarity

- Annotate with the similarity category beneath each image (1)

Reminder:

| | |
|--|----------------|
| | Exact Match |
| | Partial Match |
| | Semantic Match |
| | No Match |
| | Personal Match |

- You can then choose to hide or display categories on the Side Menu (2)
- Paste [the reference of an image](#) in the text box to add it to the similarity results (3)
- Click the red cross to indicate “no similarity for this image”(4)



New Data and New Treatments

Add Witnesses / Series

EIDA Gérer les utilisateurs Gérer les documents

EIDA Admin Home

Bienvenue sur l'interface Admin Clara Grometto !

| Entités | | |
|---------|-------------------------|------------------------|
| Témoin | Ajouter | Lister |
| Série | Ajouter | Lister |
| Œuvre | Ajouter | Lister |

- On the homepage, you can access the Witness Form through the side menu and the central block (“Add”).

- Enter Metadata
- Click Submit

Launch Treatments

The set

- Select the Witnesses, Series and Works in the lists
- Click on “Selection” on the bottom right
- Manage your Document Set

Regions sélectionnées (4)

Témoin

- #1 Bibliothèque nationale de France | gr. 2389
- #4 Bibliothèque nationale de France | Lat. 7214
- #5 Bibliothèque nationale de France | Lat. 16211
- #8 Bibliothèque nationale de France | gr. 2925

Vider la sélection **Sauvegarder la sélection**

You can consult your sets in Manage documents > Document Set

⚠ To manage a previously saved Set, go to Manage documents > Document Set and click on the “Load to selection” button

Document set #7 **MathieuSavAlmagest #7 (6 documents)**

maubry 2024-08-07 11:11

similarity #fb2de534-ba91-4ff4-be1e-31eb3ad39712

| | | | |
|------------|---|------------|--|
| Witness #1 | Bibliothèque nationale de France gr. 2389 | Witness #2 | Biblioteca del Monasterio de El Escorial O II 10 |
| Witness #3 | Bibliothèque nationale de France Chinois 4957 | Witness #4 | Bibliothèque nationale de France Lat. 7214 |
| Witness #5 | Bibliothèque nationale de France Lat. 16211 | Witness #6 | Universitätsbibliothek F II 7 |

Load to selection

Your set will appear in the “Selection” tab on the bottom right. You can add or delete documents through the list interfaces.

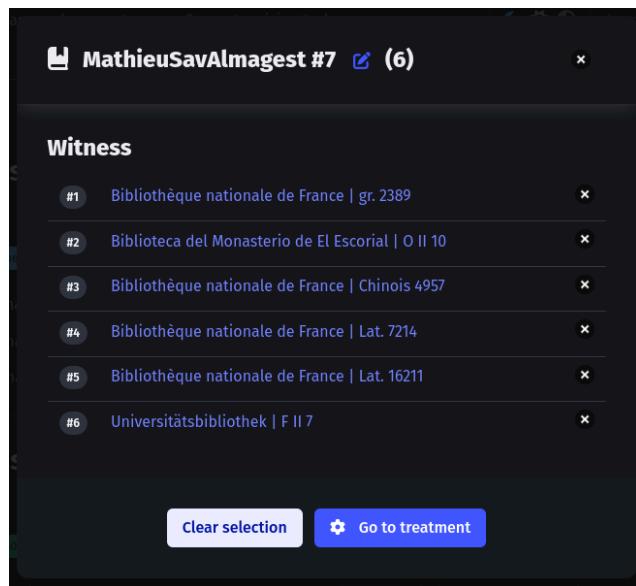
Treatment Form

⚠ Computer vision tasks are not launched automatically anymore.

⚠ To launch a Treatment on an individual document, you still need to select it and save your selection – the document will then be available for treatment (its title will be displayed in the set list in the form).

Once your set is saved, you can perform a Treatment on it:

- In the selection popup, click on “Go to treatment”



- OR go to Computer vision tasks > Treatment > Add

The screenshot shows a form titled "Add new treatment". It has several input fields: a dropdown menu for "Task type", a search bar for "Document set", and a checked checkbox for "Notify by email" with the sub-instruction "Send an email when the task is finished". At the bottom is a large blue "Submit" button.

- In the form, select the **task** you want to perform (Regions extraction, similarity search, vectorization), and the **document set** you want the task to be performed on.
- You can ask to be notified by email when your task is done running.

 You can only request tasks on your own document sets

 You need to have regions extracted for the similarity and vectorization to launch

Liste des acronymes

- ALTO *Analysed Layout and Text Object.* 40, 41, 89
- API *Application Programming Interface.* 18, 29–34, 38, 41, 45, 109, 112–114, 116–118, 122, 123, 135, 138, 145, 148, 157–160, 170, 216
- ARK *Archival Resource Key.* 41
- BnF *Bibliothèque Nationale de France.* 26, 27, 37–39, 41, 45
- CITE *Collections, Indices, Texts, and Extensions.* 95
- CNN *Convolutional Neural Network.* 60–62, 67, 215
- CPU *Computing Processing Unit.* 61, 121–123
- DISHAS *Digital Information System for the History of Astral Sciences.* 12, 13, 28, 97, 98, 100, 107
- DOM *Document Object Model.* 137, 139, 140, 193
- EIDA *Editing and analysing hIstorical astronomical Diagrams with Artificial intelligence.* iii, v, xxvii–xxix, 3, 5, 6, 11–24, 26–29, 34, 35, 41–43, 51, 53, 56, 60, 62–65, 71, 73, 74, 76, 77, 80, 81, 85–87, 89, 93–98, 105, 109, 116, 121, 123, 124, 126, 128, 136–138, 147, 149, 154, 215, 216
- FAIR *Findable Accessible Interoperable Reusable.* 29, 145
- GPU *Graphics Processing Unit.* 18, 114, 117, 118, 121–123, 148
- HN *Humanités Numériques.* 38, 105, 107
- HTML *HyperText Markup Language.* xxvi, 136, 138, 139, 143, 173, 177, 182, 193, 198
- HTR *Handwritten Text Recognition.* xxv, xxvi, 40–43, 45, 82
- HTTP *Hypertext Transfer Protocol.* 116–118, 136, 137, 157, 173
- IA *Intelligence Artificielle.* iii, xxvii, xxviii, 14, 22, 38, 39, 43, 45, 46, 51, 54, 55, 73, 87, 90, 97, 107, 108, 118, 119, 121–123, 127, 129, 145, 148–150
- IIIF *International Image Interoperability Framework.* 15, 23, 29–35, 109–112, 114, 117, 122, 134, 147, 150, 174, 175, 215
- IMAGINE *Laboratoire d’Informatique Gaspard Monge.* 12
- ISCD *Institut des sciences du calcul et des données.* 116, 216

- JPEG *Joint Photographic Experts Group.* 34, 35, 46, 110, 133, 174, 183, 184, 193
- JSON *JavaScript Object Notation.* 30, 89, 114, 117, 157, 159, 160, 173
- MPA *Multiple Page Application.* 138
- OCR *Optical Character Recognition.* xxv, xxvi, 40, 43, 45, 82
- ODD *One Document Does it all.* 41
- RANSAC *RANdom SAmple Consensus.* 66
- SAS *Simple Annotation Server.* 32, 35, 110–112, 114, 122, 123
- SHS *Sciences Humaines et Sociales.* iii, 40, 150
- SI *Systèmes d'Information.* 107, 128, 130, 145, 147
- SPA *Single Page Application.* 138, 139, 141, 216
- SRU *Search/Retrieve via URL.* 41
- SVG *Scalable Vector Graphics.* 41, 64, 67, 68, 87–89, 91, 94, 96, 99–101, 117–120, 133, 135, 148, 150, 157, 163, 183, 184, 193, 198, 216
- SYRTE *Systèmes de Référence Temps-Espace.* xxix, 12, 122
- TAL *Traitemet Automatique du Langage.* 40, 42–45
- TEI *Text Encoding Initiative.* xxvi, 39–41, 45, 92
- UI *User Interface.* 105, 128, 141
- URI *Uniform Resource Identifier.* 32
- URL *Uniform Resource Locator.* 30–32, 114, 117, 136, 157, 158, 160, 161, 170, 173, 174, 176, 177, 220
- URN *Uniform Resource Name.* 95
- UX *User eXperience.* 105, 128, 141
- VHS *Vision artificielle et analyse Historique de la circulation de l'illustration Scientifique.* 3, 12–24, 28, 43, 71, 73, 74, 80, 81, 105, 109, 126, 147, 149, 153, 154, 216
- w3C *World Wide Web Consortium.* 40
- XML *eXtensible Markup Language.* xxvi, 16, 33, 39–41, 64, 67, 88, 89, 92, 94
- YOLO *You Only Look Once.* 61, 62, 77, 78

Table des figures

| | | |
|-----|--|----|
| 1.1 | Impact des labels sur la recherche de similarité. | 20 |
| 2.1 | Structure d'un <i>Manifest IIIF</i> . ¹ | 31 |
| 2.2 | Usage de 'Canvas' dans un <i>Manifest IIIF</i> par Bibliothèque Universitaire de Dresde. | 33 |
| 2.3 | Usage de 'Items' dans un <i>Manifest IIIF</i> par Bibliothèque Universitaire de Yale. | 34 |
| 4.1 | Modèle linéaire et modélisation de la <i>loss</i> | 57 |
| 4.2 | Réduction de la <i>loss</i> et ajustement des poids en fonction. | 57 |
| 4.3 | Fonctionnement d'un CNN. ² | 60 |
| 4.4 | Fonctionnement de la méthode de vectorisation, basée sur les transformers. ³ | 66 |
| 4.5 | Exemple de contenu d'un fichier NPZ | 67 |
| 4.6 | Résultat de la vectorisation. | 68 |
| 5.1 | Détection erronée des marginalia. | 77 |
| 5.2 | Données synthétiques générées par SynDoc. ⁴ | 79 |
| 5.3 | Données synthétiques générées pour l'entraînement du modèle de vectorisation. ⁵ | 80 |
| 5.4 | Données d'entraînement du modèle Segswap. | 80 |
| 6.1 | Édition et comparaison de diagrammes des <i>Éléments d'Euclide</i> par Greg De Young ⁶ , utilisant DRaFT. | 87 |

1. M. Appleby, T. Crane, R. Sanderson, J. Stroop et S. Warner, *Presentation API 3.0*, en, URL : <https://iiif.io/api/presentation/3.0/> (visité le 01/09/2024).

2. An *Intuitive Explanation of Convolutional Neural Networks*, en, août 2016, URL : <https://ujjwalkarn.me/2016/08/11/intuitive-explanation-convnets/> (visité le 11/04/2024).

3. S. Kallel, S. Trigg, S. Albouy, M. Husson et M. Aubry, *Historical Astronomical Diagrams Decomposition in Geometric Primitives*, arXiv :2403.08721 [cs], mars 2024, DOI : 10.48550/arXiv.2403.08721.

4. J. Norindr, *Le traitement des sources historiques par la vision artificielle : l'exemple des manuscrits d'astronomie de tradition ptoléméenne*, fr, mémoire de master, École Nationale des Chartes, 2023, URL : <https://dumas.ccsd.cnrs.fr/dumas-04255677> (visité le 02/02/2024), p.46.

5. Figure issue de la présentation de Syrine Kalelli à l'occasion de la conférence EIDA 2024 : *EiDA Conference 2024*, en-US, URL : <https://eida.hypotheses.org/conferences/conference-2024> (visité le 26/05/2024).

6. Gregg De Young, « Diagrams in ancient Egyptian geometry : Survey and assessment », *Historia*

| | | |
|-----|--|-----|
| 6.2 | Préparation de l'édition d'un diagramme mathématiques paléo-babylonien d'une tablette cunéiforme avec le logiciel DRaFT ⁷ | 87 |
| 6.3 | Extraction de représentations sémantiquement riches et exploitation. | 90 |
| 6.4 | Modèle de données hypothétique. | 90 |
| 7.1 | Fonctionnement de l'envoi d'un traitement à l'API.* ⁸ | 116 |
| 7.2 | Fonctionnement du système de <i>Threading</i> de l'API. | 118 |
| 7.3 | <i>Workflow</i> schématique.* | 121 |
| 7.4 | Organisation et communication des infrastructures. | 122 |
| 7.5 | Représentation schématique des fonctionnalités et processus de la plate-forme.* | 126 |
| 8.1 | Capture d'écran du formulaire de recherche des <i>Series</i> | 131 |
| 8.2 | Capture d'écran de la visualisation des extractions type <i>dump</i> , les onglets permettent de naviguer entre les interfaces de visualisation des résultats des traitements. | 132 |
| 8.3 | Sélection des <i>Region Sets</i> | 132 |
| 8.4 | Capture d'écran de l'interface de visualisation des similarités. | 133 |
| 8.5 | Capture d'écran de l'interface d'affichage des vectorisations. | 133 |
| 8.6 | Récapitulatif des catégories de similarité et leurs icônes associées. | 134 |
| 8.7 | L'interface permet la validation des extractions. | 135 |
| 8.8 | Fonctionnement schématique d'une SPA. | 139 |
| A.1 | Modèle de données de l'application VHS avant refonte.* | 153 |
| A.2 | Nouveau modèle de données de EIDA appliqué à VHS.* | 154 |
| A.3 | Nouveau modèle de données en vue de la refonte de l'application AIKON.* . | 155 |
| B.1 | Capture d'écran du formulaire d'envoi d'un traitement dans AIKON. | 159 |
| C.1 | Affichage des extractions, vue <i>dump</i> et avec la page de manuscrit en regard. | 183 |
| C.2 | Interface de manipulation des fichiers SVGS, avec exemple d'image modifiée téléchargée. | 193 |
| C.3 | Page d'accueil de la plateforme AIKON | 199 |

Mathematica, 36–4 (nov. 2009), p. 321-373, DOI : 10.1016/j.hm.2009.02.004, p.193.

7. Cette capture d'écran est issue du travail de thèse de A. Reynaud, *Les diagrammes mathématiques paléo-babylonien : catalogue, propriétés, matérielles, rôles dans les raisonnements*, These en préparation, Université Paris Cité, 2017, URL : <https://theses.fr/s197780> (visité le 26/08/2024), p.293.

8. Les schémas suivis d'un astérisque sont prélevés ou inspirés de la documentation technique interne créée par les développeur.ses du projet EIDA et de l'ISCD.

Table des matières

| | |
|---|--------|
| Résumé | iii |
| Remerciements | v |
| Introduction | xxiii |
| 0.1 Mise en contexte | xxiv |
| 0.1.1 Un changement de paradigme | xxiv |
| 0.1.2 De nouveaux outils | xxv |
| 0.1.3 Perspectives ouvertes pour l'image | xxvi |
| 0.1.4 Structurer l'information | xxvi |
| 0.1.5 Les données pour des ponts interdisciplinaires | xxvii |
| 0.1.6 <i>Pipelines</i> et <i>workflows</i> : spécificités de l'implémentation de l'IA | xxviii |
| 0.2 Mission de stage | xxviii |
| 0.3 Problématisation | xxix |
| 0.4 Annonce du plan | xxix |
| I Chaîne de traitement de la donnée visuelle : enjeux technologiques et disciplinaires | 1 |
| 1 EIDA : Contexte institutionnel et scientifique | 5 |
| 1.1 Contexte disciplinaire | 5 |
| 1.1.1 Ptolémée : modèle et transmission | 6 |
| 1.1.2 Le diagramme vecteur de connaissances | 8 |
| 1.2 Genèse et écosystème du projet | 12 |
| 1.2.1 Contributeurs et partenaires | 12 |
| 1.2.2 Projets précédents | 13 |
| 1.3 Objectifs | 14 |
| 1.3.1 Des outils d'exploration et d'analyse d'un large corpus | 14 |
| 1.3.2 Développement d'une interface commune | 17 |
| 1.3.3 Corpus | 20 |
| 1.3.4 Modèle de données | 22 |

| | |
|--|-----------|
| 2 Open-data et enjeux d'interopérabilité : la standardisation technique à grande échelle | 25 |
| 2.1 Ouverture des données : accès aux sources et partage des résultats de la recherche | 26 |
| 2.1.1 Enjeux de la récupération des sources numérisées | 26 |
| 2.1.2 Projet d'une plateforme web | 28 |
| 2.2 IIIF | 29 |
| 2.2.1 Les APIs de IIIF | 29 |
| 2.2.2 Anatomie d'un <i>Manifest</i> | 30 |
| 2.2.3 Avantages et limites du cadre IIIF | 32 |
| 2.2.4 Particularité de l'utilisation de IIIF au sein de la plateforme | 35 |
| 3 État de l'art : IA et traitement du volume | 37 |
| 3.1 Formats standards | 40 |
| 3.2 La spécialisation des modèles | 42 |
| 3.3 La normalisation des données d'annotation | 43 |
| II Mise en œuvre et exploitation de la <i>Computer Vision</i> | 49 |
| 4 Les modèles de <i>Computer Vision</i> | 53 |
| 4.1 Les réseaux de neurones, ou la généralisation prise au sens mathématique . | 54 |
| 4.2 Des traitements et des architectures diverses | 59 |
| 4.2.1 YOLOv5 pour la détection : performance et légèreté | 60 |
| 4.2.2 Similarité | 62 |
| 4.2.3 La vectorisation : les transformers | 64 |
| 5 La donnée : générique ou caractéristique, réelle ou synthétique | 71 |
| 5.1 Partir d'un modèle généraliste | 72 |
| 5.1.1 Utilisation de modèles off-the-shelf | 72 |
| 5.1.2 fine-tuner le modèle | 74 |
| 5.1.3 Construction d'une vérité de terrain | 74 |
| 5.1.4 Choisir et évaluer le modèle | 75 |
| 5.2 Quelles données pour le spécialiser? | 77 |
| 5.2.1 <i>datasets</i> synthétiques | 78 |
| 5.2.2 Les données réelles | 80 |
| 5.2.3 L'oeil de la machine : avantages et limites | 83 |
| 6 Un outils d'édition des diagrammes pour la cohésion des pratiques | 85 |
| 6.1 Perspectives ouvertes par EIDA | 85 |
| 6.1.1 Transcription automatique des diagrammes | 86 |

| | | |
|-------------------|--|------------|
| 6.1.2 | Idiomes et layers : des formats computables et des couches sémantiques | 88 |
| 6.1.3 | Une plateforme qui porte une méthode | 90 |
| 6.2 | Une problématique d'unification des pratiques | 91 |
| 6.2.1 | Sortir des pratiques aléatoires | 91 |
| 6.2.2 | L'édition numérique : une solution ? | 94 |
| 6.2.3 | Exigences et dissensus | 96 |
| III | Une plateforme pour une méthode reproductible et transposable | 103 |
| 7 | Processus et Fonctionnalités | 109 |
| 7.1 | Module de base | 109 |
| 7.1.1 | Description des données | 110 |
| 7.1.2 | Principe du Traitement | 111 |
| 7.1.3 | Extraction manuelle des zones d'image | 112 |
| 7.2 | Des tâches complémentaires : parallélisation du <i>workflow</i> | 113 |
| 7.2.1 | Modules optionnels | 113 |
| 7.2.2 | Alternance d'étapes automatiques et d'étapes manuelles | 118 |
| 7.3 | (Infra)structures | 121 |
| 7.3.1 | Hardware : une API sur le GPU | 121 |
| 7.3.2 | Software : des modules séparés | 123 |
| 8 | Interfaces | 127 |
| 8.1 | Faire le lien avec les chercheur.ses | 128 |
| 8.1.1 | Avantages des interfaces graphique pour la recherche | 129 |
| 8.1.2 | Construction logique de l'interface utilisateur.rice | 129 |
| 8.2 | Choix techniques | 136 |
| 8.2.1 | Interface Admin Django vs. vues personnalisées | 136 |
| 8.2.2 | Utilisation d'un <i>framework front</i> | 137 |
| 8.2.3 | Le design d'interface | 141 |
| Conclusion | | 147 |
| Annexes | | 153 |
| A | Évolution du modèle de données | 153 |
| A.1 | Modèle de données initial de l'application VHS | 153 |
| A.2 | Modèle de données de l'application VHS/EIDA | 154 |

| | |
|--|------------|
| A.3 Modèle de données de l'application AIKON | 155 |
| B Module vectorisation : description des développements | 157 |
| B.1 Workflow | 157 |
| B.2 Settings | 158 |
| B.2.1 AIKON | 158 |
| B.2.2 Discover-Demo | 158 |
| B.3 Envoi d'un traitement de vectorisation | 158 |
| B.3.1 Initialisation de la requête | 158 |
| B.3.2 Création du JSON | 159 |
| B.4 Discover-Demo : module vectorization | 162 |
| B.4.1 Réception à l' <i>endpoint</i> de l'API : | 162 |
| B.4.2 Configuration de la tâche de fond | 163 |
| B.4.3 Inférence avec le modèle | 163 |
| B.5 Réception des résultats dans AIKON | 170 |
| C Interfaces : développement des vues personnalisées | 173 |
| C.1 <code>views.py</code> pour traiter la logique des requêtes | 173 |
| C.2 <code>url.py</code> pour le routage des URLs | 176 |
| C.3 Les <i>templates</i> | 177 |
| C.4 Autres développements | 183 |
| C.5 Manipuler le format SVG | 193 |
| C.6 Nouvelles interfaces | 199 |
| D Interfaces : documentation | 201 |
| Acronymes | 213 |