

ÉCOLE NATIONALE DES CHARTES

Ségolène Albouy

licenciée ès lettres

Médiation des données de la recherche

**Élaboration d'une plateforme en ligne pour
une base de tables astronomiques anciennes**

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

2019

Résumé

Ce mémoire a été réalisé à l'occasion d'un stage à l'Observatoire de Paris du mois d'avril à août 2019, dans le cadre du Master « Technologies numériques appliquées à l'histoire » de l'École des chartes. Il porte sur les procédés de médiation pouvant être mis en œuvre dans la valorisation d'un corpus de recherche numérique. Il expose les étapes de conception d'une plateforme publique pour le projet de recherche DISHAS, depuis l'analyse du corpus jusqu'à la réalisation technique des interfaces. Le corpus de recherche de DISHAS porte sur les ressources de l'astronomie de tradition ptoléméenne ; il comporte une grande variété de données – ayant trait autant aux mathématiques qu'à la codicologie – pour lesquelles il a fallu concevoir des interfaces et visualisations spécifiques. Ce mémoire vise à apporter une analyse critique des enjeux, méthodes et résultats liés à la conception d'une plateforme publique de recherche.

Mots-clés : histoire des science ; numérique ; tables astronomiques ; expérience utilisateur ; médiation ; visualisation de données ; interface publique ; *front office* ; accessibilité ; corpus de recherche ; Symfony ; PHP ; JavaScript ; ElasticSearch.

Informations bibliographiques : Ségolène Albouy, *Médiation des données de la recherche*, mémoire de master « Technologies numériques appliquées à l'histoire », dir. Thibault Clérice, École nationale des chartes, 2019.

Remerciements

En préambule de ce travail, je tiens à remercier toutes les personnes qui m'ont accompagnée au cours de cette deuxième année de Master, qui fut riche en apprentissages. Ma reconnaissance va en premier lieu à Galla Topalian, qui m'a suivie avec bienveillance toute la durée de ce stage et qui n'a cessé de me prodiguer d'excellents conseils. Je remercie également Matthieu Husson et Galla Topalian de m'avoir fait confiance en me laissant la possibilité d'expérimenter quant aux réalisations qui m'ont été données de faire. Je souhaite aussi adresser mes remerciements à l'équipe d'Histoire des sciences de l'Observatoire de Paris pour m'avoir si bien accueillie. Enfin, je remercie Malcolm Hamelin d'avoir été un si bon camarade pour la durée de ce stage.

Toute ma gratitude va également à l'ensemble de la promotion 2019 du Master « Technologies numériques appliquées à l'histoire » de l'École des chartes, pour le soutien infaillible et collectif au cours de l'écriture de ce mémoire. Toute ma gratitude, enfin, à Lucence Ing pour sa patiente relecture et ses encouragements constants.

Histoire de l'astronomie

About DISHAS Project – Astronomical Tables in Database, URL : <https://dishes.observatoire.fr/about> (visité le 28/07/2019).

BESSOU (Gilles) et LAMY (Jérôme), « Lumières sur l’Univers – Histoire de l’astronomie », dans *Astrophysique sur Mesure - Observatoire de Paris*, URL : https://media4.observatoire.fr/public/ressources_lu/pages_antiquite-moyen-age/ha-epicycles-excentriques.html (visité le 29/06/2019).

CHARMETANT (Églantine), *Modélisation, encodage et valorisation d’un manuscrit astronomique : l’exemple du Latin 7281*, Mémoire de Master, École nationale des chartes, 2018.

EVANS (James), *Histoire et pratique de l’astronomie ancienne*, avec la coll. de Michel-Pierre Lerner, Concetta Luna et Denis Savoie, trad. par Alain Philippe Segonds, Les Belles Lettres, 2016.

EVANS (James) et WULF FRIEDLANDER (Michael), « History of Astronomy », dans *Encyclopedia Britannica*, URL : <https://www.britannica.com/science/astronomy> (visité le 17/06/2019).

GIESEN (Jürgen), « Heliocentric and Geocentric Motion », dans *Physics and Astronomy*, URL : <http://jgiesen.de/geocentric/index.html> (visité le 25/04/2019).

History of Astronomical Sciences in India, URL : <http://www.hamsi.org.nz/> (visité le 05/08/2019).

HUSSON (Matthieu), « L’astronomie alphonsoine dans l’*Expositio intentionis regis Alfonsii circa tabulas ejus de Jean de Murs* », *Archives d’histoire doctrinale et littéraire du Moyen Age*, Tome 78–1 (2011), p. 229-245.

HUSSON (Matthieu) et KREMER (Richard L.), « How Do Writings in the Early Astral Sciences Reveal Mathematical Practices and Practitioners? », *Centaurus*, 58–1–2 (2016), p. 1–5.

« Le Monde en sphères », dans *Expositions numériques – BnF*, URL : <http://expositions.bnf.fr/monde-en-spheres/index.html> (visité le 16/08/2019).

MERCIER (Raymond), *Studies on the Transmission of Medieval Mathematical Astronomy*, Ashgate Publishers, Burlington, VT, 2004.

MONIER (Williams), *A Sanskrit-English Dictionary*, Clarendon, Oxford, 1872.

- MONTELLE (Clemency), *Chasing Shadows : Mathematics, Astronomy, and the Early History of Eclipse Reckoning*, Johns Hopkins University Press, 2011.
- POULLE (Emmanuel), « Les instruments astronomiques de l'Occident latin aux XIe et XIIe siècles », *Cahiers de Civilisation Médiévale*, 15–57 (1972), p. 27-40.
- *Les astronomes parisiens au XIV^e siècle et l'astronomie alphonosine*, F. Paillart, Abbeville, France, 2005.
- POULLE (Emmanuel) et GINGERICH (Owen), « Les positions des planètes au Moyen Age : application du calcul électronique aux tables alphonsoines », *Comptes rendus des séances de l'Académie des Inscriptions et Belles-Lettres*, 111–4 (1967), p. 531-548, URL : https://www.persee.fr/doc/crai_0065-0536_1967_num_111_4_12173 (visité le 27/06/2019).
- Ptolemaeus Arabus et Latinus*, URL : <https://ptolemaeus.badw.de/start> (visité le 10/04/2019).
- SELIN (Helaine), SUN (Xiaochun) et KING (David A.), *Astronomy across Cultures : The History of Non-Western Astronomy*, Kluwer Academic Publishers, Dordrecht ; Boston, 2000.
- Shaping a European Scientific Scene : Alfonsine Astronomy*, URL : <https://alfa.hypotheses.org/> (visité le 08/06/2019).
- VALLERIANI (Matteo), KRAÜTLI (Florian) et WINTERGRÜN (Dirk), « De Sphaera : Modeling Data and Analyzing Diffusion Using Early Modern Cosmological Treatises », dans *Max-Planck-Institut Für Wissenschaftsgeschichte*, 2017, URL : <https://www.mpiwg-berlin.mpg.de/project/sacrobosco-sphaera> (visité le 12/04/2019).

Sciences Humaines & numérique

BARCENILLA (Javier) et MAURICE CHRISTIAN BASTIEN (Joseph), « L'acceptabilité des nouvelles technologies : quelles relations avec l'ergonomie, l'utilisabilité et l'expérience utilisateur ? », *Presses Universitaires de France*, Le travail humain 72 (2009), p. 311-331.

BERTRAND (Paul), « La révolution souterraine des Digital Humanities : pour une refondation de la critique des sources », dans *Médiéviz*, 2012, URL : <https://medievizmesblog.wordpress.com/2012/12/08/573/> (visité le 05/08/2019).

BERTRAND (Paul), MAIREY (Aude), GUYOTJEANNIN (Olivier), GUERREAU-JALABERT (Anita), EDDÉ (Anne-Marie) et BURGHART (Marjorie), « L'historien médiéviste et la pratique des textes : les enjeux du tournant numérique », *Actes des congrès de la Société des historiens médiévistes de l'enseignement supérieur public*, 38–1 (2007), p. 273-301, DOI : 10.3406/shmes.2007.1958.

CAVALIÉ (Étienne), CLAVERT (Frédéric) et LEGENDRE (Olivier), « Expérimenter les Humanités numériques : des outils individuels aux projets collectifs », dans *DesLibris*, Montréal, 2018 (Parcours numériques), URL : <https://www.deslibris.ca/ID/479645> (visité le 04/08/2019).

CLAVERT (Frédéric), « Design & Digital Humanities : le design comme méthode pour les Humanités Numériques ? », dans *THATCamp Paris*, URL : <https://tcp.hypotheses.org/849> (visité le 10/08/2019).

CLAVERT (Frédéric), HEIMBURGER (Franziska) et BERRA (Aurélien), « L'historien programmeur ? », dans *THATCamp Paris 2012 : Non-actes de la non-conférence des humanités numériques*, Paris, 2012 (La Non-Collection), URL : <http://books.openedition.org/editionsmsn/305> (visité le 04/08/2019).

DACOS (Marin) et MOUNIER (Pierre), *État des lieux et positionnement de la recherche française dans le contexte international*, Laffont, Paris, 2013.

FARGE (Arlette), *Le goût de l'archive*, Éditions du Seuil, Paris, 1997 (Points Histoire, 233).

FLÜELER (Christoph), *Digital Manuscripts as Critical Edition*, 30 juin 2015, URL : <https://schoenberginstitute.org/2015/06/30/digital-manuscripts-as-critical-edition/> (visité le 08/06/2019).

- GENET (Jean-Philippe), « Histoire, Informatique, Mesure », *Histoire & Mesure*, 1–1 (1986), p. 7-18.
- HÉGARAT (Thibault Le), « Apprendre à vulgariser ? », dans *Devenir historien-ne*, URL : <https://devhist.hypotheses.org/3413> (visité le 16/08/2019).
- HEIMBURGER (Franziska) et RUIZ (Émilien), « Has the Historian's Craft Gone Digital ? Some Observations from France », *Diacronie. Studi di Storia Contemporanea* (, 10[2012]), URL : <http://journals.openedition.org/diacronie/2795> (visité le 04/08/2019).
- HOOLAND (Seth van), GILLET-GOINARD (Florence), HENGCHEN (Simon), DE WILDE (Max) et EBERLE-SINATRA (Michael), *Introduction aux humanités numériques : méthodes et pratiques : sciences humaines et sociales*, De Boeck supérieur, Louvain-la-Neuve, 2016 (Méthodes en sciences humaines).
- JÄNICKE (Stefan), FRANZINI (Greta), CHEEMA (Muhammad Faisal) et SCHEUERMANN (Gerik), « On Close and Distant Reading in Digital Humanities : A Survey and Future Challenges », dans *EuroVis*, 2015.
- Le concept d'archivage numérique pérenne / CINES*, URL : <https://www.cines.fr/archivage/un-concept-des-problematiques/le-concept-darchivage-numerique-perenne/> (visité le 19/08/2019).
- LEIMGRUBER (Matthieu), « L'informatique a-t-elle transformé les pratiques des historien·ne·s ? », *Histoire en pratique(s)* UNIL (2003), URL : https://www.unil.ch/files/live/sites/hist/files/shared/histoirepratique/enpratique_outilsinfos.pdf.
- MASSOT (Marie-Laure), « Dessiner les acteurs en humanités numériques : la recherche en dialogue », *HAL – Archives ouvertes* (, juin 2018), URL : <https://hal.archives-ouvertes.fr/hal-01794185> (visité le 10/08/2019).
- MILLERAND (Florence), « La science en réseau », *Revue d'anthropologie des connaissances*, Vol. 6, n° 1–1 (25 avr. 2012), p. 163-190, URL : <https://www.cairn.info/revue-anthropologie-des-connaissances-2012-1-page-163.htm> (visité le 04/08/2019).
- NAKALA par Huma-Num*, URL : <https://www.nakala.fr/> (visité le 05/08/2019).
- OBERHAUSER (Pierre-Nicolas), « Des collaborations équivoques », *Revue d'anthropologie des connaissances*, Vol. 10, n° 4–4 (23 déc. 2016), p. 557-586, URL : <https://www.cairn.info/revue-anthropologie-des-connaissances-2016-4-page-557.htm> (visité le 04/08/2019).
- Plan de gestion de données – Recommandations à l'ANR*, 2019, URL : <https://www.ouvrirlascience.fr/plan-de-gestion-de-donnees-recommandations-a-lanr> (visité le 06/08/2019).
- POTIN (Yann), « Institutions et pratiques d'archives face à la « numérisation ». Expériences et malentendus », *Revue d'histoire moderne contemporaine*, n° 58-4bis-5

- (2011), p. 57-69, URL : <https://www.cairn.info/revue-d-histoire-moderne-et-contemporaine-2011-5-page-57.htm> (visité le 04/08/2019).
- POUYLLAU (Stéphane), GINOUVÈS (Véronique) et GRAS (Isabelle), « La diffusion numérique des données en SHS - Guide de bonnes pratiques éthiques et juridiques », *Presse Universitaires de Provence* (, oct. 2018), URL : <https://hal-amu.archives-ouvertes.fr/hal-01974527> (visité le 13/08/2019).
- Projet Mirador*, URL : <https://projectmirador.org/> (visité le 06/08/2019).
- « Règlement (UE) 2016/679 Du Parlement Européen et Du Conseil Du 27 Avril 2016 », dans *CNIL – Règlement Général sur la Protection des Données*, URL : <https://www.cnil.fr/fr/reglement-europeen-protection-donnees> (visité le 13/08/2019).
- RENNEVILLE (Marc) et POUYLLAU (Stéphane), « Huma-Num. La très grande infrastructure de recherche pour les humanités numériques », *Lettre de l'InSHS* (, sept. 2013), URL : https://www.huma-num.fr/sites/default/files/extrait_infoinshs_25.pdf (visité le 22/07/2019).
- RYGIEL (Philippe), « L'ordinateur, le réseau et l'écriture de l'histoire », *Materiaux pour l'histoire de notre temps*, N° 82–2 (2006), p. 75-79, URL : <https://www.cairn.info/revue-materiaux-pour-1-histoire-de-notre-temps-2006-2-page-75.htm> (visité le 04/08/2019).
- SEALE (Yvonne), *A Beginner's Guide to Digital Palaeography of Medieval Manuscripts*, 26 mars 2016, URL : <https://yvonneseale.org/blog/2016/03/26/dh-palaeography/> (visité le 08/06/2019).
- SIMANOWSKI (Roberto), *Digital Humanities and Digital Media : Conversations on Politics, Culture, Aesthetics and Literacy*, Open Humanities Press, 2016 (FibreCulture Books).
- SINATRA (Michaël E.) et VITALI-ROSATI (Marcello), « Histoire des Humanités numériques », dans *Pratiques de l'édition Numérique*, dir. Michael E. Sinatra, Montréal, 2014 (Parcours Numérique), chap. 3, p. 49-60, URL : <http://books.openedition.org/pum/317> (visité le 04/08/2019).
- Textual / EADH - The European Association for Digital Humanities*, URL : <http://eadh.org/textual> (visité le 08/06/2019).
- VALENCE (David), « L'historien et les chiffres », *La Vie des idées* (, 13 mars 2008), URL : <http://www.laviedesidees.fr/L-historien-et-les-chiffres.html> (visité le 08/06/2019).
- WILMS (Lotte), DERVEN (Caleb), O'DWYER (Liam), LINGSTADT (Kirsty) et VERBEKE (Demmy), *Europe's Digital Humanities Landscape : A Study From LIBER's Digital Humanities & Digital Cultural Heritage Working Group*, Zenodo, 2019, URL : <https://zenodo.org/record/3247286#.XR55YpMzb0Q> (visité le 05/07/2019).
- ZYSBERG (André), « Impact de l'informatique sur la recherche historique », *Matériaux pour l'histoire de notre temps*, 10–1 (1987), p. 35-37.

Visualisations de données

- « Adams' Synchronological Chart », dans *David Rumsey Historical Map Collection*, URL : <https://www.davidrumsey.com/luna/servlet/detail/RUMSEY~8~1~226099~5505934:Composite--Adams--Synchronological-> (visité le 28/08/2019).
- « Cartes », dans *Musée National de l'histoire de l'immigration*, URL : <http://www.histoire-immigration.fr/questions-contemporaines/cartes> (visité le 19/08/2019).
- CÔME (Etienne), *Portfolio / Étienne Côme*, URL : <https://www.comeetie.fr/#portfolio> (visité le 18/08/2019).
- FRY (Ben), *Visualizing Data*, O'Reilly Media, Sebastopol, California, 2008.
- GROLEMUND (Garrett) et WICKHAM (Hadley), *R for Data Science*, O'Reilly Media, Sebastopol, California, 2017, URL : <https://r4ds.had.co.nz/> (visité le 26/08/2019).
- « Let Viewers Change Chart Type », dans *Blog amCharts*, 2018, URL : <https://www.amcharts.com/dataviz-tip-30-let-viewers-change-chart-type/> (visité le 16/08/2019).
- MEEKS (Elijah), « Visualizing Databases », dans *Digital Humanities Specialist*, 2011, URL : <https://dhs.stanford.edu/spatial-humanities/visualizing-databases/> (visité le 29/04/2019).
- PLANTIN (Jean-Christophe), « L'avènement de la carte comme médiation. Généalogie des rencontres entre cartographie et théories de l'information », *Questions de communication n°25* (, 31 août 2014), p. 309-326, URL : <http://journals.openedition.org/questionsdecommunication/9050> (visité le 22/08/2019).
- POPOVA (Maria), « How to Map Time : A Visual History of the Timeline », dans *The Atlantic*, 2012, URL : <https://www.theatlantic.com/entertainment/archive/2012/02/how-to-map-time-a-visual-history-of-the-timeline/252825/> (visité le 18/08/2019).
- « Research – Data Visualization », dans *Digital Humanities*, URL : <https://dh.fbk.eu/research> (visité le 16/08/2019).
- « The Timemap of World History : The Online Atlas and Encyclopedia », dans *TimeMaps*, URL : <https://www.timemaps.com/> (visité le 18/08/2019).
- Wappalyzer - JavaScript Graphics*, URL : <https://www.wappalyzer.com/categories/javascript-graphics> (visité le 26/08/2019).

« World History Maps & Timelines », dans *GeaCron*, URL : <http://geacron.com/home-en/> (visité le 19/08/2019).

YAU (Nathan), *Visualize This : The FlowingData Guide to Design, Visualization, and Statistics*, Wiley Publisher, Indianapolis, Ind, 2011.

Technologies et méthodes

Documentation de la bibliothèque AmCharts 4, URL : <https://www.amcharts.com/docs/v4/> (visité le 23/06/2019).

DROUILLAT (Benoît), *Le design interactif : du Web design aux objets connectés*, Dunod, Paris, 2016.

Elastic Stack and Product Documentation, URL : <https://www.elastic.co/guide/index.html> (visité le 04/08/2019).

Elastica Documentation, URL : <https://elastica.io/> (visité le 04/08/2019).

« Etudiez le fonctionnement d’Elasticsearch », dans *OpenClassrooms*, URL : <https://openclassrooms.com/fr/courses/4462426-maitrisez-les-bases-de-donnees-nosql/4474691-etudiez-le-fonctionnement-d-elasticsearch> (visité le 20/08/2019).

GARRETT (Jesse James), « Additional Resources », dans *The Elements of User Experience*, URL : <http://www.jjg.net/elements/resources/> (visité le 13/08/2019).

— *Les éléments de l’expérience utilisateur : Placer l’utilisateur au cœur de la conception des produits web et mobiles*, Pearson, Paris, 2011.

HAMELIN (Malcolm), *Observatoire de Paris – Projet DISHAS*, Rapport de stage, Paris, 2019, p. 33.

KING (Rochelle), CHURCHILL (Elizabeth F.) et TAN (Caitlin), *Designing with Data : Improving the User Experience with A/B Testing*, O'Reilly Media, Inc, Sebastopol, CA, 2017.

LAUGIER (Anne-Sophie) et FRANCE (Louis-Baptiste), « Les bases NoSQL orientées colonnes », dans *IllustraData*, 2017, URL : <http://www.illustradata.com/bases-nosql-orientees-colonnes-quest-cest/> (visité le 25/08/2019).

PHP : Documentation, URL : <https://www.php.net/docs.php> (visité le 04/08/2019).

« Reinventing the Wheel », dans *Wikipedia*, URL : <https://en.wikipedia.org/wiki/Reinventing-the-wheel> (visité le 10/08/2019).

SANNIÉ (Colette), « Prendre en compte l’expérience utilisateur : une opportunité au delà de la contrainte » (, 12 sept. 2017), URL : <https://dumas.ccsd.cnrs.fr/dumas-01804558> (visité le 06/08/2019).

Stack Overflow – Where Developers Learn, Share, & Build Careers, URL : <https://stackoverflow.com/> (visité le 04/08/2019).

Symfony 4.3 Documentation, URL : <https://symfony.com/doc/current/index.html> (visité le 04/08/2019).

WANYOIKE (Michael), « History of Front-End Frameworks », dans *LogRocket Blog*, URL : <https://logrocket.com/blog/history-of-frontend-frameworks/> (visité le 11/04/2019).

Introduction

Le recours aux technologies numériques en sciences humaines a modifié les procédés d'analyse des sources primaires de la recherche : les documents, qui auparavant fournissaient seuls la matière de l'étude, doivent subir une étape liminaire de numérisation en vue d'un traitement informatique. Le processus de numérisation transforme le matériel documentaire, parfois complexe et hétérogène, en données manipulables par l'ordinateur. Il constitue un état intermédiaire du document, intelligible pour un algorithme, mais souvent hermétique au public. Cette étape entre source primaire et production scientifique ajoute en effet un niveau d'abstraction à l'objet étudié, abstraction tant du point de vue de sa nature, que de sa structuration : si elle est nécessaire à la computation par ordinateur, l'abstraction rend également le matériel documentaire davantage polyvalent.

La segmentation de l'information en données permet notamment de rendre les corpus plus accessibles et réutilisables : accessibles, car ils peuvent ainsi transiter par les canaux de diffusion d'Internet, et réutilisables, car un même document numérisé pourra subir autant de traitements informatiques qu'il existe de volontés de l'analyser. Rendre ainsi partageables les données donne l'occasion d'accroître conjointement la visibilité et les dimensions des corpus de recherche. Cet enjeu de valorisation des données brutes est au cœur du mouvement de l'*open data* qui prône l'ouverture des données pour leur libre accès et exploitation. Il vise à la diffusion des sources numériques auprès du grand public autant qu'à l'enrichissement des ressources de la communauté scientifique de manière à favoriser l'innovation ouverte. Le recours à des standards et normes permet en outre à des projets variés de collaborer autour de corpus communs par la production de données interopérables.

En établissant 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. Les données standardisées constituent un terrain d'entente, tout particulièrement profitable à l'étude de corpus transverses et aux champs de recherche pluridisciplinaires¹. En histoire des sciences en

1. « [...] c'est à un déplacement des frontières que nous assistons : frontières entre les disciplines, dont le recours commun aux instruments informatiques peut faciliter le rapprochement ; frontières entre

particulier, la recherche recouvre une expertise dans une grande variété de domaines, allant des mathématiques à la codicologie, dont le dialogue peut être facilité par le médium informatique : l'exposition des données de recherche et la mise en commun de sources constituent, en ce sens, des moyens pour mettre au jour des interactions entre des phénomènes complexes. Dans le domaine de l'histoire de l'astronomie dans une plus forte mesure, où les pratiques des astronomes ont été largement influencées par la circulation continue des savoirs depuis l'Antiquité, la mutualisation des systèmes d'information numériques peut être d'une utilité majeure si l'on veut envisager de considérer le développement de cette science dans son entièreté.

Dans l'océan du réseau, de nouveaux territoires de données sont ainsi apparus. Il s'agit maintenant de créer les conditions de possibilité de construire des échanges entre eux. Pour passer au stade de la compétition coopérative, les données de la recherche doivent être ouvertes, documentées et contextualisées, interopérables et — autant que possible — libres d'accès dans le respect du droit.²

Ainsi, le processus de numérisation engendre un nouvel état des sources, un état susceptible d'être le support d'une analyse scientifique, mais aussi d'être valorisé en tant que tel. La valorisation de la donnée repose sur plusieurs qualités qui œuvrent ensemble : l'accessibilité, la réutilisabilité et l'interopérabilité sont trois des grands principes liés aux FAIR (*Findable Accessible Interoperable Reusable*) *data* et sont complétés par « trouvabilité » de la donnée. Cette dernière propriété repose sur une multitude d'aspects – architecture de l'information, *design* des interfaces utilisateur, optimisation pour les moteurs de recherche – qui assurent la visibilité des corpus numérisés. C'est un effort de médiation qui doit donc être fourni : médiation autour de la donnée qui, isolée, est dépouillée de sa signification, médiation autour des bases les contenant, qui, par leur degré d'abstraction au regard de la source, doivent être rendue explicites et contextualisées, et médiation autour des questions de recherche, qui, bénéficiant de la plateforme de diffusion qu'offre le numérique, peuvent fructifier de l'intérêt porté par le public. En plus de produire un discours sur le document, la recherche dans les humanités numériques s'ingénie donc à produire un discours sur les données, qui constituent un vivier d'analyse autant qu'une manne à valoriser.

Cette attention portée à la valorisation de l'état numérique de la source œuvre, au-delà de la simple mise à libre disposition des corpus, à leur pérennité. Sans un travail de

chercheurs en sciences humaines et ingénieurs, puisqu'un dialogue fécond se noue entre enjeux scientifiques et possibilités techniques. » Marin Dacos et Pierre Mounier, *État des lieux et positionnement de la recherche française dans le contexte international*, Laffont, Paris, 2013, p. 3

2. Marc Renneville et Stéphane Pouyllau, « Huma-Num. La très grande infrastructure de recherche pour les humanités numériques », *Lettre de l'InSHS* (, sept. 2013), URL : https://www.huma-num.fr/sites/default/files/extrait_infoinshs_25.pdf (visité le 22/07/2019)

médiation, la masse considérable des données produites par la communauté scientifique en SHS peut rester inaccessible et perdue dans des systèmes d'informations obsolètes. La diffusion des contenus numériques opère d'un double mouvement qui irrigue à la fois les travaux des spécialistes, mais renouvelle l'accès de tous au patrimoine culturel. En élaborant des plateformes sur mesure pour exposer les données de la recherche, il est possible de faire surgir de nouvelles questions de recherche en même temps que de créer des vecteurs d'analyse alternatifs.

C'est dans cette lignée que s'inscrit DISHAS (*Digital Information System for the History of Astral Sciences*) qui tâche de fédérer les projets de recherche en astronomie ancienne³. L'astronomie est issue d'une tradition continue vieille de près de 4000 ans qui transcende les cultures et les langues. Cette richesse rend toute étude globale malaisée, compte tenu de la diversité des sources et de la variété des approches possibles ; un effort d'exposition des données dans une perspective de mutualisation des travaux de recherche était donc nécessaire. C'est pour répondre à cette problématique d'inclusivité qu'a été conçu DISHAS : le projet est né d'une volonté mise en commun des outils et des ressources pour l'étude non seulement de l'astronomie de tradition ptoléméenne, mais aussi des sources en histoire des sciences. Son corpus est pluriel et nécessite d'être abordé avec des méthodologies au carrefour de la bibliographie matérielle, de l'histoire et des mathématiques. Plusieurs ambitions découlent de cette étude, en particulier l'examen des pratiques des astronomes et la description de milieux intellectuels liés à la transmission des textes astronomiques. Les sources étudiées couvrent des matériaux aussi divers que des instruments, des tables astronomiques, des documents iconographiques ou encore des textes théoriques : toutefois, en tant que projet pilote, DISHAS se concentre sur les tables astronomiques qui, par leur grande complexité, constituent des ressources d'une richesse inégalée pour l'étude⁴.

Avec une équipe dispersée sur plusieurs continents et s'intéressant à des aspects différents des sources astronomiques, il a été important de définir des intérêts communs pour façonner des outils profitables à toute la communauté de recherche. Si dans un premier temps, l'attention a été portée à développer des outils aidant l'intégration des données dans la plateforme, en particulier grâce à une interface facilitant la saisie de tables astronomiques, à présent, ce sont les dispositifs permettant la valorisation des données numérisées qui font l'objet de l'attention. L'une des réalisations principales qui en découle

3. « *DISHAS relies on a network of international projects in Chinese, Sanskrit, Arabic, Latin and Hebrew sources.* » *About DISHAS Project – Astronomical Tables in Database*, URL : <https://dishes.obspm.fr/about> (visité le 28/07/2019)

4. « *Thus in addition to their individual contents, astronomical tables viewed more generally can provide unmatched sources for studying the transmission of computational know-how, writing technologies and layouts, theoretical models, and numerical parameters.* » *Ibid.*

est la création d'une plateforme en ligne agrégeant des fonctionnalités développées par les différents membres du projet et exposant les données de recherche de la manière la plus adaptée possible.

C'est pour remplir cette mission que l'équipe chargée du projet DISHAS au sein de l'Observatoire de Paris a fait appel à l'École nationale des chartes, pour son expertise dans la transformation du document vers la donnée. C'est à cette occasion que j'ai participé à la construction de l'application Web du projet, en qualité de stagiaire ingénierie numérique d'avril à août 2019. Les objectifs de mon stage portaient sur l'élaboration des interfaces publiques de la plateforme DISHAS, tant du point de vue de la constitution d'un parcours utilisateur que de la réalisation de visualisations de données sur mesure. L'enjeu résidait dans les différents niveaux de lecture nécessaires à la construction d'une interface à destination autant d'un public de chercheurs que de simples amateurs. Par quelles formes de médiation les données de recherche peuvent-elles être exposées à un public varié, en appréhendant de concert la navigation dans son ensemble et les visualisations ponctuelles ? Comment concevoir la médiation autour de données très spécialisées dans le respect des principes FAIR tout en fournissant au public des outils de compréhension sur la donnée elle-même et d'analyse sur l'interaction des données entre elles ?

Ce mémoire, écrit en vue de la validation de la seconde année du Master TNAH de l'École nationale des chartes, a pour but de présenter les réalisations accomplies lors de ce stage dans une perspective globale et analytique. Une première partie sera consacrée à la présentation du contexte de création de la donnée au sein du projet DISHAS : pour comprendre quelles modalités d'exposition sont adaptées à un objet aussi spécifique que les données astronomiques, il faut se pencher sur les institutions productrices de ces données, sur la nature du corpus produit, aussi bien que sur les méthodes de saisie qui en ont permis la numérisation et sur les structures d'accueil de la donnée en fin de parcours. Dans une deuxième partie, la réflexion se portera les différentes manières de présenter la donnée à l'utilisateur dans un objectif d'en favoriser la consultation, tout d'abord en analysant les enjeux spécifiques liés aux données de recherche en astronomie ancienne, puis en explorant comment les interfaces et visualisations peuvent jouer un rôle dans les missions d'accessibilité, de contextualisation et d'analyse. Enfin, une ultime partie portera sur les questionnements techniques corrélés au déploiement de telles interfaces, notamment le traitement technique à appliquer à la donnée pour la réalisation de visualisations ou encore en assurer l'interopérabilité et la réutilisation.

Première partie

Production, nature et méthode d'acquisition de la donnée

Chapitre 1

Le projet DISHAS

1.1 Présentation du projet

1.1.1 Objectifs du projet

Génèse de DISHAS

La pratique actuelle de l'astronomie est issue du perfectionnement de savoirs transmis de manière continue depuis l'Antiquité : la recherche qui s'attache à étudier le développement de cette science représente ainsi une entreprise vaste qui transcende les cultures et les époques. Pour tenter de mieux comprendre les interactions et influences entre les différentes traditions astronomiques, un corpus suffisamment large et englobant doit donc être constitué ; de même, des outils et méthodes susceptibles d'analyser cette grande variété de sources doivent être mis en œuvre. C'est dans cette perspective d'élargissement des champs de recherche et de mutualisation des ressources qu'a été élaboré le projet DISHAS¹.

Répartie entre un large éventail de disciplines et proposant une grande diversité dans l'approche des sources, la recherche en astronomie ancienne requiert un projet d'envergure pour rassembler les travaux réalisés sur les différentes traditions astronomiques. DISHAS est né de la volonté de fédérer l'ensemble des projets de recherche du domaine, afin de disposer d'une vue d'ensemble sur le développement de cette science et de produire des outils innovants profitables à toute la communauté de recherche : DISHAS est une entreprise collective soutenue par la CHAMA (*Commission for the History of Ancient and Medieval Astronomy*), comité qui a pour but de favoriser l'étude dans le domaine de l'histoire des sciences astreales. Le projet s'appuie notamment sur un réseau de projets internationaux portant sur des corpus chinois, sanskrit, arabe, latin et hébreu. Dans le but de rassembler, plusieurs pistes ont été envisagées pour le développement de DISHAS, notamment celle d'une plateforme d'*open access* pour les publications scientifiques en his-

1. /diśā/ signifie « direction », « point cardinal » en sanskrit.

toire de l'astronomie : la mutualisation des sources de la recherche (*open data*) plutôt que de ses productions a finalement été préférée.

Toutefois, les sources primaires astronomiques revêtent d'une grande variété – textes théoriques et poétiques, diagrammes, instruments de mesures, entre autres – qu'il n'était pas possible d'inclure tout à la fois, les infrastructures numériques à concevoir pour l'étude de ces objets étant trop différentes. Si de futurs développements prévoient l'intégration d'autres sortes de contenus, il a été décidé de ne se concentrer dans un premier temps que sur un seul type de document. Ce sont les tables astronomiques qui ont été choisies pour constituer l'unité fondamentale du projet DISHAS. Les tables – sorte de tableaux de nombres aidant au calcul de positionnement des astres – sont en effet porteuses d'informations incomparables pour les chercheurs. Elles sont une ressource exceptionnelle pour l'étude de la transmission du savoir-faire mathématique, des techniques d'écriture et de mise en page ainsi que de la construction des modèles théoriques. Ces tables ont été construites, compilées et copiées à travers les siècles et les cultures pour répondre à des besoins religieux, politiques ou encore de compréhension du monde. Grâce à des instruments adaptés, elles peuvent offrir d'inépuisables possibilités d'exploration.

Réalisations du projet

La première mission de DISHAS a donc été de constituer une base de données pour la mise en commun des ressources, condition nécessaire à la création un corpus abondant, permettant d'envisager une étude à grande échelle. La modélisation intellectuelle de la donnée est impliquée pour une grande part sur la versatilité d'un corpus numérique : un équilibre a dû être trouvé entre fort besoin de normalisation dans le but de traitements automatisés et importante flexibilité afin d'accueillir différentes formes de données tabulaires².

Pour garantir la compatibilité des dispositifs d'analyse avec les données, la conception de la base et celle des outils doivent être pensées de concert. En regard du modèle de données, plusieurs réalisations techniques ont été élaborées pour le traitement et l'examen du matériel documentaire de la base : on peut citer la création d'une interface graphique pour assister à la transcription de tables, la constitution d'outils pour le calcul dans des systèmes numériques et calendaires anciens, l'élaboration d'un dispositif permettant de générer automatiquement des éditions critiques de tables ou encore le développement d'une bibliothèque pour l'exploration par intelligence artificielle des sources tabulaires. Ces outils ont pour but de gérer tout à la fois, les différents types de nombres, différentes mises en page tabulaires, différentes formes de variantes et différents modes d'édition, garantissant ainsi un maximum d'inclusion au projet.

2. La forte dissemblance des tables astronomiques chinoises et cunéiformes a entraîné leur mise de côté à ce stade du projet.

L'ensemble de ces outils, une fois développés, a vocation à constituer un progiciel qui sera disponible sur la plateforme en ligne de DISHAS. Cette plateforme vise à proposer en outre des interfaces pour la gestion et la consultation des données, constituant une manière alternative de publier la recherche³. L'application Web a été conçue en deux parties : une interface administrateur, uniquement destinée aux chercheurs, pour la saisie et la gestion des données de la base ; et une interface publique de découverte de ces données. L'élaboration et le développement de cette interface publique ont représenté les missions qui m'ont été confiées lors de mon stage.

À l'issue de la réalisation de toutes ces fonctionnalités, le progiciel de DISHAS doit contribuer au renouvellement des productions scientifiques en astronomie ancienne, comme en histoire des sciences. *In fine*, la mise en commun des ressources vise à favoriser l'apparition de techniques de travail innovantes et à amener à de nouvelles façons de mener, de connecter et de découvrir la recherche. Le déploiement d'une plateforme commune renforce également la collaboration, nécessaire à la production de données partagées, et incite à la reformulation des enjeux de recherche au prisme d'intérêts communs. En bout de parcours, une telle entreprise doit aider à faire surgir de nouvelles analyses quant à la circulation des savoirs astronomiques dans l'histoire, grâce à un corpus documentaire suffisamment conséquent.

1.1.2 Collaboration au sein du projet

Les acteurs du projet

Le projet DISHAS est mené en coopération par des équipes provenant d'universités et de laboratoires de recherche situés en Europe, en Amérique du Nord, en Asie et en Océanie : le projet associe à présent quatorze chercheurs issus de différents champs de l'astronomie ancienne. Pour l'instant, la gouvernance de DISHAS est essentiellement accordée à l'équipe du projet ALFA (*Alfonsine Astronomy*) situé à l'Observatoire de Paris, au sein de la société d'histoire des sciences. D'autres projets de recherche ainsi que des chercheurs indépendants constituent le reste des partenaires. Parmi les projets les plus importants, il est possible de citer PAL (*Ptolemaeus Arabus et Latinus*) et HAMSI (*History of Astronomical Sciences in India*).

Le projet PAL, porté notamment par Benno Van Dalen et David Juste, est consacré à l'édition et à l'étude de versions arabes et latines des textes de Ptolémée et de leurs documents connexes⁴ ; le corpus étudié couvre ainsi les œuvres authentiques de Ptolémée – l'*Almageste*, le *Tetrabiblos* et d'autres œuvres mineures – les textes qui lui ont été

3. La transcription de tables manuscrites relève d'un travail de recherche en lui-même, tant la part d'interprétation du document peut être importante. La mise en ligne de tables transcrrites est ainsi une forme de publication scientifique.

4. *Ptolemaeus Arabus et Latinus*, URL : <https://ptolemaeus.badw.de/start> (visité le 10/04/2019)

attribués, mais aussi les commentaires de ces deux catégories. L’ambition du projet est de fournir un inventaire des manuscrits de ces textes et d’en réaliser des éditions. En dernier lieu, des outils pour l’étude sont également développés, en particulier un glossaire Grec-Arabe-Latin et des programmes informatiques d’édition et d’analyse des sources astronomiques. Certains de ces outils pourront à terme intégrer les fonctionnalités proposées par la plateforme DISHAS.

Le projet HAMSI, coordonné par ses membres fondateurs Clemency Montelle et Kim Plofker, est basé en Nouvelle-Zélande et s’intéresse à l’apparition des sciences mathématiques et astronomiques en Inde⁵. Le matériel documentaire étudié est principalement constitué de tables de calcul sanskrites, composées au cours du deuxième millénaire de notre ère. Les membres d’HAMSI s’attachent à décrypter l’évolution stylistique dans les textes Jyotisa⁶, à identifier les interactions entre auteurs, scribes et commentateurs et à révéler les potentielles influences entre ces écrits astronomiques et d’autres genres de littérature sanskrite.

Les membres de ces équipes partenaires viennent régulièrement présenter leurs travaux et conduire des recherches au sein de l’Observatoire de Paris. En outre, ils participent aux débats relatifs à la conception et à l’implémentation de certaines fonctionnalités au sein de la plateforme DISHAS. L’ensemble des partenaires a en particulier été sollicité pour les décisions fondatrices du projet concernant le choix d’une granularité pour la base de données et l’élaboration du modèle conceptuel.

Équipe au sein de l’Observatoire de Paris

DISHAS est né de l’impulsion initiale de Matthieu Husson, chercheur au CNRS, à présent chargé du projet de recherche ALFA consacré à l’astronomie alphonse : il a donc été logique dans un premier temps, de confier la gouvernance de DISHAS à l’équipe localisée à l’Observatoire de Paris. Une importante partie des infrastructures matérielles employée par le projet est donc implantée à Paris. Ainsi, les serveurs utilisés pour la base de données sont pour l’instant situés à l’Observatoire et sont partagés avec les données du projet ALFA. Il a été envisagé faire héberger les données de DISHAS sur des serveurs appartenant à la TGIR Huma-Num, mais des questions de proximité et de stabilité ont favorisé le choix d’un serveur plus local. Le projet DISHAS ne bénéficie pas de financement propre, mais a pour but d’être porté alternativement par les différents projets partenaires, c’est pourquoi certains partages d’équipements sont nécessaires. En outre, la localisation parisienne du projet a contribué à entamer un partenariat avec l’acteur institutionnel

5. *History of Astronomical Sciences in India*, URL : <http://www.hamsi.org.nz/> (visité le 05/08/2019)

6. Science du suivi et de la prévision des mouvements des astres afin de conserver le temps. Williams Monier, *A Sanskrit-English Dictionary*, Clarendon, Oxford, 1872, p. 353

français Huma-Num, permettant l'accès à des services additionnels, dont le versement définitif des données de recherche produites, dans la base RDF⁷ de NAKALA⁸.

L'équipe chargée du projet est partagée entre ingénieurs et chercheurs : Matthieu Husson est le référent parmi les chercheurs. Il fait également le lien avec l'équipe technique, portée par Galla Topalian, cheffe de projet numérique et tutrice lors de mon stage. Durant mon stage, le pôle de technique du projet fut composé de trois autres membres : un ingénieur chargé du développement de l'interface de saisie des tables, Antonin Penon, un stagiaire s'occupant de l'implémentation d'ElasticSearch et de la création d'une API⁹, Malcolm Hamelin, et moi-même. Les interfaces et fonctionnalités numériques de DISHAS sont construites au fur et à mesure, par la consultation régulière de chacune des équipes, afin d'ajuster au mieux les réalisations techniques aux volontés des chercheurs.

1.2 Projets de recherche en histoire de l'astronomie

1.2.1 Le projet ALFA

Approche vis-à-vis des sources

Le projet ALFA, mené par Matthieu Husson, est un organe de l'équipe d'histoire des sciences, elle-même faisant partie du SYRTE (Systèmes de Référence Temps-Espace) ; le projet est financé par l'ERC (*European Research Council*) sur une période allant de 2017 à 2022. Le corpus étudié par ALFA est celui de l'astronomie alphonse, telle que pratiquée entre la seconde partie du XIII^e et le milieu du XVI^e d'abord en Castille puis à Paris ; les sources alphonsoines sont à ce titre un matériau privilégié pour étudier les interactions intellectuelles à l'œuvre dans l'établissement de l'une des premières réalisations scientifiques européennes¹⁰. Plusieurs aspects de l'étude des textes découlent de cet état de fait, notamment une attention particulière portée à la notion de milieu intellectuel¹¹ et à la matérialité du texte en tant que vecteur d'informations sur les pratiques des astronomes anciens.

7. Le RDF (*Resource Description Framework*) est un modèle de base de données en graphe sur lequel s'appuie le Web sémantique

8. « NAKALA permet à des équipes de recherche, qui en font la demande, de déposer leurs données numériques (fichiers texte, son, image, vidéo) dans un entrepôt sécurisé qui assure à la fois l'accessibilité aux données et leur citabilité dans le temps. » *NAKALA par Huma-Num*, URL : <https://www.nakala.fr/> (visité le 05/08/2019)

9. Une API (*Application Programming Interface*) est, pour ce cas d'usage précis, une manière de communiquer avec une base de données sans passer par une interface graphique. Une API peut être utile pour ajouter récupérer de grandes quantités de données brutes.

10. « *Alfonsoine astronomy is arguably among the first European scientific achievements. It shaped a scene for actors like Regiomontanus or Copernicus.* » *Shaping a European Scientific Scene : Alfonsoine Astronomy*, URL : <https://alfa.hypotheses.org/> (visité le 08/06/2019)

11. Espace géographique, sociologique ou chronologique dans lequel un ensemble de pratiques intellectuelles peuvent être identifiées. *Ibid.*

Cet objectif de compréhension des mécanismes qui sous-tendent la transmission des savoirs fait d’ALFA un acteur particulièrement indiqué pour mener les premiers développements du projet DISHAS, qui ambitionne de mettre au jour l’influence des différentes traditions astronomiques entre elles. Néanmoins, il existe un risque d’orienter l’évolution du projet collaboratif vers les hypothèses scientifiques d’ALFA : si la mise en lumière de milieux intellectuels est l’une des volontés de DISHAS, il est important de discerner les choix scientifiques faits au sein des deux projets.

Toutefois, certaines exigences qualitatives imposées au projet ALFA quant aux données sont prises en compte dans le développement de DISHAS. Le caractère mixte des données des deux projets, comprenant des valeurs numériques comme des informations de catalogage, exige une grande rigueur quant aux productions numériques. En outre, le financement de l’ERC constraint à la création de données conformes aux principes FAIR : elles doivent ainsi être « retrouvables », accessibles, interopérables et réutilisables. Ces impératifs de qualité sont appliqués de la même manière aux données du projet DISHAS.

Missions du projet

L’une des principales missions que s’est donné ALFA est l’établissement d’un inventaire de toutes sources primaires contenant des tables alphonines. En retrouvant les tables similaires dans différents manuscrits ou incunables, il est possible d’étudier les provenances et de retracer le parcours qu’a subi une même production intellectuelle. Pour ce faire, les chercheurs s’attachent à répertorier les lieux de conservation et les cotes de ces sources, mais aussi à signaler précisément la présence de tables astronomiques en leur sein : l’ensemble de ces informations est ensuite stocké dans la base de données d’ALFA. Une partie de cet inventaire a été réalisé antérieurement au projet par les scientifiques au cours de leurs travaux ; cela implique que les informations récoltées ne sont pas standardisées et répondent à différentes logiques de recherche. Une forme de « rétro-encodage » a donc été réalisé lors de l’élaboration de la base de données ; le formalisme dans la modélisation de la base n’a pas pu être aussi poussé que lors de la conception de la base de données de DISHAS. En outre, un effort d’uniformisation de données doit en plus être fourni lors de leur numérisation.

Le recensement de ces sources a permis en parallèle de développer, avec la collaboration notamment de Biblissima et du consortium IIIF¹², une interface de visualisation Mirador¹³ qui rassemble toutes les numérisations disponibles sur les bibliothèques numériques des institutions conservant des ouvrages de tradition alphonine. Par ailleurs, le projet ALFA prévoit l’établissement de notices et d’éditions en XML TEI d’extraits d’une

12. Le consortium IIIF (*International Image Interoperability Framework*) vise à faciliter l’accès aux images numérisées faites des sources matérielles.

13. « *Mirador is a configurable, extensible, and easy-to-integrate image viewer, which enables image annotation and comparison of images from repositories dispersed around the world.* » Projet Mirador, URL : <https://projectmirador.org/> (visité le 06/08/2019)

sélection de manuscrits représentatifs du corpus alphonsins. Une partie de ce travail a été entamée l'année dernière lors du stage d'Églantine Charmetant, alors étudiante de Master 2 à l'École des chartes.

En dernier lieu, le suivi du travail scientifique est présenté grâce à un carnet de recherche sur la plateforme en ligne Hypotheses, ce à quoi s'ajoute la construction d'une application Web dans le but de rendre accessible l'inventaire des sources alphonsines, ainsi que les différentes réalisations numériques les concernant. Cet effort de médiation des productions de la recherche est au cœur de projet ALFA, qui s'attache à rendre accessibles les travaux des chercheurs à un large public¹⁴. Toutefois, les fonctionnalités proposées en ligne par ALFA ne permettent pas l'analyse et la manipulation des sources et se cantonnent à une simple mise à disposition.

1.2.2 Autres projets : description des ressources en ligne

Parmi les projets de recherche en astronomie ancienne, certains sont dotés de plateformes en ligne qui proposent à l'utilisateur d'explorer leur fonds numérique et de découvrir leurs travaux. Il est important dans la conception de l'application Web du projet DISHAS de s'assurer de proposer un contenu numérique innovant qui offre au public des fonctionnalités complétant celles des plateformes déjà existantes. Dans cette mesure, l'examen des visualisations proposées, ainsi que des potentielles insuffisances de ces sites s'intéressant à l'histoire de l'astronomie, peut être éclairant pour discerner les besoins dans le cadre d'un projet collaboratif ; la diversité des approches vis-à-vis des sources est un élément à considérer également.

HAMSI Le site internet du projet HAMSI comporte des représentations abondantes de textes astronomiques sanskrit : tant par le biais d'illustrations que par les transcriptions numériques de certaines tables de sinus. De nombreuses ressources permettent également d'éclairer l'utilisateur quant à la nature de ces tables et fournissent des guides pour la transcription des chiffres et l'examen approfondi du matériel documentaire. Des textes introductifs assurent une certaine clarté malgré la complexité du sujet abordé : la plateforme de HAMSI met en contexte et illustre les enjeux de recherche du projet. Des documents annexes, à l'instar de glossaire ou de bibliographie, participent à ouvrir le projet à d'autres chercheurs.

PAL L'application Web de PAL propose essentiellement un accès à la base de données des œuvres ptoléméennes en arabe et en latin réalisée dans le cadre du projet. Chaque enregistrement de cette base est lié à une notice consultable par le biais d'une interface

14. « *Efforts will be devoted to bring these results not only to the relevant scholarly communities but also to a wider audience as a resource in the public debates around science, knowledge and culture.* » *Shaping a European Scientific Scene : Alfonsine Astronomy...*

de recherche : les notices sont extrêmement complètes et les fonctionnalités pour les parcourir permettent une grande finesse dans l'exploration. L'interface de recherche autorise notamment l'utilisation d'expressions régulières et propose un système de recherche facettée très abouti. Un glossaire ainsi qu'un mode d'emploi pour utiliser au mieux toutes ces modalités permettent d'accompagner l'internaute dans une utilisation optimale de l'outil. Toutefois, le dépouillement de l'interface et la technicité requise pour exécuter des recherches peuvent rendre quelque peu aride l'utilisation de la plateforme aux publics les plus inexpérimentés.

De Sphaera

Le projet *De Sphaera*, intégré au Max-Planck-Institut de Berlin, porte sur le traité composé au XIII^e siècle par Johannes de Sacrobosco intitulé *Tractatus de sphaera*. Ce texte élabore une synthèse de la cosmologie géocentrique de Ptolémée et d'Aristote tout en s'ancrant dans les réflexions de son temps, par la citation d'érudits anciens et nouveaux : depuis l'invention de l'imprimerie jusqu'à la fin du XVII^e siècle, on dénombre quelques 350 éditions du traité ; éditions qui ont été collectées au sein de la base de données du projet¹⁵.

Dans le cadre du projet, un important choix technique a été fait pour modéliser au mieux la diffusion de ce texte et l'interconnexion qui en découle : la base de données du projet est une base de données RDF. En effet, les éditions et les acteurs historiques de la base ne sont pas définis dans une relation bilatérale, mais grâce à des triplets, plus proches de la structure sujet/verbe/complément. Ce choix de modélisation a fortement déterminé la construction du portail *Sphaera CorpusTracer* qui donne accès aux données de la base. Le RDF permet par exemple d'ouvrir les systèmes d'information et de rattacher le portail aux données de Wikidata pour en enrichir les notices ; en outre, un système de recherche intuitif a pu être mis en place, reposant sur la similitude des relations en graphe de la base de données avec le langage naturel.

L'utilisation d'une technologie peut donc fortement influencer la constitution d'une interface publique. Sur la base d'un recensement étendu des sources historiques au sein d'une base de données, il est possible de fournir des interfaces d'une grande diversité. L'intelligibilité de telles plateformes pour des publics variés est garantie par la présence de contenus entourant et traitant la donnée présente au sein du corpus, effet encore renforcé par l'ajout de ressources annexes et par l'attention portée à la lisibilité de l'interface. La plateforme en ligne de DISHAS vise à intégrer toutes ces notions, en particulier pour la

15. Matteo Valleriani, Florian Kraütl et Dirk Wintergrün, « De Sphaera : Modeling Data and Analyzing Diffusion Using Early Modern Cosmological Treatises », dans *Max-Planck-Institut Für Wissenschaftsgeschichte*, 2017, URL : <https://www.mpiwg-berlin.mpg.de/project/sacrobosco-sphaera> (visité le 12/04/2019)

partie publique de l'application, de manière à rendre accessible le contenu scientifique en même temps qu'explicite.

Chapitre 2

Corpus de recherche

2.1 Bornes historiques du corpus

2.1.1 Ptolémée et l'*Almageste*

Le corpus du projet DISHAS a pour ambition de rassembler toutes les sources astronomiques de tradition ptoléméenne ; en effet, l'œuvre de Ptolémée a essaimé dans les productions des astronomes anciens de toute l'Eurasie jusqu'à l'arrivée de la théorie héliocentrique de Copernic. L'astronomie est l'une des premières sciences à se développer et à atteindre un niveau élevé de sophistication¹ : la grande régularité du passage des objets célestes, ainsi que la capacité à quantifier ces mouvements permettent de concevoir dès le I^e millénaire avant notre ère² des systèmes mathématiques pour prévoir le comportement des astres.

Les théories pour prédire les trajets des planètes se complexifient en particulier avec l'apport des Grecs à partir de 500 avant notre ère³. Afin de prévoir avec justesse les événements célestes, les Grecs fondent leurs calculs sur des modélisations géométriques des cieux. Différents modèles se développent pour expliquer les irrégularités dans le mouvement des astres⁴, parmi lesquels les concepts de cercles excentriques⁵ et d'épicycle⁶.

1. James Evans et Michael Wulf Friedlander, « History of Astronomy », dans *Encyclopedia Britannica*, URL : <https://www.britannica.com/science/astronomy> (visité le 17/06/2019)

2. *Ibid.*

3. Les Babyloniens calculaient les mouvements de planètes grâce à des théories arithmétiques à partir du deuxième millénaire avant notre ère. *Ibid.*

4. À l'instar des rétrogradations, apparents reculs d'une planète par rapport à son mouvement habituel.

5. « le corps céleste se déplace le long de la circonférence d'un cercle « excentrique », c'est-à-dire un cercle dont le centre ne coïncide pas avec la Terre. » Gilles Bessou et Jérôme Lamy, « Lumières sur l'Univers – Histoire de l'astronomie », dans *Astrophysique sur Mesure - Observatoire de Paris*, URL : https://media4.obspm.fr/public/ressources_lu/pages_antiquite-moyen-age/ha-epicycles-excentriques.html (visité le 29/06/2019)

6. « la Terre est au centre d'un cercle nommé déférent. Le centre C de l'épicycle se meut sur le déférent autour de la Terre [...], tandis que la planète parcourt son épicycle. » *Ibid.*

Les astronomes grecs s'ingénient à fournir une vision cohérente et uniforme du cosmos, fondée sur des principes à la fois géométriques et philosophiques : avec Ptolémée, au premier siècle de notre ère, la pratique de l'astronomie grecque trouve son apogée. Ptolémée rassemble et synthétise l'œuvre de ses prédécesseurs, en particulier Hipparche dont on ne connaît les idées qu'à travers le résumé qui en est fait dans l'*Almageste*. L'œuvre de Ptolémée devient une référence, de sorte que les écrits astronomiques antérieurs cessent d'être copiés. L'*Almageste* est une œuvre riche qui contient des comptes rendus d'observations, des opérations trigonométriques, des théories nouvelles sur les mouvements lunaires, mais aussi des descriptions de procédures mathématiques permettant de dresser des tables pour le calcul de position des planètes. Une de ses principales innovations est l'introduction du point équant⁷ qui permet à la théorie grecque des planètes d'atteindre une grande précision prédictive quant au calcul des longitudes⁸.

À la suite de Ptolémée, la Grèce connaît un certain déclin dans la pratique de l'astronomie⁹. Toutefois, ses écrits continuent d'être transmis et adaptés aux cultures qu'ils traversent : cette diffusion en s'opérant façonne les productions des astronomes anciens. L'œuvre ptoléméenne ne reçoit pendant presque treize siècles que quelques corrections marginales, en revanche, les valeurs utilisées pour les calculs sont modifiées pour correspondre à d'autres lieux et époques. 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 certaines erreurs sont autant de signes qui témoignent des échanges culturels qui ont façonné la pratique de l'astronomie.

2.1.2 Propagation du modèle ptoléméen

Diffusion vers l'Orient

La transmission des sources grecques vers l'Est s'est notamment opérée grâce au déplacement des Perses jusqu'en Inde. En Inde, de nombreuses indications révèlent l'incorporation des théories grecques aux pratiques astronomiques autochtones. On relève des termes empruntés au grec – par exemple le titre du canon *Romaka Siddhanta* datant du début du v^e et qui marque les origines de la science astronomique sanskrite¹⁰ – ainsi que des modèles épicycliques et des méthodes de calculs requérant des paramètres numériques

7. L'équant représente le point à partir duquel la vitesse angulaire du centre de l'épicycle est constante ; il rompt avec le principe de mouvement circulaire uniforme prédominant jusqu'alors. *Ibid.*

8. Coordonnées permettant de repérer la position d'un point par rapport à l'écliptique (axe de rotation du Soleil).

9. Raymond Mercier, *Studies on the Transmission of Medieval Mathematical Astronomy*, Ashgate Publishers, Burlington, VT, 2004, p. 6

10. *Ibid.*, p. 7

hérités d'Hipparque ; les écrits indiens fournissent ainsi parfois des indices sur la pratique pré-ptoléméenne de l'astronomie grecque.

L'astronomie chinoise s'est développée de manière relativement indépendante à partir du deuxième millénaire avant notre ère avec des observations d'éclipses solaires et lunaires. À l'inverse de l'Occident, la cosmogonie chinoise distingue les procédures de calcul de sa théorie concernant l'organisation des cieux¹¹. Il est possible néanmoins de déceler des traces de l'influence de l'astronomie grecque, dont les théories auraient été diffusées par le biais de l'Inde ; on peut relever notamment l'utilisation du système sexagésimal pour le calcul des jours, également employé par les astronomes islamiques. À partir du XVI^e cependant, des missions jésuites mettent les pratiques astronomiques chinoises et européennes en contact direct.

Monde arabe et Europe occidentale

Au VIII^e siècle, les astronomes musulmans accèdent aux textes grecs grâce au retour des Perses vers l'ouest. L'*Almageste* est traduit à quatre reprises en arabe ce qui en permet la diffusion et le perfectionnement. Le regain d'intérêt pour les textes grecs s'observe en particulier pendant la dynastie des Abbassides qui court de 750 à 1258. Des auteurs comme Al-Khwârizmî montrent bien la mixité des influences en astronomie : son œuvre est un mélange de théories provenant tout à la fois de Perse, de Grèce et d'Inde, mais n'est pas exempte d'innovations¹². Il invente le genre du *zīj*¹³ ; ensemble de tables permettant de retrouver la position des astres. Grâce à la maîtrise de l'astronomie par les Arabes pendant le haut Moyen Âge, les Européens accèdent à partir du X^e siècle aux textes antiques et commencent à développer à leur tour leurs propres théories.

Les tables tolédanes, composées en Castille vers 1080 par un groupe d'astronomes musulmans et juifs, furent prépondérantes dans la propagation de la pratique astronomique d'abord en Espagne puis dans le reste de l'Europe¹⁴. Les scientifiques occidentaux n'accèdent en premier lieu à l'astronomie qu'au travers des livres produits en Al-Andalus, mais les écrits arabes ne se diffusent réellement en Occident qu'avec le concours d'acteurs tels que Gérard de Crémone, qui apprend l'arabe afin de traduire l'œuvre de Ptolémée en latin. Vers 1270, un nouvel ensemble de tables est composé à l'initiative du roi Alphonse X de Castille, reprenant le modèle ptoléméen avec quelques adaptations : un demi-siècle plus tard, l'astronomie alphonquine est récupérée par un groupe d'astronomes parisiens, avant de se répandre dans l'ensemble du monde latin. L'effervescence dans les milieux

11. Helaine Selin, Xiaochun Sun et David A. King, *Astronomy across Cultures : The History of Non-Western Astronomy*, Kluwer Academic Publishers, Dordrecht ; Boston, 2000, p. 438

12. J. Evans et M. Wulf Friedlander, « History of Astronomy »...

13. « The Arabic word *zīj* is the generic name for those astronomical handbooks which consist of groups of tables representing all the functions required in calculations of planetary longitudes and other quantities of interest. » R. Mercier, *Studies on the Transmission of Medieval Mathematical Astronomy...*, p. 451

14. Avant cela, l'étude de l'astronomie était principalement fondée sur l'œuvre de Pline l'Ancien.

savants parisiens s'explique en partie par la nécessité de refonte du calendrier¹⁵ et dans ce cadre, les figures de Jean de Murs, Jean de Lignères et Jean de Saxe se distinguent. L'astronomie alphonosine se perpétue jusqu'au développement de la théorie héliocentrique copernicienne, elle-même conçue à partir de tables de tradition alphonosine.

Les astronomes anciens n'examinent que très rarement le ciel pour tenter de confirmer ou infirmer leurs calculs¹⁶ : ils se fondent sur les écrits théoriques composés par leurs prédecesseurs, accentuant ainsi l'interconnexion qui existe entre toutes les productions astronomiques anciennes.

2.2 Nature documentaire du corpus

2.2.1 Les sources primaires astronomiques

Variété du matériel documentaire

Le matériel documentaire produit par les astronomes anciens peut être divisé en différentes catégories, que l'on prenne en compte leur forme ou bien leur utilisation. Les sources disponibles aux chercheurs sont constituées par des écrits aussi bien que des instruments ; ces documents pouvaient servir au calcul de positions des planètes, comme à la représentation géométrique du cosmos. Dans la plupart des cas, ces sources étaient employées dans le but de réaliser des horoscopes ; la pratique de l'astrologie ayant été longtemps confondue avec celle de l'astronomie.

De nombreux instruments ont été inventés pour aider à déterminer les positions des planètes et à repérer des coordonnées sur la sphère céleste : on peut citer pour cela les astrolabes¹⁷, les sphères armillaires¹⁸, ou encore les « atlas » du ciel comme l'*Astronomicum Caesareum* de Peter Apianus, permettant, grâce à un système complexe de volvelles, d'estimer la localisation d'un astre à une date donnée.

Un des genres les plus importants dans l'histoire de l'astronomie est celui des canons : les canons sont des guides donnant des instructions pour la détermination d'une donnée astronomique, que ce soit la position d'une planète, la date d'une éclipse ou la valeur

15. L'astronomie arabe ayant introduit l'idée de trépidation des équinoxes au lieu de la théorie grecque de la précession, l'établissement d'un calendrier fiable et la détermination des dates de fêtes religieuses sont remis en question. Emmanuel Poulle, *Les astronomes parisiens au XIV^e siècle et l'astronomie alphonosine*, F. Paillart, Abbeville, France, 2005

16. Seuls de rares exemples comme l'almanach de la fin du XIII^e de Guillaume de Saint-Cloud témoignent de réelles observations astronomiques « *It is sad to note that his fine effort, far from marking the start of observational work in the Latin world, was disregarded.* » R. Mercier, *Studies on the Transmission of Medieval Mathematical Astronomy...*, p. 22

17. Instrument reposant sur une projection plane de la voûte céleste. Il est perfectionné par les Arabes et aide notamment à calculer la latitude d'un corps céleste à une heure précise.

18. Instrument qui modélise la sphère céleste, utilisé depuis les Grecs, le plus souvent à des fins pédagogiques.

d'une date dans un type de calendrier. Ces textes intègrent souvent une forte dimension poétique, laissant ainsi une grande part à l'interprétation. Les canons sont composés de plusieurs chapitres, chacun guidant dans la réalisation d'un calcul, les derniers chapitres s'appuyant sur la maîtrise des précédents ; le travail des chercheurs consiste notamment à transcrire ces instructions dans un langage mathématique moderne. Les canons aident à dresser les tables astronomiques : ce sont des matériaux essentiels pour comprendre les raisonnements mathématiques sur lesquels elles s'appuient.

Bien souvent, le recours à une représentation diagrammatique peut aider à la compréhension des canons. Les diagrammes, en général absents des textes originaux, se trouvent parfois ajoutés dans les marges ou tracés dans les notes d'un étudiant. Ils constituent des indices précieux pour apprécier comment les astronomes concevaient géométriquement les textes théoriques et quels publics les consultaient.

Les sources tabulaires

Les tables constituent un des matériaux les plus fondamentaux dans les pratiques des astronomes anciens. On en distingue notamment deux types : les tables de calculs et les éphémérides. Leurs fonctions ne sont pas les mêmes et l'établissement d'un éphéméride se fonde sur des tables de calcul. D'après Richard Kremer, les tables de calcul constituent l'un des socles sur lequel s'appuie l'ensemble des pratiques des astronomes de l'époque : la philosophie naturelle du cosmos, héritée de Ptolémée, amène à la création de modèles théoriques qui trouvent leur expression dans les tables de calcul. À partir de ces tables, les astronomes prévoient à l'avance les positions des planètes¹⁹, qu'ils consignent ensemble dans des éphémérides. En examinant dans les éphémérides l'arrangement prévu du ciel à une date précise, les astrologues peuvent interpréter la disposition des astres et ainsi composer des horoscopes. En comparaison des tables de calculs, on ne conserve que peu d'éphémérides, en raison probablement de l'obsolescence rapide des informations qu'ils contiennent.

Les tables astronomiques de calcul constituent ainsi un matériau indispensable pour l'étude : elles ont par conséquent été choisies pour former le cœur du corpus de recherche de DISHAS. Les tables, par la dimension d'abstraction mathématiques qu'elles requièrent, sont particulièrement marquées par les différentes influences intellectuelles de ceux qui les ont composées. Plutôt que d'être recalculées entièrement, les astronomes les adaptaient à leurs besoins ; ainsi, elles sont révélatrices du milieu intellectuel de leur créateur.

Pour le chercheur moderne, une table astronomique représente une fonction mathématique (par exemple : $f(x) = ax + b$), dont les valeurs a et b sont des constantes et dont

19. E. Poulle et Owen Gingerich, « Les positions des planètes au Moyen Age : application du calcul électronique aux tables alphonsoines », *Comptes rendus des séances de l'Académie des Inscriptions et Belles-Lettres*, 111–4 (1967), p. 531–548, URL : https://www.persee.fr/doc/crai_0065-0536_1967_num_111_4_12173 (visité le 27/06/2019)

0 20 4 80 2 40 2 0 70 1 1 33 32	1 20 8 20 11 12 11 2 126 8 160 3	2 20 7 80 11 10 12 1 2 3 23 10
0 21 4 39 3 16 2 0 79 1 38 3	1 21 8 39 11 9 12 2 10 8 161 26	2 21 3 39 11 9 12 1 2 3 23 12
0 22 4 38 3 10 2 0 712 1 32 3	1 22 8 38 11 14 22 2 28 8 162 24	2 22 3 78 10 43 13 1 46 3 14 12
0 23 4 31 3 23 2 0 742 1 22 3	1 23 8 31 11 21 23 2 29 8 163 28	2 23 3 31 10 23 13 1 46 3 10 40
0 28 4 36 6 1 2 0 41 1 41 21	1 28 8 36 11 6 23 2 30 8 164 21	2 28 3 30 10 33 13 1 41 3 6 38
0 24 4 34 3 10 2 0 749 1 44 42	1 24 8 34 11 37 24 2 31 8 165 18	2 24 3 34 10 22 13 1 49 3 2 24
0 26 4 32 5 9 2 1 1 2 0 74	1 26 8 32 11 39 20 2 32 8 166 28	2 26 3 38 10 11 13 1 46 2 40 11
0 20 4 33 3 14 1 1 2 2 0 72	1 27 8 33 11 24 24 2 33 8 167 11	2 27 3 33 10 0 13 1 43 2 43 23
0 28 4 32 2 6 1 1 2 2 0 71	1 28 8 32 11 40 24 2 34 8 168 40	2 28 3 32 0 28 13 1 21 2 28 31
0 29 4 31 8 14 3 1 2 2 13 18	1 29 8 31 11 14 26 2 34 8 169 24	2 29 3 31 0 34 14 1 38 2 28 14
0 30 4 30 2 23 3 1 2 2 11 17	1 30 8 30 12 0 26 2 35 8 170 48	2 30 3 30 0 22 14 1 34 2 39 34
0 31 4 29 2 23 3 1 2 2 21 89	1 31 8 29 12 6 21 2 36 8 171 18	2 31 3 29 0 18 16 1 32 2 38 12
0 32 4 28 2 21 3 1 2 2 24 44	1 32 8 28 12 10 21 2 37 8 172 12	2 32 3 28 0 14 16 1 32 2 38 6
0 33 4 24 2 22 3 1 2 2 30 14	1 33 8 21 12 4 28 2 38 8 173 11	2 33 3 21 0 18 16 1 32 2 38 16
0 32 4 26 2 18 3 1 2 2 32 12	1 32 8 26 12 20 28 2 38 8 174 44	2 34 3 26 0 22 16 1 32 2 38 23
0 34 4 21 4 8 1 2 2 38 11	1 34 8 24 12 20 29 2 38 8 175 0	2 34 3 24 0 14 16 1 32 2 38 26
0 26 4 20 4 16 1 2 2 36 8		

FIGURE 2.1 – BnF, Département des manuscrits – Latin 14481, f. 81v – Équation de la Lune

le résultat est calculé de manière incrémentale pour chaque valeur de x . Une table de multiplication est en ce sens comparable aux tables de calculs : sans disposer d’information supplémentaire, l’enchaînement des nombres peut suffire à déterminer de quelle table de multiplication il s’agit. Ainsi, la disposition des nombres dans les cases, la mise en page du tableau et les nombres en eux-mêmes portent des informations sur les pratiques des savants : il est possible d’y déceler des variations dans les procédures ou dans les formules, ou encore d’observer des variations dans les types de nombres et dans les usages arithmétiques.

L’intégralité des sources primaires astronomiques a vocation à être progressivement intégrée aux interfaces et outils développés par DISHAS, et certains documents réalisés par les chercheurs de l’équipe pourraient d’ores et déjà être mis à disposition au sein de la plateforme en tant que contenu explicatif. Toutefois, dans un souci d’inclusion, il est important que les productions scientifiques des projets partenaires soient également exposées au sein de l’interface de DISHAS. Les tables de calcul sont donc, par leur contenu essentiellement numérique, un élément fédérateur des différentes cultures représentées dans le projet ; il faut cependant envisager dès à présent, dans la conception de la plateforme, les espaces et les fonctionnalités qui pourront être dédiées aux sources non-tabulaires.

2.2.2 Les paramètres

Type de paramètres

Les paramètres correspondent à la granularité la plus fine du corpus de DISHAS : ils représentent les données numériques qui interagissent avec la fonction sous-jacente d’une table. Ces paramètres sont tantôt visibles au sein même de la table, tantôt présents implicitement dans le rapport des valeurs numériques entre elles. Dans DISHAS, on distingue trois types de paramètres : les paramètres astronomiques (ou *set* de paramètres), les paramètres mathématiques et les paramètres « contextuels ». Ces derniers sont les valeurs numériques liées à un lieu ou une période, une table astronomique étant généralement

calculée pour un endroit en particulier. Les paramètres mathématiques sont des artifices mis en place par les astronomes pour faciliter les calculs. Ils se divisent en deux groupes : les *shift*, décalage des nombres d'une colonne, et les *displacement*, ajout d'une constante à tous les nombres de la table pour éviter les opérations avec des valeurs négatives.

Les paramètres astronomiques sont les nombres – visibles ou non dans une table – qui se retrouvent dans des dispositions analogues au sein des tables de calcul du même type. De façon similaire qu'il est possible de déduire de l'enchaînement des nombres qu'il s'agit d'une table de multiplication, la présence de certains paramètres astronomiques indique qu'une table a trait à un objet astronomique en particulier : l'apparition du paramètre 4,56 est, par exemple, révélateur du fait qu'il s'agit d'une table concernant la Lune. Le plus souvent, ces paramètres sont combinés en *sets* : le rôle des chercheurs est, entre autres, de comprendre comment ces assemblages numériques ont été composés et adaptés par les acteurs de l'époque.

Spécificité et intérêt des paramètres

Le projet DISHAS avait initialement été envisagé comme une base de données de paramètres astronomiques : en effet, pour un grand nombre de chercheurs, les paramètres astronomiques sont des marqueurs forts pour déterminer la provenance d'une table. L'étude de la répartition de l'apparition de certains paramètres permet notamment de reconstituer les flux d'échanges intellectuels, le long de la route de la soie par exemple. Les paramètres sont un instrument pour l'identification d'une tradition astronomique : dans le cas de l'histoire de l'astronomie, la donnée mathématique s'associe au contenu textuel pour révéler le contexte culturel et l'environnement historique d'un document.

La description de ces paramètres est centrale pour la recherche en astronomie ancienne. DISHAS doit avant tout assister les chercheurs du projet dans leur étude : une attention toute particulière doit donc leur être portée au sein de l'interface. La nature de ces paramètres a déjà fortement façonné les réalisations construites dans le cadre du projet ; ces paramètres, bien que parfois sous-entendus par les scribes, doivent être transcrits explicitement dans la base de données afin d'être pris en compte dans des traitements statistiques. Ils sont la clef de compréhension des tables, nécessaires à l'analyse du corpus. Les paramètres constituent assurément une donnée centrale, mais, par leur nature, ils peuvent rester hermétiques pour de nombreux publics. En ce sens, un effort de médiation doit être fourni pour assurer la valorisation de cette donnée.

Chapitre 3

Numérisation et accueil de la donnée

3.1 Modélisation de la donnée

La modélisation de données est une étape fondamentale pour tout projet en humanités numérique : elle est l'occasion de définir précisément les enjeux du travail de recherche¹ et de considérer les sources documentaires étudiées dans ce qu'elles ont de commun. La modélisation est également le lieu privilégié du dialogue entre ingénieur et chercheur, les compétences de chacun étant sollicitées pour la création d'une donnée rigoureuse scientifiquement et utilisable informatiquement². Dans la perspective de construire une interface en adéquation avec la donnée, il est primordial de maîtriser le modèle dans les subtilités qu'il peut présenter. Les intentions comme les implications de celui-ci doivent être intégrées à la réflexion ; en fonction de la conceptualisation des données, les possibilités de visualisations ne sont pas équivalentes.

Dans le cadre du projet DISHAS, les exigences quant à la polyvalence des données et à l'intégration culturelle ont présidé à l'établissement du modèle : il était important que la structure de la base de données soit flexible, dans le but de publier et d'interroger des sources de nature diverses. Un des choix de modélisation principaux concerne la granularité donnée à la base : comme il a déjà été mentionné, la table constitue l'unité fondamentale du modèle. Autour d'elle s'articulent les entités *Original item*, le témoin, et *Edited text*, l'édition. Le modèle de données se fonde sur cette dichotomie : une partie du modèle est orientée vers les enjeux scientifiques liés à l'édition d'une table et à son contenu numérique, l'autre partie portant sur la description historique des sources. Ces

1. « Objet construit, le modèle, par la formalisation qu'il nécessite, oblige l'historien à une plus grande rigueur dans l'expression des hypothèses interprétatives qui le sous-tendent. » Arlette Farge, *Le goût de l'archive*, Éditions du Seuil, Paris, 1997 (Points Histoire, 233), p. 25

2. Frédéric Clavert, Franziska Heimburger et Aurélien Berra, « L'historien programmeur ? », dans *THATCamp Paris 2012 : Non-actes de la non-conférence des humanités numériques*, Paris, 2012 (La Non-Collection), URL : <http://books.openedition.org/editionsmsmsh/305> (visité le 04/08/2019), § 17-18

deux parties se rejoignent grâce au lien existant entre document original et édition de ce document. Cette bipartition est caractéristique du travail de recherche des membres de DISHAS, touchant à des disciplines diverses, comme la codicologie et les mathématiques anciennes.

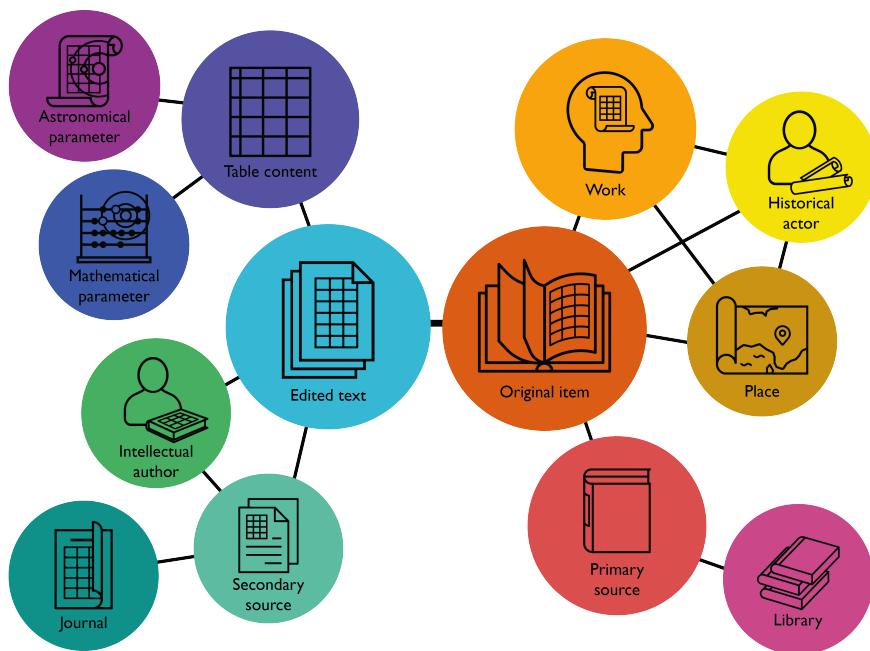


FIGURE 3.1 – Modèle conceptuel de la base de données de DISHAS

La souplesse du modèle a été garantie grâce à la distinction faite entre contenu intellectuel et métadonnées de catalogage. Cette structure s'inspire notamment des spécifications du FRBR (*Functional Requirements for Bibliographic Records*), où sont différenciés œuvre, expression, manifestation et item. Cette séparation implique notamment que l'édition soit partitionnée entre informations relatives au contexte éditorial d'une part, et valeurs numériques transcrrites dans l'édition d'autre part. De la même manière, une source historique est découpée entre sa manifestation matérielle (incunable ou manuscrit), le contenu intellectuel qu'elle contient (les différentes œuvres qu'on peut y trouver) et les différents exemplaires de tables présents dans son corps. Cette finesse de description permet par exemple d'analyser en détail la composition intellectuelle d'un manuscrit donné, ou encore d'associer plusieurs ensembles tabulaires à une même édition. Par la généralisation du modèle, il est ainsi possible d'intégrer toutes les particularités du matériel à numériser.

3.1.1 Les sources historiques

L’item original

L’*original item* représente une partie du contenu de la source primaire qu’il est possible de considérer indépendamment. Il peut être constitué d’un diagramme, d’un texte ou encore d’une table ; dans la base, il est associé à une métadonnée de type – tabulaire, diagrammatique ou textuel – qui en précise la nature. Toutefois, les interfaces et outils développés sont pour l’instant uniquement destinés aux tables astronomiques.

Afin d’expliciter la place de l’*original item* au sein la base de données, un rapprochement peut être fait avec le genre du poème : un poème, en effet, constitue une unité textuelle qui possède sa propre cohérence. Il a été composé – généralement par un auteur unique – pour apparaître à l’intérieur d’un recueil réunissant plusieurs poèmes : ce recueil est l’œuvre intellectuelle d’où provient le poème. Il peut faire l’objet de plusieurs éditions, où la mise en page et même le texte peuvent varier. Ces éditions correspondent aux sources primaires du modèle conceptuel ; celles-ci peuvent contenir ou non l’intégralité des poèmes du recueil. Les éditions, quand elles prennent la forme d’une anthologie, mêlent les poèmes de plusieurs auteurs et d’œuvres différentes ; deux poèmes d’un même auteur provenant de recueils séparés pouvant ainsi se trouver rassemblés.

L’*original item* est à la fois lié à une œuvre, c’est-à-dire à une production intellectuelle intangible dont il est une manifestation, et à une source primaire, objet matériel dont il est une partie. L’œuvre et la source ne sont pas associées dans la base de données, une source, de nature souvent composite, pouvant contenir plusieurs œuvres différentes. En outre, les informations concernant la langue et l’écriture sont également liées uniquement à l’item, parce qu’une œuvre peut par exemple se manifester dans différentes langues, ou encore qu’un manuscrit peut être copié dans plusieurs types d’écriture³. De la même manière, le scribe et le lieu de copie d’une source ne sont liés dans la base de données qu’à l’*original item*, permettant ainsi de signaler avec précision quels passages d’une source primaire ont été rédigés par un même acteur historique ou à un même endroit. Ainsi, l’item centralise les informations au niveau le plus fin de la description, permettant de cette manière de caractériser l’ensemble des sources historiques avec davantage de finesse.

Entités associées

Deux entités majeures se distinguent dans leurs corrélations à l’*original item* : ce sont l’œuvre et la source primaire. La source primaire dans DISHAS correspond à un manuscrit ou un incunable, dans lequel peut être trouvé un *original item*. Le rôle de cette entité dans la base de données est d’identifier de manière univoque le document en lui-même, de manière à pouvoir retrouver facilement son lieu de conservation, dans l’éventualité

3. La normalisation des langues et des écritures dans la base se fonde sur le standard ISO facilitant ainsi un travail d’alignement ultérieur.

d'une étude approfondie. Une attention particulière a été portée à la matérialité du texte, chaque item étant repéré précisément au sein de sa source, de manière à pouvoir mener des investigations sur la composition d'un manuscrit et retracer l'élaboration progressive de l'objet. Comme dit précédemment, une source primaire n'est pas associée à un lieu dans la base, il est possible de connaître l'endroit de sa copie uniquement par le biais des *original items* : ce choix de modélisation a constitué une difficulté à contourner dans l'élaboration de certaines visualisations.

L'œuvre au sein de la base de données correspond à une création intellectuelle distincte : ce sont généralement des écrits théoriques contenant des textes, des diagrammes et des tables astronomiques. Une œuvre est identifiée grâce à son titre, ou à défaut, son incipit. En outre, l'œuvre est reliée dans la base à un acteur historique – son créateur – et au lieu de sa conception. Comme il a déjà été mentionné, le contenu d'une œuvre se retrouve de manière plus ou moins complète à l'intérieur des sources primaires qui la contiennent. De même qu'il est intéressant, dans une perspective de bibliographie matérielle, de considérer l'agencement des œuvres au sein d'une même source, les variations de complétude d'une œuvre au sein des sources primaires peuvent faire l'objet d'un examen approfondi. Grâce à une modélisation centrée sur l'*original item*, ces informations sont contenues dans la base de données et peuvent être mises au jour.

Par ailleurs, le concept d'œuvre pose plusieurs difficultés et peut être sujet de débat, compte tenu des différentes approches de recherche : quels contours exacts peut-on donner à sa définition ? La traduction d'un écrit constitue-t-elle une nouvelle œuvre ? Comment considérer les *marginalias* qui entourent le texte ou encore l'ajout d'une colonne à une table préexistante ? La réponse à ces interrogations est à la charge du chercheur ; le plus souvent, c'est l'importance des ajouts et des modifications qui distingue une œuvre originale, d'une simple copie ou transposition. Une traduction qui admettra des adaptations du document sur lequel elle s'appuie – par exemple en changeant des paramètres en fonction du lieu de rédaction – sera ainsi encodée dans la base en tant qu'œuvre indépendante. Cela peut occasionner des situations quelque peu paradoxales où une œuvre est associée dans la base à un traducteur, alors que seul l'*original item* porte l'information quant à la langue de cette œuvre : la souplesse du modèle de données rentrant ici en concurrence avec une représentation plus logique de l'information.

3.1.2 Structuration des données d'édition

Le texte édité constitue le pendant de l'*original item* au sein de la partie éditoriale du modèle de données. L'édition en tant que telle est divisée entre son contenu tabulaire (*Table content*) et les métadonnées éditoriales (*Edited text*) : cette distinction correspond entre autres à la nature de la donnée, l'entité *Table content* ayant les caractéristiques d'une base de données orientée documents.

Le contenu tabulaire rassemble les données numériques de la table, définissant notamment les paramètres utilisés ; les métadonnées éditoriales concernent quant à elles l'auteur intellectuel de l'édition, le journal dans lequel elle est parue et le titre de la publication.

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 interprétation⁴, 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, voir ci-dessous pour la définition).

Données éditoriales

L'application Web du projet DISHAS a été pensée de prime abord pour être une plateforme d'éditions de table⁵, cet aspect de la recherche étant particulièrement important en astronomie ancienne. Toutefois, le modèle de données relatif aux éditions n'accueille en réalité que le contenu de la table, et non l'intégralité de l'article susceptible de l'accompagner. Le travail de modélisation s'est donc concentré sur la description des types d'éditions de table ; il en existe trois sortes au sein de la plateforme :

- Édition de type A : s'appuyant sur un unique *original item*, ce sont les éditions 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 ;
- Édition de type B : ces éditions reposent sur plusieurs éditions de type A mais aussi potentiellement de type C ; elles figurent donc un *apparat critique*.
- Édition de type C : pour réaliser ce type d'édition, le chercheur récupère les paramètres d'une table témoin et recalcule l'ensemble de ses valeurs grâce à des modèles mathématiques contemporains. Ce type peut servir notamment à comparer les techniques arithmétiques anciennes et modernes, pour observer par exemple les variations de certaines constantes mathématiques.

S'il est prévu de pouvoir éditer des tables de manière nativement numérique grâce à l'interface de DISHAS, une partie des éditions qui ont vocation à être intégrées à la base ont été déjà publiées sur papier. C'est pour cette raison que les entités *Journal* et *Secondary source* ont été ajoutées au modèle de données ; elles signalent d'où provient une édition, permettant de retrouver le matériau de départ pour l'étudier plus avant. Pour les

4. Certaines tables, compte tenu du chevauchement des nombres ou encore de certaines graphies, comportent des incertitudes de lecture ; cette ambiguïté était probablement déjà présente chez les lecteurs médiévaux, il s'agit de ne pas l'effacer lors de la numérisation.

5. Projet qui a été écarté au profit d'une plateforme d'*open data* pour ne pas rentrer en concurrence avec d'autres plateformes de publication en ligne.

sources primaires et secondaires, DISHAS a donc destination à servir de portail, plus que de plateforme de consultation.

Données astronomiques

Les données liées aux aspects mathématiques des tables sont encodées dans la partie éditoriale de la base, car elles constituent une interprétation moderne des raisonnements arithmétiques des acteurs de l'époque ; seul le type de table est associé aux enregistrements d'*edited texts* comme d'*original items*. On dénombre 70 types de tables au sein de la base ; ils représentent la fonction sous-jacente à une table, son objectif de calcul (par exemple déterminer le mouvement moyen d'un astre). Le type de table est l'entité qui centralise les informations relatives au contenu mathématique et astronomique d'une table, que ce soient les paramètres ou les définitions de modèles.

Les paramètres Les paramètres, en particulier les paramètres astronomiques, constituent les entités les plus complexement décrites dans le modèle de données ; le schéma ci-dessous représente les tables de la base de données utilisées pour la modélisation des paramètres astronomiques. La multiplicité de ces tables permet notamment d'effectuer des traitements mathématiques automatisés sur les données numériques de la base, mais cette finesse de description est également utile aux chercheurs dans leurs travaux et doit être exposée au mieux dans les interfaces graphiques de la plateforme.

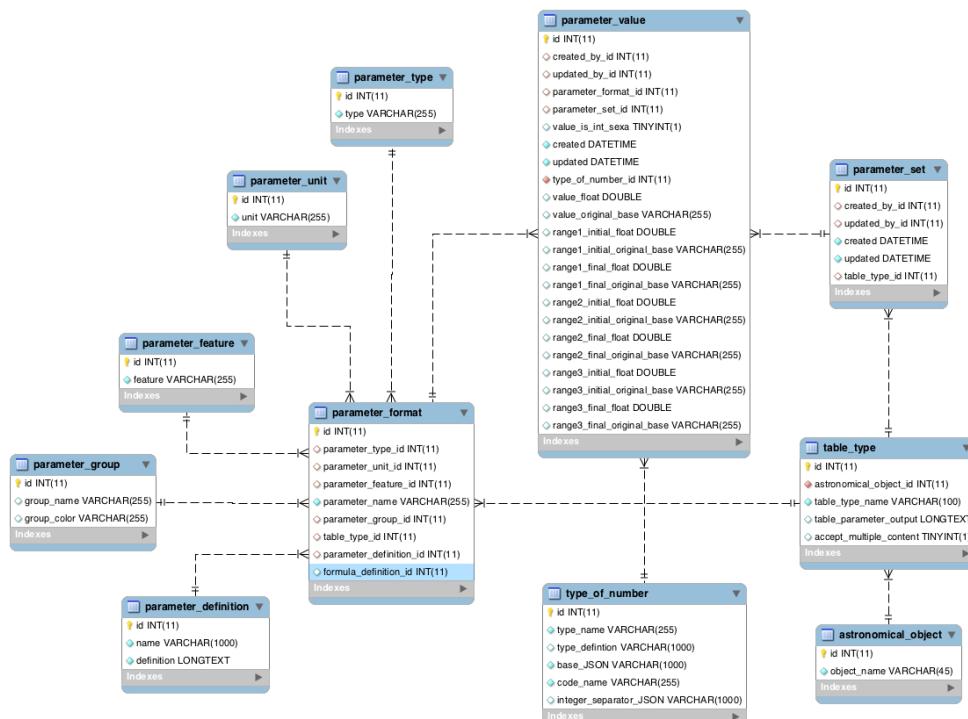


FIGURE 3.2 – Visualisation de l'architecture des données de paramètres astronomiques dans la base

Toutes les tables représentées ci-dessus fonctionnent ensemble : la table *Parameter type* par exemple, renseigne sur le mode d'apparition d'un paramètre dans une table (explicite, implicite ou mixte⁶), tandis que la table *Parameter feature* indique le nombre d'arguments admis par le paramètre (simple, double, parfois triple). La table *Parameter group* réunit quant à elle les paramètres fonctionnant ensemble, là où la table *Parameter set* rassemble les paramètres permettant de décrire un type de table⁷. Enfin, la table *Parameter format* agrège toutes ces informations pour les associer à un enregistrement de la table *Parameter value* qui instancie une valeur à un paramètre, conservée à la fois sous sa forme originale et sous sa forme fractionnaire afin d'effectuer des calculs.

Si les tables de paramètres sont aussi nombreuses, c'est pour assurer l'encodage exhaustif toutes les informations implicitement contenues dans une table astronomique. Il est important de rendre accessibles l'ensemble de ces données au public de la plateforme, en s'assurant de fournir une interface adaptée pour leur présentation.

Les modèles de tables Les modèles – contenus dans la table *Formula definition* de la base de donnée⁸ – reposent en grande partie sur la définition de ces paramètres. Ils sont une forme modernisée des canons, résumant à l'aide d'une formule mathématique la fonction sous-jacente à une table. Un modèle est défini pour un type de table en particulier – par exemple les tables de vitesse solaire – et requiert un set de paramètre précis – par exemple la distance Terre/Soleil ainsi que la valeur de l'excentricité solaire. En revanche, un même type de table peut admettre plusieurs modèles ; autrement dit, pour le calcul d'une même donnée astronomique, plusieurs méthodes mathématiques sont possibles.

La définition de ce genre de modèles est très utile pour comprendre la manière dont ont été composées les tables astronomiques : ils permettent notamment de recalculer les valeurs d'une table et de les comparer avec les valeurs telles que présentes dans une source, afin de révéler les différences de conception mathématiques entre nos méthodes modernes de calcul et les techniques anciennes de computation. Ils peuvent également être utilisés pour aider à la saisie automatique de table, de manière à faciliter la transcription tout en révélant les zones où les formules arithmétiques contemporaines ne correspondent pas à la logique des savants passés. Enfin, en intégrant ces modèles à des réseaux de neurones entraînés, il est possible d'imaginer pouvoir aider à résoudre les équations sous-tendues à

6. Visible au sein de la table (explicite) ou contenu implicitement dans le rapport des valeurs de la table entre elles (implicite). Le type mixte correspond généralement à un cas où le paramètre est exprimé explicitement, mais où la disposition bouleversée de la table en rend l'identification malaisée

7. Le terme « set de paramètre » est également utilisé pour signifier « paramètre astronomique » : ces deux significations désignent des réalités distinctes. Un paramètre astronomique, à l'instar de $4,56$, correspond à une valeur définie ; un set de paramètre est l'ensemble des variables nécessaires au calcul d'une table selon un modèle précis, les valeurs n'en sont pas définies.

8. Pour une représentation de la base de données présentant toutes les tables concernant l'histoire de l'astronomie, voir annexe 119.

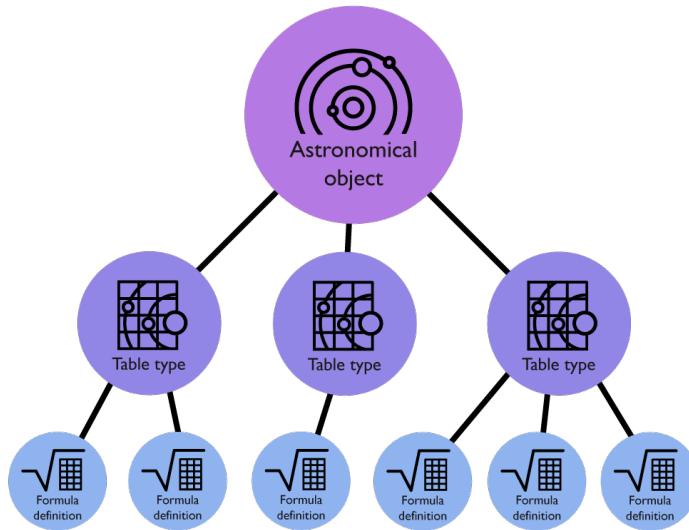


FIGURE 3.3 – Exemple de structuration de modèles mathématiques dans la structure de données

certaines tables, qui n'ont été élucidées jusqu'à présent⁹.

Cette donnée, fruit de l'expertise des chercheurs du projet, constitue une porte d'entrée pour aborder l'histoire de l'astronomie. Seuls les administrateurs de la plateforme sont autorisés à rentrer au sein de la base de données des modèles de tables : ils représentent donc une des valeurs ajoutées de DISHAS, une ressource qu'il faut absolument valoriser auprès du public.

3.1.3 Développements futurs de la base de donnée

Certaines évolutions ont été prévues pour la base de DISHAS, en premier lieu l'ajout au modèle de données de tables décrivant les paramètres contextuels (liés au contexte historique et géographique d'une table) : il existe d'ores et déjà dans la base les entités *Calendar* et *Era*, tâchées de décrire le système temporel utilisé par une table. Ces entités vont être rattachées au reste des tables dans des développements futurs.

En outre, les diagrammes et les textes pourront à terme être ajoutés en tant qu'enregistrements de la base de données mais ces modifications seront probablement réalisées par un autre groupe de recherche en histoire de l'astronomie, extérieur à l'Observatoire. Le modèle de données n'est donc pas entièrement arrêté, chose à prendre en compte lors

9. Les réseaux de neurones devrait réussir en réalité à « craquer » les tables, c'est-à-dire trouver toutes les fonctions, aussi complexes et invraisemblables soient-elles, qui reproduisent la même progression numérique que celle observée dans le contenu tabulaire. Le travail du chercheur sera de raffiner ces résultats pour trouver des estimations qui semblent les plus plausibles

des développements des interfaces.

Pour davantage d'informations sur le modèle de données, un glossaire des principales entités de la base de DISHAS est disponible en annexes B, à la page 121.

3.2 Données de gestion et encodage de l'information

3.2.1 Données structurelles

Données non-manageables

Les tables de la base de données peuvent être rangées en deux catégories : les tables de données modifiables par l'utilisateur, et les tables contenant des informations n'étant pas vouées à être remaniées. Ce dernier type assure, entre autres, la cohésion de l'encodage des métadonnées. C'est le cas par exemple des entités *Script* et *Language* qui centralisent tous les enregistrements de langues et d'écritures pouvant être utilisées au sein d'un *original item* de la base. L'exhaustivité est assurée par le recours à un standard – le standard ISO en l'occurrence – qui garantit de décrire tous les arrangements possibles de graphies et de vocabulaires, tout en facilitant un travail d'alignement ultérieur.

L'utilisation de normes permet en effet de renforcer la cohérence des métadonnées ; le risque peut être cependant de restreindre les chercheurs quant à l'encodage qu'ils souhaitent appliquer à leur sources. Il aurait ainsi été possible de calquer les données de lieux sur la base géographique de GeoNames, mais cela n'aurait pas permis, par exemple, d'ajouter des noms de lieux fictifs. Les données auxquelles l'utilisateur n'a pas accès doivent donc couvrir toutes les éventualités qu'il est possible de rencontrer dans les sources.

En outre, certains choix d'organisation du corpus ont été fixés dans la base de données, bien que leur modélisation eut pu être différente : c'est le cas en particulier de la table *Astronomical object* qui classifie tous les types de tables astronomiques en 11 groupes. Ces groupements ont été constitués au moment de la création de la base de données ; il s'agit donc de justifier et d'expliquer ces options de modélisations au sein même de l'interface afin qu'elles n'apparaissent pas comme arbitraires aux utilisateurs.

Données d'interface

Un autre type de données structurelles a été intégré directement à la base, il concerne les éléments utilisés pour la construction de l'interface. Deux tables en particulier sont utilisées dans ce but : la table *User interface text* et la table *Definition*. La première rassemble les textes destinés aux utilisateurs lors de leur navigation dans la plateforme, le texte d'introduction au projet comme les messages d'annonce pour la partie administrateur. Ces textes sont stockés sous format HTML de manière à pouvoir être mis en forme

dans une interface graphique et affichés de manière identique : cette pratique s'inspire des CMS (*Content Management System*), à l'instar de Drupal ou Omeka, qui offrent ainsi une grande liberté de configuration, tout en conservant une interface graphique plus facile à manier par les utilisateurs.

La table *Definition* quant à elle réunit les informations relatives à chaque entité du modèle de données concernant l'astronomie ancienne : l'entité *Original item* y est par exemple associée à une définition, à une couleur et à plusieurs variations dans l'écriture de son intitulé¹⁰. En effet, y sont distingués le véritable nom de la table dans la base de données SQL – *Original text* – de celui utilisé au sein de la plateforme. Ceci permet une grande modularité des options de présentation et aide à la mise à jour pour le changement de la charte graphique entre autres.

Données utilisateurs

Le dernier type de données qu'il est important d'aborder concerne les données générées à la création de comptes utilisateur : elles constituent des renseignements sensibles, qui tombent sous le coup du RGPD (Règlement Général sur la Protection des Données) et qu'il faut traiter avec attention. Bien que les informations conservées sur les utilisateurs soient minimales – nom, prénom, email et mot de passe crypté – il est du rôle des ingénieurs de la plateforme de garantir la transparence quant à leurs utilisations, en même temps que d'en assurer le caractère privé.

Parallèlement, les concepteurs de DISHAS doivent veiller à la qualité des données qui sont ajoutées ; en ce sens, il est essentiel de garder une traçabilité vis-à-vis des entrées et modifications dans la base. Ainsi, toutes les tables concernant les entités manipulables par les utilisateurs disposent de champs précisant la date et l'auteur de l'ajout, ainsi que les mêmes informations en cas de remaniement. Les données utilisateurs sont donc centrales dans la plateforme et doivent en conséquence être considérées attentivement dans leur utilisation.

3.2.2 Encodage et typologie de la donnée

Permissivité du modèle

L'élaboration d'un modèle de données ayant vocation à être inclusif peut entraîner des difficultés : comment établir un cadre pour contrôler la qualité des données, sans pour autant dénaturer le matériau sur lequel elles s'appuient ? D'une part, il est important de pouvoir ajouter des données n'ayant pas le même niveau de précision descriptive, d'autre part, une certaine régularité et consistance est nécessaire pour traiter l'information, dans les interfaces comme dans leur exploitation scientifique.

10. En *camel case* pour l'utilisation PHP (« originalText ») ou en *snake case* pour l'intitulé dans la base de données (« original_text »).

Le caractère obligatoire d'une information n'est pas spécifié dans le modèle la base ; il a été stipulé en revanche dans la construction des interfaces de saisies. Cela signifie que certaines données ne peuvent être insérées, tant que certains champs n'ont pas été remplis. Par exemple, il est indispensable de renseigner les bornes temporelles (*Terminus Post Quem* et *Terminus Ante Quem*) d'une œuvre lors de son ajout. L'éventuelle absence d'information exacte est prise en compte par la possibilité d'indiquer une période et non une donnée ponctuelle. Cela permet *a posteriori* de faire des estimations sur les périodes les plus florissantes intellectuellement. De la même manière, si le lieu de copie exact d'un *original item* n'est pas connu, il est possible d'indiquer, pour le renseigner, le nom d'un pays ou d'une région plutôt que celui d'une ville¹¹.

En revanche, il ne faut pas que les exigences de visualisations prennent le pas sur l'exigence scientifique : les modalités d'encodage et les objectifs de traitement de la donnée s'influencent, mais ce sont les ambitions de recherche qui doivent avoir la haute main sur la nature des données. Dans la conception des interfaces, il faut donc savoir composer avec l'absence, tout en incitant les chercheurs à produire la donnée la plus complète possible.

Encodage de la donnée en fonction de son utilisation

Le format de stockage qui a été choisi pour DISHAS est celui de la base de données relationnelle. Ce genre de base de données a l'avantage d'éviter les écueils de répétition de l'information et de pouvoir être maintenu facilement, dans la mesure où il est bien connu des directions de systèmes d'information. Pourtant, il ne constitue pas le format le plus adapté pour tous les types de matériau à intégrer à une base de données. Les formats XML (*eXtensible Markup Language*) et JSON (*JavaScript Object Notation*) offrent par exemple plus de flexibilité et d'autonomie pour la constitution d'une base ; ils sont également plus indiqués pour la description de matériel textuel. Plusieurs solutions ont été déployées dans le cadre de DISHAS pour pallier l'insuffisance du format relationnel de la base de données.

Les tables astronomiques, tout particulièrement, ont reclamé un traitement spécifique pour leur incorporation dans la base de données. Comment encoder un tableau de nombres, mis en page de façon complexe, à l'intérieur d'une ou plusieurs chaînes de caractères ? Le format JSON a été préféré au XML, notamment pour sa facilité de manipulation et sa compatibilité avec JavaScript, ce langage étant largement utilisé pour la construction de la partie frontale d'un site Web. La distinction dans le format de données a donc participé à la division conceptuelle des éditions de table : le contenu numérique encodé en JSON à l'intérieur de *Table content*, et les métadonnées éditoriales au format plus classique dans la table *Edited text*.

La conceptualisation du lien entre *original item* et *edited text* a également nécessité des ajustements. En effet, une édition pouvant elle-même être basée sur une édition, il était

11. Il n'est cependant pas obligatoire de renseigner un lieu lors de l'ajout d'un *original item*.

important de prévenir le risque de récursivité. L'architecture en graphe permet d'éviter ce genre d'aberration ; ainsi, la bibliothèque *GraphTree*, développée pour les besoins du projet, a été employée pour vérifier qu'une édition ne fait pas référence à elle-même et du même coup, permettre de réaliser des visualisations de la parenté intellectuelle d'une édition.

3.3 Saisie de la donnée

3.3.1 Interface administrateur

La plateforme en ligne de DISHAS dispose d'ores et déjà d'interfaces entièrement fonctionnelles ; c'est le cas de l'interface administrateur. Dans un site Web dynamique, l'interface administrateur – ou *back office* – est la partie dédiée à la gestion des ressources. Les chercheurs disposant d'un compte¹² peuvent s'y connecter pour ajouter des données à la base, sans prérequis de connaissances en SQL (*Structured Query Language*) ou autre langage de requête de base de données. Elle s'oppose à l'interface utilisateur – *front office* – qui est ouverte au public et qui fut l'objet de mon stage. Le rôle de l'interface administrateur est avant tout de faciliter la gestion des données pour les administrateurs, qu'il s'agisse de leur ajout, de leur modification ou de leur retrait. Elle constitue le socle sur lequel s'appuie la partie publique de l'application.

Rôle du *back office*

L'interface vise à assister les administrateurs dans la gestion de leurs données : en ce sens, les pages de vues doivent identifier clairement les différentes entités de la base. Les chercheurs n'étant pas nécessairement familiers des bases de données, l'intuitivité des interfaces est l'une des priorités ; néanmoins, la simplicité d'utilisation ne doit pas se faire aux dépens de la compréhension du modèle de données sous-jacent. En effet, le *back office* a aussi pour mission de garantir la qualité des données, en assistant les chercheurs dans le processus de numérisation ; l'explicitation du modèle conceptuel participe à cette qualité.

Il existe deux types de plateformes en ligne pour les données de recherche : les plateformes où les données ont été moissonnées à partir de multiples systèmes d'information, et celles – à l'instar de DISHAS – où l'intégralité du contenu numérisé a été créé par des chercheurs : l'interface administrateur est l'instrument même de cette numérisation. L'objectif du premier type est le regroupement des ressources, l'ambition du second est la création d'un corpus qualitatif. La normalisation et le contrôle des données est donc du cœur du projet DISHAS et l'interface administrateur doit aider dans cette mission.

12. La sélection des chercheurs ayant accès au *back office* est faite par Matthieu Husson, Benno Van Dalen et Clemency Montelle.

Plusieurs fonctionnalités ont été mises en place dans ce but : l'appel à des bases de référentiels comme le VIAF (*Virtual International Authority File*) ou l'ISNI (*International Standard Name Identifier*) (pour l'ajout d'acteurs historiques et de bibliothèques)¹³, des procédures de vérification à la soumission de formulaires ou encore l'obligation au remplissage de certains champs. La manière dont l'interface est construite accompagne le chercheur dans l'insertion des données ; pour l'ajout d'une nouvelle édition par exemple, les informations à renseigner sont réparties dans plusieurs formulaires successifs, chacun conditionnant les champs à remplir dans le suivant.

Une des finalités de la création de ces données est leur publication au sein de l'interface publique ; la mise en ligne étant un objectif du travail de recherche en soi¹⁴. Toutefois, certaines données ne sont prêtes à la publication qu'à l'issue d'un examen attentif de la part du chercheur ; deux états de travail ont donc été créés, un état « public » et un état « ébauche ». Les administrateurs de la base ont ainsi la possibilité de différer la mise en ligne d'un document, s'ils estiment que le processus de numérisation n'est pas terminé.

Structure de l'interface

La structure du *back office* est calquée sur celle de la base de données. Sur la page d'accueil, l'utilisateur peut trouver la liste de toutes les entités auxquelles il a accès. Ces entités sont rassemblées dans des groupes cohérents, reprenant le modèle de données :

- Entités centrales : *Original item*, *Edited text* et *Table content* ;
- Celles liées aux sources historiques : *Primary source*, *Work*, *Historical actor*, *Place* et *Library* ;
- Celles liées à l'édition : *Intellectual author*, *Secondary source* et *Journal* ;
- Celles liées au contenu tabulaire : *Astronomical parameter* et *Mathematical parameter*.

Chacune de ces entités donne accès à une page de liste de l'ensemble de ses enregistrements, ainsi qu'à des pages dédiées à chacun d'entre eux. La partie administrateur constitue en ce sens une interface graphique pour la base de données, les administrateurs pouvant accéder et modifier les ressources qu'ils ont ajoutées et parcourir celles ajoutées par d'autres ; l'autre fonctionnalité principale étant bien sûr l'insertion de données.

Aucune simplification vis-à-vis de la structure de données n'a été effectuée. Les ajustements relatifs à l'interface ont été concentrés sur la cohérence dans la navigation.

13. Ces référentiels servent également à valider la cohérence des *inputs*.

14. « Il est ainsi désormais entendu que l'obtention du financement d'un programme de recherche collectif temporaire, notamment de l'ANR, dépend fréquemment de la mise en œuvre, et plus encore de la mise en ligne, de « données » et de « corpus », destinées à la fois à être les instruments et les résultats de l'opération de recherche elle-même. » Yann Potin, « Institutions et pratiques d'archives face à la « numérisation ». Expériences et malentendus », *Revue d'histoire moderne contemporaine*, n° 58-4bis-5 (2011), p. 57-69, URL : <https://www.cairn.info/revue-d-histoire-moderne-et-contemporaine-2011-5-page-57.htm> (visité le 04/08/2019), § 2

Plusieurs niveaux de détails sont disponibles pour la visualisation des ressources : une vue d'ensemble – sur la page d'accueil – listant toutes les entités importantes de la base, des pages de listes détaillant tous les enregistrements pour une entité donnée, et enfin, des pages individuelles pour la visualisation des métadonnées relatives à un enregistrement unique. Ces différentes granularités de vues offrent plusieurs manières d'aborder les données : pour aider encore à leur appréhension, des visualisations graphiques ont été incorporées, sous forme de carte et de graphe¹⁵.

3.3.2 Saisie assistée de tables astronomiques : présentation de DTI

L'édition numérique de tables a fait partie, dès les prémisses du projet, du cahier des charges. En effet, les tables astronomiques constituent à de nombreux points de vue des objets complexes, dont l'édition, autant électronique que papier, soulève de nombreuses problématiques. La table constitue l'unité fondamentale du corpus de recherche ; à ce titre, l'élaboration d'outils dédiés à leur numérisation est une des conditions à l'obtention d'un corpus abondant. La création d'une interface pour assister à leur saisie fut donc l'un des objectifs de l'équipe technique du projet.

Choix du développement d'un outil sur mesure

Plusieurs exigences techniques se sont dessinées des volontés de l'équipe de recherche. Elles visaient d'une part à faciliter la saisie du contenu tabulaire – quelque soit son format et sa provenance – et d'autre part à intégrer des outils pour le traitement de ce contenu : remplissage semi-automatique grâce à des formules mathématiques intégrées, aide à la vérification des valeurs tabulaires, instruments de conversion et manipulation numérique, génération d'apparat critique.

De nombreux outils, existants antérieurement au projet, répondaient en partie aux exigences de développement d'une interface de saisie de table. Dans le domaine des humanités numériques en effet, de nombreuses applications sont dédiées à la transcription de sources historiques¹⁶, toutefois la plupart s'appuient sur des versions numérisées des sources. Dans DISHAS en revanche, les tables du corpus ne disposent pas nécessairement de numérisation, et une importante partie des ressources à transcrire comprend des tables déjà éditées sous format papier. Pour ces dernières, l'interface de saisie a principalement vocation à rendre la numérisation plus rapide. D'autres outils, pour l'édition ou la conversion de nombres, préexistaient également au projet, mais le format tabulaire des ressources et la présence de systèmes numériques très particuliers¹⁷ rendaient impossible leur utili-

15. Voir en annexes C les captures d'écran de l'interface administrateur.

16. On peut citer Transkribus, Recogito, *From The Page* et bien d'autres.

17. Notamment des systèmes « décimaux historiques » *fen* et *miao* chinois.

sation dans le cadre de DISHAS. Les spécificités du corpus ont donc amené à la création d'une interface sur mesure, plutôt qu'à l'agrégation d'outils existants.

La conception en interne de cette interface – appelée DTI (*DISHAS Table Interface*) – a également permis de réaliser une application intégrable directement au site Web de DISHAS. Dans un contexte où la plupart des services sont disponibles en ligne, un logiciel à télécharger aurait constitué une difficulté supplémentaire à la numérisation du corpus. De plus, DTI, par sa portabilité et sa relative facilité d'utilisation¹⁸, offre la possibilité à des non-spécialistes de saisir des tables déjà publiées, afin de décharger quelque peu les chercheurs du travail de numérisation.

Fonctionnalités de DTI

DTI rassemble aujourd’hui de nombreuses fonctionnalités, au sein d’une interface conçue pour être le plus intuitive possible. Après avoir renseigné les dimensions de la table (nombre de colonne et de rangées), un tableau vide – semblable à une feuille de tableur – est généré.

Le remplissage des cellules du tableau peut être réalisé de plusieurs façons. La méthode la plus simple est de recopier individuellement la valeur de chaque cellule, mais il est possible de ne renseigner que certaines valeurs et de laisser les outils d’interpolation remplir les cases intermédiaires. Les modèles de tables peuvent également être utilisés pour des modalités d’auto-complétion : en donnant les valeurs des arguments, DTI est ensuite capable de calculer les valeurs des celulles de la tables, selon la fonction définie dans le modèle. Chaque cellule doit ensuite être validée individuellement par l’utilisateur, de manière à assurer la qualité de la transcription. Tout état d’avancement peut être sauvegardé en tant que brouillon, tant que l’utilisateur estime que le travail de numérisation n’est pas achevé.

En outre, un module de comparaison de tables intitulé CATE permet de générer automatiquement des apparats critiques à partir de plusieurs transcription de tables. Les cases comportant des variantes sont signalées grâce à un fond de couleur distincte. D’autres repères colorés permettent d’indiquer la présence de commentaires liés à une case en particulier. L’utilisateur a également la possibilité d’exporter ou d’importer une table dans différents formats comme LATEX, JSON ou PDF. Enfin, différents traitements informatiques peuvent être réalisés à partir de ce matériau numérisé, notamment des conversion numériques et la génération de graphiques pour la visualisation des valeurs.

DTI représente l’une des importantes réalisations numériques de DISHAS. Les tables numérisées disposent, grâce à cet outil, de nombreuses informations et fonctionnalités satellites, qui doivent faire l’objet d’une médiation toute particulière auprès du public. Il

18. L’interface s’inspire largement du logiciel Excel.

est du rôle du *front office* d'œuvrer pour constituer une interface mettant en valeur cette richesse, tout en simplifiant l'appréhension d'un tel matériel documentaire.

Deuxième partie

Modalités d'exposition de la donnée

Chapitre 4

Analyse des besoins

Cette partie vise à présenter le travail de recherche qui a précédé la réalisation des interfaces du *front office* : elle expose les formes de médiation de données qui ont été mises en place pour la constitution du parcours de navigation. Ma réflexion s'est orientée autour des différents publics de la plateforme – amateurs, chercheurs, concepteurs – chacun mis en regard avec un niveau de compréhension du corpus numérique – modèle conceptuel, interactions entre les données et base en elle-même. Les chapitres suivants sont le fruit des questionnements menés lors de la partie liminaire de mon stage, circonscrivant dans un premier temps les besoins liés au développement de telles interfaces, puis présentant les solutions élaborées pour la valorisation des ressources du projet.

4.1 Définition des publics

La définition des *persona* est l'une une étape primordiale à l'élaboration de toute plateforme à destination du public : elle permet d'en cerner les attentes et d'établir en conséquence un cahier des charge et de concevoir des parcours de navigation adaptés. L'objectif est d'accompagner le public, pour l'amener à trouver ce qu'il cherche le plus aisément possible.

4.1.1 Chercheurs

La communauté de chercheurs en astronomie ancienne constitue naturellement la partie la plus importante des futurs utilisateurs de la plateforme. Leurs besoins vis-à-vis de l'interface sont ceux à prioriser. De plus, l'élaboration des pages du *front office* a été réalisée au sein d'une équipe de recherche : ce sont donc les suggestions et retours des chercheurs qui ont avant tout été pris en compte.

Administrateurs de la plateforme

Les administrateurs de DISHAS disposent d'un compte sur la plateforme, leur ouvrant ainsi l'accès au *back office*. Le *front office* doit en conséquence être pensé en termes de complémentarité avec l'interface administrateur. Cette dernière étant avant tout dédiée à la gestion des données ; l'interface utilisateur, quant à elle, doit être vouée à la présentation de ces données. Le projet DISHAS ayant vocation à fédérer les ressources, les pages publiques doivent veiller à leur mise en commun au sein de l'interface, ainsi qu'à fournir une mise en perspective de corpus de recherche très spécialisés.

Dans cette mesure, les visualisations graphiques sont des *media* privilégiés pour la valorisation des données. Elles donnent l'occasion aux chercheurs de considérer d'un seul regard les ressources ajoutées par l'ensemble des administrateurs. Les chercheurs de DISHAS mènent en effet leurs travaux au sein de projets aux enjeux variés : le *front office* doit permettre de faire le lien dans la diversité des sources produites. Pour que le *front office* s'adapte à tous les partenaires du projet, les visualisations doivent donc être orientées autant vers la recherche en SHS que vers les disciplines davantage mathématiques. En outre, il faut concevoir des visualisations pouvant aussi bien illustrer les travaux de recherche que faire surgir de nouvelles interrogations.

Afin d'assister le travail des administrateurs, la partie publique de la plateforme doit aider à trouver et mettre en rapport les ressources. Des fonctionnalités de recherche avancées doivent donc être mises en place afin d'aider à identifier et rassembler les données pour l'examen d'hypothèses scientifiques. De plus, les liens qui existent entre les données de la base doivent être rendus visibles pour aider à établir des rapprochements entre les sources du corpus. L'interface vise à faire découvrir, autant qu'à permettre de retrouver la donnée.

En dernier lieu, la présentation des données très spécialisées, à l'instar des tables astronomiques, doit être particulièrement soignée, afin d'offrir aux administrateurs une interface valorisant leur travail de recherche. Il est possible d'imaginer pour cela que la page publique dédiée à une édition de table puisse être citée au sein d'un article scientifique, ou servir de support à des travaux de recherche. Ainsi, la constitution d'interfaces offrant une approche intéressante et utile sur la source peut aussi être une motivation pour le chercheur à partager ces données. Une médiation efficace participe donc de l'utilisation optimale de la plateforme.

Étudiants en histoire de l'astronomie

Le site internet de DISHAS a également été pensé pour constituer une plateforme à vocation didactique : il doit pouvoir servir à des étudiants en histoire des sciences, aussi bien qu'à être utilisé par des élèves plus jeunes, dans le cadre d'une première introduction

à l'astronomie ancienne. Pour ce genre de public, un travail de vulgarisation doit être produit, en veillant toutefois à ne pas dévoyer le discours scientifique. Les visualisations de données peuvent, à ce titre, constituer un moyen efficace pour appréhender aisément la complexité d'un corpus.

Dans une perspective pédagogique, la plateforme doit également former un portail de ressources pour une étude approfondie, en même temps que d'offrir une introduction générale à l'histoire de l'astronomie. Les pages publiques de DISHAS visent, en ce sens, à centraliser les informations relatives à la recherche historique en astronomie et à fournir des outils dédiés à sa compréhension. Des pages de ressources annexes – glossaire, téléchargements, bibliographies – peuvent répondre à ce genre de demande de la part du public, voire constituer le matériau de potentiels supports de cours.

4.1.2 Non-spécialistes

Le projet développé par DISHAS vise en priorité un public avisé, principalement constitué d'universitaires et d'amateurs éclairés : néanmoins, dans une perspective d'ouverture, il ne faut pas négliger les utilisateurs moins avertis. La mise en contexte et la médiation du travail de recherche doivent être effectuées avec particulièrement de soins, une portion du public n'ayant pas forcément reçu d'introduction à l'histoire de l'astronomie. En effet, une partie des utilisateurs est susceptible de provenir de sites de bibliothèques en ligne, où les notices de sources primaires de l'interface seront référencées.

Amateurs

Afin de s'adresser à un public d'amateurs, il est essentiel de faire comprendre la nature de la donnée aux utilisateurs. En l'absence de discours explicatifs, il est difficile d'appréhender ce que représente l'astronomie ancienne, et plus encore ce que sont les tables de calculs. Dans cette mesure, la priorité doit être donnée à l'ajout de documentation iconographique et de contenu éditorialisé, qui participe à rendre les ressources de la plateforme plus accessibles.

L'expérience utilisateur – ergonomie des interfaces, architecture de l'information, cohérence graphique, etc. – guide, en outre, les publics les plus néophytes dans la circulation au sein du site et dans la navigation à travers son contenu. Des visualisations de données dynamiques, par leur aspect didactique et interactif, peuvent également accentuer l'intérêt vis-à-vis des pages de l'interface. En ce sens, l'attention portée à l'enveloppe visuelle du site constitue un atout pour la médiation auprès de public inexpérimenté : elle permet notamment à des utilisateurs curieux de saisir plus aisément les objectifs scientifiques de la plateforme et la nature de l'objet de recherche.

Professionnels des bibliothèques

L’interface utilisateur de DISHAS ayant vocation à mettre à disposition du savoir de spécialistes à propos de manuscrits et d’incunables, les professionnels de bibliothèques sont un public à considérer dans la réalisation des pages publiques. Par conséquent, les métadonnées relatives aux sources primaires doivent être particulièrement soignées : les notices pourront ainsi être référencées sur les plateformes Web des institutions de conservation, de manière à enrichir leurs catalogues numériques. Les données de DISHAS constituent un corpus qualitatif, s’appuyant sur des standards et des normes¹ : l’exposition de ces métadonnées contribue à établir des notices riches, susceptibles de compléter les informations bibliographiques des institutions de conservation n’ayant pas nécessairement l’expertise pour identifier le contenu scientifique de ces sources primaires.

D’autre part, afin d’assister au travail de catalogage, l’interface publique de DISHAS veut proposer des fonctionnalités de recherche au sein même d’une table. En renseignant une partie du contenu numérique d’une table, il serait alors possible d’identifier des manuscrits similaires, ou bien de remonter jusqu’à l’œuvre intellectuelle d’où elle provient. Le soin porté aux fonctionnalités à destination des bibliothécaires peut ainsi aider à valoriser les fonds de bibliothèques en enrichissant leurs notices numériques, mais aussi à améliorer la visibilité du projet, en disposant de relais au sein des plateformes en ligne des institutions de conservation.

4.1.3 Ingénieurs numériques

Le dernier public à considérer dans l’élaboration du *front office* est formé par les utilisateurs intéressés par la donnée brute, que ce soit les concepteurs de la plateforme ou bien des développeurs souhaitant exploiter le corpus numérique. Bien souvent, les interfaces graphiques invisibilisent les ressources numériques sur lesquelles elles s’appuient. En accord avec les logiques prônées par le mouvement de l’*open data*, il est important de donner un accès direct à la base de données, ou du moins à une partie. Le déploiement d’une API, couplée à la construction d’une interface de requête, est l’un des moyens privilégiés d’interaction avec la base de données pour les utilisateurs extérieurs.

D’autres fonctionnalités peuvent être implémentées au sein du *front office* dans le but d’exposer les données brutes : on peut citer notamment les exports dans des formats largement répandus comme le JSON. Les visualisations de données peuvent également revêtir un intérêt pour un public technicien : des représentations graphiques de la base peuvent, par exemple, donner à voir l’architecture de données et éventuellement révéler des corrélations inattendues entre les entités du corpus. En dernier lieu, l’ouverture du code et la documentation des interfaces représentent un point essentiel de la médiation

1. Les bibliothèques sont alignées sur l’ISNI, les acteurs historiques sur le VIAF et les sources secondaires disposent d’un identifiant ISBN, ISSN ou DOI.

pour les ingénieurs. Il est donc important d'expliciter les traitements de la donnée dans la plateforme DISHAS, pour éventuellement que des utilisateurs puissent réemployer et valoriser le corpus numérisé.

Ainsi, l'interface publique doit faire cohabiter des parcours de navigation différenciés : de multiples discours sur la base de données se mêlent pour constituer une plateforme ouverte sur le public et adaptée aux besoins spécifiques des chercheurs. Chacun des utilisateurs doit trouver au sein du *front office* les fonctionnalités qui répondent à ses attentes, de même que les visualisations qui éclairent sa compréhension des enjeux de recherche. Enfin, différents degrés d'appréhension de la donnée doivent être mobilisés, allant de la conceptualisation abstraite à la donnée brute et réutilisable.

4.2 Enjeux liés à la donnée

L'examen de la nature des données à exposer est une étape indispensable de l'élaboration d'interfaces publiques. De la même manière que la modélisation tient compte de la forme du matériel documentaire et des hypothèses de recherche, la construction d'une plateforme publique doit prendre en considération la manière dont la donnée est modélisée, ainsi que le cadre technique à disposition pour sa réalisation. L'interface utilisateur doit correspondre au mieux aux données et à leur mise en valeur, en accord avec les principes du *fitness for use* – aptitude à l'emploi – souvent évoqués pour évaluer la qualité des données.

4.2.1 Contraintes qualitatives

Principes FAIR

L'incorporation de DISHAS au projet ALFA implique que les exigences liées au financement par l'ERC s'appliquent avec la même rigueur aux deux corpus numériques produits. Le plan de gestion de données d'ALFA inclut donc dans ses prévisions le contrôle qualité de la base de données de DISHAS.

Les prérogatives quant à la donnée prodiguées par l'ERC peuvent être résumées par les principes FAIR. Ils définissent quatre propriétés de la donnée : la « retrouvabilité », l'accessibilité, l'interopérabilité et la « réutilisabilité ». Ces qualités forment un écosystème qui assure l'exploitation optimale des corpus de recherche ; leur prise en compte dans l'interface publique constitue donc un impératif.

Plusieurs fonctionnalités peuvent ainsi concourir au respect des principes FAIR : la réutilisation et l'interopérabilité, par exemple, peuvent être facilitées par la possibilité d'export des données dans de multiples formats. L'exposition de la base par le biais d'une API ou encore l'implémentation d'un moteur de recherche efficace favorisent quant à eux

l'accessibilité des ressources. La visibilité des données enfin, peut être assurée par divers moyens, allant de la SEO (*Search Engine Optimization*) – trouvabilité sur les moteurs de recherche – à l'UX *design*² – trouvabilité auprès de son public – en passant par la création d'identifiants pérennes – trouvabilité des contenus. Le travail de médiation en accord avec les principes FAIR revêt donc de multiples aspects.

Contraintes liées à la protection des données

La base de données de DISHAS conserve certaines données à caractère personnel³ pour le bon fonctionnement de la plateforme : nom, prénom, adresse email et manipulations effectuées dans la base. Ces informations tombent sous le coup du RGPD⁴, leur caractère sensible rend leur utilisation dans les interfaces publiques particulièrement délicate. Le *front office* doit donc, en même temps que garantir la visibilité du travail des chercheurs en mentionnant leur nom dans les notices, assurer la protection des données personnelles et informer le public de l'utilisation qui en est faite. Ainsi, les informations nécessaires à l'identification de l'auteur doivent être affichées sur les pages d'enregistrements, mais aucune autre donnée personnelle ne doit être divulguée, dans la mesure où cela n'est pas essentiel. Par exemple, pour éviter de publier les adresses email des administrateurs, une interface de messagerie intégrée au site pourra être mise en place pour mettre en contact le public et les chercheurs.

4.2.2 Spécificités des données de DISHAS

Données comme matériel de la recherche

Les données du projet DISHAS sont le produit comme le support du travail des chercheurs : la qualité du corpus en fait un objet qu'il est intéressant de valoriser au-delà de la seule communauté de recherche en histoire de l'astronomie. C'est pour cette raison qu'en fin de parcours, les ressources du projet seront versées dans les entrepôts numériques d'Huma-Num⁵ dans le but d'être conservées sur le long terme. Le *front office* constitue

2. L'UX (*User eXperience*) *design* ou expérience utilisateur en français est au carrefour de plusieurs pratiques complémentaires du *design* : *design* d'interface, d'interaction, architecture de l'information et ergonomie des interfaces. Benoît Drouillat, *Le design interactif : du Web design aux objets connectés*, Dunod, Paris, 2016

3. Informations se rapportant à une personne vivante identifiée ou identifiable. « Règlement (UE) 2016/679 Du Parlement Européen et Du Conseil Du 27 Avril 2016 », dans *CNIL – Règlement Général sur la Protection des Données*, URL : <https://www.cnil.fr/fr/reglement-europeen-protection-donnees> (visité le 13/08/2019)

4. Les données à caractère personnel doivent être « collectées pour des finalités déterminées, explicites et légitimes, et ne pas être traitées ultérieurement d'une manière incompatible avec ces finalités ; le traitement ultérieur à des fins [...] de recherche scientifique [...] n'est pas considéré, conformément à l'article 89, paragraphe 1, comme incompatible avec les finalités initiales » – Article 5.1-b) du RGPD. *Ibid.*

5. À l'intérieur du triplestore de Nakala.

la première étape d'exposition de la donnée, étape d'élaboration et d'identification des besoins en terme d'interfaces.

L'exposition des données sur la plateforme DISHAS ne procède pas des mêmes contraintes que la publication des données sur une base collective comme Nakala : en effet, les ressources sur lesquelles s'appuie l'interface de DISHAS sont des données en construction ; les corpus de recherche déposés dans Nakala en revanche, sont constitués de données ayant fini leur cycle de vie. La conception de la plateforme doit prendre en compte la fluctuation de la masse de données, mais aussi les potentielles évolutions du modèle conceptuel. En outre, les données de DISHAS ont vocation à être utilisées, et leur manipulation dans les travaux de recherche peut faire surgir de nouveaux besoins vis-à-vis de l'interface ; la flexibilité et la polyvalence sont par conséquent des qualités déterminantes pour la pérennité des pages de la plateforme.

Hétérogénéité du corpus

Un autre aspect à prendre en compte dans la construction des pages publiques est constitutif de l'hétérogénéité du corpus : la plateforme de DISHAS réunit des projets traitant de sources très diverses et provenant d'horizons variés. La pluralité des cultures représentées dans le projet ne doit pas être réduite au sein de ses interfaces ; chacune des visualisations réalisées pour le *front office* doit donc être élaborée dans une perspective d'inclusion. De la même manière, si un contenu scientifique concernant une tradition astronomique est exposé sur la plateforme, il devra être mis en regard de contenus similaires traitant des autres cultures présentes dans DISHAS. De même, une attention particulière devra être portée à la représentation des différentes cultures du projet de manière équivalente dans l'iconographie, ainsi qu'à l'emploi de polices typographiques comprenant des signes étrangers, allant des idéogrammes chinois aux caractères arabes.

Hétérogénéité du corpus signifie également d'éventuelles disparités au niveau de l'information. En effet, selon la source et selon le besoin d'analyse, la finesse de description du matériel documentaire peut amener à varier. Les données de DISHAS, bien que qualitatives, ne sont pas nécessairement homogènes : ainsi, dans l'encodage, un même manuscrit pourra être indiqué comme ayant été copié à Paris ou bien au collège de Sorbonne, selon le contributeur. De la même façon, une œuvre pourra être désignée dans la base par son incipit comme par un titre forgé. L'interface doit donc pouvoir composer avec les différents niveaux de description, en proposant notamment des visualisations de données mettant sur le même plan des encodages dissemblables, à l'instar des cartes, où sont physiquement rapprochés des lieux qui sont désignés diversement dans la base de données. Le matériau de la base étant protéiforme, la plateforme publique de DISHAS se doit d'être adaptable et modulaire.

4.3 Constitution d'une interface publique

4.3.1 Objectifs d'une interface publique

L'interface publique de DISHAS doit ainsi répondre à de nombreuses exigences. Les données de recherche, constituant le cœur de la plateforme, sont de natures diverses et nécessitent d'être traitées de manière particularisée. Les publics, ayant différentes demandes vis-à-vis de la plateforme, doivent trouver au cours de leur navigation les éléments correspondant à leurs attentes. Le *front office* doit donc pouvoir servir de catalyseur au travail de recherche, mais également façonner un cadre propice à l'exploration libre.

Plusieurs objectifs se dessinent de ces impératifs. Ils seront discutés plus en détail dans les prochains chapitres de ce mémoire :

- la contextualisation de la donnée, qui a pour but de reconstituer l'environnement dans lequel elle trouve son sens, et ainsi présenter le corpus de recherche et le projet plus largement ;
- l'illustration et l'analyse du matériel numérique, qui permettent sa mise en valeur comme son examen approfondi ;
- l'exposition des ressources – ressources de recherche comme ressources techniques – qui vise à les rendre trouvables sur le site par des modalités de recherche, mais également de manière plus large sur le Web.

Tous ces objectifs ne sont bien entendu pas imperméables, et ils se mêlent pour la constitution d'interfaces riches. Chaque page doit receler en elle d'éléments liés à chacune de ces missions, de manière à pourvoir l'utilisateur de points d'accès multiples à la donnée. La plateforme publique doit de plus proposer des interfaces innovantes, pensées en complémentarité du *back office*, mais également des autres plateformes de projet en histoire de l'astronomie : en ce sens, elle peut également être un portail vers les ressources extérieures, enrichissant ainsi le réseau dans lequel elle s'inscrit.

L'interface publique doit être modelée à partir des hypothèses scientifiques sous-jacentes au projet ; elles constituent un fil conducteur à la navigation, guidant l'internaute dans sa circulation à travers les pages. Le *front office* constitue ainsi une manière nouvelle de médiatiser le travail du chercheur : il est une forme de discours sur la base de données, en même temps qu'un discours sur les sources. En ce sens, il peut être considéré comme faisant partie des produits du travail scientifique⁶ : tous les aspects de son élaboration

6. « La monographie, avec sa linéarité, n'est plus la seule forme possible d'écriture de l'histoire, qui pourrait se faire multimédiale, incorporer une multiplicité de parcours possibles, ou devenir polyphonique, sous l'effet de hypertexte, de l'accroissement des capacités de stockage et de l'augmentation constante des capacités de calcul. » Philippe Rygiel, « L'ordinateur, le réseau et l'écriture de l'histoire », *Materiaux pour l'histoire de notre temps*, N° 82-2 (2006), p. 75-79, URL : <https://www.cairn.info/revue-materiaux-pour-l-histoire-de-notre-temps-2006-2-page-75.htm> (visité le 04/08/2019), § 4

doivent donc être considérés avec attention, dans un dialogue constant entre technique et recherche⁷.

4.3.2 Notions d'expérience utilisateur

Principes généraux

Le *front office*, plus encore que le *back office*, se doit d'être doté d'interfaces simples d'apprehension. En effet, comme le public visé comporte des utilisateurs n'étant pas forcément familiers de l'astronomie ancienne, il faut veiller à la fluidité de la navigation ainsi qu'à l'intuitivité des fonctionnalités. En outre, les pages publiques sont une vitrine pour le projet DISHAS, et la limpideur du discours et l'attractivité des interfaces participent à la communication du travail de recherche⁸.

L'*UX design* ou « expérience utilisateur » est un concept né sous la plume de Donald Norman, professeur émérite en sciences cognitives de l'Université de Californie à San Diego, dans son livre *The Design of Everyday Things* paru en 1988. Il se définit comme la conception fondée sur l'utilisateur : à chaque étape du processus de création d'un produit, les besoins des utilisateurs sont placés au centre de la réflexion. L'expérience utilisateur vise à l'ergonomie des interfaces par rapport au besoin du public, de manière à réaliser un produit à la fois fonctionnel et signifiant.

L'application concrète de ces principes se situe au carrefour de la technique (par le développement des fonctionnalités), de la stratégie (par l'attention portée aux besoins des utilisateurs) et du graphisme (par l'élaboration d'une charte visuelle). Elle prend en compte tous les aspects du développement, depuis l'analyse des besoins jusqu'au choix d'une palette de couleur. L'expérience utilisateur procède souvent par processus itératifs, comprenant une phase de réflexion, de construction puis de test.

Méthode Garrett

J.J. Garrett propose dans son livre *The Elements of User Experience*⁹ une méthode globale pour la constitution de plateformes en accord avec les principes de l'expérience utilisateur. L'idée est d'amener le concepteur à interroger chaque étape du processus de

7. Voir en annexes D les diapositives présentées lors d'un séminaire numérique en présence de l'équipe de recherche.

8. « Dès lors, si l'on ne veut pas oublier d'humaniser les humanités numériques elles-mêmes, [...] il convient de travailler à atteindre dans les projets la plus haute qualité d'expérience utilisateur, non seulement grâce au design d'information et au design de données [...], mais plus encore en faisant appel systématiquement à la culture globale du projet en design et aux méthodologies les plus récentes qu'elle a fait émerger. » F. Clavert, « Design & Digital Humanities : le design comme méthode pour les Humanités Numériques ? », dans *THATCamp Paris*, URL : <https://tcp.hypotheses.org/849> (visité le 10/08/2019)

9. Dans sa version française : Jesse James Garrett, *Les éléments de l'expérience utilisateur : Placer l'utilisateur au cœur de la conception des produits web et mobiles*, Pearson, Paris, 2011

création pour faire émerger les solutions qui seront le plus à même de guider l'utilisateur vers ce qu'il cherche.

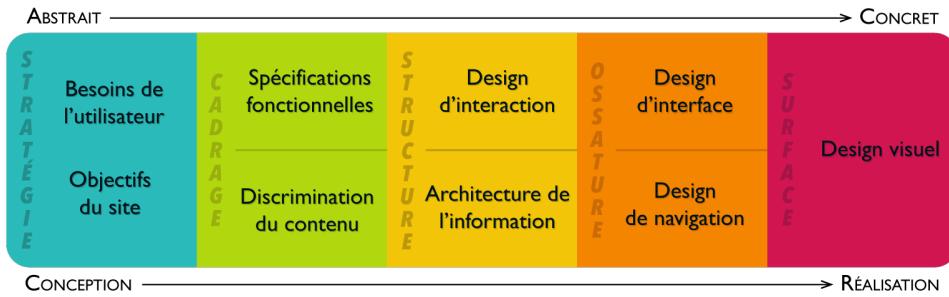


FIGURE 4.1 – Schéma inspiré du diagramme de l'UX *design* conçu par J.J. Garrett

Le modèle de Garrett¹⁰ propose une élaboration en cinq étapes, correspondant à cinq niveaux de conception de l'interface, allant du plus abstrait au plus concret. Il s'agit tout d'abord de définir les enjeux et de circonscrire les besoins, en explicitant en quoi la construction de la plateforme est profitable. La deuxième étape de cadrage délimite le périmètre technique du projet, par l'établissement notamment d'un cahier des charges fonctionnel¹¹. L'étape suivante concerne la structure globale de la plateforme ; y sont définis les scénarios de navigation et les interactions entre les différentes parties du site. La réflexion porte ensuite sur l'agencement de l'interface à l'intérieur de l'écran : la composition générale et la cohérence structurelle des pages sont alors établies. Enfin, une dernière étape vient fixer la charte graphique de l'ensemble, dans le but de caractériser l'identité visuelle de la plateforme. Chacune de ces étapes sert de socle à celle qui suit. En effet, si l'apparence du site est la composante la plus remarquée de l'UX *design* par l'utilisateur, chacun des choix visuels se doit d'être justifié par les décisions structurelles précédentes.

Datant des années 2000, la méthode développée par J.J. Garrett nécessite quelques ajustements, mais elle reste pertinente pour guider le processus d'élaboration d'une plateforme dans ses grandes lignes. L'attention portée à chaque étape de la conception permet d'éviter des écueils qui pourraient causer le délaissage de la plateforme par son public.

4.3.3 Constitution d'un parcours de navigation

Modalités de circulation dans l'interface

Le parcours navigation s'inscrit dans les réflexions liées à l'architecture de l'information : sa définition constitue une étape structurelle importante, qui détermine comment l'utilisateur va se repérer au sein de la plateforme. Plusieurs modèles de navigation peuvent être utilisés à l'intérieur d'une application Web.

10. Id., « Additional Resources », dans *The Elements of User Experience*, URL : <http://www.jjg.net/elements/resources/> (visité le 13/08/2019)

11. Voir le tableau récapitulatif des interfaces à développer à l'annexe E.

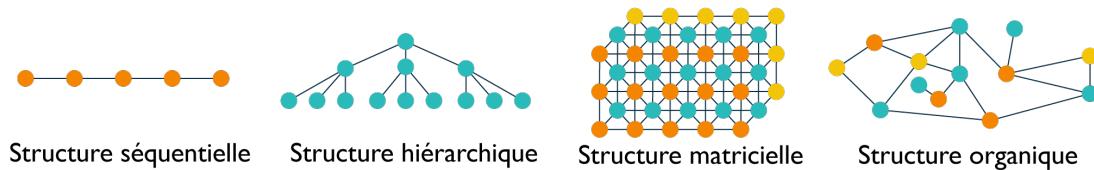


FIGURE 4.2 – Schéma inspiré du livre *The Elements of User Experience*

La navigation séquentielle repose sur le déploiement successif des pages : elle se rapproche de la logique linéaire du livre-codex. On l'utilise généralement lorsque l'ordre de déroulement est essentiel, comme c'est le cas pour un tutoriel. La structure hiérarchique correspond à une formation en arborescence des pages : ce type de structure, s'appuyant sur des relations parent-enfant, est bien compris par les utilisateurs, puisque c'est l'architecture utilisée pour l'organisation des fichiers informatiques. La structure matricielle superpose plusieurs plans, entre lesquels l'internaute peut se déplacer. Elle est utile lorsque les utilisateurs ont des besoins divergents, mais que les allers-retours doivent être possibles entre les différentes couches du site. La navigation organique ou hypertextuelle enfin, est la plus spécifique des modes de circulation du Web. Elle est plus intuitive, dans la mesure où elle suit le désir de l'utilisateur, mais induit un accès davantage discontinu à l'information. Cette structuration ne se fonde pas sur un modèle défini et donne ainsi l'impression d'une navigation libre dans les contenus, toutefois, elle rend aussi plus difficile de s'y repérer.

Ces structures entraînent différents rapports à l'information : leurs usages peuvent être combinés afin de laisser plus de liberté à l'internaute. La navigation influence profondément la compréhension du contenu ; elle est un mode d'expression du corpus, mais n'use pas des mêmes canaux que les supports plus classiques de l'exposition du travail scientifique¹². Pour la médiation des données de recherche, il faut donc repenser entièrement les modalités de présentation et d'exploration.

Structure du *front office*

Ainsi, un choix éditorial doit être effectué pour la constitution du *front office* : il résulte entre autres, de la restructuration de l'information numérique. De manière à décrire précisément les sources, le modèle conceptuel de DISHAS (figure 3.1) admet une grande complexité, mais un tel raffinement constitue un frein à la compréhension. En effet, la distinction entre contenu tabulaire (*Table content*) et texte édité (*Edited text*) n'a, par exemple, pas de sens dans une perspective explicative. Ainsi, certaines simplifications par rapport à l'interface administrateur ont été opérées pour définir la structure de la

12. « Les dispositifs électroniques élaborés par les historiens sont puissamment structurés par les imaginaires de leurs créateurs et en particulier par les métaphores dont ils usent lorsqu'ils tentent de se les approprier. » P. Rygiel, « L'ordinateur, le réseau et l'écriture de l'histoire »..., § 14

plateforme publique ; en particulier, le retrait des pages de notices¹³ d'entités secondaires. Les données relatives à ces tables – comme par exemple celle des *Historical actors* – ne sont pas supprimées des interfaces, mais considérées comme des métadonnées des entités principales.

Deux grands axes de navigation ont ensuite été déterminés, correspondant à la bipartition présente dans le modèle de données. La polarisation des parcours dans l'interface publique a été l'occasion de penser deux voies d'accès à la donnée : une davantage historique, tournée vers les sources anciennes de l'astronomie ; et une autre orientée vers les aspects « scientifiques » du corpus, à savoir les éditions de tables et les paramètres astronomiques. Chacune de ces voies mène *in fine* à l'entité centrale de la base de données, à savoir la table astronomique. Ces deux parcours composent en ce sens, deux matrices parallèles qui se font écho à travers de multiples correspondances : l'utilisateur, en fonction de ce qu'il cherche – un manuscrit ou un paramètre – sera dirigé vers l'une ou l'autre partie du *front office*. Néanmoins, de nombreuses passerelles permettront de rebondir d'un côté à l'autre, assurant ainsi le décloisonnement de l'information.

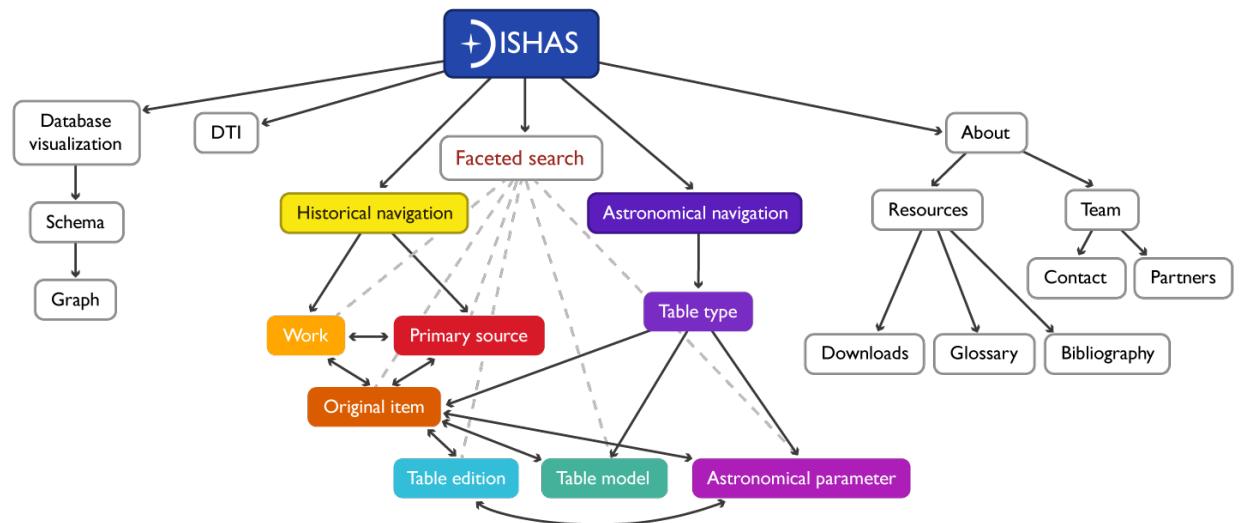


FIGURE 4.3 – Diagramme d'architecture de l'interface publique

L'abondante présence de liens est le support de l'interconnexion entre les ressources de l'interface : par les redirections, il est possible de constituer un réseau sous-jacent qui se fait l'écho aux relations modélisées dans la base de données. La navigation se construit ainsi de manière individuelle, chaque utilisateur dessinant sa propre séquence d'interactions¹⁴. Toutefois, une structure plus hiérarchique de l'information est maintenue, pour garantir un cadre dans lequel l'utilisateur peut de repérer.

13. L'expression « pages de notice » est utilisée ici pour désigner les pages qui concernent un unique enregistrement dans la base de données : par exemple, une page qui présente un manuscrit en particulier.

14. « Il s'agit, en d'autres termes, de créer des liens entre des textes et d'autres types de documents pour pouvoir organiser de façon dynamique nos parcours de lecture et de recherche et donc notre accès

Une organisation selon le niveau de granularité de la donnée permet ainsi de conserver une cohérence dans les contenus, en les ordonnant du plus général au plus spécifique. Dans cette logique, les pages de listes agrégeant les données conduisent aux notices d'enregistrement unique. De la même façon, les entités de nature mixte amènent vers les entités qui les composent : pour accéder à un *original item*, l'utilisateur devra donc transiter par la notice d'une source primaire ou d'une œuvre, dans la mesure où elles constituent un assemblage de tables. En dernier lieu, les fonctionnalités de recherche amènent un autre mode de navigation, celui de l'accès direct à la ressource à l'opposé de la libre circulation dans les interfaces. Enfin, des pages annexes (en gris sur le schéma) apportent en parallèle des contenus additionnels, utiles à la compréhension du projet de recherche.

Ainsi, ce plan permet à la plateforme publique de développer de nombreuses ramifications, tout en conservant une ossature relativement simple à comprendre. À chaque nœud de l'arborescence, il est possible de se déplacer en profondeur comme par translation, le but étant d'amener l'utilisateur le plus rapidement vers ce qu'il cherche. Le parcours de navigation pourra ainsi donner à l'internaute un accès local autant que des perspectives de longue portée sur la donnée. Chaque élément de la base y est replacé dans son environnement. Le *front office* compose ainsi un discours sur la base de données, un discours qui met au centre la table astronomique tout en mettant en valeur les données qui y sont corrélées.

Les chapitres suivants portent sur les différentes médiations de données qui ont été élaborées et conçues dans le cadre de mon stage : certaines précisions doivent être énoncées avant de présenter mes réflexions et réalisations. Les différentes interfaces et visualisations réalisées ont fait l'objet de discussion avec l'équipe de recherche ; toutefois, le développement du *front office* n'ayant pas été achevé, les interfaces conçues n'ont pas pu bénéficier de véritable retour d'expérience. De plus, la base sur laquelle s'appuient les visualisations accomplies est une base de données de test, ne présentant ni l'exactitude, ni les dimensions de la base de données finale. Du reste, la durée de mon stage ne m'ayant permis de ne travailler que sur la navigation historique, l'accent sera donc mis sur l'élaboration des interfaces de cette partie. Le travail d'élaboration des interfaces, notamment pour le parcours astronomique, n'a pas été achevé ; les solutions présentées au sein de ce mémoire sont donc sujettes à modifications.

aux contenus. » Michaël E. Sinatra et Marcello Vitali-Rosati, « Histoire des Humanités numériques », dans *Pratiques de l'édition Numérique*, dir. Michael E. Sinatra, Montréal, 2014 (Parcours Numérique), chap. 3, p. 49-60, URL : <http://books.openedition.org/pum/317> (visité le 04/08/2019), § 6

Chapitre 5

Présentation de la structure de données : médiation pour les publics non-spécialistes

En tant que vitrine, la plateforme publique de DISHAS a pour vocation de présenter les missions de recherche du projet. Un utilisateur se rendant pour la première fois sur le site doit pouvoir identifier les ambitions scientifiques portées par DISHAS, ainsi que les données exposées sur la plateforme. Plus encore que la donnée individuelle, c'est l'ensemble du modèle conceptuel qui doit être rendu explicite à l'internaute pour l'aider à comprendre la nature des ressources. En ce sens, la mission du *front office* est de rendre les contenus accessibles intellectuellement à un public profane, en proposant une mise en forme éclairante de l'information.

5.1 Garantir l'ouverture à des publics non-spécialistes

5.1.1 Organisation de l'information comme vecteur de compréhension

L'architecture de la plateforme est l'un des véhicules de la clarification du propos scientifique¹ ; l'effort de médiation doit donc se porter à y faire transparaître l'organisation du corpus de recherche. La structure conceptuelle qui accueille le matériel documentaire doit être le support de la navigation ; elle participe à la compréhension de la nature des sources comme des enjeux scientifiques. En distinguant deux parcours vers la donnée, la

1. « la vulgarisation se pense à différents niveaux : au niveau de l'écriture [...], dans le choix des sujets, dans la construction du texte, dans le fait de travailler avec un public précis en tête » Thibault Le Hégarat, « Apprendre à vulgariser ? », dans *Devenir historien-ne*, URL : <https://devhist.hypotheses.org/3413> (visité le 16/08/2019)

plateforme offre deux points de vue sur le corpus. Ces deux parcours se calquent sur le modèle de données², lui-même traduisant le travail des chercheurs sur les ressources.

Dans le *front office* de DISHAS, deux modes de médiation du modèle conceptuel ont été explorés dans le parcours de navigation. Une organisation hiérarchique suivant les relations d'inclusion entre les différentes entités de la base et une organisation hypertextuelle traduisant les corrélations entre les éléments de l'architecture.

Structuration hiérarchique

Contrairement aux bases orientées document, les liens hiérarchiques à l'intérieur d'une base de données relationnelle ne sont pas directement apparents. Ce sont les cardinalités qui traduisent les formes d'inclusions entre les entités ; ce type de connexion entre les données, simple d'appréhension pour l'utilisateur, est important à rendre visible dans l'interface. Structurer les pages de la plateforme en reproduisant ces relations entre entités permet d'accompagner l'utilisateur dans sa compréhension du site.

La tripartition conceptuelle du document manuscrit – production intellectuelle, témoin matériel, et élément tabulaire – peut par exemple constituer une difficulté initiale pour le public. Toutefois, dans la base de données, il est possible de voir qu'un *original item* n'est toujours relié qu'à une œuvre et à une source primaire ; en ce sens, ces entités peuvent être considérées comme des « parents » de l'*original item*. En effet, une œuvre comme une source primaire rassemble plusieurs tables ; en présentant l'*original item* dans un second temps de la navigation – toujours par le biais de la page de notice de son « parent » – l'utilisateur est plus à même de saisir l'interaction entre ces trois entités.

De la même manière, il est possible d'architecturer l'information au sein d'une même page. Ainsi, la page de navigation astronomique se déplie à partir de l'entité centrale de l'objet astronomique. Le concept d'objet astronomique – Lune, Soleil, Saturne, etc. – est relié dans la base de données avec de nombreuses entités, qu'il est possible de classifier en fonction du corps céleste auquel elles font référence. Chaque type de table³ est ainsi lié à un objet astronomique, chaque édition, set de paramètres⁴ et modèle⁵ étant eux-mêmes liés à un type de table. La conception de la page *Astronomical navigation* s'est développée autour de cette structure ; l'objet astronomique y sert de point d'accès aux autres entités de la base.

2. Voir figure 3.1.

3. L'entité *Table type* fait référence à la fonction d'une table, par exemple, le calcul de la vitesse solaire est un type de table.

4. Un paramètre astronomique correspond à une valeur numérique signifiante au sein d'une table : Voir 3.1.2.

5. Un modèle de table est la définition grâce à une syntaxe mathématique moderne de la fonction sous-jacente à une table : Voir 3.1.2.

Réseau hypertexte

La manière hiérarchique de dérouler l'information permet d'ordonner les entités de manière à les rendre plus signifiantes. Néanmoins, l'exploration libre des contenus peut aussi aider l'utilisateur à faire l'expérience des corrélations entre les données : en rebondissant de page en page, l'internaute tisse un réseau d'interactions. Chaque page de la navigation a donc été conçue comme un point névralgique, duquel rayonne une multiplicité de liens menant vers les contenus connexes. Ainsi, chaque élément de l'interface a été conçu pour être cliquable et rediriger vers l'entité à laquelle il fait référence. Les visualisations de données en particulier, illustrant le rapport entre les ressources, sont élaborées pour servir de portails : chaque composant graphique constitue un lien sur lequel l'utilisateur peut cliquer, afin d'explorer davantage l'entité qu'il représente⁶.

L'objectif de l'interface est de retrouver la cohérence d'une donnée profondément éclatée par sa numérisation : chaque page de notice doit composer une représentation intelligible du matériel documentaire de départ, ne se résumant pas à une simple collection de métadonnées. Avec l'aide des liens et des visualisations de données, l'interface publique doit traduire un enregistrement de la base en une représentation digitale du matériel documentaire.

5.1.2 Nécessité du contenu éditorialisé

Création de contenu décrivant la donnée

Cependant, la donnée ne se suffit pas forcément à elle-même pour être explicite. Afin de rendre les ressources de l'interface publique intelligibles, il faut parfois recourir à l'ajout de contenus explicatifs. Ces contenus peuvent avoir vocation à enrichir la donnée, à résigner son contexte de création, à la replacer dans une perspective de recherche scientifique ou encore simplement à la définir au sein du modèle de données.

Le contenu explicatif et la donnée fonctionnent de manière complémentaire et vise à fournir un accès optimal à l'information : l'exemple de Wikipédia et Wikidata est en ce sens révélateur. Selon ses besoins, l'utilisateur utilisera l'une ou l'autre de ces plateformes pour accéder à ce qu'il souhaite. S'il cherche à se renseigner sur un sujet, il privilégiera l'encyclopédie en ligne ; s'il désire en revanche extraire des données pour leur appliquer un traitement, il préférera la base de Wikidata. La donnée brute est utile dans un cadre informatique, mais en termes de compréhension, les contenus éditorialisés sont préférables. Plusieurs types de textes explicatifs ont été intégrés à l'interface utilisateur de DISHAS.

En premier lieu, il est évidemment essentiel à toute plateforme publique de recherche de disposer d'un texte introduisant les problématiques soulevées par son projet scientifique. Plusieurs versions de ce texte peuvent être composées : pour DISHAS, un texte court

6. Par exemple, sur un diagramme en colonne représentant les sources primaires contenant une œuvre, le clic sur une colonne occasionne la redirection vers la notice de la source primaire concernée.

sur la page d'accueil sert de présentation ; tandis qu'une page « À propos » développe quant à elle une version étendue. Non moins primordiales, les définitions des différentes entités de la base doivent être rendues accessibles. Elles ont été dupliquées dans DISHAS, à la fois au sein la page de chaque notice, mais également rassemblées à l'intérieur d'un glossaire. D'autres textes introductifs sont destinés à intégrer la plateforme ; lors de la conception des interfaces, il a semblé à plusieurs endroits nécessaires l'adjonction de contenus explicatifs. Dans l'attente de leur rédaction par l'équipe de recherche, des espaces ont été aménagés lors du développement des vues à l'aide de texte de remplissage.

En effet, la création de contenus éditorialisés est le fruit d'une collaboration entre chercheurs et ingénieur ; l'ingénieur aidant à identifier quelles productions scientifiques contribuent à une meilleure médiation, et le chercheur contrôlant par son expertise scientifique la pertinence des interfaces. En outre, si l'application Web de DISHAS n'a pas pour mission d'être une plateforme de publication des productions des chercheurs, une sélection de ressources diverses peut être mise à disposition de l'internaute s'il souhaite approfondir sa connaissance en astronomie ancienne. Cette documentation pourra se composer d'une bibliographie, de diapositives présentées lors de séminaires, de liens vers des ressources en ligne, etc.

L'ensemble de ces contenus façonne un contexte dans lequel la métadonnée informatisée trouve son sens. Cette éditorialisation de l'interface peut être distillée au sein des pages de visualisation de données – grâce aux définitions et textes introductifs – mais également dans les pages périphériques à la navigation, où peut être centralisée la documentation utile à un internaute curieux souhaitant pousser son exploration.

Donner à voir par la visualisation

L'édition se transforme en éditorialisation : l'ensemble des pratiques d'organisation et de structuration de contenus sur le web. La différence principale entre le concept d'édition et celui d'édition éditorialisation est que ce dernier met l'accent sur les dispositifs technologiques qui déterminent le contexte et l'accessibilité d'un contenu, ainsi que sur la réflexion autour de ces dispositifs. » M. E. Sinatra et M. Vitali-Rosati, « Histoire des Humanités numériques »..., § 20

7. « L'édition se transforme en édition éditorialisation : l'ensemble des pratiques d'organisation et de structuration de contenus sur le web. La différence principale entre le concept d'édition et celui d'édition éditorialisation est que ce dernier met l'accent sur les dispositifs technologiques qui déterminent le contexte et l'accessibilité d'un contenu, ainsi que sur la réflexion autour de ces dispositifs. » M. E. Sinatra et M. Vitali-Rosati, « Histoire des Humanités numériques »..., § 20

les œuvres, plusieurs formats ont été expérimentés pour représenter le lien entre œuvre, source primaire et *original item*.

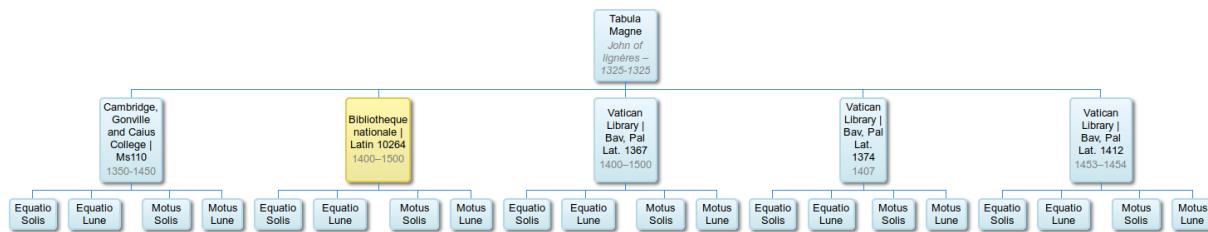


FIGURE 5.1 – Organigramme réalisé avec *Google charts* lors de phase d’élaboration des interfaces

Par exemple, la représentation sous forme d’organigramme (cf Figure 5.1) d’une œuvre pose plusieurs problèmes : en premier lieu, la disposition horizontale de la visualisation est susceptible d’excéder rapidement la largeur d’un écran, si le jeu de données est trop important. En outre, l’organigramme est un type de visualisation qui évoque une hiérarchie entre les éléments, alors que le modèle de données établit l’œuvre et la source primaire comme étant deux versants de l’*original item*. Si l’œuvre préexiste au manuscrit, elle n’en est pas la parente. Ainsi, la visualisation, bien que juste, véhicule de fausses impressions sur la donnée et n’aide pas l’utilisateur dans sa compréhension des sources.

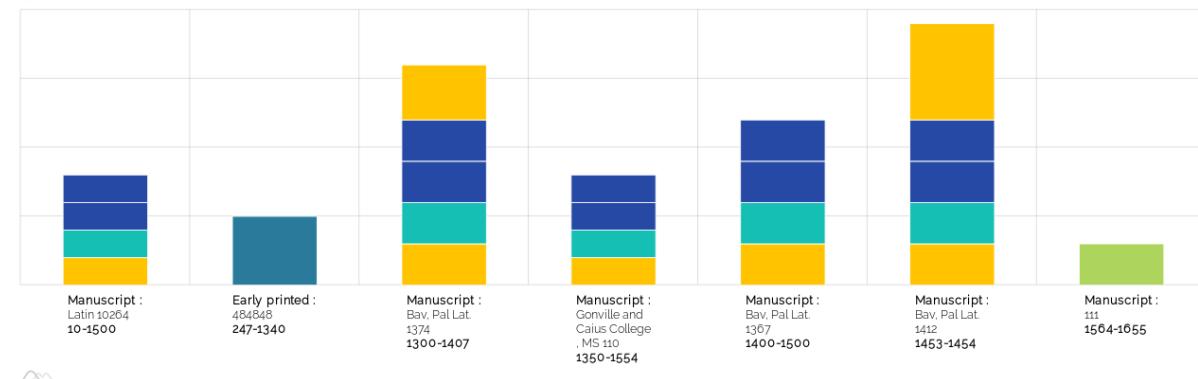


FIGURE 5.2 – Diagramme en colonnes conçu avec *AmCharts* pour les notices d’œuvres

La visualisation élaborée pour la page de notice d’une œuvre évite les écueils de l’organigramme. Ici, chaque colonne correspond à une source primaire contenant l’œuvre de la *Tabula Magna*. Chacune de ces colonnes est composée de blocs, représentant chacun un *original item* contenu dans la source primaire. La hauteur des blocs est proportionnelle au nombre de folios de l’*original item* en question, la couleur faisant référence à l’objet astronomique dont il traite. Ainsi, d’un seul coup d’œil, l’utilisateur peut repérer

la complétude de l'œuvre dans chacune des sources primaires, pour se faire une idée générale de leurs contenus. En peu d'espace, de nombreuses informations sont transmises à l'utilisateur, et la relation entre les données est rendue visible sans être dénaturée.

5.2 Attractivité des contenus

L'usage du numérique permet de concevoir de nouvelles médiations en regard de celles employées pour le format imprimé⁸ : les modes de présentation de l'information doivent être repensés pour prendre en compte la plus vaste diffusion des ressources permise par l'informatique. Afin de susciter l'intérêt du public, la seule mise à disposition du contenu n'est généralement pas suffisante. Si l'architecture de l'information aide à le rendre cohérent, sa mise en forme graphique participe à son attractivité : en effet, le succès d'une plateforme est en partie occasionné par le soin apporté à ses visuels. Plusieurs éléments sont à prendre en compte dans la création d'une plateforme : l'aspect graphique bien sûr⁹, mais également l'intuitivité et l'interactivité des interfaces, qui impliquent l'internaute dans sa navigation.

5.2.1 Cohérence des interfaces

Définition d'une charte graphique

La définition d'une charte graphique pour la plateforme publique de DISHAS vise deux objectifs : rendre les interfaces plus attractives visuellement, mais également assurer la cohérence entre les pages. La navigation hypertextuelle a tendance à rendre le propos décousu ; la reprise d'éléments graphiques le long du parcours de l'utilisateur peut être utile à l'intelligibilité du discours. L'enveloppe visuelle de la plateforme ajoute une couche de signification au contenu, en mettant en avant les informations principales et en tissant des liens entre les composants de même nature. En repérant clairement les entités de la base à l'aide de couleurs et de pictogrammes distincts, il est possible de rendre le modèle de données plus apparent au travers de la plateforme. Par ailleurs, l'interaction entre les vues doit répondre à une logique unique, chaque élément devant réagir de la manière la plus adéquate aux attentes de l'utilisateur¹⁰.

L'identité visuelle des pages publiques a été élaborée en contraste de l'apparence du *back office*. L'ensemble a été pensé pour être ludique et facile d'appréhension : les

8. « Nous trouvons trace également de dispositifs savants nouveaux, sans équivalents papiers [...]. » P. Rygiel, « L'ordinateur, le réseau et l'écriture de l'histoire »..., § 8

9. « La création artistique et tout particulièrement le design numérique (art appliqué au numérique), est sans conteste un acteur clé de ce renouveau scientifique. » Marie-Laure Massot, « Dessiner les acteurs en humanités numériques : la recherche en dialogue », *HAL – Archives ouvertes* (, juin 2018), URL : <https://hal.archives-ouvertes.fr/hal-01794185> (visité le 10/08/2019)

10. Le principe d'adéquation de l'apparence et du comportement dans le domaine du design est nommé par Donald Norman « affordance ».



FIGURE 5.3 – Pictogrammes élaborés pour les principales entités de la plateforme de DISHAS

visualisations de données ont par exemple été placées au centre des pages de notices. En outre, des pictogrammes pour chaque entité ont été conçus et utilisés comme des repères dans l’arborescence du site ; des couleurs leur ont été assignées pour renforcer encore l’impression de cohésion entre les parties du *front office*. Enfin, le comportement des éléments a reçu un soin particulier, à l’instar du mouvement de déploiement et de repli de la barre latérale¹¹ sur les pages de notices¹², de manière à renforcer l’ergonomie de la plateforme.

Construction des pages de notice

Toutes les pages de notices ont été conçues sur le même modèle : en haut de l’écran, une aide à la navigation rappelle à quel endroit de l’arborescence se trouve l’utilisateur. En dessous de ce « fil d’Ariane »¹³, un titre offre des renseignements quant à l’entité en train d’être visualisée, et introduit la visualisation graphique qui occupe l’essentiel de l’espace, chaque entité disposant d’une visualisation différente. Enfin, sur la gauche, une barre latérale peut être déployée afin d’afficher à l’écran les métadonnées relatives au présent enregistrement, mais aussi d’en donner une définition ou encore de proposer des exports des informations contenues dans la page.

De manière à enrichir les notices, chacune des pages agrège les informations relatives aux entités connexes à un enregistrement. En croisant les données des différentes tables de la base, l’interface compose de nouvelles métadonnées pour l’entité à visualiser. De cette façon, l’utilisateur n’a pas seulement accès aux métadonnées d’un unique enregistrement : il lui est aussi par exemple possible de connaître les langues présentes dans un manuscrit, alors que la table *Language* n’est reliée qu’à la table *Original item*¹⁴. Le choix des métadonnées à afficher dans une notice doit être l’objet de considération, certaines données n’étant pas pertinentes à faire figurer : par exemple, il serait trompeur d’indiquer une liste de bibliothèques à l’intérieur d’une notice d’œuvre, l’œuvre étant un concept intellectuel, plutôt qu’un objet susceptible d’être conservé.

Enfin, il est possible d’inscrire les notices de DISHAS dans une structure plus large que la simple base de données du projet. Dans la phase de modélisation, le recours à des normes

11. Des captures d’écrans vidéos ont été prises des interfaces et ajoutées dans les fichiers annexes.

12. Voir la fonction pour le comportement de la barre latérale à l’annexe G.2.2 : l’ouverture est déclenchée au clic sur un pictogramme et la fermeture au clic en dehors de la barre latérale.

13. Pour plus d’informations concernant la réalisation des *breadcrumbs*, voir à l’annexe G.4.1.

14. Il s’agit en effet de métadonnées de métadonnées au sein de la base.

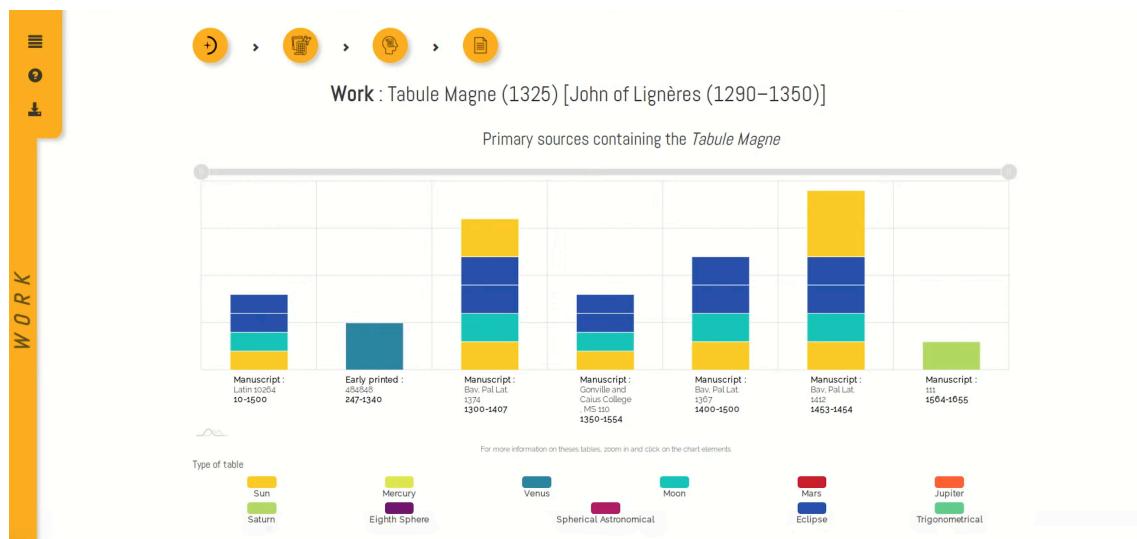


FIGURE 5.4 – Exemple de page de notice pour l’œuvre de la *Tabule Magne*

permet d’ouvrir les silos informationnels ; lors de la conception d’interfaces, la redirection vers les notices connexes des plateformes en ligne de ces normes aide au décloisonnement de l’application Web. Les notices de tables astronomiques, bien que ne répondant pas aux modèles de description bibliographique classiques, peuvent ainsi s’ancrer dans un écosystème de ressources numériques ; les pages d’enregistrement de DISHAS pourront ainsi à l’avenir être référencées en retour par les institutions de conservation, de manière à enrichir les notices de leurs collections en ligne.

5.2.2 Intégration de modules interactifs

Implication de l’internaute

L’exemple des expositions en ligne nous renseigne sur les différences entre exhibition matérielle – davantage linéaire et dirigée – et virtuelle des sources – transversale et interactive. Le cas de l’exposition « Le monde en sphères »¹⁵ est significatif, puisque les ressources exposées comprenaient de nombreux objets, notamment des globes, dont la corporéité pose la question de la numérisation. La modélisation 3D des globes a été choisie pour remplacer le dispositif en vitrine, laissant l’utilisateur libre de la manipulation pour le choix de l’angle de vue le plus adéquat¹⁶. Cette forme de médiation aide donc l’internaute à s’emparer des ressources digitales, sans occasionner de perte notable vis-à-vis du document original. La présentation numérique d’un corpus implique donc d’adapter les méthodes d’appréhension des sources : si la dématérialisation des documents implique une

15. « Le monde en sphères » Bibliothèque nationale de France (16 avril - 21 juillet 2019). Commissaires d’exposition : Catherine Hofmann & François Nawrocki.

16. Des écrans tactiles permettant aux visiteurs de manipuler les globes dans leur version 3D ont été intégrés au parcours d’exposition de la BnF.

distanciation et une restriction¹⁷, l’implication de l’utilisateur à leur exploration permet une réappropriation du contenu.

Les visualisations de données statiques véhiculent l’information de manière unilatérale : l’ajout d’interactivité aux interfaces permet de laisser l’utilisateur manipuler la donnée pour lui donner la forme qu’il trouve la plus signifiante¹⁸. La dimension dynamique d’une plateforme, tout en participant à son attractivité, propose à chaque utilisateur de façonner des visualisations sur mesure ; en ce sens, elle adapte le discours, sans simplifier le propos.

Visualisations de données modulables

Plusieurs formes d’interactivité ont été intégrées aux contenus Web de la plate-forme publique. La forme la plus élémentaire d’interaction avec le contenu graphique est le zoom. Chaque visualisation développée pour les interfaces dispose d’un système de zoom en conformité avec son type : l’utilisation de la molette sur une carte entraînera son agrandissement, tandis que la sélection d’une zone sur une barre de déroulement (*scrollbar*) occasionnera l’étalement horizontal du diagramme.

Par ailleurs, la visualisation élaborée pour la notice de source primaire dispose de deux états entre lesquels l’utilisateur peut alterner. Cette visualisation vise à synthétiser l’ensemble des *original item* contenus dans la source, en les classant par œuvre. Un premier état représente tous les *original item* de manière accumulée, tandis que dans le second état, les *original item* sont répartis sur le graphique selon la même disposition qu’au sein de la source primaire. Si l’utilisateur veut estimer la portion de chaque œuvre dans un manuscrit, il préfèrera la première configuration ; s’il s’intéresse en revanche à l’aspect composite du document, la seconde configuration lui conviendra davantage¹⁹.

Enfin, une troisième forme de modularité a été ajoutée aux interfaces : elle repose sur la mise en relation de plusieurs formes de représentations graphiques de l’information. L’assemblage de deux différents types de visualisations de données permet, grâce aux systèmes de sélection et zoom de chacun de ces types, de croiser et de raffiner les données à visualiser. Il est ainsi possible à l’utilisateur de n’afficher à l’écran que les ressources qui l’intéressent, confectionnant de cette manière une visualisation sur mesure. Son élabora-

17. Cf. « Les numérisations de documents originaux » Paul Bertrand, Aude Mairey, Olivier Guyotjeannin, Anita Guerreau-Jalabert, Anne-Marie Eddé et Marjorie Burghart, « L’historien médiéviste et la pratique des textes : les enjeux du tournant numérique », *Actes des congrès de la Société des historiens médiévistes de l’enseignement supérieur public*, 38-1 (2007), p. 273-301, DOI : 10.3406/shmes.2007.1958

18. « For web applications, on the other hand, the shape and form of live data coming from some data source are not always known upfront. So, whatever you imagine about the data in the design phase may turn out slightly or majorly wrong in production. Even in the least extreme cases, the chart type you chose to visualize the data may not be the best one from viewers perspective considering both the data itself and the insights they are trying to gain from it. » « Let Viewers Change Chart Type », dans *Blog amCharts*, 2018, URL : <https://www.amcharts.com/dataviz-tip-30-let-viewers-change-chart-type/> (visité le 16/08/2019)

19. Pour une description plus précise de la visualisation, voir section 6.2.1.

tion sera évoquée plus en détail dans la section 6.2.2.

La mission de présentation du projet DISHAS par le *front office* s'appuie donc avant tout sur une mise en forme attentive du contenu. L'architecture de l'information tout d'abord, guide l'internaute dans sa découverte du corpus de recherche, en s'enracinant sur le modèle de données. L'adjonction de contenus éditorialisés ensuite, fournit le socle explicatif sur lequel la donnée peut trouver son sens ; elle assure également une introduction scientifique de qualité au projet de recherche. Enfin, la définition graphique de l'interface, autant vis-à-vis de l'apparence de ses composants que du comportement de ses visualisations, achève de composer une plateforme ouverte à un public varié.

Chapitre 6

Mise en valeur des données et de leurs interactions : participation au travail de recherche

Les chercheurs constituent la part la plus importante du public de la plateforme publique de DISHAS : importante en termes de nombre, mais également du point de vue de leurs attentes. La constitution du corpus de DISHAS permet aux chercheurs de travailler sur des ensembles de sources considérables ; il est du rôle de la plateforme en ligne d'aider à embrasser plus facilement leur diversité. En fournissant des outils pour examiner la donnée dans ce qu'elle a de spécifique, ainsi que des instruments pour illustrer des assemblages importants de ressources, les chercheurs peuvent combiner des méthodes de *close* et *distant reading*¹ dans leur analyse du corpus.

La construction du *front office* est l'occasion de créer des médiums nouveaux pour l'exploitation scientifique des sources². Le travail de création des interfaces s'est ainsi concentré sur l'élaboration de visualisations signifiantes ainsi que d'interfaces sur mesure pour l'accueil de la donnée.

1. « *Close reading is the thorough interpretation of a text passage by the determination of central themes and the analysis of their development.* » / « *[Distant reading] aims to generate an abstract view by shifting from observing textual content to visualizing global features of a single or of multiple text(s).* » Stefan Jänicke, Greta Franzini, Muhammad Faisal Cheema et Gerik Scheuermann, « On Close and Distant Reading in Digital Humanities : A Survey and Future Challenges », dans *EuroVis*, 2015

2. « En effet, l'arrivée d'Internet est un moment clé dans le développement des humanités numériques, car il n'est pas un simple outil supplémentaire pour la recherche : il devient aussi un objet de recherche et, finalement, une technique qui modifie l'ensemble de nos pratiques au-delà de la communauté savante et, plus généralement, notre façon de voir le monde. » M. E. Sinatra et M. Vitali-Rosati, « Histoire des Humanités numériques »..., § 11

6.1 Décloisonnement des corpus spécialisés

6.1.1 Enrichir les données par l'interconnexion

Connecter les ressources de la base de données

L'analyse des sources numériques exige de rendre manifestes les interactions entre les données. En effet, un enregistrement isolé dans une base constitue difficilement une ressource suffisante à l'analyse ; c'est dans le réseau qui lie les différentes entrées de la base que se trouve le matériau sur lequel peut s'appuyer le travail scientifique. Le *front office* doit révéler cette trame d'interconnexions, afin d'aider les chercheurs à reconstituer les liens entre des ressources éloignées. La réunion de tables de différentes traditions dans un même corpus doit être l'occasion de créer des interfaces riches, où il est possible à l'utilisateur de déceler par exemple les échanges intellectuels qui se sont opérés.

Le premier instrument de cette mise en connexion est l'interface de recherche. Afin de parcourir efficacement les collections, le *front office* soit se dôter de fonctionnalités de recherche avancées, permettant au chercheur de filtrer parmi les ressources, pour réunir dans les résultats les données qui l'intéressent. La page de recherche de la plateforme publique ne dispose pas pour l'instant de système de filtres abouti ; la seule recherche plein-texte est disponible à l'utilisateur. Néanmoins, ce filtrage permet déjà d'effectuer des recherches assez précises. L'implémentation en outre du moteur de recherche ElasticSearch³ apporte différentes fonctionnalités qui renforcent l'efficacité de la recherche : la « *fuzziness* », entre autres, autorise un certain degré d'imprécision quant à l'orthographe des termes. Cela permet par exemple, de retrouver un auteur même en utilisant une graphie différente de celle encodée dans la base. Par ailleurs, ElasticSearch permet d'exécuter des requêtes, non seulement sur les champs liés à une entrée du corpus numérique, mais également sur les champs des entités qui lui sont associés : cela signifie qu'il est possible de retrouver tous les manuscrits contenant l'œuvre de la *Tabule Magne*, alors que l'œuvre n'est liée qu'à un *original item* dans le modèle de données.

Ces fonctionnalités extensives de recherche sont particulièrement utiles à l'identification de ressources connexes, car il a été observé que les utilisateurs sont parfois rétifs à l'utilisation de modalités avancées de recherche. En outre, l'utilisation de caractères génériques (*wildcards*) et d'opérateurs booléens – fonctionnalités prévues par ElasticSearch – dans les termes de requêtes renforce encore l'acuité des résultats d'une simple recherche par mots-clés.

Ouverture grâce aux métadonnées

L'interconnexion des données a constitué un enjeu principal dans la construction des interfaces publiques. Il était important de reconstituer dans chacune des pages le réseau

3. Pour une présentation plus développée du fonctionnement d'ElasticSearch, voir chapitre 9.2.

des données à l'intérieur de la base : les pages de notices ont donc été élaborées comme des nœuds, desquels se propage une série de liens vers les entités apparentées. Comme il a déjà été mentionné⁴, les composants graphiques des visualisations permettent *via* le clic de rediriger l'utilisateur vers les entités représentées ; mais la barre latérale de métadonnées a également fait l'objet d'aménagements particuliers pour réinscrire la donnée isolée dans son contexte relationnel. En premier lieu, grâce à l'enrichissement des notices par le croisement des données, mais également par la mise à disposition d'ouvertures relatives aux métadonnées d'un enregistrement. En effet, à côté de chacun des champs décrivant une notice, une icône de loupe a été placée : elle donne accès aux entités de même type partageant la même métadonnée. Il est par exemple ainsi possible de retrouver toutes les œuvres conçues par un même auteur (par Jean de Lignières dans le cas de la *Tabule Magne*) : en cliquant sur cette icône⁵, l'utilisateur est envoyé sur la page de recherche où les résultats de la requête correspondante apparaissent.

Incipit	Multiplicis philosophie
Creator	John of Lignères (1290–1350)
Date of conception	1325
Place of conception	Paris, France
Author of the record	Matthieu Husson

FIGURE 6.1 – Barre latérale des métadonnées de l'œuvre de la *Tabule Magne*

Ce mode de circulation dans les données permet d'anticiper les potentielles questions de recherche des utilisateurs : l'ajout systématique de ces « loupes » aux métadonnées connecte la notice à tous les enregistrements partageant une même caractéristique. Les loupes permettent de décrire des ensembles cohérents dans les ressources de la base. Elles constituent une fonctionnalité alternative de recherche, correspondant en réalité à

4. Voir la section 5.1.1.

5. Au survol, les termes de la requête exprimés en langage naturel apparaissent à l'écran : par exemple, « Toutes les œuvres conçues par Jean de Lignières ».

une forme de requête facettée⁶. Quelque soit l'angle d'approche du chercheur, cette voie d'accès à la donnée offre des perspectives pour élargir les corpus de recherche.

6.1.2 Constitution d'une interface spécialisée : le cas des éditions de tables

Les éditions de tables sont un objet central de la base de données de DISHAS ; leur mise en valeur au sein de l'interface publique constitue l'un des enjeux cruciaux de la plateforme en ligne. En effet, les contenus tabulaires sont un matériau essentiel de la recherche en astronomie ancienne et le projet DISHAS a été envisagé dès le départ comme un moyen de leur mise en commun. Si l'interface numérique offre des perspectives renouvelées pour l'édition critique, elle peut également dérouter lorsqu'elle n'est pas conçue correctement⁷.

Problématiques de l'édition critique de tables

Préalablement à la conception d'une interface, il est primordial d'identifier les problématiques spécifiques liées à l'édition de tables. Au premier chef, le format tabulaire impose de réfléchir à un moyen de repérer l'apparat critique : la pratique des chercheurs varie sur ce point. Pour certaines éditions, la table est encadrée par deux axes numérotés permettant de désigner une case à l'aide de coordonnées ; d'autres distinguent chaque variante à l'aide d'un nombre⁸, tandis que d'autres encore, dupliquent le cadre de la table pour n'y reproduire que les variantes. D'autres procédés existent pour baliser l'apparat critique ; ils disposent tous d'avantages comme d'inconvénients. Toutefois leur complexité rend parfois la lecture d'une édition critique particulièrement laborieuse⁹.

Une seconde difficulté est liée à la mise en page. Dans les manuscrits souvent, les tables sont accolées les unes aux autres sur la page, et parfois se poursuivent sur plusieurs folios à la fois. Il peut ainsi être ardu de transposer ces configurations à l'intérieur d'une édition, en particulier lorsque le format de la page impose ses bornes. Pour assister au travail d'édition, l'outil DTI offre de nombreuses fonctionnalités pour la saisie des tables¹⁰, la plateforme doit quant à elle fournir une interface adaptée à la présentation de ces tables.

6. Le clic de l'utilisateur envoie à ElasticSearch une requête formulée préalablement dans le *back end*, la page de recherche met ensuite en forme les résultats reçus. L'implémentation des fonctionnalités d'ElasticSearch dans DISHAS est plus amplement discuté à la section 9.2.

7. « L'extraordinaire possibilité de faire des allers-retours entre document original numérisé et texte édité (qui décuple les possibilités de contrôle et d'appropriation de l'édition par le lecteur), de naviguer entre plusieurs pages (plus vite et plus pertinemment que dans le volume papier) [...] arrive pourtant assez vite à saturation ; les actes, les sections de texte se trouvent largement coupés de leur contexte » P. Bertrand, A. Mairey, O. Guyotjeannin, *et al.*, « L'historien médiéviste et la pratique des textes... »

8. Cette méthode est la plus proche de la manière dont sont encodées les tables dans la base ; chaque variante est indiquée à l'endroit même de la case. La syntaxe est semblable à celle-ci :

`[{"valeur": 12, "apparat-critique": "Ms-1:11, Ms-2:13"}]`.

9. Selon les propos recueillis lors des séminaires organisés avec l'équipe de recherche de DISHAS.

10. Voir la section 3.3.2.

Propositions de fonctionnalités

Afin d'aider les chercheurs à accéder aux sources, l'interface de visualisation d'une édition doit intégrer les questionnements liés à la publication imprimée : un affichage ergonomique de l'information doit résulter de ces réflexions. Dans le but de pouvoir être éventuellement mentionnée au sein d'un article, la notice d'édition de table dans la plate-forme publique a été particulièrement soignée ; son élaboration a fait l'objet de multiples discussions avec l'équipe de recherche. Cependant, la construction de l'interface n'a pas été entamée. Les solutions présentées ci-après feront l'objet de développements futurs, la mise en forme de l'interface est donc également susceptible d'être corrigée¹¹.

Dans le but de prendre en compte les différentes perspectives de recherche, la page de notice d'une édition doit se déployer sur quatre onglets : historique, mathématique, éditorial et contenu tabulaire. Ces quatre onglets, dédiés à différents aspects de l'édition, pourront être chargés séparément afin d'optimiser les temps d'attente pour l'utilisateur. Le contenu tabulaire sera placé en position liminaire, de manière à être visualisé en premier.

L'onglet historique doit porter sur le contexte de production de la source primaire éditée ; il sera donc disponible uniquement pour les éditions de type A et B, dans la mesure où les types C sont recalculés entièrement. Cet onglet comprendra les informations relatives à l'œuvre, la source primaire et l'*original item* d'où provient l'édition. L'onglet mathématique concertera la tradition astronomique dans laquelle s'inscrit la table, en présentant les informations relatives aux paramètres utilisés en son sein. L'onglet éditorial présentera l'auteur et le contexte de publication, l'édition étant potentiellement tirée d'une version imprimée auparavant. Chacun de ces onglets comportera une visualisation de données différente, sous forme de carte pour les deux premiers, et sous forme de graphe pour le dernier.

L'onglet présentant le contenu tabulaire édité sera présenté sous forme de tableau de nombres. Dans le cas d'une édition critique (type B), un encart latéral permettra d'afficher l'apparat critique au clic sur une case ; il est possible d'imaginer un dispositif pour faire apparaître l'apparat critique de multiples cases simultanément. Un système d'intensité de couleur indiquera la densité des variantes pour les différentes cellules du tableau, identifiant ainsi les passages problématiques de la computation. Une autre sorte de repère visuel signalera la présence de commentaires. Enfin, un lien permettra d'ouvrir l'édition dans l'interface de saisie de table, afin de pouvoir la manipuler grâce aux outils implementés dans DTI. Plusieurs détails de l'interface restent encore à préciser : doit-on permettre à l'utilisateur de visualiser plusieurs tables de manière concomitante ? Quels formats d'export sont les plus adaptées à ce genre de contenu ? Comment concevoir la mise en page de tables à double argument ? Ou encore, comment afficher l'apparat critique à l'intérieur

11. La cahier des charges relatif à la réalisation des interfaces publiques donne des renseignements plus précis quant au développement de la notice d'édition : voir en annexe E.

du panneau latéral ?

Le développement de cette interface occasionnera assurément d'importantes modifications par rapport à cette première ébauche : la réalisation technique en effet devra être effectuée en collaboration étroite avec l'équipe de recherche. Les éditions de tables constituent la donnée pivot du projet DISHAS : la page de notice publique doit faire le lien entre sources historiques et paramètres astronomiques. Par la conception d'une interface sur mesure, les chercheurs disposeront d'un support adapté à la présentation et la consultation des tables éditées. Cette interface constituera en ce sens un atout du projet, pouvant inciter les chercheurs *a posteriori* à ajouter davantage de données de manière à enrichir encore la plateforme en ligne.

6.2 Analyse par la visualisation

Contrairement à l'infographie où sont organisées visuellement des données connues à l'avance, la visualisation de données dynamique s'appuie sur des données mouvantes, dont il n'est pas possible de prédire avec exactitude la manifestation graphique. En ce sens, la visualisation révèle des informations sur les ressources qu'elle représente, elle les ordonne de manière à les faire surgir sous une forme signifiante¹². Dans le but d'accompagner le travail de recherche, les visualisations de données sont particulièrement indiquées pour l'examen de larges corpus de données ; l'interface publique doit donc employer aux mieux les données de la base, de manière à proposer aux chercheurs des visualisations propre à faire émerger l'analyse.

6.2.1 Des manuscrits composites : montrer l'agencement d'un source primaire

La recherche en astronomie ancienne s'intéresse de près à l'histoire des textes ; la matérialité des documents sources constitue un terrain d'étude pour les chercheurs. La visualisation de données pour la notice de source primaire s'est ainsi construite de manière à en révéler la composition. Sans accès à la numérisation des documents originaux¹³, il est parfois difficile de fournir une représentation satisfaisante des sources. Néanmoins, le degré d'abstraction induit par cette lacune est une occasion d'aborder le matériel docu-

12. « *Visualization tools and techniques are crucial to the analysis of digital humanities data, especially in case of large amounts of data. Current visualization techniques allow now a better communication of ideas and analysis results than verbal communication. Therefore, the exploration and implementation of novel visualization techniques for displaying processed material in graphical format is an important research topic, helping to mediate a message for different types of audience.* » « Research – Data Visualization », dans *Digital Humanities*, URL : <https://dh.fbk.eu/research> (visité le 16/08/2019)

13. Numérisation au sens de reproduction numérique scannée du document.

mentaire d'un point de vue plus analytique. Adoptant un point de vue davantage distant, la représentation graphique permet de visualiser la source primaire dans son ensemble.

La source primaire astronomique – manuscrit ou incunable – est en effet souvent constituée de différentes parties, réunies plus ou moins artificiellement en son sein. Plusieurs œuvres intellectuelles peuvent s'y mélanger, parfois même s'y chevaucher ; les tables issues de ces œuvres sont parfois réparties conformément au document d'origine, mais elles peuvent également se trouver imbriquées sur un même folio. Selon les traditions, les pratiques des copistes varient, toutefois il est difficile d'évaluer les différences de compositions dans les sources¹⁴, les chercheurs s'intéressant généralement à un corpus cohérent.

Afin de révéler la configuration des œuvres et des tables au sein d'une source primaire, les métadonnées de foliation des *original item* ont été mobilisées pour l'élaboration de la visualisation. Deux états de la visualisation figurent deux façons de représenter le contenu. Le premier décrit la quantité de folios accumulée de tous les *original item* de la source, répartie par œuvre ; cet état est analogue à la visualisation conçue pour les notices d'œuvres. Il permet d'évaluer la proportion de chaque œuvre dans une source primaire. Un système de couleurs correspondant aux différents objets astronomiques identifie le sujet des *original item* contenus dans la source.

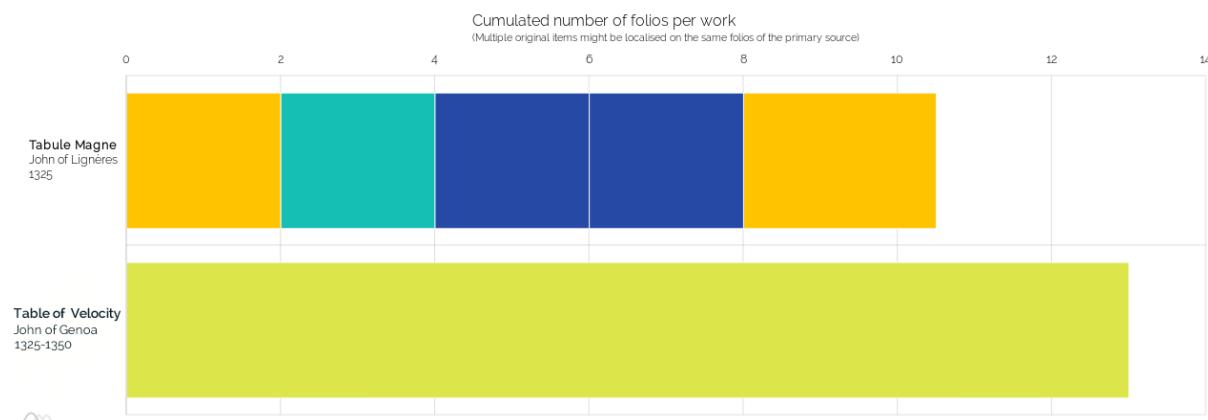


FIGURE 6.2 – Visualisation du nombre de folios accumulé dans une source primaire

Le deuxième état de la visualisation présente les *original items*, non plus comme accolés les uns aux autres, mais distribués horizontalement en fonction de leur foliation. Il est ainsi possible de repérer si plusieurs tables ont été réunies sur les mêmes feuillets, ou encore si une même œuvre a été fractionnée à divers endroits de la source primaire. Cette visualisation révèle également si le copiste ou l'éditeur a entremêlé des tables provenant de plusieurs œuvres, de manière à composer un contenu original¹⁵.

14. Les manuscrits européens semblent par exemple présenter un contenu plus composite que les manuscrits arabes.

15. Des versions simplifiées des visualisations en format HTML ont été ajoutées aux documents accompagnant ce mémoire.

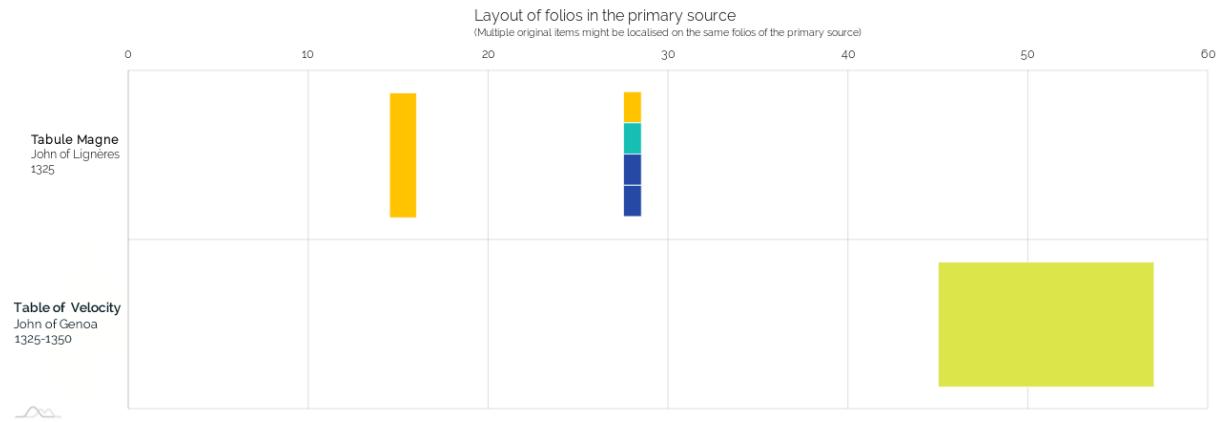


FIGURE 6.3 – Visualisation de la disposition des folios dans une source primaire

La visualisation ici sert à embrasser d'un seul regard des informations qui ne pourraient être obtenues ailleurs que par un examen attentif des sources primaires. Néanmoins, cette forme de représentation visuelle ne pourra être tout à fait significative qu'une fois que seront intégrés les autres types de contenus des sources astronomiques – diagrammes et textes – à la base de DISHAS ; il sera alors possible aux chercheurs de déceler certains motifs structurels susceptibles d'apporter un nouvel éclairage sur les sources.

6.2.2 Représenter un milieu intellectuel : le cas de la carte historique

Les visualisations ponctuelles, employées pour les pages de notices, ne représentant qu'un seul enregistrement de la base, sont utiles pour résumer graphiquement une donnée. Elles apportent un éclairage complémentaire aux seules métadonnées, mais ne peuvent remplacer une observation méthodique du matériel documentaire. La visualisation numérique accède à son véritable potentiel, lorsqu'elle agrège un grand nombre de données pour en dresser une synthèse visuelle. Plus les corpus représentés sont importants, plus ce genre de visualisation trouve son intérêt, donnant à voir ainsi des informations qu'il serait extrêmement ardu à obtenir autrement.

Représentation spatiale et chronologique

La mise en évidence de milieux intellectuels est un enjeu de recherche central pour le projet DISHAS. Le concept de milieu intellectuel peut être fondamentalement décrit par l'articulation d'un lieu et d'une époque ; leur représentation conjointe a souvent fait l'objet de réflexion, tant dans une perspective historique que pour des emplois plus éclectiques¹⁶.

16. GoogleMaps propose par exemple une fonctionnalité de *timeline* permettant d'explorer les trajets réalisés avec l'application. ; par ailleurs, la bibliothèque de visualisation en carte *Leaflet* propose huit *plugins* pour la création de « carte temporelle », généralement pour figurer des itinéraires.

De nombreux historiens se sont attachés à proposer des solutions pour résumer au sein d'une même forme graphique le temps et l'espace : on peut citer des initiatives aussi diverses que le *Chronological Chart of Ancient, Modern and Biblical History* de Sebastian C. Adams publié en 1871 qui mêle histoire profane et religieuse sur une période de près de 5000 ans¹⁷ ou encore que l'encyclopédie et atlas digital *TimeMaps* qui propose une série de cartes thématiques sur différentes époques de l'humanité.

Si la carte et la frise chronologique sont des moyens d'expression privilégiés de l'étude en SHS, leur articulation pose plusieurs problèmes, notamment dans une perspective de synchronicité. Bien souvent, le problème a été éludé en proposant une juxtaposition de cartes, chacune représentant une période définie¹⁸, ou en figurant à l'aide de symboles visuels (flèches, zones colorées, etc.) des mouvements ou événements particuliers¹⁹. L'irruption des techniques numériques dans le champ de la cartographie a ouvert les possibilités de visualisations : en laissant l'utilisateur interagir avec les variables à représenter, il est possible de construire des cartes sur mesure^{20 21}, mettant au jour des relations entre les données qui n'auraient pu être prévues avant la conception de la visualisation.

Réalisation d'une carte temporelle

Afin de représenter la diffusion des textes et la circulation des paramètres astronomiques, l'idée de construire une visualisation combinant carte et frise chronologique s'est donc rapidement imposée. La visualisation en carte a en outre l'avantage d'harmoniser les différentes granularités de description de la donnée : les entrées de la table *Place* « Paris » et « Collège de Sorbonne » apparaîtront ainsi côté à côté, bien qu'ils constituent deux enregistrements tout à fait disjoints dans la base de données. Toutefois, d'autres interrogations surgissent en parallèle, en particulier liées à la représentation d'entités placées sur un même lieu.

17. « Adams' Synchronological Chart », dans *David Rumsey Historical Map Collection*, URL : <https://www.davidrumsey.com/luna/servlet/detail/RUMSEY~8~1~226099~5505934:Composite--Adams--Synchronological-> (visité le 28/08/2019)

18. Comme la plateforme *GeaCron* créée en collaboration avec l'Université Rey Juan Carlos (« World History Maps & Timelines », dans *GeaCron*, URL : <http://geacron.com/home-en/> (visité le 19/08/2019)) ou l'atlas *TimeMaps*

19. À l'instar des nombreuses cartes du Musée national de l'histoire de l'immigration : « Cartes », dans *Musée National de l'histoire de l'immigration*, URL : <http://www.histoire-immigration.fr/questions-contemporaines/cartes> (visité le 19/08/2019).

20. À propos d'un projet pionnier en la matière, l'atlas élaboré par Éric Guichard en 1999 : « L'atlas [...] permet à l'utilisateur de choisir tant les variables qu'il souhaite voir projetées sur la carte, que d'influencer sur le mode de discréétisation des données. Il devient alors possible de générer une multitude de cartes, adaptées aux besoins du lecteur ou du chercheur, ce qui revient à créer un outil dont les fonctions ne sont pas reproductibles par un atlas papier, qu'il serait de toute façon impossible de publier. » M.L. Massot, « Dessiner les acteurs en humanités numériques : la recherche en dialogue »...

21. Les cartes interactives créées par Étienne Côme pour visualiser les flux liés aux transports en communs sont très éclairantes à ce sujet : Etienne Côme, *Portfolio / Étienne Côme*, URL : <https://www.comeetie.fr/#portfolio> (visité le 18/08/2019).

L'ajout de la dimension temporelle soulève en outre des questionnements spécifiques : comment figurer le temps sur une carte ? Est-il préférable de pouvoir sélectionner une période ou une date ponctuelle ? Comment symboliser des objets – sources primaires, paramètres – n'étant pas directement (ou parfois même aucunement) liés à un lieu ? Par quels moyens peut-on représenter des données temporelles et spatiales parfois imprécises, sans induire l'utilisateur en erreur ? La visualisation qui a été élaborée tente de répondre au mieux à ces interrogations.

Cette visualisation en carte temporelle a été conçue pour être polyvalente et s'adapter à la représentation de différentes variables. Pour l'instant, elle a été implantée sur la page de notice des *original items* et sur la page d'accueil de la navigation historique. Il est prévu qu'elle puisse servir à la présentation des paramètres astronomiques. Nous prendrons pour sa description la version de la page *Historical navigation*.

La visualisation se divise en deux parties fonctionnant de manière synchrone : une carte géographique sur le dessus où figure l'ensemble des œuvres et des manuscrits de la base de données ; sur le dessous, une carte de chaleur représente l'ensemble de la période couverte par les entités de la base. Une barre de sélection comportant deux curseurs sert de lien entre les deux cartes. La saturation de la couleur sur le dessous symbolise les périodes les plus prolifiques ; le jaune correspondant aux créations d'œuvres, le rouge aux copies/éditions de sources primaires. Les repères sur la carte représentent quant à eux les lieux de création et copie/édition de ces mêmes entités. Le rouge et le jaune gardent les mêmes références, tandis que l'orange indique un mélange entre œuvres et sources. Le diamètre des points quant à lui signale un plus ou moins grand nombre d'entités à représenter. L'utilisateur peut, au survol sur ces points et sur les bandes de la carte de chaleur, voir le nombre d'œuvres et de sources primaires correspondant au composant concerné.

La première remarque qui doit être faite concerne la visualisation des sources primaires. En effet, la source primaire n'est pas spatialisée dans la base de données, seul l'*original item* est associé à un lieu. Dans la mesure où il est plus logique d'accéder à la table astronomique par le biais de son contenant, il a été choisi de ne représenter sur cette page que des sources primaires, bien que ce ne soient en réalité que des *original item* qui apparaissent sur la carte²². Cela signifie qu'une source primaire dispose potentiellement de plusieurs points sur la carte, dans le cas où elle n'aurait pas été produite à un unique endroit.

Étant donné que cette visualisation a été pensée comme un portail, chaque repère sur la carte est cliquable de façon à faire apparaître au-dessous de la visualisation des liens vers les entités concernées. Il est également prévu de rendre cliquables les différentes zones de la carte de chaleur, permettant ainsi d'accéder à la donnée par le biais chronologique comme

22. Pour la visualisation des paramètres astronomiques, le même procédé pourra être repris de manière à faire apparaître sur une carte les sets de paramètres associés à des *original items*.

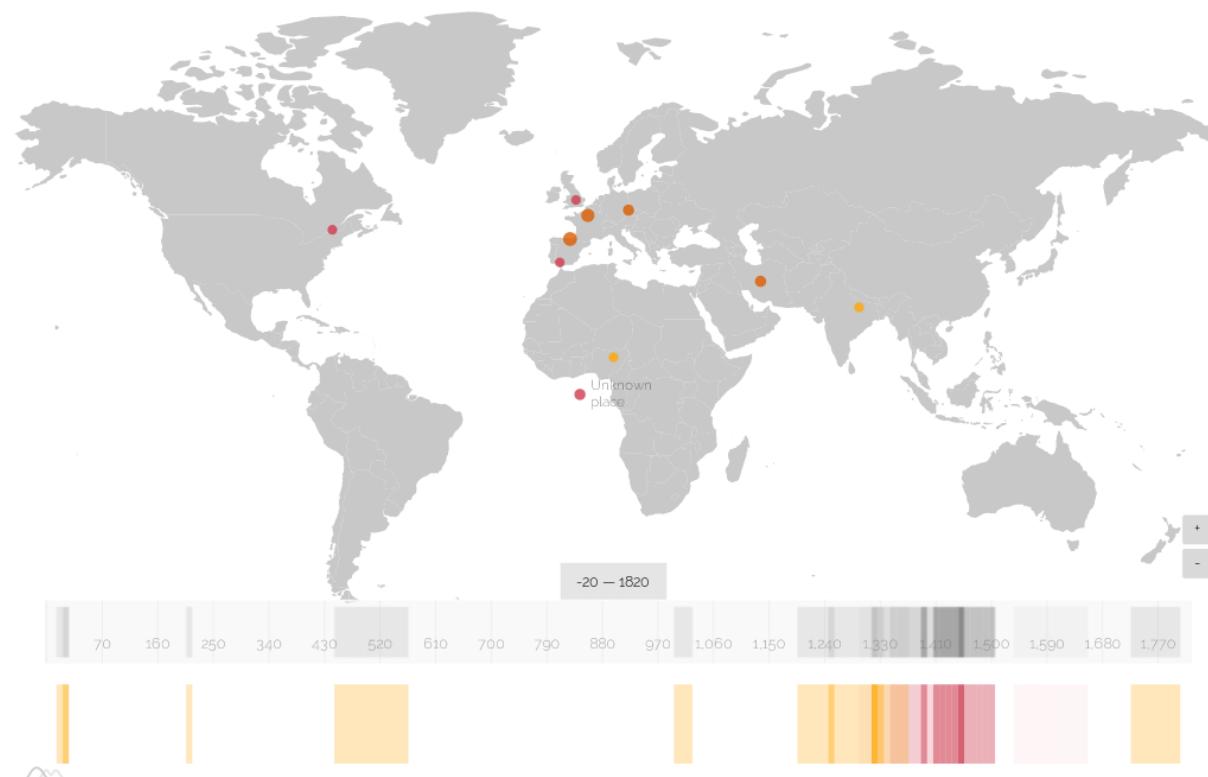


FIGURE 6.4 – Carte temporelle de la page *Historical navigation*

spatial. Cependant, certaines entités ne sont associées à aucun lieu dans la base, mais elles doivent tout de même être rendues accessibles. Pour pallier cette difficulté, tous les enregistrements ne disposant pas d'information géographique ont été artificiellement situés aux coordonnées 0,0 ; une annotation sur la carte précise qu'il s'agit d'une « *Unknown place* » pour éviter toute confusion.

Lorsque l'utilisateur effectue une sélection grâce aux curseurs mis à sa disposition, les données affichées sur la carte sont mises à jour : le diamètre, la couleur et les informations au survol des points évoluent²³ en même temps qu'un agrandissement est effectué sur la frise au-dessous. En revanche, le zoom sur la carte n'entraîne pas d'actualisation de la carte de chaleur pour des raisons d'impossibilités liées à la technologie utilisée pour la visualisation²⁴.

Cette visualisation illustre les possibilités permises par la visualisation numérique. En représentant simultanément l'ensemble des données de la base, les corrélations entre elles sont rendues manifestes, alors qu'elles restaient auparavant inaccessibles. Différents choix éditoriaux ont été faits pour la figuration des données ; ils conditionnent la manière dont sont appréhendées les informations. Il est en ce sens du rôle du concepteur de tâcher

23. Pour le détail du code permettant ces modalités d'affichage, voir en annexes G.1.2.

24. La bibliothèque JavaScript *AmCharts* ne dispose pas de moyen d'accéder aux informations relatives à la portion de carte affichée à l'écran de l'utilisateur.

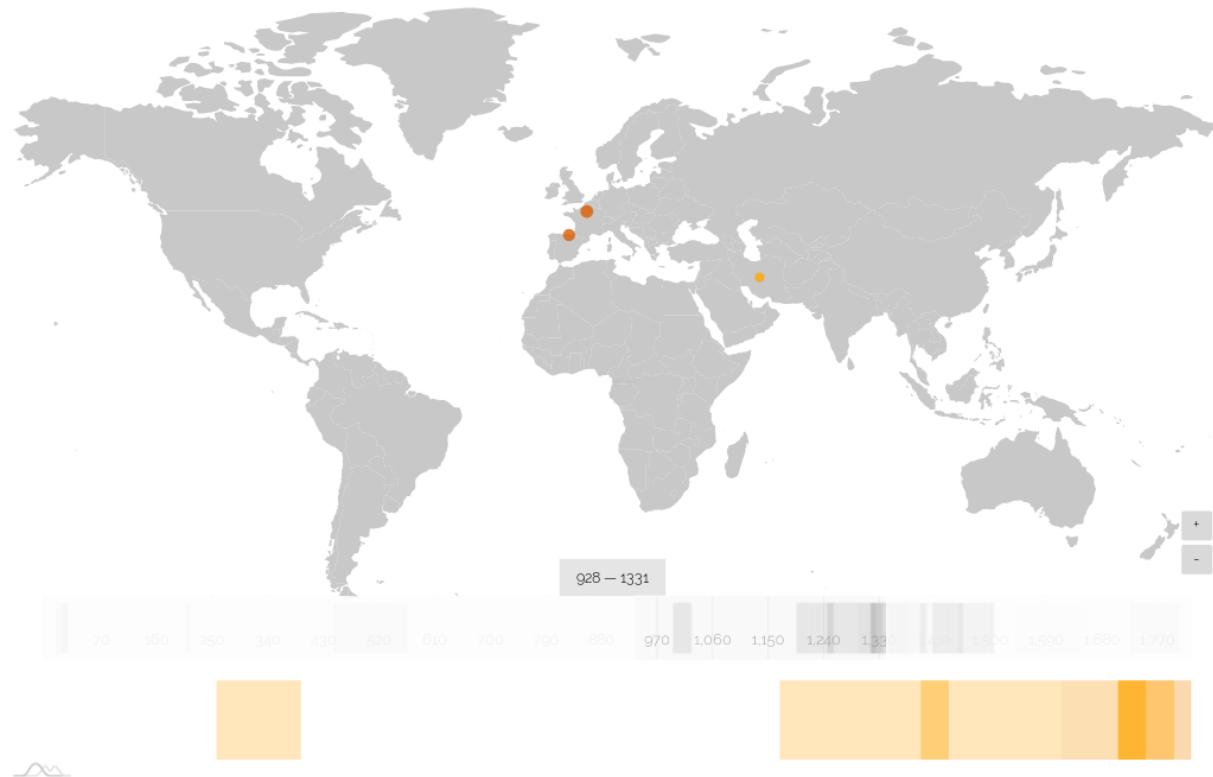


FIGURE 6.5 – Sélection d’une période sur la carte temporelle

de biaiser le moins possible les représentations, pour permettre aux chercheurs de tirer un maximum de profit de ce genre de réalisations numériques.

Ainsi, la mission de la plateforme publique vis-à-vis des chercheurs de DISHAS constiste en grande partie à révéler les interactions contenues dans la base de données, au travers de visualisations analytiques, comme de fonctionnalités permettant de réunir des ensembles cohérents dans le corpus numérisé, à l’instar des modalités de recherche et de l’ouverture des métadonnées au sein des notices. Si le *back office* propose des interfaces fonctionnelles pour la gestion des données, le *front office* doit offrir des interfaces propres à manifester leurs correspondances et interconnexions au sein de la base, dans le but d’accompagner au mieux le travail de recherche.

Chapitre 7

Exposition des données brutes : assurer la réutilisation du corpus numérique

Les deux précédents chapitres se sont attachés à décrire comment présenter et mettre en relation les données au sein du *front office*. Si ces formes de médiatisation sont nécessaires pour donner du sens au corpus numérique, elles le rendent également univoque. En effet, ces modes d'exposition constituent un accès indirect à la donnée ; l'élaboration d'une plateforme en ligne offre en cela une vision réductrice de la donnée. Les ressources numériques, une fois incorporées à un discours qui les décrit et intégrées à une interface graphique, sont rendues en grande partie inutilisables. Dans une perspective de conformité avec les principes FAIR, un effort de médiation doit être fourni pour rendre accessible la donnée brute, de manière à permettre la réexploitation du corpus numérisé.

7.1 Accessibilité des données brutes

La plateforme de DISHAS a l'ambition de servir de portail de partage de données dans l'objectif de favoriser la pratique de l'*open data* en histoire de l'astronomie. La mise à disposition du matériau numérique non transformé offre au public la possibilité de réutiliser les ressources scientifiques : cette exposition des données peut s'opérer à plusieurs niveaux.

7.1.1 Modalités d'accès à la donnée

Entrepôts de données pérennes

En France, plusieurs infrastructures s'attachent à accompagner les chercheurs dans l'exposition et la pérennisation des données de recherche. Le CINES (Centre Informatique

National de l’Enseignement Supérieur), établissement public placé sous la tutelle du Ministère de la Recherche et de l’Innovation, s’est fixé depuis sa création en 1999, une mission d’archivage des documents électroniques produits par la communauté scientifique. L’archivage pérenne concerne non seulement la gestion de l’obsolescence des formats et des environnements logiciels, mais également l’assurance de l’accessibilité des ressources¹. De toute évidence, les missions de conservation et de communication sont complémentaires et doivent être envisagées ensemble, l’une n’ayant pas de sens sans l’autre. Le CINES insiste en particulier sur l’importance de la description des contenus numériques à l’aide de métadonnées : en effet, sans métadonnée, la finalité d’un document peut être perdue, celui-ci devenant ainsi rapidement inexploitable.

La TGIR Huma-Num est le second acteur majeur pour la valorisation des données de recherche, elle travaille par ailleurs en collaboration avec le CINES. Afin d’assurer un accès persistant et interopérable aux ressources numériques, Huma-Num a mis en œuvre un système d’exposition des données intitulé Nakala. Nakala est une base de données triplestore² où peuvent être entreposés les corpus de recherche ; le SPARQL *endpoint* de Nakala permet de récupérer ces données numérisées, chacune étant signalée grâce à un identifiant pérenne, assurant leur trouvabilité dans le temps. Le recours aux standards du Web de données (lié à l’utilisation d’un triplestore), reposant sur des technologies et méthodes standardisées, permet de faciliter à la fois la récupération des données, mais aussi leur réutilisation. En effet, de nombreuses formes de valorisations s’appuient sur le format RDF, notamment le moissonnage des métadonnées par des services spécialisés comme Isidore ou Gallica, ou encore la réalisation de visualisation de données³.

Exposition de données courantes

Cependant, cette forme d’exposition des données brutes implique l’établissement d’un corpus défini, exploitable en l’état. La base de données de DISHAS est encore sujette à de nombreuses modifications et le travail de numérisation des sources est encore loin d’être achevé. Il est donc nécessaire de trouver des voies d’exposition de cette donnée « intermédiaire » ; avant la pérennisation définitive du corpus dans Nakala, l’interface publique de DISHAS doit servir de plateforme de diffusion pour les données en train d’être façonnées par les chercheurs. L’application Web du projet ne se substitue pas à Nakala, leurs missions sont complémentaires et les développements prévus pour DISHAS se doivent de ne pas entrer en concurrence avec les services proposés par Huma-Num.

1. « L’archivage numérique pérenne n’est pas non plus l’ultime étape du stockage des données avant l’oubli ou la perte définitive » *Le concept d’archivage numérique pérenne / CINES*, URL : <https://www.cines.fr/archivage/un-concept-des-problematiques/le-concept-darchivage-numerique-perenne/> (visité le 19/08/2019)

2. Un triplestore est une base de données graphe utilisée pour le stockage et la récupération des triplets RDF ; le langage de requêtage d’un triplestore est le langage SPARQL (*SPARQL Protocol and RDF Query Language*).

3. *NAKALA par Huma-Num...*

Dans le but de donner un accès direct aux données de recherche, différentes fonctionnalités ont été implantées : Malcolm Hamelin, stagiaire du projet d'avril à juillet 2019, a été chargé de mettre en place une API ainsi que le moteur de recherche ElasticSearch sur la base de données⁴. L'API permet aux utilisateurs et aux machines de communiquer directement avec la base de données de DISHAS ; par le biais d'une interface en ligne de commandes, il est possible de récupérer, de modifier ou d'ajouter des données dans la base⁵. L'API est à destination principale des chercheurs, leur permettant d'intégrer massivement des données en contournant l'interface graphique : en effet, l'utilisation de l'API requiert un *token* d'identification, défendant au grand public de rentrer n'importe quelle donnée dans la base.

ElasticSearch en revanche, fonctionne sur un duplicat de la base de données ; il n'est donc pas possible d'ajouter des données à la base relationnelle par le biais d'ElasticSearch. Son utilisation principale concerne la récupération des données de DISHAS et leur exposition à un public plus large que la seule équipe de recherche du projet. En effet, la base d'ElasticSearch peut être interrogée à la manière d'une API, de façon à ouvrir l'accès aux données brutes de DISHAS à des utilisateurs extérieurs comme à d'autres plateforme en ligne. Les données récupérées par le biais de l'API et d'ElasticSearch sont encodées en JSON, format largement employé, facilitant ainsi leur réutilisation.

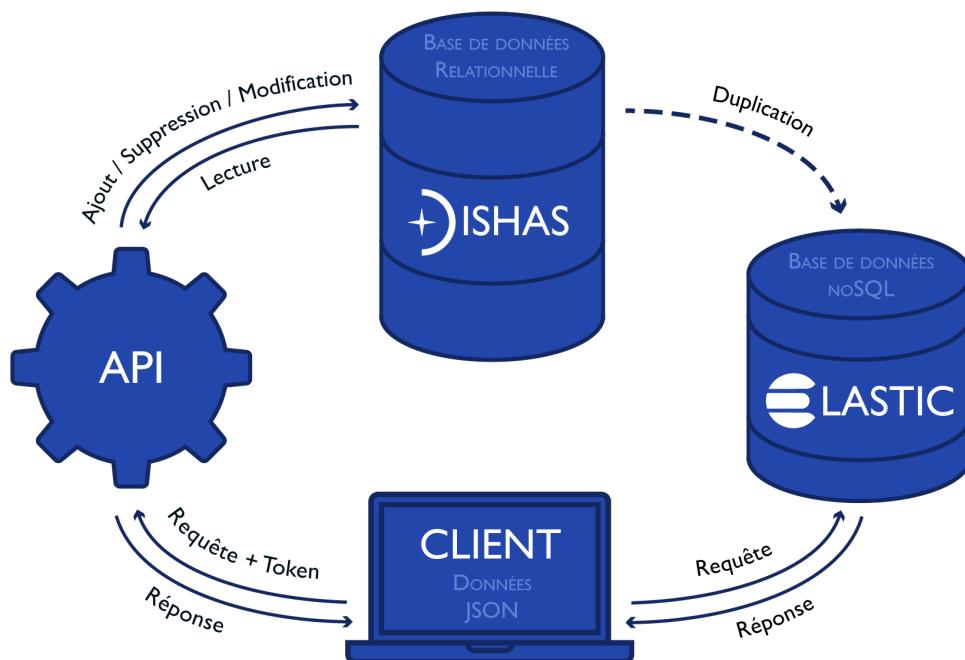


FIGURE 7.1 – Interaction entre la base de données de DISHAS, l'API et ElasticSearch

4. Pour davantage de détails concernant l'implémentation, voir section 9.2.

5. Quatre actions peuvent être effectuées par l'API. Ce sont les opérations CRUD (*Create Read Update Delete*), correspondant aux différentes façons d'interagir avec une base de données : ajout, modification, suppression et lecture de données.

7.1.2 Fonctionnalités implémentées dans l'interface publique

La plateforme publique de DISHAS n'intègre pour l'instant que peu d'accès à la donnée brute : seuls deux endroits du *front office* exposent une partie des données nécessaires à leur constitution.

La barre latérale présente sur les pages de notice, comporte plusieurs icônes à la disposition de l'utilisateur. L'un d'entre eux permet le téléchargement en JSON de l'ensemble des métadonnées affichées. Toutefois, il s'agit d'un objet qui a été formaté pour répondre aux besoins de mise en forme imposés par le *template* de la barre latérale, il serait donc difficile à un utilisateur d'exploiter véritablement ce matériau. Il est prévu de réviser cette fonctionnalité.

L'interface de recherche quant à elle, repose sur la base de données d'ElasticSearch ; lorsque l'utilisateur soumet une recherche plein-texte, une requête est expédiée au serveur d'ElasticSearch⁶. Les informations renvoyées par le serveur sont ensuite disposées dans un tableau de résultats pour une présentation graphique satisfaisante ; cependant, un bouton sous la barre de recherche propose à l'utilisateur de copier la requête envoyée à ElasticSearch dans son presse-papier. En collant cette requête dans la barre d'adresse d'un navigateur, la réponse en format JSON telle que renvoyée par le serveur est rendue disponible.

Ces fonctionnalités restent relativement rudimentaires et peu documentées. De futurs développements du *front office* devront s'attacher à rendre les données de recherche davantage accessibles ; néanmoins, ce travail se concentrera principalement sur la documentation des outils déjà existants. En effet, il existe d'ores et déjà des interfaces tout à fait fonctionnelles pour le requêtage de données, l'enjeu est avant tout d'expliquer au public comment utiliser ces outils. Ainsi, il est possible d'imaginer que des tutoriels pour l'utilisation d'ElasticSearch puissent être intégrés aux pages publiques, sous forme de ressource à télécharger ou de contenu à part entière. La question des exports des données devra également être réfléchie plus avant, notamment vis-à-vis du choix des informations à extraire et de leurs modélisation à l'intérieur d'une chaîne JSON.

7.2 Exposer les méthodes : documentation et maintien de code

La seule mise à disposition des données au public ne suffit pas à favoriser leur réutilisation. En effet, les corpus numériques, isolés de leur contexte de production, sont dénués de signification propre : pour promouvoir leur usage, il est important que les concepteurs de la donnée et des outils qui l'accompagnent, fournissent des documents

6. Pour plus de précision sur les méthodes de requêtes asynchrones à ElasticSearch, voir section 9.2.2.

explicitant leurs pratiques. Ce contenu est destiné aux utilisateurs extérieurs au projet, mais plus encore aux futurs et actuels développeurs de l'application de DISHAS. En outre, cette documentation doit s'appuyer sur du code de qualité, de manière à ne pas nécessiter de modifications majeures lors de la reprise du projet par une nouvelle équipe d'ingénieurs.

Les développements réalisés au cours de mon stage s'efforcent de respecter les bonnes pratiques énoncées ci-après, de multiples ajustements seront toutefois probablement nécessaires en vue de leur mise en production.

7.2.1 Lisibilité et documentation

Intelligibilité du code

Une des premières exigences des développements pour la constitution d'applications solides et réutilisables concerne la lisibilité du code. L'indentation, la suppression des redondances ou l'utilisation de sucre syntaxique⁷ sont autant de moyens de rendre la lecture du code plus aisée. Pour le développement des interfaces de DISHAS, plusieurs relectures du code ont permis d'en alléger la structure et de retirer des pans entiers de procédures inutiles. Des noms de variables signifiants ont été choisis de manière attentive, dans le but de participer à l'intelligibilité du code. L'emploi de la bibliothèque jQuery enfin, disposant d'une syntaxe plus compréhensible que le JavaScript simple⁸, a en outre aidé à la clarté des développements, grâce aux nombreux raccourcis et méthodes intégrés à la bibliothèque.

```
1  /* VANILLA JAVASCRIPT */
2  if (document.getElementById("sidebar").getAttribute("isOpen") === "false") {
3      document.getElementById("toolbar").classList.add("open-toolbar");
4 }
```

```
1  /* JQUERY */
2  if ($("#sidebar").attr("isOpen") === "false") {
3      $("#toolbar").addClass("open-toolbar");
4 }
```

FIGURE 7.2 – Différence de syntaxe entre Javascript simple et jQuery

Exhaustivité de la documentation

La documentation des développements est d'une importance capitale pour leur maintenance et leurs réemplois ultérieurs. Le code, par-delà la question de la lisibilité, requiert d'être explicité pour être rendu véritablement accessible, et par là même, réutilisable.

7. Méthode de clarification de la syntaxe programmatique à l'aide de raccourcis et tournures de code allégées.

8. Aussi appelé JavaScript *vanilla*.

8.7.2. EXPOSER LES MÉTHODES : DOCUMENTATION ET MAINTIEN DE CODE

La documentation doit s'insérer à différents niveaux du développement pour assurer des explications tant du point de vue de l'architecture globale de l'application, que des plus petits composants de la syntaxe programmatique.

```
1  /* EXEMPLE D'EXPLICITATION DU CODE */
2  if ($("#sidebar").attr("isOpen") === "false") {
3    // Si la valeur de l'attribut "isOpen" de l'élément HTML avec pour identifiant "sidebar" est
4    // égale à "false"
5    $("#toolbar").addClass("open-toolbar");
6    // alors ajouter la classe "open-toolbar" à l'élément HTML avec pour identifiant "toolbar"
7 }
```

FIGURE 7.3 – Exemple de commentaires de code intralinéaires

Au cours de la conception du *front office*, des commentaires intralinéaires ont été ajoutés pour expliciter les passages les plus ardu斯 du code. Chacune des fonctions élaborées a été en outre documentée : leur utilité, les arguments nécessaires à leur exécution et leur fonctionnement ont systématiquement été notifiés. Afin de garder des traces des différentes réalisations, chaque *commit*⁹ a été associé à un message détaillant précisément les modifications et ajouts effectués sur le code. Enfin, des notes plus extensives ont été rédigées à l'achèvement du travail, exposant avec précision la nature des développements et les instructions pour les reproduire au besoin¹⁰.

7.2.2 Gestion de *bugs*

Pour assurer une expérience positive liée à l'utilisation de la plateforme, les erreurs d'exécution du code doivent être évitées au maximum. Le chargement correct et rapide des pages, le bon fonctionnement des composants de l'interface ou encore le comportement approprié des différents modules sont des facteurs influençant non seulement l'expérience utilisateur, mais également l'exposition générale des données du projet. L'efficacité des interfaces est la conséquence directe de l'anticipation et la correction d'erreurs informatiques. Le maintien régulier du code est souvent un argument retenu par les développeurs pour attester de la stabilité d'un projet ; la gestion des *bugs* doit donc faire partie intégrante du processus de développement.

Prévenir les erreurs

Plusieurs méthodes peuvent être mises en place pour anticiper la survenue d'erreurs. En premier lieu, l'utilisation de bibliothèques et d'infrastructures logicielles¹¹ éprouvées

9. Sauvegarde d'un état d'avancement du travail.

10. Ces documents sont reproduits aux annexes F, G et H.

11. Aussi appelés *frameworks*. Un *framework* est un ensemble cohérent de composants logiciels servant de cadre à la conception d'une partie ou de la totalité d'une application.

simplifie le travail de correction tout en assurant un environnement stable pour l'implémentation de nouvelles fonctionnalités¹². Elles garantissent que les différentes briques logicielles constituant l'application bénéficient de mises à jour régulières, ou *a minima* d'un support de la communauté de programmeurs utilisant le même outil.

Les modèles – ou *template* – aident également à la maintenabilité du code. Les modèles permettent d'appliquer des traitements similaires à un ensemble de pages partageant des structures communes. Contentrant dans un seul fichier l'ensemble des propriétés de multiples interfaces, ils en assurent la cohérence mais facilitent également leurs modifications. De manière générale, la mutualisation du code allège le travail de développement et fournit un environnement de travail plus stable. La réunion de morceaux de code cohérents au sein de fonctions enfin, simplifie leur réutilisation et participe à construire une application plus robuste. Tous ces éléments façonnent une logique globale pour la plateforme, qui prévient un maximum de dysfonctionnements programmatiques.

Recettes fonctionnelles

La conduite de tests est également essentielle à la construction d'une application. Différentes méthodes permettent d'éprouver le fonctionnement du code : l'intégration continue, par exemple, consiste en la conduction de tests à chaque nouvelle version du code, de manière à repérer et corriger immédiatement les éventuelles régressions ou erreurs. Cette approche vise à réduire les problèmes d'intégration. Les tests unitaires quant à eux, permettent d'évaluer le bon fonctionnement d'une partie précise du code. Il est préconisé de les rédiger simultanément au code, de manière à les faire correspondre parfaitement et d'éviter ainsi la propagation d'erreur.

La mise en œuvre de telles procédures requiert de nombreuses heures de développement, et l'équipe technique chargée du projet DISHAS ne compte, à l'heure actuelle, que deux membres ; le déploiement systématique de tests n'a donc pas pu être encore intégré au processus de travail des ingénieurs. De manière à assurer la viabilité de l'application, l'équipe technique a en revanche recours aux recettes fonctionnelles. Les recettes fonctionnelles consistent à évaluer la conformité du produit fini avec le cahier des charges initial. Un ensemble de points est identifié ; une série de tests est définie pour chacun de ces points, afin de vérifier l'adéquation de leurs fonctionnements avec les spécifications originales. Par exemple, pour chaque contenu texte des interfaces, il sera testé si l'utilisation de caractères spéciaux ou d'alphabets étrangers ne corrompt pas l'affichage. Toutefois, ces vérifications ne sont pas encore automatisées et leur déploiement nécessite d'être optimisé dans le futur.

12. « *As it has already been invented, and is not considered to have any operational flaws, an attempt to reinvent it would be pointless and add no value to the object, and would be a waste of time, diverting the investigator's resources from possibly more worthy goals.* » « Reinventing the Wheel », dans *Wikipedia*, URL : <https://en.wikipedia.org/wiki/Reinventing-the-wheel> (visité le 10/08/2019)

8.2. EXPOSER LES MÉTHODES : DOCUMENTATION ET MAINTIEN DE CODE

L'exposition des données brutes et des méthodes permettant leur utilisation influe sur la diffusion des sources scientifiques du projet. La description technique du corpus numérique doit être mise en avant au sein de la plateforme publique, mais également à l'intérieur même du code sur lequel elle s'appuie, sous la forme de documentation rigoureuse. Si le public susceptible d'être intéressé par cet aspect de la donnée ne constitue qu'une part minoritaire des utilisateurs, l'importance de la communication sur les ressources numériques et sur leurs outils est capitale. En effet, cette forme de médiation assure aux données de recherche un cycle de vie dépassant le simple cadre de DISHAS, en préservant leur caractère polyvalent. La médiation des données brutes représente également une valorisation du travail des ingénieurs et des chercheurs ; l'exposition du processus d'élaboration de ces ressources participe à faire fructifier les réalisations numériques du projet.

Troisième partie

Traitement technique de la donnée et management des interactions

Chapitre 8

Gestion de projet

8.1 Dialogue entre recherche et ingénierie

Le développement du projet DISHAS s'inscrit à de nombreux égards dans la pratique des humanités numériques ; son aboutissement est le fruit que d'un partenariat étroit entre équipe de recherche et équipe technique. De nombreux enjeux peuvent être soulevés par une telle collaboration : afin d'assurer le déroulement optimal des réalisations du projet, il est intéressant de comprendre quelles problématiques peuvent émerger du dialogue entre ingénieur et chercheur.

8.1.1 Coopération dans le domaine des humanités numériques

La place des humanités numériques au sein de la communauté de recherche en sciences humaines a souvent été l'objet de débat. Le caractère protéiforme de leur pratique et la multiplicité de leurs applications rendent leur définition malaisée : cette difficulté à circonscrire les humanités numériques a parfois été l'origine d'incompréhension parmi les acteurs des SHS. Incompréhension notamment vis-à-vis du rôle des ingénieurs numériques au sein des projets de recherche : tantôt perçus comme des simples exécutants¹, tantôt considérés comme ayant la solution à tout problème, le positionnement des ingénieurs

1. « La contribution des techniciens est rarement analysée en tant que telle (du point de vue d'une sociologie du travail scientifique), mais presque toujours en fonction de leur contribution aux processus de production des connaissances scientifiques (du point de vue d'une sociologie de la connaissance) » Florence Millerand, « La science en réseau », *Revue d'anthropologie des connaissances*, Vol. 6, n° 1–1 (25 avr. 2012), p. 163–190, URL : <https://www.cairn.info/revue-anthropologie-des-connaissances-2012-1-page-163.htm> (visité le 04/08/2019), p. 163

peut ainsi être difficile à définir². Ces facteurs tendent à rendre complexe l'établissement d'un dialogue fécond entre chercheur et « technicien »³.

Plutôt que de renforcer la dichotomie entre recherche et numérique, la mise en commun des savoirs peut être le point de départ d'un enrichissement mutuel. D'une part, les chercheurs, en s'appropriant les techniques numériques, ouvrent l'horizon de leurs travaux scientifiques ; d'autre part, les ingénieurs élargissent le champ de leurs réalisations par l'ouverture aux sciences humaines. Une distinction étanche entre les domaines d'expertise nuit à l'émergence de profils mixtes, susceptibles de surmonter le cloisonnement disciplinaire et d'être vecteurs d'innovation.

L'ingénieur en humanités numériques, par son interdisciplinarité, est ainsi au centre du processus de recherche ; il doit à la fois comprendre les enjeux scientifiques, en même temps qu'évaluer les contraintes techniques. Les réalisations qu'il met en œuvre doivent, dans cette mesure, être le fruit d'une collaboration étroite avec le chercheur, qui doit appréhender de son côté les enjeux techniques sous-jacents⁴. Cette coopération permet de prévenir des choix techniques dommageables – lorsque les enjeux scientifiques ne sont pas bien explicités ou compris – comme d'éviter l'élaboration d'instruments qui finissent inutilisés – lorsque les chercheurs ne s'approprient pas les outils.

Si le chercheur ou l'enseignant-chercheur ont la responsabilité d'impulser et de porter des axes de recherche novateurs, en véritables initiateurs du dialogue entre discipline et technique, l'ingénieur est le dialogue en acte. En véritable traducteur, il transmet des techniques, des méthodes, mais aussi conçoit et expérimente de nouveaux outils et guide le chercheur dans ses découvertes. Il est à l'interface entre les humanités, l'informatique et le design, en dialogue constant entre disciplines et techniques. Il sert de pont entre l'idée du chercheur et l'innovation technique, la recherche et son application, son incarnation concrète.⁵

Le rôle de l'ingénieur numérique n'est donc pas tant celui d'un intermédiaire – mettant en œuvre des moyens pour répondre à une demande – que d'un médiateur, accom-

2. « *In these digital times, IT tools are no longer an auxiliary science to history. Integrating them into the “established procedures” has got to be the major challenge facing the discipline in the coming years.* » F. Heimburger et Émilien Ruiz, « Has the Historian's Craft Gone Digital ? Some Observations from France », *Diacronie. Studi di Storia Contemporanea* (, 10[2012]), URL : <http://journals.openedition.org/diacronie/2795> (visité le 04/08/2019), § 40

3. « Le paysage universitaire français a en effet pour particularité d'être structuré sur une séparation radicale entre recherche d'un côté, et services de l'autre. Cette dichotomie se retrouve au niveau des individus avec un cloisonnement étanche entre différents corps de la fonction publique, et au niveau des structures avec l'opposition bien connue entre unités de recherche et unités de service. » M. Dacos et P. Mounier, *État des lieux et positionnement de la recherche française dans le contexte international...*, p. 44

4. « En résumant ces différents éléments, on peut suggérer que l'enjeu pour les spécialistes en informatique est de parvenir à faire voir les limites qu'ils fixent aux possibilités d'action de l'équipe comme indépendantes de leurs propres capacités. » Pierre-Nicolas Oberhauser, « Des collaborations équivoques », *Revue d'anthropologie des connaissances*, Vol. 10, n° 4–4 (23 déc. 2016), p. 557-586, URL : <https://www.cairn.info/revue-anthropologie-des-connaissances-2016-4-page-557.htm> (visité le 04/08/2019), § 21

5. M.L. Massot, « Dessiner les acteurs en humanités numériques : la recherche en dialogue »..., p. 3

pagnant les pratiques scientifiques dans leurs mutations liées à l'apport des technologies numériques⁶. Dans cette perspective, l'ingénieur se doit de ne pas être un simple exécutant, mais aussi un acteur décisionnel, afin d'aider à définir quels développements seront les plus adaptés au travail du chercheur.

8.1.2 Processus de conception des interfaces

La conception des interfaces publiques de la plateforme de DISHAS est le fruit de nombreux allers-retours entre exigences scientifiques et contraintes techniques. À la suite d'un examen approfondi de la base de données et de nombreuses discussions à propos du corpus de recherche, une première ébauche des pages du *front office* a pu être dressée. Aucun cahier des charges préalable n'a orienté ou déterminé le choix des réalisations ; au contraire, c'est l'appropriation progressive des enjeux scientifiques sous-tendant au projet, qui a présidé à l'élaboration de la plateforme publique. L'évolution graduelle des développements s'est ensuite opérée au travers de deux phases complémentaires, répétées de manière itérative : une phase projective de description détaillée des interfaces à réaliser, et une phase démonstrative de confrontation des interfaces (concrétisées ou anticipées) avec les membres de l'équipe pour ajuster les développements.

Les procédés exposés ci-après sont employés au sein de l'équipe de DISHAS dans le processus de production numérique. Ces méthodes visent à favoriser la communication entre les chercheurs et les ingénieurs pour qu'en émanent des interactions fructueuses. Elles ont été implantées au fur et à mesure de l'avancée du projet, dans le but d'éviter de cloisonner le travail des membres de chaque équipe. Les paragraphes suivants constituent un retour d'expérience concernant l'élaboration de l'interface publique.

Séminaires numériques

Chaque mois est organisé au sein de l'Observatoire un séminaire d'équipe orienté DH (*Digital Humanities*) : ce séminaire est alternativement animé par les membres du pôle numérique de DISHAS et par des intervenants extérieurs. L'objet de ces séminaires varie, il s'agit autant d'exposer l'avancement de travaux de l'équipe d'ingénierie, que de présenter des instruments informatiques intéressants. À titre d'exemple, les quatre derniers séminaires ont porté sur : l'assistant bibliographique Zotero, les interfaces du *front office*⁷, les différentes applications de l'édition avec XML TEI et l'implémentation des outils d'ElasticSearch. Ces séminaires sont l'occasion de faire se rencontrer les chercheurs

6. « Loin d'être un simple développement technologique ayant un impact sur le processus de recherche et de visualisation des données en sciences humaines et sociales, les humanités numériques nous amènent à repenser le sens même de la recherche et, par conséquent, l'ensemble du modèle de production et de circulation du savoir à l'époque de l'édition numérique. » M. E. Sinatra et M. Vitali-Rosati, « Histoire des Humanités numériques »..., § 3

7. Le document de présentation utilisé lors du séminaire est reproduit en annexe D.

et les ingénieurs pour échanger à propos des réalisations numériques, mais aussi pour accompagner une première prise en main de ces outils⁸.

L'ingénieur doit tâcher de concentrer la discussion sur les points qu'il estime être les plus pertinents. Certains aspects trop techniques peuvent ainsi être éludés, tandis que les points concernant la représentation de l'information peuvent être davantage développés. Ainsi, les échanges lors de la présentation du *front office* se sont notamment orientés sur le raffinement des visualisations de données : les suggestions de chacun ont été évaluées et débattues, tant du point de vue de leur faisabilité que de leur intérêt, aboutissant à l'élaboration d'interfaces plus riches. Certaines propositions vis-à-vis de l'architecture de l'information dans le parcours de navigation ont par ailleurs été entièrement repensées, donnant lieu par exemple à la fusion des pages concernant les éditions et de celles relatives aux modèles mathématiques. Il a été évoqué enfin, la part active que doivent prendre les chercheurs dans la conception de la plateforme publique, en particulier dans la rédaction de textes introductifs pour les différentes pages de la navigation⁹.

Spécifications fonctionnelles

L'établissement d'un document détaillant l'ensemble des développements planifiés a constitué une seconde phase dans le travail d'élaboration du *front office*. La définition écrite des réalisations prévues vise plusieurs objectifs. En premier lieu, l'évaluation globale des moyens à mettre en œuvre pour leur concrétisation, mais aussi la constitution d'un document maître auquel se référer. Ce document¹⁰ est un cahier des charges fonctionnel où est décrite en détail chaque page de la plateforme publique. Il se présente sous forme de tableau où chaque rangée correspond à une page de l'interface, chaque colonne concernant quant à elle un aspect précis de cette page :

- une première colonne présente l'interface de manière générale : son objectif, son architecture globale avec la description de ses différents composants ainsi qu'éventuellement des interrogations vis-à-vis de sa conception ;
- une deuxième colonne expose l'ensemble des données qui seront récupérées de la base pour l'affichage, en particulier pour les métadonnées requises dans les pages de notices ;

8. « La conclusion à en tirer est que les chercheurs doivent d'abord connaître les logiciels développés pour eux, pas nécessairement qu'ils sachent programmer. Des logiciels puissants existent, il n'est pas toujours nécessaire de réinventer la roue. L'ignorance informatique des historiens doit être évitée, mais il est nécessaire de disposer d'une bonne culture informatique, autorisant un bon apprentissage des logiciels. » F. Clavert, F. Heimburger et A. Berra, « L'historien programmeur ? »..., § 21

9. « Nous abordons la question de la planification des tâches à accomplir et des objectifs à atteindre dans le cadre du projet, en montrant comment les efforts de coordination déployés par les chercheurs s'appuient de manière décisive sur les informations transmises par les spécialistes en informatique. » P.N. Oberhauser, « Des collaborations équivoques »..., § 5

10. Ce document, réalisé durant le premier mois de mon stage, a été reproduit en annexe E.

- une troisième colonne s'attache à expliciter précisément la visualisation de donnée apparaissant sur la page (si tel est le cas), tant dans son aspect visuel que dans ses enjeux intellectuels ;
- une quatrième colonne concerne les données nécessaires à la constitution de cette visualisation ;
- une cinquième colonne dresse la liste de tous les liens de redirection présents à l'intérieur de la page, composant ainsi le réseau dans lequel elle s'inscrit ;
- une dernière colonne détaille les corpus de données qui pourront être établis à partir de cette page. Cette partie concerne essentiellement les pages de notice, et fait référence aux « loupes » disposées sur la barre latérale de métadonnées¹¹.

Ce document, en plus d'aider au cours du développement de la plateforme, fournit un support pour les remarques potentielles des chercheurs ; il doit donc s'efforcer de viser l'exhaustivité des descriptions, énoncées dans un style le plus limpide possible. Il constitue ainsi un socle commun sur lequel peut s'appuyer le dialogue entre utilisateurs et concepteurs. En s'assurant en amont que le projet numérique est en adéquation avec les attentes de l'équipe de recherche, il est possible mettre de côté les réalisations inutiles et d'identifier celles auxquelles il faut accorder la priorité.

Le cahier des charges fonctionnel, ainsi que les différentes étapes d'avancement des interfaces ont pu être présentés et discutés lors de rencontres DH plus informelles entre le chef du projet de recherche, Matthieu Husson, et la cheffe de projet numérique, Galla Topalian. Ces réunions, organisées hebdomadairement, permettent d'ajuster les demandes des chercheurs par rapport aux contraintes techniques ; si la conception de l'interface d'édition de table revêtait par exemple d'une utilité immédiate plus importante, il a été décidé de se concentrer dans un premier temps sur les pages de « navigation historique » car la réalisation de visualisation de tables exigeait des compétences techniques plus approfondies¹².

8.2 Organisation du travail intra-équipe

8.2.1 Instruments de gestion du travail

Au cours des mois d'avril à août 2019, l'équipe technique du projet DISHAS a compté quatre membres chargés du développement de différentes parties de la plateforme en ligne. Galla Topalian, cheffe du pôle numérique, accompagne le projet depuis trois ans : elle a conçu la partie la plus importante de la plateforme, en particulier l'ensemble des interfaces du *back office* pour l'insertion de données. Antonin Penon s'est occupé de la réalisation

11. Leur fonctionnement est décrit plus en détail à la section 6.1.1.

12. En particulier par rapport à la complexité de l'encodage des tables astronomiques dans la base de données.

de DTI, outil dédié à la saisie de table, qui a été intégré aux pages administrateur. Malcolm Hamelin s'est quant à lui consacré à la création d'une API et à l'implémentation d'ElasticSearch sur la base de données de DISHAS. Enfin, le développement des interfaces du *front office* a constitué l'essentiel de mon travail.

Gestionnaire de versions

Afin de superviser l'avancement de ces différentes tâches, plusieurs logiciels sont utilisés au sein de l'équipe. En premier lieu, l'outil de *versionning* Git, qui est indispensable à tout travail de développement informatique à plusieurs. Les gestionnaires de versions comme Git permettent de sauvegarder différents états d'avancement d'un dossier, de manière à pouvoir rappeler la version antérieure d'un fichier en cas de problème. Git assiste également dans le travail en groupe grâce à la création de « branches ». Une branche est un état parallèle du dossier de travail ; chaque collaborateur peut ainsi effectuer des modifications sur sa propre version du dossier, sans entrer en concurrence avec le travail des autres. Git prend ensuite en charge la fusion, lorsqu'il faut intégrer les développements d'une branche à la version-maître de l'application (branche *master*). Une description de la branche dédiée aux interfaces du *front office* est disponible en annexe G.

Git fonctionne de manière complémentaire avec les services Web que sont GitHub et GitLab : ils permettent d'héberger en ligne les dossiers gérés à l'aide de Git, tout en proposant de multiples fonctionnalités additionnelles destinées à la collaboration, comme le suivi de *bugs*, la gestion des tâches ou encore la création de Wiki¹³. Le système des tickets permet par exemple de consigner les suggestions, comme de notifier la survenue d'erreurs ou encore de soulever des interrogations relatives à un point particulier du développement. Ces tickets peuvent ensuite être assignés à un membre du projet ; ainsi, lorsque des questionnements sont intervenus lors de l'élaboration des visualisations de données, un ticket a pu être créé afin d'en rapporter le contenu à Matthieu Husson à l'occasion d'une réunion DH.

Le Wiki du projet a aussi été largement utilisé, principalement dans un objectif de documentation. Y sont consignées à la fois les instructions pour l'installation de l'application DISHAS sur un serveur, mais aussi la description de l'arborescence des dossiers ou encore l'explication du fonctionnement des différentes réalisations numériques. Ces documents visent à faciliter la reprise des développements de DISHAS par une autre équipe d'ingénieurs, issue d'un projet de recherche partenaire.

Supervision du travail

D'autres outils plus spécifiquement dédiés au *management* de projet sont utilisés par les membres de l'équipe technique. L'instrument de gestion de projet en ligne Trello

13. Application Web collaborative dont le contenu est rédigé et modifié par les internautes autorisés.

est par exemple employé pour visualiser de manière synthétique l'avancement du travail. Chaque tâche y est identifiée par une « carte » et catégorisée selon son état d'avancement. L'ensemble des cartes des membres de l'équipe sont figurées concomitamment, permettant de suivre simultanément la progression des différentes réalisations. Cet outil est également utile pour lister au fur et à mesure les éléments nécessaires à la collaboration entre deux ingénieurs ; cela a particulièrement été utile lors de la prise en main des fonctionnalités d'ElasticSearch pour leur utilisation dans le *front office*. Trello a alors permis par exemple d'aider à formaliser des requêtes dans le langage spécifique d'ElasticSearch.

Ces instruments participent à optimiser le travail des ingénieurs numériques en établissant de bonnes pratiques dans le processus de réalisation¹⁴. La gestion efficace de chaque étape de développement est capitale pour mener à bien un projet numérique ayant vocation à être pérenne. L'utilisation d'outil de *management* de projet permet de garder un historique de la progression des développements, simplifiant les échanges entre les collaborateurs et contribuant à l'établissement de méthodes durables.

8.2.2 Méthodes de développement

Le développement des interfaces publiques s'est fait au travers de phases de travail individuel, comme de phases de programmation en binôme. La programmation en binôme – aussi appelé *peer coding* – et la revue de code à plusieurs a permis d'améliorer sensiblement la qualité du code. Cette pratique a donné l'occasion à la cheffe de projet numérique de suivre les développements effectués au fur et à mesure des mois de stage, donnant ainsi la possibilité de proposer des optimisations, ainsi que de suggérer des solutions déjà implémentées dans l'application. Par exemple, pour la génération de tableaux de résultats sur la page de recherche, le recours à une classe PHP déjà utilisée dans certaines pages de la plateforme a largement participé à accroître la robustesse des développements¹⁵.

La revue du travail en binôme donne également l'opportunité d'expliquer la démarche logique sous-jacente à un morceau de code, de manière à en révéler les incohérences et les maladresses. Le fait d'expliciter le processus programmatique à voix haute permet souvent d'y déceler des redondances et ainsi d'éliminer les passages inutiles. En outre, cette pratique comporte une forte dimension pédagogique : en identifiant des solutions pour pallier à certaines faiblesses, il est possible d'améliorer très efficacement la qualité de son code. Cette forme de coopération rend possible de conserver une cohérence dans les développements applicatifs tout en renforçant fiabilité de l'application.

14. « *Mailing lists, collaborative writing tools, blogs and wikis are only a small number of the tools which make collaborative work easier. They encourage methodological exchange and collective procedures.* » F. Heimburger et É. Ruiz, « Has the Historian's Craft Gone Digital?... », § 32

15. Pour le détail de l'utilisation de cette classe, voir en annexe G.3.1.

Chapitre 9

Environnement technique

9.1 Architecture de l'application Web

9.1.1 Modèle-Vue-Contrôleur

À l'instar de nombreuses applications Web, la plateforme en ligne de DISHAS est structurée selon le schéma MVC (*Model-View-Controller*). Ce motif d'architecture logicielle se divise en trois types de modules gérant différentes parties de l'application. Pour synthétiser les interactions entre ces trois composants de manière schématique ; le modèle s'occupe de traiter la donnée de la base, le contrôleur manipule cette donnée et la renvoie à la vue, vue qui la met ensuite en forme visuellement dans un *template*.

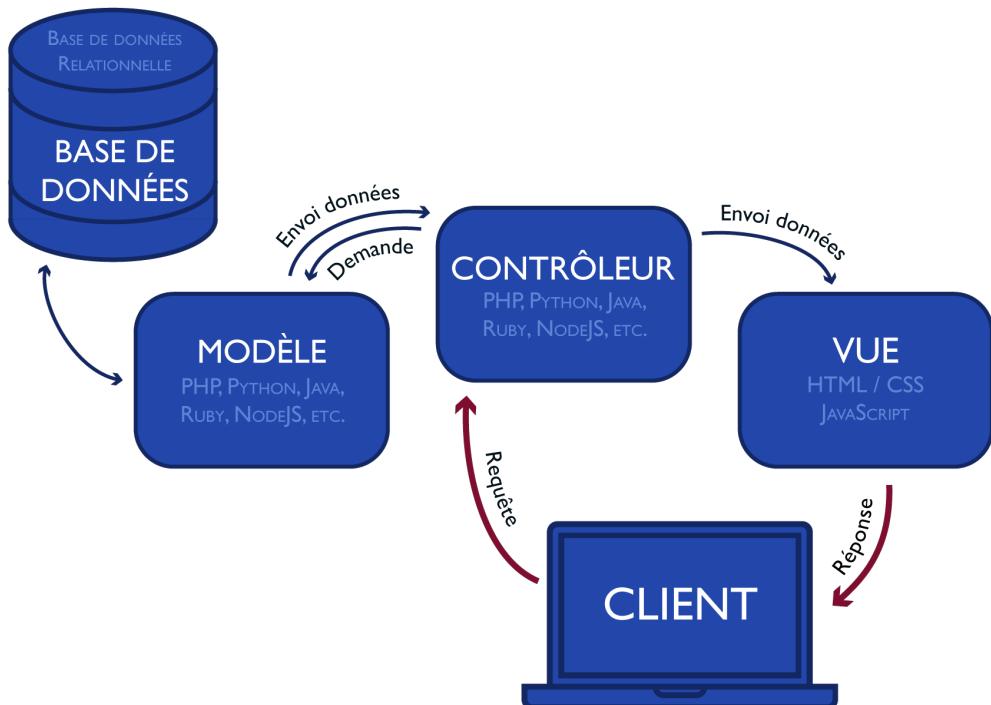


FIGURE 9.1 – Fonctionnement simplifié de l'architecture logicielle MVC

Présentation du *framework* Symfony

L’application Web de DISHAS a été élaborée grâce au *framework* MVC Symfony, dans sa version 3.4. Symfony est une infrastructure logicielle pour la création de sites internet en PHP. Le langage de programmation PHP est largement répandu dans le domaine du développement Web ; il est donc bien connu des directions de systèmes d’information, ce qui facilite la maintenance des projets utilisant ce langage. Il existe de multiples *frameworks* PHP et Symfony est l’un de ceux proposant le plus de fonctionnalités. Il est d’ailleurs le socle sur lequel se fondent d’autres *frameworks* à l’instar de Laravel, lui aussi communément utilisé. Symfony compose un environnement de travail cohérent, utilisé pour de larges applications à l’instar de Dailymotion ; sa robustesse est également un atout pour les projets plus réduits comme DISHAS, dans la mesure où il offre une grande fiabilité, en même temps qu’une importante adaptabilité. La documentation et la communauté soutenant cet outil en font un instrument robuste pour la conception d’une application Web ayant vocation à être pérenne.

Structure de l’application L’arborescence des fichiers contenant le code nécessaire à l’exécution de l’application se divise en différentes branches, chacune chargée du fonctionnement d’un module précis. Il ne sera évoqué ici qu’un petit nombre de ces principaux composants.

Dans le dossier *Entity* sont reproduites les tables de la base de données : pour chacune d’entre elles, une classe reprenant chacun des champs de la table est définie. Un enregistrement de la base sera ainsi transformé en objet PHP utilisable dans le code ; à cet objet seront associées des méthodes spécifiques, afin de manipuler les propriétés qui lui sont associées. Pour récupérer le titre d’une entrée de la table *Work*, on pourra par exemple employer la méthode `getTitle()` ; pour définir la valeur de la propriété *title* de cet objet, on utilisera en revanche la méthode `setTitle()`. Le dossier *Repository* quant à lui, définit pour chacune des entités des méthodes supplémentaires, davantage complexes, faisant jouer le rapport entre les données ; par exemple, l’entité *Work* y dispose d’une méthode `getPrimarySources()` qui a pour but de récupérer toutes les sources primaires issues d’une œuvre.

Le dossier *Controller* réunit quant à lui tous les contrôleurs, sortes de chefs d’orchestre de l’application. Les contrôleurs sont des fonctions qui agrègent les données nécessaires à la constitution d’une vue pour les transmettre au fichier de *template*. Si l’on a par exemple besoin du titre et des sources primaires reliés à une œuvre au sein d’une page de l’interface, on stockera la valeur des deux méthodes évoquées précédemment à l’intérieur de variables ; ces variables seront ensuite ajoutées à la valeur de retour de la fonction, de manière à les faire parvenir aux fichiers de vues.

Gestion des vues Le développement *front end* – portant sur ce que l'utilisateur voit à l'écran (utilisant notamment les langages HTML, CSS et JavaScript¹) – concerne la partie « vue » de l'architecture MVC. La conception des interfaces publiques de DISHAS s'est donc concentrée en grande partie sur l'élaboration de *templates* pour l'affichage des pages. Symfony dispose du moteur de *templates* Twig, qui s'intègre directement dans le code d'une page HTML. Twig sert notamment à transmettre les données définies dans la partie « modèle » de l'application au code qui sera envoyé au navigateur.

Pour afficher par exemple le titre d'une œuvre sur une page de notice, l'information contenue dans la base de données doit d'abord être transposée en objet PHP grâce au fichier d'entité définissant les caractéristiques d'une œuvre. L'information est ensuite récupérée par le contrôleur dédié aux pages de notices d'œuvre, puis transmise à la vue par le biais de Twig².

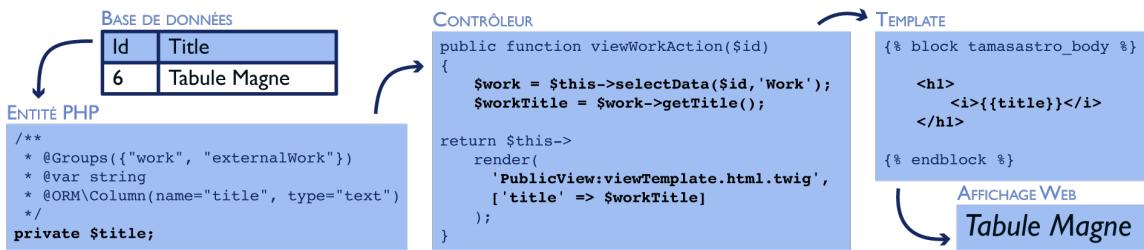


FIGURE 9.2 – Étapes pour l'affichage d'une donnée de la base au sein d'une page Web

Stockage et récupération des données

L'application de DISHAS s'appuie sur une base de données relationnelle : cela signifie que les informations utilisées par la plateforme sont stockées à l'intérieur de tables, connectées entre elles par des relations définies à l'aide de cardinalités³. Pour accéder aux données, l'application doit utiliser un intermédiaire logiciel que l'on nomme ORM (*Object-Relational Mapping*). L'ORM de Symfony est appelé Doctrine ; il permet de faire le lien entre les tables de la base de données et les classes du programme applicatif. Son rôle consiste donc à synchroniser les enregistrements de la base avec les objets PHP de l'application.

Les fonctionnalités de Doctrine permettent d'interroger et récupérer de manière simple les données de la base. En convertissant en objets les données issues de la base

1. De manière générale, le langage HTML définit la structure d'une page Web, le CSS son apparence graphique et JavaScript le comportement interactif des composants de la page.
2. Les passages Twig dans le template HTML sont repérables par leur balisage à l'aide d'accolades : `{{ }}`.
3. La cardinalité désigne le nombre minimum et maximum de possibilités de relation entre deux entités. Par exemple, un *original item* ne peut être contenu que dans une source primaire. Une source primaire en revanche peut contenir plusieurs *original items*. La cardinalité entre ces deux entités est donc de 1 à n.

relationnelle, l'ORM modifie la morphologie de l'information de manière à la rendre compatible avec la logique des langages orientés objet comme PHP. Néanmoins, Doctrine constitue en cela une couche logicielle supplémentaire qui peut nuire aux performances générales de l'application.

9.1.2 Le paradigme client/serveur

De nombreux systèmes informatiques reposent sur l'architecture client/serveur. Dans le cas des applications Web, elle désigne le mode de communication entre deux programmes : le client est un programme qui envoie des requêtes et le serveur est un programme qui répond à ces requêtes. Dans une plateforme en ligne, le serveur est la machine qui envoie les pages apparaissant sur le navigateur de la machine client.

Distinction *back end/front end*

Cette bipartition dans le fonctionnement de l'application fait écho à la distinction entre développement *back end* et développement *front end*. Le *back end* fait référence à la partie du code – rédigée en PHP dans le cas de DISHAS – qui n'est pas visible à l'utilisateur. Le *front end* en revanche représente la partie qui aura une influence directe sur ce qui est affiché sur l'écran du navigateur : les langages utilisés dans le *front* sont HTML, CSS et JavaScript.

Cette distinction se retrouve également dans la manière dont sont exécutés les différents langages utilisés dans la plateforme. L'intégralité du code écrit en PHP est traité par le serveur, alors que le HTML, CSS et JavaScript sont interprétés par le navigateur sur l'ordinateur du client. Dans le cas des *templates*, la partie Twig est exécutée par le serveur, tandis que le reste est à la charge de la machine client. Il existe ainsi différentes méthodes pour le traitement de la donnée : une même information pourra manipulée et formattée à l'intérieur du PHP, comme au sein du JavaScript.

Ces méthodes de traitement de la donnée ont chacune leurs implications. En premier lieu, la puissance de calcul du serveur est généralement plus importante que celle du client ; cela signifie que des procédures programmatiques lourdes sont effectuées plus rapidement si elles sont rédigées en PHP. Il est ainsi préférable de circonscrire au maximum l'utilisation de JavaScript pour traiter les ressources numériques. Par exemple, si une interface a besoin d'un jeu de données formatté de manière spécifique, il est conseillé de le mettre en forme dans le *back end*, plutôt que de transmettre des données brutes au *front end* pour ensuite les agencer de manière adéquate. Ainsi, pour remplir le *template* de barre latérale de métadonnées, une méthode propre à chaque entité a été définie dans les *Repositories* afin de générer des tableaux associatifs⁴ contenant les métadonnées d'un enregistrement.

4. Un tableau associatif en PHP est une manière d'organiser la donnée, où chaque valeur est associée à une clef. Ce format est comparable aux objets JavaScript ou aux dictionnaires Python.
Exemple : `$manuscrit = ["cote" => "Latin 14420", "folios" => 32, "langue" => "latin"]`.

Contraintes du développement *front*

Toutefois, il n'est pas toujours possible de traiter les ressources numériques au sein du *back end* de l'application. En effet, lors de l'utilisation d'ensembles de données importants au sein d'une même interface, il est parfois préférable d'avoir recours à des méthodes de programmation asynchrone⁵ au sein du *front end*, de façon à réduire les durées de chargement à l'ouverture de la page⁶. En effet, JavaScript permet de récupérer certaines informations du serveur, sans entraver l'exécution du reste du code, de manière à éviter de soumettre l'utilisateur à d'importants temps d'attente.

Ce procédé est nommé Ajax (*Asynchronous JavaScript and XML*) : le principe consiste à envoyer une requête au serveur, indépendamment du chargement de la page. Le code asynchrone est exécuté comme tâche de fond, sans occasionner de temps de latence lié à l'actualisation entière de la page. Cette technique est employée notamment pour charger de façon incrémentale un fil d'actualité, ou encore pour suggérer des termes d'autocomplétion dans un champ de saisie⁶ : cela permet de fluidifier la navigation, de manière à rendre transparent le processus de requêtage de données. La méthode Ajax est notamment employée dans le *front office* pour récupérer les données nécessaires à la composition de la carte chronologique : cette technique permet d'afficher certains éléments de la page avant que l'intégralité de la visualisation ne soit constituée.

Ces formes d'optimisation du code sont nécessaires à considérer dans le processus de développement des interfaces, en particulier pour les visualisations de données qui sont très coûteuses en termes de puissance de calcul. Le code *front end* étant interprété par l'ordinateur du client, la rapidité d'exécution du programme dépend de la puissance de sa machine, ainsi que des capacités de compilation de son navigateur. S'il n'existe plus de sensibles différences entre les navigateurs Web, il subsiste quelques variations concernant certaines fonctionnalités implémentées par les langages *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 utilisateurs des interfaces optimales et ergonomiques.

5. L'exécution du code est généralement linéaire, chaque ligne étant traitée l'une après l'autre. Dans le cas de l'asynchrone, la portion de code concernée est placée dans une file d'attente et n'est exécutée qu'une fois le reste du code effectué.

6. La suggestion de termes se fondant sur une liste de mots accessible depuis une base distante.

9.2 Implémentation d'ElasticSearch

9.2.1 Présentation d'ElasticSearch

Fonctionnement d'ElasticSearch

ElasticSearch est un logiciel libre⁷ servant à l'indexation et la recherche de données ; il s'appuie sur une base de données NOSQL (*Not Only SQL*) ainsi que sur le moteur de recherche Lucene⁸. Les bases de données NOSQL se définissent en opposition aux bases relationnelles, elles recouvrent des formes aussi diverses que les bases graphes⁹ ou les bases orientées documents¹⁰. La base ElasticSearch s'organise autour d'un système complexe de *clusters* et de nœuds très flexible, pouvant s'adapter à toute sorte de donnée. Ce type d'architecture a l'avantage d'être résilient, performant et de bien supporter la montée en charge. Les données insérées et récupérées de la base ElasticSearch sont au format JSON, elles ne sont en revanche pas stockées ainsi.

La base de données d'ElasticSearch est le support de son moteur de recherche. Les fonctionnalités proposées par ElasticSearch permettent à la fois d'y extraire un grand nombre de résultats à l'aide de requêtes imprécises, comme d'effectuer des recherches extrêmement détaillées, grâce à des possibilités de filtres extensives, tout cela avec une grande rapidité. Le moteur d'ElasticSearch est également fortement orienté texte, les résultats de recherche étant classés selon leur pertinence vis-à-vis des mots-clefs renseignés par l'utilisateur. La vitesse de requêtage ainsi que les fonctionnalités de recherche avancées sont deux atouts qui ont amené l'équipe technique de DISHAS à implémenter cette solution en complément de l'API. L'API a vocation à servir principalement à interagir avec la base de données relationnelle sans passer par l'interface graphique, tandis qu'ElasticSearch doit être utile à l'extraction d'importants ensembles de données et à la construction de requêtes complexes. ElasticSearch constitue en cela un élément essentiel de l'exposition des données de la base de DISHAS.

L'implantation d'un serveur ElasticSearch est relativement aisée, la principale difficulté de l'implémentation de cet outil réside dans la constitution du *mapping* définissant l'architecture conceptuelle de la base. Cette étape permet de transposer les données de la base relationnelle à l'intérieur de la base NOSQL d'ElasticSearch ; elle a été facilitée par l'utilisation de l'extension Symfony FOSElasticaBundle. Certaines tables de la base n'ont pas été reproduites dans le *mapping* d'ElasticSearch, seules les données intéressantes pour les chercheurs ont été intégrées.

7. ElasticSearch est un logiciel *open source* sous licence libre Apache.

8. Lucene est un moteur de recherche modulable spécialisé dans la recherche textuelle

9. Pour le stockage de triplets RDF notamment.

10. À l'instar des bases de données constituées de documents au format XML ou JSON.

Visualiser et explorer avec Kibana

L’implantation d’ElasticSearch sur une base de données donne en outre accès aux différents outils de la suite Elastic. Kibana est l’un de ces outils : il permet de créer par le biais d’une interface graphique des visualisations sur mesure, à partir des données indexées dans ElasticSearch. Les fonctionnalités de Kibana peuvent être étendues par l’ajout de *plug-ins*, permettant l’utilisation de divers procédés d’analyse sur les données ainsi que de multiples types de visualisations. Certains des modules de Kibana sont par ailleurs payants, à l’instar de l’extension Graph développée par Elastic, qui permet de représenter les relations entre les enregistrements de la base sous forme de graphe.

Kibana a été originellement créé pour composer des tableaux de bord, aidant les administrateurs d’une application à visualiser les *logs*¹¹. S’il est possible de partager les visualisations réalisées avec Kibana et de les intégrer dans un fichier HTML, elles ne sont pas conçues pour être diffusées sur une plateforme publique. Pour le projet DISHAS, il a ainsi été décidé que l’accès à Kibana serait réservé aux chercheurs, de manière à leur permettre de créer des représentations graphiques sur mesure des données qu’ils souhaitent voir interagir.

9.2.2 Utilisation dans le *front office*

Les fonctionnalités de recherche proposées par ElasticSearch ont été largement utilisées au cours de l’élaboration des interfaces du *front office*. La mise au jour des interactions entre les données étant au cœur de l’architecture de la plateforme publique, les modalités de recherche d’ElasticSearch ont permis de tisser des liens entre les ressources de manière plus efficace et précise que les méthodes de requêtage offertes par l’ORM Doctrine. ElasticSearch est ainsi utilisé à la fois dans la partie *back end* de l’application, au sein du PHP, mais également dans la partie *front end* par le biais de méthodes de requêtage asynchrone.

Requêtage en *back end*

La barre latérale de métadonnées présente sur les pages d’enregistrements dispose d’un système de « loupes » qui rattache chaque notice aux enregistrements connexes de la base. Si une œuvre a par exemple été créée par Jean de Lignières, l’icône de loupe permet à l’utilisateur d’avoir accès à toutes les œuvres créées par le même auteur. Ainsi, pour chaque champ de la barre latérale est associée une requête à la base de données correspondant à l’ensemble des entités partageant la même caractéristique.

Pour chaque entité de la base qui dispose d’une page de notice, a été définie une méthode spécifique permettant de générer les données à afficher dans cette barre latérale¹².

11. Les *logs* sont des fichiers d’historiques des événements.

12. Cette méthode est décrite plus en détail à l’annexe G.2.

Cette méthode définit non seulement les métadonnées d'un enregistrement mais également les requêtes qui seront envoyées à ElasticSearch. Les requêtes faites à ElasticSearch sont formulées en JSON, toutefois, cette technologie évolue rapidement et la syntaxe des requêtes est modifiée régulièrement. Afin d'aider à l'élaboration de ces requêtes et d'assurer une compatibilité accrue au cours du temps, il a été décidé d'utiliser le client PHP Elastica qui permet faire le lien avec le serveur ElasticSearch dans le *back end*. Elastica permet de formuler à l'aide d'une classe PHP des requêtes qui seront ensuite traduites en JSON.

Les requêtes à élaborer pour la constitution des barres de métadonnées conservent souvent des structures similaires et il aurait été redondant de reproduire systématiquement des procédures semblables pour chaque requête. Il a donc été décidé de créer une extension de la classe `Query` d'Elastica afin de systématiser la rédaction de requêtes.

Par exemple, voici la formulation d'une requête au format JSON pour récupérer dans la base ElasticSearch tous les *original item* élaborés entre 1450 et 1500, suivi de son équivalent, formulé à l'aide du client Elastica :

```

1 GET original_text/_search
2 {
3     "query": {
4         "bool": {
5             "must": [
6                 {
7                     "range": {
8                         "tpq_date": {"gte": "1450-01-01"}
9                     }
10                },
11                {
12                    "range": {
13                        "taq_date": {"lte": "1500-01-01"}
14                    }
15                }
16            ]
17        }
18    }
}

```

FIGURE 9.3 – Formulation dans le langage de requête d'ElasticSearch

```

1 // Nouvelle instance de la classe Query
2 $query = new Elastica\Query();
3
4 // Création des filtres
5 $greaterThan = new \Elastica\Query\Range("tpq_date", ["gte" => "1450-01-01"]);
6 $lessThan = new \Elastica\Query\Range("taq_date", ["lte" => "1500-01-01"]);
7
8 $bool = new \Elastica\Query\BoolQuery();
9 $bool->addMust([$greaterThan, $lessThan]);
10
11 // Attribution de ces filtres à la requête
12 $query->setQuery($bool);

```

FIGURE 9.4 – Formulation dans le langage de requête du client PHP Elastica

Avec l'extension de la classe `Query`, il est possible de créer la même requête en une ligne (`newTimeQuery(1450, 1500)`) de manière à alléger le code et faciliter la maintenance. Ces requêtes sont ensuite insérées dans le tableau associatif servant à générer la barre latérale des pages de notices. Lors du clic sur une icône de loupe, la requête est envoyée à la page de recherche par le biais du contrôleur.

Chargement asynchrone des données

Le moteur de recherche ElasticSearch est employé pour deux utilisations distinctes dans la partie *front end* de l'application : pour la visualisation en carte ainsi que pour la page de recherche.

Toutes requêtes qui transitent par la page de recherche sont envoyées en Ajax à ElasticSearch. L'utilisation de méthodes asynchrones pour la récupération de résultats est particulièrement adaptée, dans la mesure où elles permettent d'effectuer plusieurs requêtes les unes à la suite des autres, sans nécessiter de réactualisation de la page.

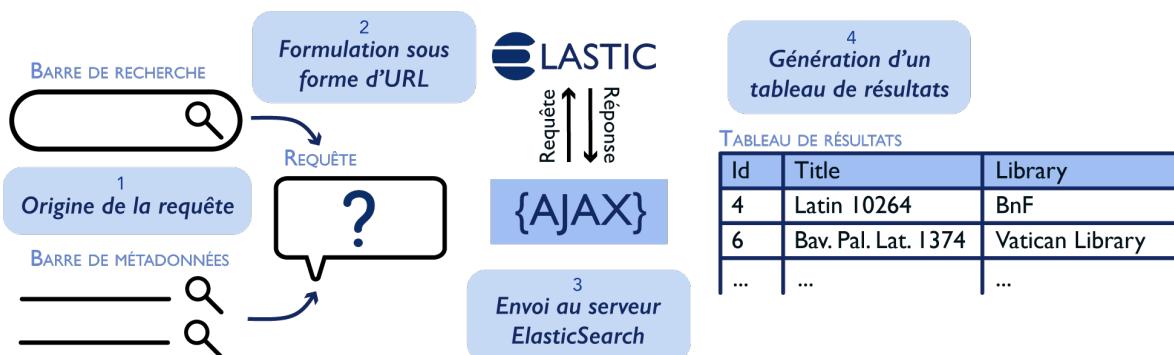


FIGURE 9.5 – Processus de requête asynchrone à ElasticSearch pour la page de recherche

Le processus d'exécution d'une recherche suit toujours les mêmes étapes. Deux types de requêtes peuvent advenir ; soit la requête est envoyée depuis une page de notice, soit elle est formulée à partir des termes renseignés dans la barre de recherche. La requête est ensuite envoyée à ElasticSearch ; le résultat reçu est alors formaté de manière à apparaître au sein d'un tableau de résultats. Cette méthode permet de conserver un temps d'attente minimal pour la réception de résultat, tout en profitant des fonctionnalités avancées de recherche du moteur ElasticSearch. Cependant, seule la recherche plein-texte a été pour l'instant mise en place dans l'interface de recherche, il est prévu toutefois d'y ajouter la possibilité d'utiliser des filtres.

L'utilisation d'ElasticSearch pour la visualisation de carte chronologique répond à d'autres motivations. Cette visualisation en effet rassemble de nombreuses données, puisqu'elle vise à représenter tous les enregistrements d'œuvres et de sources primaires de la base. Le chargement de l'ensemble de ces données par l'intermédiaire de l'ORM Doctrine

est supérieur à l'utilisation d'ElasticSearch. Il est toutefois difficile d'estimer exactement le gain de temps occasionné, dans la mesure où l'exécution du code asynchrone est postérieure au chargement de la page. De plus, la base de données de DISHAS reste pour l'instant assez réduite et les temps de requêtage restent courts. Néanmoins, pour la même quantité de résultats, le recours à ElasticSearch divise quasiment par deux le temps de chargement de la page (de 46,44 ms à 24,33 ms¹³).

Ainsi, l'utilisation d'ElasticSearch est complémentaire au requêtage de la base de données par un ORM ; si elle induit parfois la nécessité de traiter la donnée dans la partie *front end* de l'application, elle offre des avantages de performance et participe à alléger les interfaces pour des visualisations plus fluides. ElasticSearch est également un outil indispensable pour explorer les données dans le détail, afin de permettre aux utilisateurs de réunir des données qu'il aurait été difficile de connecter en utilisant des langages de requêtes plus élémentaires¹⁴.

13. Ces résultats ont été obtenus grâce au moniteur de performance de Symfony.

14. ElasticSearch permet notamment d'effectuer des recherches de distance par rapport à point particulier, par exemple pour retrancher tous les enregistrements créés dans un rayon de 100km. Reproduire le même résultat à l'aide d'un ORM ou du langage SQL est bien plus complexe.

Chapitre 10

De la conception à l'implémentation des visualisations

10.1 Technologies de visualisation de données

10.1.1 Choix du moyen de réalisation

Les visualisations de données sont une partie centrale de la plateforme publique : elles constituent un moyen d'expression flexible et modulable pour exposer les données de recherche du projet DISHAS. De nombreuses technologies numériques permettent de réaliser des visualisations, cependant elles ne sont pas toutes adaptées aux mêmes types d'application, et n'exigent pas le même niveaux de compétence en programmation.

Logiciels de visualisations

Il existe une grande variété de logiciels permettant de composer des visualisations de données directement au sein d'une interface graphique. Ils n'exigent aucune connaissance préalable, si ce n'est une compréhension minimale des données à traiter. Les logiciels comme Excel, Tableau ou les plateformes comme Palladio et Kibana permettent à des utilisateurs novices de concevoir des visualisations riches et diversifiées. Néanmoins, les possibilités de configuration sont réduites aux fonctionnalités admises dans le logiciel, restreignant ainsi la variété des réalisations potentielles. En outre, ce genre d'outil n'est pas toujours adapté à la publication sur le Web : s'il est souvent possible d'insérer ce type de visualisations à l'intérieur du code HTML, il peut subsister des difficultés d'affichage. Dans le cas de Kibana notamment, l'intégration de visualisations à l'intérieur d'une page publique peut entraîner des failles de sécurité ; en ce qui concerne Tableau, toutes les réalisations sont conservées sur le serveur de l'application, occasionnant parfois des délais de chargement supplémentaires.

Les logiciels de visualisations sont particulièrement indiqués pour aider des chercheurs à concevoir des graphiques sur mesure : pour des utilisateurs connaissant bien leur corpus de données, ces outils offrent des ressources intéressantes pour formuler graphiquement des hypothèses de recherche. En revanche, ces instruments ne sont souvent pas conçus pour être employés sur un site internet. Leur utilisation est à privilégier pour la création de visualisations ponctuelles, indépendantes d'une interface Web.

Visualisation de données avec R

R est un langage de programmation statistique avec lequel il est possible de générer des visualisations de données. Il est extrêmement modulable et permet aux développeurs de créer une multitude de graphiques, grâce aux nombreuses extensions qui y peuvent être ajoutées. C'est un outil particulièrement employé dans le monde académique, notamment car il permet de manipuler les données à l'aide de fonctions statistiques et mathématiques avancées. Il est souvent utilisé en parallèle du langage de programmation Python, qui offre également des possibilités de visualisations et d'analyses statistiques extensives, grâce à des bibliothèques comme MatPlotLib ou encore Seaborn. Si R est plus généralement utilisé pour analyser des jeux de données en CSV, il existe des *packages* pour parser du JSON, de l'XML ou encore des fichiers de tableurs.

R n'est pas réputé pour être un langage difficile d'apprehension au premier abord, toutefois il semble que certains types de diagrammes nécessitent une connaissance accrue du langage pour être réalisés. En outre, R est plutôt dédié à la conception de graphiques statiques ; les bibliothèques additionnelles qui proposent d'ajouter de l'interactivité utilisent généralement JavaScript pour les éléments dynamiques¹.

Bibliothèques JavaScript

En ce sens, il semble plus indiqué de n'avoir recours qu'à JavaScript pour créer les visualisations de données dans le *front office*. Il existe un grand nombre de bibliothèques JavaScript spécialement dédiées à l'intégration de visualisations de données aux pages Web. Ces bibliothèques disposent toutes de spécificités ; certaines sont extrêmement versatiles et polyvalentes – à l'instar de D3.js ou Chart.js² – tandis que d'autres sont spécialisées dans un unique type de graphique – comme Leaflet pour les cartes ou Chartist pour les diagrammes simples. La richesse d'une bibliothèque JavaScript est souvent corrélée à la difficulté de son apprentissage ; en outre, plus les visualisations à réaliser sont complexes, plus une connaissance approfondie de la bibliothèque est nécessaire.

Il existe des bibliothèques JavaScript adaptées à tous les usages ; si cette richesse est profitable, elle rend le choix d'une technologie particulièrement ardu. Plusieurs questions

1. Notamment pour bibliothèques Plotly, DT et Shiny qui intègrent toutes des fonctionnalités d'interactivité à l'aide de JavaScript.

2. D3.js et Charts.js sont à la tête du marché des visualisations sur le Web d'après ??.

sont donc à considérer : quels types de visualisations sont prévus ? Quelle est la taille du jeu de données à traiter ? Quel environnement de développement est utilisé au sein du projet³ ? L'application doit-elle être disponible en version mobile ? Les fonctionnalités de la bibliothèque sont-elles bien supportées par les différents navigateurs ? Quel degré de personnalisation de l'interface est nécessaire ? Chacune des réponses à ces interrogations conditionne le choix d'une bibliothèque.

10.1.2 Choix d'une bibliothèque JavaScript

Critères de choix

Ainsi, le choix d'une bibliothèque JavaScript doit prendre en compte de nombreux paramètres. En premier lieu, le régime de tarification est important à considérer : en effet, le projet DISHAS ne bénéficie pas de financements suffisants pour payer la licence d'une bibliothèque JavaScript ; or, si la plupart proposent une partie de leurs fonctionnalités gratuitement, il existe souvent certaines conditions d'utilisation payantes. Il faut donc veiller à sélectionner une bibliothèque qui accepte les usages prévus pour le *front office* de la plateforme.

Un second critère essentiel de choix concerne la documentation et la communauté liées à une technologie. En effet, une bibliothèque extrêmement efficace et flexible ne peut être utilisée au maximum de son potentiel, s'il n'existe pas d'exemple ou de documentation auxquels se référer pour l'utiliser. La communauté de développeurs peut également être de grand secours ; une question relative à une bibliothèque largement utilisée sera résolue très rapidement sur des sites comme StackOverflow. Un bon indicateur de la vivacité d'une communauté peut être trouvé sur les *repositories* GitHub des différentes technologies : le nombres d'étoiles⁴ et la date du dernier *commit* peuvent être ainsi révélateurs du dynamisme d'un projet.

Un autre élément à considérer est évidemment les fonctionnalités offertes par une bibliothèque. La recherche de technologies de visualisations de données s'est fait ultérieurement à l'élaboration des graphiques à implémenter dans le *front office* : ainsi, il était important que la bibliothèque choisie comporte une grande variété de visualisations, notamment des visualisations en carte. La flexibilité de la bibliothèque est particulièrement importante dans la mesure où DISHAS est un projet aux effectifs réduits et que l'utilisation de grand nombre de dépendances augmente la difficulté à maintenir l'application. En outre, la partie *back office* de la plateforme disposait déjà de visualisations ; le choix d'une technologie s'est donc fait de manière à compléter les bibliothèques préalablement implémentées dans l'application.

3. L'emploi de *frameworks* comme React ou Angular peuvent conditionner le choix vers des bibliothèques compatibles.

4. Les étoiles indiquent sur GitHub le nombre d'utilisateurs ayant mis en favori un projet en particulier.

La bibliothèque de visualisations pour DISHAS devait donc être à la fois robuste, versatile et fortement personnalisable. Une première sélection de bibliothèques a donc été effectuée en fonction des critères énoncés ci-dessus pour être mise à l'épreuve. Cinq bibliothèques⁵ ont été testées dans la réalisation du graphique envisagé pour la notice d'œuvre : il s'agit d'un graphique en colonnes relativement simple, idéal pour une première évaluation des possibilités d'une bibliothèque. Cette expérimentation a permis de déceler certaines différences de modélisations des données, ainsi que les avantages et inconvénients de ces technologies. Par exemple, D3.js est une bibliothèque remarquablement développée, ce qui rend sa première appréhension particulièrement difficile. Par ailleurs, Chart.js s'appuie sur l'élément HTML `canvas`, là où les autres bibliothèques emploient l'élément `svg` : cette différence a une influence sur la rapidité de chargement d'une visualisation. En effet, les bibliothèques utilisant le format SVG (*Scalable Vector Graphics*) offrent davantage de flexibilité, car chaque élément du graphique correspond à un nœud unique du DOM (*Document Object Model*)⁶ qu'il est possible de configurer. En revanche cela signifie également que plus nombreuses seront les données à afficher, plus le temps de chargement sera important ; l'élément `canvas` *a contrario* est traité comme un assemblage de pixel par le navigateur, et donc est comparable à une simple image en termes de chargement.

Présentation de AmCharts

La bibliothèque AmCharts a finalement été retenue pour le développement des interfaces publiques, en premier lieu pour la variété des visualisations qu'elle propose : AmCharts offre plus de quinze types de graphiques différents, comprenant notamment des cartes, alors que ces dernières font généralement l'objet de bibliothèques spécialisées. Amcharts est une bibliothèque utilisée par de grandes entreprises, comme Amazon, Apple ou la NASA ; cela garantit la fiabilité de cette technologie, ce qui est important pour un projet comme DISHAS. Son utilisation est gratuite à condition de conserver le logo d'AmCharts dans le coin inférieur des visualisations de données.



FIGURE 10.1 – Logo de la bibliothèque JavaScript AmCharts

5. Il s'agit des bibliothèques Chart.js, D3.js, Google graphs, ApexCharts et AmCharts.

6. Le DOM correspond à l'architecture en arbre d'un document HTML.

Amcharts dispose en outre d'une impressionnante documentation et de très nombreux exemples de visualisations, ce à quoi s'ajoute un blog où sont régulièrement postées des astuces pour la confection de graphiques complexes et modulables. Un avantage de cette bibliothèque réside dans la possibilité de configurer tous les éléments d'une visualisation, de manière à s'adapter au plus près des envies des concepteurs. Elle propose par ailleurs de nombreuses fonctionnalités pour concevoir des visualisations interactives et dynamiques, que l'utilisateur peut manipuler. Pour une première approche de la bibliothèque Amcharts, un tutoriel succinct a été mis à disposition à l'annexe H ; des fichiers HTML simplifiés, reproduisant les visualisations déployées dans le *front office*, ont été également ajoutés aux appendices documentaires du mémoire.

10.2 Constitution d'une visualisation

10.2.1 Dialogue entre conception et réalisation

La constitution d'une visualisation consiste souvent en un travail itératif, où phases d'élaboration et de développement se répètent jusqu'à obtenir à une forme graphique satisfaisante. La visualisation permet de révéler des informations sur les données qui n'étaient pas prises en compte au départ, donnant lieu parfois à des incohérences ou des incompatibilités graphiques ; il est donc souvent nécessaire d'effectuer de multiples aller-retours entre réflexion et réalisation, pour parvenir à la visualisation la plus pertinente.

Le diagramme en barre prévu initialement pour la page de notice des sources primaires a particulièrement évolué entre sa forme liminaire et son aspect final. Cette visualisation a initialement été conçue en miroir de la notice d'œuvres : l'enjeu était d'y représenter chaque *original item* contenu dans une source primaire, classé en fonction de l'œuvre dont il est tiré. À la suite de discussions avec les chercheurs du projet au cours d'un séminaire DH, il a été décidé d'intégrer une composante de foliation au graphique, pour proposer deux modes de visualisation : un présentant la quantité d'*original items* par œuvre et un autre montrant leur disposition au sein de la source primaire. Une seconde version de la visualisation a donc été réalisée ; il est alors apparu que plusieurs *original items* pouvaient être accolés les uns aux autres sur une même page. Sur la visualisation, les éléments représentant les *original items* étaient ainsi superposés les uns sur les autres sans qu'il soit possible de les distinguer. Une seconde étape de programmation a donc été mise en œuvre, pour faire en sorte de distinguer les items situés sur les mêmes folios.

C'est alors sur la visualisation représentant la quantité accumulée de folios – montrant les *original item* juxtaposés les uns aux autres – que se sont portées les réflexions : n'est-il pas trompeur de figurer comme étant un nombre total de folios, ce qui est en réalité l'accumulation des folios de chaque *original item* pris indépendamment ? En effet, cette configuration peut donner l'impression qu'une œuvre compose la majeure partie

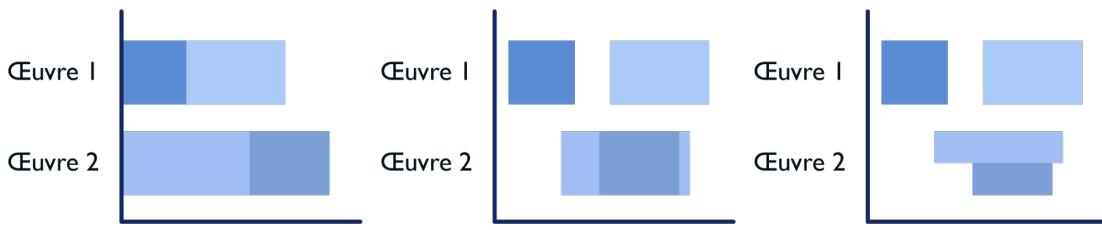


FIGURE 10.2 – Évolution de la visualisation pour la page de notice de sources primaires

d'une source primaire, alors qu'elle comptabilise moins de pages qu'une autre. Plusieurs solutions ont été envisagées pour pallier à cet effet : ne représenter par exemple que le pourcentage d'*original item* par œuvre, ou encore signifier par différentes épaisseurs la proportion prise par l'item au sein du folio. Il a finalement été décidé de conserver une visualisation avec des *original items* accolés les uns aux autres⁷, car elle permet d'évaluer la quantité de tables issues de chaque œuvre, ce qui est aussi significatif qu'un nombre de folios. En outre, un diagramme circulaire a été ajouté au sein de la barre latérale afin de représenter la part véritable de chaque œuvre au sein d'une source. Il reste cependant des interrogations vis-à-vis de cette visualisation, notamment la question du mélange entre folio et page (dans le cas de manuscrits recomposés) qui a pour l'instant été éludée⁸.

10.2.2 Anticiper les obstacles à la visualisation

Il existe donc de nombreux paramètres qu'il n'est pas toujours possible de connaître préalablement, mais qu'il faut essayer de prendre en compte, pour assurer l'adéquation de la représentation graphique avec la donnée à figurer. La base de données sur laquelle s'est appuyée le travail d'élaboration des interfaces est encore profondément lacunaire, et comporte de plus de nombreuses données de test, n'ayant aucune véracité scientifique ; en outre, si la base ne sera ensuite alimentée qu'avec des données contrôlées, il subsistera inévitablement des variations dans la description des enregistrements. Toutes ces composantes sont à anticiper pour proposer des visualisations flexibles, susceptibles de conserver leur pertinence, malgré une incertitude vis-à-vis de la donnée.

Données lacunaires

Le premier élément auquel une interface modulable doit faire face est l'inconsistance des données. L'absence d'une donnée peut être envisagée de deux manières différentes dans une visualisation : soit ne faire figurer l'information manquante, soit la représenter, mais en la distinguant des autres. Dans les pages de notice, il est fréquent qu'un enregistrement ne dispose pas de toutes les métadonnées qui puissent lui être associées ; il est

7. En plus d'une visualisation des *original items* répartis tels que dans la source primaire.

8. Il semble qu'une telle configuration sera rarement rencontrée ; pour l'instant, les deux systèmes de pagination sont traités de manière quasi équivalente.

important de discerner parmi les informations qui font défaut, celles qui sont signifiantes, de celles accessoires. Par exemple, au sein de la barre de métadonnées, certains champs sont éludés s'ils ne sont pas renseignés, alors que d'autres sont conservés pour souligner l'absence de l'information⁹. Par ailleurs, cette absence peut être le signe que l'information n'a simplement pas été renseignée par l'administrateur, ou qu'il n'est pas possible de la donner dans la mesure où le document originel ne permet pas de statuer. Ainsi, l'absence de donnée est un point délicat qu'il n'est pas possible de résumer aisément au sein d'une visualisation ; chaque métadonnée doit être traitée différemment.

Les visualisations de données de la plateforme publique servent, dans la plupart des cas, de portails vers les autres pages de l'interface publique : la carte chronologique par exemple est le seul moyen d'accès – mis à part l'interface de recherche – aux notices d'œuvres et de sources primaires. Chacune de ces entités doit donc figurer sur la carte, quand bien même elle ne serait pas associée à un lieu en particulier. Ainsi, toutes les données qui ne sont pas situées géographiquement ont été arbitrairement placées aux coordonnées 0,0.

De la même manière, il n'est pas exigé dans le modèle de données de préciser les pages d'un *original item*, ce qui met en cause la visualisation de la notice de sources primaires. Pour remédier à ce problème, il a été décidé d'obliger les utilisateurs à remplir ce champ au moment de la soumission du formulaire pour l'insertion d'un nouvel item. Des valeurs par défaut peuvent être renseignées si l'information est véritablement manquante. Il s'agit d'inciter les chercheurs à décrire le plus exhaustivement les données qu'ils ajoutent à la base ; en revanche, il ne faut pas que les visualisations prévalent sur l'exigence scientifique. C'est pourquoi chacune de ces décisions doit être attentivement évaluée en compagnie de l'équipe de recherche, de manière à encourager l'utilisation de la plateforme en proposant des visualisations éclairantes, mais en contrignant le moins possible la pratique des chercheurs.

Données trop abondantes

Si les données lacunaires peuvent poser problème, l'abondance des données peut également susciter des difficultés. La superposition des éléments graphiques et la concentration des informations peuvent en effet rendre illisible une visualisation. Certaines fonctionnalités et formes visuelles cependant permettent d'éviter ces écueils, en atténuant la complexité des données ; les cartes proportionnelles (*treetemap*) et les diagrammes circulaires notamment, figurent de manière facilement appréhendable de grandes diversités d'information en minimisant la part des données en faible proportion. Les modalités de zoom donnent également l'occasion à l'utilisateur de cadrer la visualisation à l'échelle qu'il trouve la plus cohérente.

9. Ce point est discuté en détail en annexes G.2.1.

Cependant, la principale difficulté de l'abondance de données à visualiser est avant tout la lenteur de chargement qui peut en être occasionnée. L'utilisation d'une bibliothèque JavaScript pour réaliser les cartes et les diagrammes du *front office* induit que le code sera exécuté sur l'ordinateur du client : la puissance de la machine et l'efficacité du navigateur étant inconnus, il est nécessaire d'alléger au maximum la partie programmatique du *front end*. Les méthodes de chargement asynchrone ou incrémental (*lazy loading*) permettent notamment de répartir les passages les plus coûteux en puissance de calcul, en échelonnant le chargement des composants à afficher. Il faut en effet distinguer la vitesse de chargement d'une page Web, de sa vitesse d'affichage : la première est une mesure technique, tandis que la seconde est liée à la perception de l'utilisateur. Une page peut être affichée rapidement sans que son chargement soit terminé, offrant ainsi une expérience utilisateur jugée satisfaisante¹⁰. D'autres procédés ont vocation à être intégrés au *front office* pour fluidifier encore davantage la vitesse d'exécution des scripts, notamment le regroupement des données en *clusters*, moins coûteux à traiter par le navigateur. Toutefois, les ajustements programmatiques qui ont été implémentés pour l'instant restent limités, du fait de mes compétences techniques encore modestes.

Pour évaluer les performances des pages publiques, le *bundle* de Symfony appelé Fixtures a été utilisé. Cette extension permet d'ajouter de grandes quantités de données à la base, afin de tester la réaction des interfaces. La carte chronologique a été le principal objet de cette évaluation, dans la mesure où il s'agit de la visualisation agrémentant le plus de données : cela a permis d'identifier certains points à simplifier¹¹, mais également de ne constater aucun problème majeur. Les navigateurs en outre, disposent d'outils intégrés permettant d'estimer les passages du code exigeant le plus de performance, rendant ainsi possible de caractériser les passages du script à améliorer.

L'efficacité et la vitesse d'affichage des pages publiques de DISHAS ne constitue pas uniquement des ajustements superficiels ; ces éléments sont d'une importance capitale si l'on veut inciter les chercheurs à utiliser les outils de la plateforme. L'augmentation de la performance des navigateurs a engendré un nouveau paradigme pour le développement d'applications ; ce qui pouvait être téléchargé et installé est aujourd'hui plus volontiers en ligne, accessible sur n'importe quel poste d'ordinateur. Les applications en ligne deviennent de plus en plus sophistiquées, habituant ainsi les utilisateurs à des interfaces complexes, chargeant de manière quasi instantanée. Si le *front office* de DISHAS n'offre que des outils laborieux à manipuler, les utilisateurs seront d'autant plus prompts à se désintéresser de la plateforme. La vitesse d'un site Web participe donc à l'exposition des données et doit être prise en considération, au même titre que l'ergonomie générale des

10. La vitesse d'affichage est par ailleurs prise en compte dans le système de classement des moteurs de recherche.

11. Le niveau de description de la carte de chaleur en particulier, de manière à passer d'une granularité à l'année à des tranches temporelles d'une décennie.

interfaces.

Que ce soit pour gérer l'abondance ou l'insuffisance des données, la flexibilité des visualisations aide à manier des ressources dont la forme est imprévisible. Le modèle de données induit un nombre fini de possibilités concernant le format des enregistrements de la base ; le rôle du concepteur des interfaces est de veiller à trouver un mode d'affichage pour chacune de ces éventualités, de manière à prévenir les incompatibilités techniques ainsi que les aberrations graphiques. Dans la mesure où la structure de la base de données est encore amenée à évoluer, il sera néanmoins nécessaire de revoir l'agencement des pages du *front office*¹², même si la structure de la plateforme publique a été pensée pour admettre des évolutions.

10.2.3 Limites de la visualisation

S'il est un moyen d'appréhender aisément des informations complexes, le format graphique qu'offrent les visualisations de données peut également induire en erreur. Parfois la donnée en elle-même est ambiguë et il n'existe pas de forme visuelle adaptée ; parfois la représentation graphique véhicule des concepts qui ne sont pas en adéquation avec le contenu.

Incertitude relative à la donnée

Le premier élément pouvant aboutir à une visualisation trompeuse est relatif à la forme des données sur lesquelles elle s'appuie : qu'il s'agisse de l'aspect lacunaire de certaines informations ou de l'accumulation des données, il existe de nombreux prédictats qui peuvent biaiser une visualisation dans son ensemble. Par ailleurs, plus une visualisation résume une importante quantité de données, plus elle est susceptible de transmettre une impression d'exhaustivité ; cependant, la base de données ne reflètera jamais que les documents sur lesquels des chercheurs ont porté leur attention. Le contexte de présentation d'une visualisation peut aider à dissiper cette équivoque, en faisant en sorte d'exposer les données en tant qu'enregistrement d'une base, et non comme représentation objective de la réalité historique.

Si les visualisations de notices ont une dimension plus exacte du fait de leur ponctualité, il subsiste des incertitudes quant à la représentation de certaines données. Dans le cas par exemple d'une source primaire qui mêlerait deux types de pagination – folio et page – essayer de représenter les deux formes de contenu sur la même échelle¹³ serait fautif : en effet, ce genre de configuration est généralement signe d'un ouvrage hétérogène

12. Notamment pour l'aspect *responsive* des interfaces, dimension qui n'a pas été encore abordée au sein de la plateforme DISHAS.

13. En convertissant les folios en pages par exemple.

composé de parties indépendantes. Faire figurer côté à côté ces différentes parties est donc souvent erroné. Il n'est pourtant pas possible de présumer de la disposition des feuillets d'un ouvrage, en conséquence, une visualisation de données générique ne pourra jamais être tout à fait adéquate. L'idéal serait de pouvoir s'appuyer sur la pagination d'une reproduction numérique, il n'existe cependant pas de numérisation de tous les documents à intégrer à la base, rendant le travail de conception des interfaces d'autant plus délicat.

Représentations biaisées

La seconde difficulté dans l'élaboration d'une visualisation pertinente réside dans l'aspect graphique à choisir pour représenter la donnée. L'échelle, la couleur, la forme ou encore l'interaction des différents éléments, véhiculent des informations à l'utilisateur. Chacun de ces composants doit être choisi avec soin, car il influence l'appréciation des données présentées. Par exemple, la simple fluctuation de l'échelle peut induire des conclusions tout à fait contrastées sur une même donnée ; une graduation resserrée amplifie les phénomènes de variation, l'inverse les atténue.

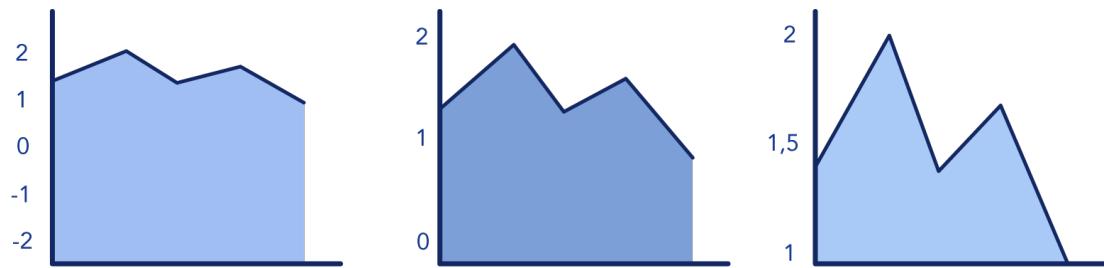


FIGURE 10.3 – Conséquences visuelles du changement d'échelle d'une visualisation

Au sein du *front office*, plusieurs visualisations ont soulevé des questionnements relatifs à la représentation idéale des informations. Par exemple, comment figurer le chevauchement des *original items* au sein d'une même page : par transparence ? par un système de rayures colorées ? Il a finalement été décidé de superposer horizontalement les blocs d'*original item* de manière à pouvoir cliquer indépendamment dessus. Néanmoins, il existe de multiples manières d'exprimer visuellement cette information ; le concepteur des interfaces doit s'efforcer de trouver celle qui est la plus adaptée, tout en restant réalisable.

La visualisation de la page de navigation historique a également été problématique : si l'utilisation d'une carte a semblé évident, le choix d'un mode de représentation de la temporalité a été l'objet d'interrogations. La frise chronologique devait faire figurer la quantité d'œuvres et de sources primaires créées par période ; dans un premier temps, un diagramme linéaire a été choisi. Cependant, les dates de création encodées dans la base indiquent en réalité une période de conception probable : une œuvre conçue entre 1400 et 1450 correspond sur un diagramme linéaire, à un plateau d'un demi siècle. Or, l'utilisateur peut comprendre de cette mise en forme qu'une œuvre a été créée toutes les années entre

1400 et 1450, ce qui est contradictoire avec la donnée de départ. C'est pourquoi, il a été jugé préférable de suggérer ces variations grâce à une carte de chaleur, où la saturation retranscrit plus justement la fluctuation dans le rythme de production.

Chacun des aspects d'une visualisation, dans la mesure où il peut être vecteur de renseignement sur la donnée, doit être choisi avec soin. Depuis le choix d'une technologie pour réaliser des diagrammes, jusqu'à la formulation des libellés, l'élaboration de visualisation mobilise de multiples modalités de médiation, à chaque étape de son processus. De plus, les données de recherche induisent un degré d'exigence scientifique supplémentaire à la conception de graphiques pour les représenter : l'élaboration des pages de la plateforme publique de DISHAS a ainsi tenté de valoriser aux mieux les ressources numériques par différents d'expression, autant visuel que conceptuel.

Conclusion

Ce mémoire s'est attaché à décrire le travail réalisé au cours du stage que j'ai effectué au sein de l'équipe de recherche de DISHAS, à l'Observatoire de Paris. La mission qui m'avait été confiée consistait en l'élaboration de l'interface publique de l'application Web du projet : il était important pour l'équipe de DISHAS de disposer d'une plateforme ouverte, qui participe à valoriser les données produites dans le cadre du projet, tout en restant accessible à des publics variés. Les réflexions qui ont accompagné mon travail se sont concentrées sur les différentes modalités de médiation numérique par lesquelles il est possible de présenter un corpus de recherche. Mes réalisations se sont attachées à fournir des solutions à chaque étape de conception de la plateforme, tant au niveau du parcours de navigation, que de l'interactivité des interfaces ou encore de l'agencement des visualisations de données.

Le processus de création de la plateforme publique du projet DISHAS s'est déroulé en plusieurs phases successives. Une première phase d'analyse du contexte a permis d'appréhender le cadre dans lequel s'inscrivait ce travail. Une seconde phase de réflexion a été consacrée aux différents modes de médiation qu'il est possible d'appliquer à un corpus de données scientifiques. Enfin, une phase de réalisation technique a été l'occasion de développer certaines des interfaces de la plateforme.

La première partie de ce mémoire porte donc sur le contexte de création des données à valoriser, de manière à mieux appréhender les besoins qui peuvent émaner de la nature de ces ressources. La variété des approches scientifiques des équipes associées au projet permet notamment de comprendre la diversité des modes d'utilisation du matériel documentaire : il est ainsi possible d'en déduire des exigences vis-à-vis des instruments informatiques à développer. La nature du corpus et les procédés de sa numérisation constituent également des axes de réflexion privilégiés, car la base de données compose véritablement le socle sur lequel peut se construire la plateforme en ligne.

En second lieu se pose la question de la mise en forme de ces données. Dans cette optique, trois modes de médiation des ressources numériques ont été établis comme objectifs de l'interface publique : la présentation du corpus, l'accompagnement du travail de recherche et l'exposition des données brutes. La présentation des ressources numériques

vise à les rendre explicites tout en introduisant les utilisateurs aux enjeux scientifiques qui sous-tendent le projet. L'accompagnement du travail de recherche consiste quant à lui à fournir des instruments pour assister les chercheurs dans leurs investigations du corpus numérique. L'exposition des données brutes enfin, a pour ambition d'aider ceux qui voudraient exploiter ces ressources : elle vise à rendre accessible non seulement les données, mais également la documentation qui leur est associée.

Enfin, la description technique des différents composants de l'interface publique permet d'exposer les moyens mis en œuvre pour réaliser cette médiation de données. Ces instruments comprennent à la fois des infrastructures logicielles fiables, permettant d'assurer le fonctionnement global de la plateforme, mais aussi des modes de stockage et requêtage alternatifs de la donnée, ou encore le recours à des bibliothèques *front end*, pour la création d'interfaces sur mesure. Par ailleurs, le travail technique réside également dans l'établissement de bonnes pratiques, à la fois pour la gestion des développements programmatiques, mais également pour assurer le dialogue entre recherche et ingénierie.

Au sein de la conception de la plateforme publique, le travail de médiation consiste donc à dresser une passerelle entre la base de données et l'utilisateur, en mobilisant à la fois une appréhension fine du corpus de recherche, une compréhension des attentes des différents publics ainsi que des connaissances des technologies numériques. L'interface publique doit proposer un discours sur la donnée qui réponde à des questions de recherche, en mettant en œuvre différents principes de médiation à l'aide de procédés informatiques. La plateforme de DISHAS vise à exposer la donnée de manière conforme aux principes FAIR, de manière à participer à un écosystème ayant pour but la valorisation des ressources numériques.

L'élaboration des interfaces du *front office* n'est cependant pas achevée et il subsiste encore de nombreuses formes d'expression la donnée à explorer. Le projet DISHAS n'en est par ailleurs qu'à son commencement et il n'est pas encore possible d'évaluer avec précision l'adéquation des développements avec la pratique des utilisateurs. Ainsi, de nombreuses perspectives se dessinent quant à l'évolution des modes de valorisation du corpus de recherche ; l'accroissement de la base de données va être notamment la première étape vers la constitution de contenus inédits, en particulier car elle va être l'occasion de voir émerger de nouvelles hypothèses de recherche.

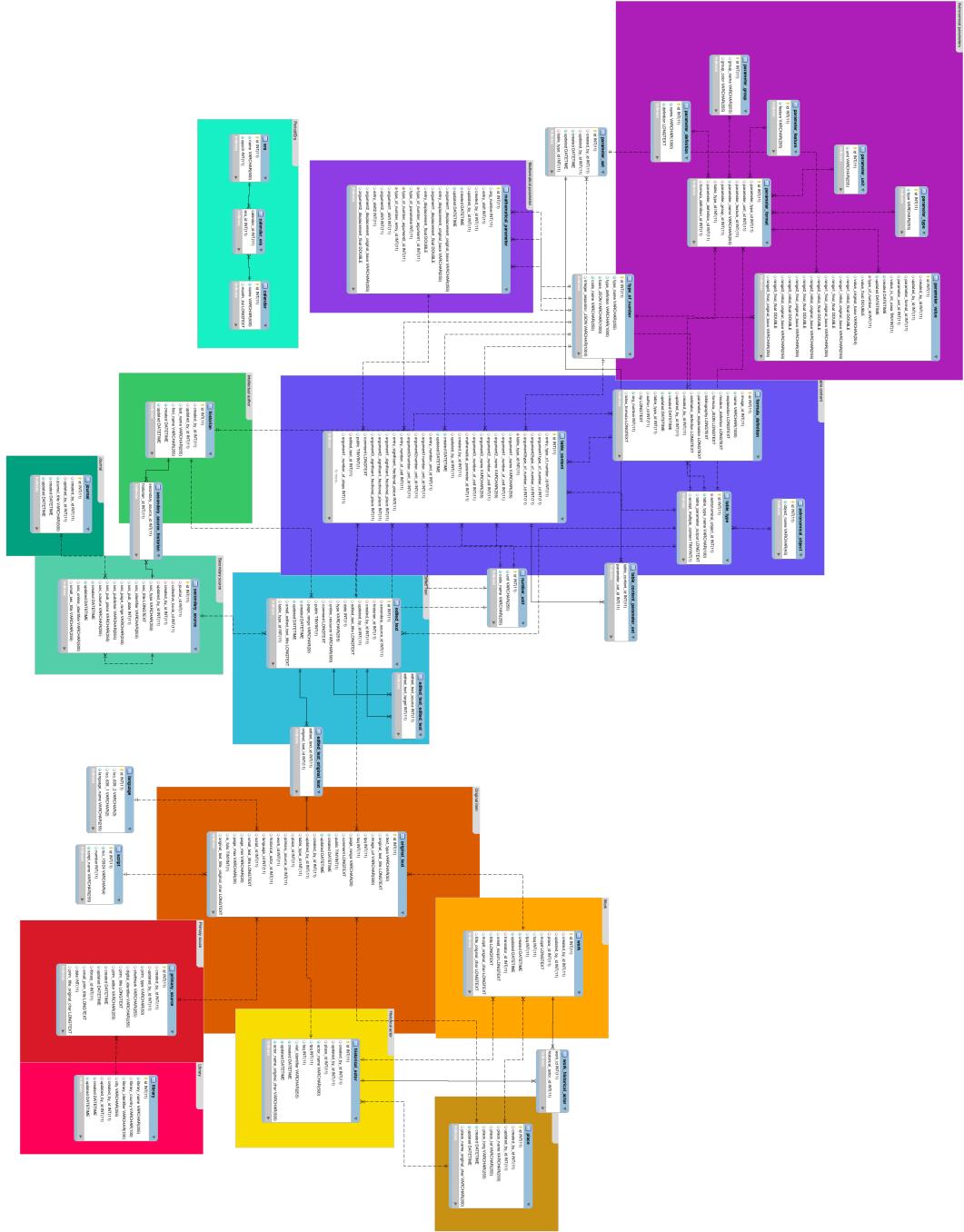
Annexes

Annexe A

Base de données du projet

Ce schéma, réalisé avec *MySQL Workbench*, exclut les tables ne concernant pas la description du corpus de recherche, c'est-à-dire les tables structurelles comme celles liées à la gestion des utilisateurs.

FIGURE A.1 – Schéma de la base de données réalisé avec *MySQL Workbench*



Annexe B

Glossaire

B.1 Astronomical object

An astronomical object is a celestial entity associated with some natural phenomena. In DISHAS, astronomical objects are used as a way to classify astronomical tables in broader categories.

B.2 Astronomical parameter set

A parameter set is a set of astronomical quantities that describes tables at the level of astronomical theories. A parameter set can be shared by several tables across different traditions. Broadly speaking, there are two kinds of parameters :

- explicit parameters that are directly read off the table, e.g., the maximum value(s) ;
- implicit parameters that need to be retrieved from the table content computationally, e.g, solar eccentricity.

In some cases, these explicit or implicit parameters are only significant in groups ; whereas in other instances, parameters can be independently significant. Parameters are a central tool in delineating and connecting astronomical traditions and analysing the mathematical and astronomical content of the tables. A parameter-set is only linked to an intellectual author through an edited text because a single parameter-set can be shared by many different sources.

B.3 Edited text

An edited text is a production credited to a contemporary intellectual author or historian. It documents one specific table, text, or diagram. This document can be from a previous edited source or a new edition created using the web interface of DISHAS.

An edited text can have various edition-types including an edition of a specific original text, a recomputation, a new production based on multiple sources, etc. An essential goal of the database is to make editions of astronomical tables that are accessible and comparable. In achieving this, edited texts, as classified in various edition-types, help identify mathematical and astronomical parameters, examine and develop critical and computational tools for creating new editions, and provide a centralised resource for studying astronomical tables from different traditions.

We collect general information about the edition, e.g., the title, the date of creation, etc., to record the metadata associated with it. This, along with the other information associated with the edited text, helps us unify and create standards for comparison in our database. We collect secondary source information for bibliographic accuracy and catalogue reference.

B.4 Historical actor

Historical actors are individuals or collectives related to the items or works stored in the database. They include author, copyist, owners, dedicatee, etc. associated to the materials of the original item. Within the framework and scope of our database, we collect certain information related to the historical actors, namely, their name, their flourit, and their location, to enable us to create a standardised lists of actors.

With the additional (and highly recommended) Virtual International Authority File (VIAF) number recorded, the database makes it possible to align names added at a latter stage with a larger, normalised, authority register of person names. We collect information about the actors in order to build prosopographic data.

B.5 Intellectual author

An intellectual author is a contemporary historian who has produced an edition that is stored in the database. This historian could be different to the person entering the data into the database. For example, in certain instances, a scholar (user) may be entering the data from an edition created by another historian (from, perhaps, an unpublished collection). The intellectual author in this case is the creator-historian and not the scholar (user).

Collecting information about intellectual authors validates the factual reliability of the database. This effort helps preserve the work of previous generation of historians as a way to foster and encourage further studies. The information collected about intellectual authors helps identify modern scholarship of the edition entered in the database.

B.6 Journal

In academic publishing, a scientific journal is a periodical publication intended to further the progress of science, usually by reporting new research.

B.7 Library

A library is a present-day or last-known location where the primary source is situated. The knowledge of the physical location of the library where the primary source is held is important for establishing the reliability of the database as a research resource. Facilitated with assistance tools like ISNI, entering the library information helps the database register and identify the correct library. For other libraries (not on ISNI records), a new entry helps create a record of its existence in the database for subsequent use.

B.8 Mathematical parameter

There are two kind of mathematical parameters :

- A table is said to be “displaced” when a constant value is added to its arguments or/and its entry. The value of these added constant are the first kind of mathematical parameters.
- A table can also be “shifted”. This happens when the ordering of the arguments and/or entry is modified according to a permutation cycle. The number of shifts in the permutation cycle is the second type of mathematical parameters.

The main purpose of identifying mathematical parameters is to compare displaced and/or shifted tables with those which are not displaced or shifted. We note that both kinds of tables may rely on the same set of astronomical parameters. Mathematical parameters can be shared by different sources. Additionally, shift and displacement are peculiar kinds of mathematical techniques linked to the manipulation of table sets and as such are worthy of investigation in their own right.

B.9 Original text

An original item is a table, text, or diagram as it appears in a given primary source. In instances where the actual state of the item is the result of several distinctive production acts, e.g., an extra column added to a table by a later actor, each differing state is a specific item. A scholar (user) may input these instances as separate original items. In the current stage of development of DISHAS, only tables are considered; texts and diagram will be included at a later stage. The original item, as it appears in the primary source, is the

fundamental unit of the database. Hence, elementary units are individual tables, diagrams, or texts that are actually witnessed in specific primary sources. This level of granularity is central to make DISHAS an efficient tool for mathematical analysis and critical editing.

We collect material information about the original item to allow source criticism. These include the date and the place of production, the title of the work, the manuscript shelfmark, etc.

B.10 Place

The definition of a place depends on the object to which it relates. For an original item, it refers to the place of production. For an author, it refers to his main place of activity where his work can be (approximately or accurately) geo-localised. For a work, the place refers to the original place of conception. This information is important to understand the diffusion and circulation of items in various works and traditions.

Collecting information about the place, including, in some instances, its latitude and longitude, helps us represent the locations on a map. This visual representation can become an important tool for textual analysis and circulation studies.

B.11 Primary source

A primary source is a manuscript or an early printed edition in which the original item is found. It is essential for the reliability of DISHAS that it provide accurate identification of its primary sources. Additionally, having accurate information about primary sources in the database can prove to be an important tool to study the history of modern collections in astral sciences.

Information about the primary source helps us identify the physical institution or library that holds the primary source. This identification allows scholars to locate the original source for further investigation.

B.12 Secondary source

A secondary source is a book, an article, a chapter, etc. that indicate a contemporary edition, or a survey of the table. Any edited document in the database can only be linked to a single secondary source for brevity. Information about secondary sources provides a reliable and tractable resource for further studies. This information will help current and future scholars use the database for their research.

B.13 Table content

The table content is the collection of numerical values in the edited text. The relation between these numerical values and those attested in the original item(s) associated with the edited text depends on the edition-type selected. We understand edited text as the type of work done by the intellectual author responsible of its creation.

Comparing numerical contents of various tables and building new computational and analytic tools is central to the aim of DISHAS. In order to achieve this, a minimum level of standardisation in the layout of tables and the format of recorded numbers is necessary. However, it is equally important to allow a certain level of flexibility to reflect the practices of historical actors faithfully. Hence, there are several possibilities about table layout, table types, and types of numbers. The choices of an intellectual author are carefully documented to ensure a meaningful comparison between the table contents of various edited texts. We collect information about the table contents in order to structure the contents for displaying purposes. The astronomical parameters are linked to the edited texts (and not to the original text) because a given parameter set is always attributed to a table by an intellectual author after analyses.

B.14 Table type

Tables are classified in different types. These different types correspond to the specific steps in astronomical computations as organised by historical actors. In modern terms, these types correspond to distinct mathematical functions which express a portion of an underlying astronomical theory.

B.15 Type of edition

Three types of edition are defined in DISHAS

Type-A : Edition Numerically consistent table

The first stage involves the rendition of the individual table (taken from individual MSS of the same work) from its diplomatic transcription onto a more numerical consistent form, i.e., a numerical format free of any data integrity errors. The following is a list ; albeit not an exhaustive one, of the different types of strategies employed in creating a numerically consistent table.

1. The first stage involves the rendition of the individual table (taken from individual MSS of the same work) from its diplomatic transcription onto a more numerical consistent form, i.e., a numerical format free of any data integrity errors. The fol-

lowing is a list ; albeit not an exhaustive one, of the different types of strategies employed in creating a numerically consistent table.

2. Numerical parsing on a local and global scale to keep regularity, e.g., correcting a number to maintain simple mathematical regularity. It should be noted here that this inspection is merely observational and not computational.
3. Using pattern recognition and visual inspection to correct numbers for constancy, uniformity, and accuracy (in reproduction).
4. Noting all numerical changes, e.g., instances where ambiguous, dubious, or inconsistent numbers are altered, alternatives or originals are duly noted as comments.

Note : Any ancillary material (paratext, scribal error marks and corrections, change in handwriting, etc.) is NOT included in this table : these remain features that are to be noted in the TEI version of the table.

Type-B : Edition Numerical Reproduced Table

The second stage involves employing the different individual tables (belonging to the same work, but taken from different MSS) and forming an EDITION table. This process can be defined based on a few different strategies :

1. Any eclectic, stemmatic, or copy-text method (in general, any philological method) of comparing the different 'witness' tables (from Part A) and establishing a critical 'base' numerical table, with all variants duly noted in comments. These comments effectively make the apparatus criticus.
2. Any mathematical method (from Part C) of establishing a critical 'base' numerical table and then employing tables (from Part A) to note the variants.
3. Any combination of the two strategies above.
4. Additionally, any previously 'critically' edited table (from a set of extant or non-existing manuscripts) belonging to the same table-type and work (as set in Part A) can also be stored in Part B here. In such cases, it is important to note any relation, if applicable, of this previous work to the different tables in Part A. Failing this, a common reference to the work and table- type becomes minimally necessary.

Several different 'critical editions' (based on different editorial strategies) of the same table may be stored in this stage ; all of which, are identified with the correct relational structure to the tables in parts A and C.

Type-C : Edition Numerically Recomputed Table(s)

The third stage involves the recomputed tables – corresponding to each individual table in Part A – that are stored relationally in the SQL database. These tables are generated based on the following possible strategies :

1. A modern mathematical recomputation of the table based on (a) choice of appropriate parameter, (b) mathematical equations governing the underlying function tabulated, (c) selection of a suitable generative or iterative algorithm.
2. A historically congruous mathematical algorithm based on editorial decisions. These tables are generated by a modern scholar based on his or her own understanding of the mathematical apparatus presumably intended to have been used by the original author in his time. These tables form a collection of our modern efforts to reconstruct numerical tables within its historical context.
3. Any other derivatives or combinations of mathematical procedures, algorithms, iterative techniques, rounding schemes, etc. used to generate the numerical entries of the table form separate mathematical tables that are also stored in this part.

B.16 Work

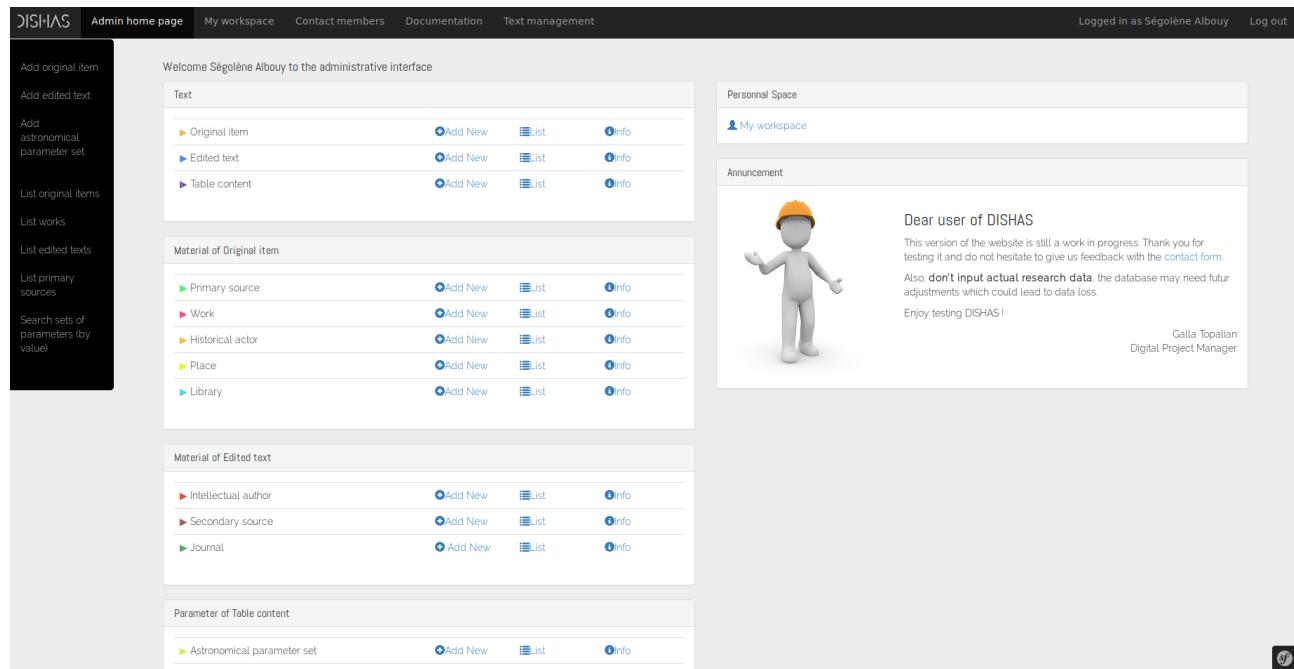
An astronomical work is a distinct intellectual creation consisting of tables, texts, or diagrams. It is often identified by a title or incipit and attributed to a given historical actor. Its content and organisation may vary depending on the particular primary source.

We can analyse the different attestations of a given work as original items rather than individual works are the fundamental unit of DISHAS. For instance, one can identify a core set of tables around which various satellites tables, texts, or diagrams may be presented depending on the primary source. With this design, the geographical and chronological evolution of these dynamical sets of original items can be analysed and studied. This ability is essential when it comes to critical editing a work. Additionally, the choice of an original item as the fundamental unit also allows to analyse the intellectual composition of a given manuscript. This is an important step in understanding how manuscript shape intellectual traditions.

A work needs to be identified by a current title or incipit. When possible supplementary information like place, dates, authors or translator is very useful for studying the circulation of the work.

Annexe C

Captures d'écrans de l'interface administrateur



The screenshot shows the DISHAS administrative interface homepage. At the top, there is a navigation bar with links: Admin home page, My workspace, Contact members, Documentation, and Text management. On the right side of the top bar, it says "Logged in as Ségolène Albouy" and "Log out". The left sidebar contains a vertical list of administrative tasks:

- Add original item
- Add edited text
- Add astronomical parameter set
- List original items
- List works
- List edited texts
- List primary sources
- Search sets of parameters (by value)

The main content area is divided into several sections:

- Welcome Ségolène Albouy to the administrative interface**: A general welcome message.
- Text**: A section containing three categories: Original item, Edited text, and Table content, each with "Add New" and "List" buttons.
- Material of Original item**: A section containing five categories: Primary source, Work, Historical actor, Place, and Library, each with "Add New" and "List" buttons.
- Material of Edited text**: A section containing three categories: Intellectual author, Secondary source, and Journal, each with "Add New" and "List" buttons.
- Parameter of Table content**: A section containing one category: Astronomical parameter set, with "Add New", "List", and "Info" buttons.
- Personnal Space**: A section titled "My workspace" which is currently empty.
- Announcement**: A section featuring a 3D white figure wearing a yellow hard hat. The text reads:

Dear user of DISHAS
This version of the website is still a work in progress. Thank you for testing it and do not hesitate to give us feedback with the contact form.
Also, **don't input actual research data**, the database may need future adjustments which could lead to data loss.
Enjoy testing DISHAS!

Galla Topalian
Digital Project Manager

FIGURE C.1 – Page d'accueil de l'interface administrateur

DISHAS Admin home page My workspace Contact members Documentation Text management Logged in as Ségolène Albouy Log out

Add original item
Add edited text
Add astronomical parameter set
List original items
List works
List edited texts
List primary sources
Search sets of parameters (by value)

WHAT IS AN EDITED TEXT

An edited text is a production credited to a contemporary intellectual author or historian. It documents one specific table, text, or diagram. This document can be from a previous edited source or a new edition created using the web interface of DISHAS. An edited text can have various edition-types including an edition of a specific original text, a recomputation, a new production based on multiple sources, etc.

An essential goal of the database is to make editions of astronomical tables that are accessible and comparable. In achieving this, edited texts, as classified in various edition-types, help (i) identify mathematical and astronomical parameters, (ii) examine and develop critical and computational tools for creating new editions, and (iii) provide a centralised resource for studying astronomical tables from different traditions.

We collect general information about the edition, e.g. the title, the date of creation, etc. to record the metadata associated with it. This, along with the other information associated with the edited text, helps us unify and create standards for comparison in our database. We collect secondary source information for bibliographic accuracy and catalogue reference. Three types of edition are defined in DISHAS:

- Type-A Edition Numerically consistent table
- Type-B Edition Numerical Reproduced Table(s)
- Type-C Edition Numerically Recomputed Table(s)

Show 10 entries Reload Column visibility Search:

#	Title	Edition type	Year	Bib. ref.	Author(s)	Original Item(s)	Table contents(s)	Parameter Sets(s)
1	Lunar velocity	a	2018		Casulleras (Josep)	CHECK (1330 - 1340): Vatican Library Vat. Pal. Lat. 1376, 57v-60r: Unknown work	(id n°62)	

Add an edited text

FIGURE C.2 – Page de liste des enregistrements des éditions de tables astronomiques

DISHAS Admin home page My workspace Contact members Documentation Text management Logged in as Ségolène Albouy Log out

Add original item
Add edited text
Add astronomical parameter set
List original items
List works
List edited texts
List primary sources
Search sets of parameters (by value)

WHAT IS AN ORIGINAL ITEM

An original item is a table, text, or diagram as it appears in a given primary source. In instances where the actual state of the item is the result of several distinctive production acts, e.g., an extra column added to a table by a later actor, each differing state is a specific item. A scholar (user) may input these instances as separate original items. In the current stage of development of DISHAS, only tables are considered; texts and diagram will be included at a later stage.

The original item, as it appears in the primary source, is the fundamental unit of the database. Hence, elementary units are individual tables, diagrams, or texts that are actually witnessed in specific primary sources. This level of granularity is central to make DISHAS an efficient tool for mathematical analysis and critical editing.

We collect material information about the original item to allow source criticism. These include the date and the place of production, the title of the work, the manuscript shelfmark, etc.

Show 10 entries Reload Column visibility Search:

#	Title	Term. Post Quem	Term. Ante Quem	Scribal agent	Place of conception	Language	Script	Work	Work creator(s)	Primary source
1	Equations of the Sun			Dinakara (1730 - 1800)	Varanasi, Uttar Pradesh, India	Sanskrit	Devanagari (Nagari)	Dinakara Sarani (1730-1800) [Unknown creator]	Benaras Sanskrit College B36	
2	العمل في الأفلاك	1000	1025	Kushyār ibn	Iran	Arabic	Arabic	الرجح الجامع	Kushyār ibn Süleymaniye Kütiphanesi Fatih	

Add an original item

FIGURE C.3 – Page de liste des enregistrements des *original items*

Input your table - DISHAS project - astronomical table in database - Mozilla Firefox

localhost:8000/adm/view-tableContent/80

File Document Enrich data Select Data viz Window Help

a \ b	0	1	2	3	4	5	6	7	8	9
00	00	01	02	03	04	05	06	07	08	09
00	05	02	03	04	05	06	07	08	09	02
00	10	03	04	05	06	07	08	09	05	04
00	15									
00	20	03	04	05	06	07	08	09	05	04
00	25	02	03	04	05	06	07	08	09	03
00	30	01	02	03	04	05	06	07	08	02
00	35									
00	40									
10	21	26								
11	17	09								
12	12	51								
13	08	34								
14	04	17								
15	00	00								

Logs Cell Commentary
Cell type: sexagesimal
Cell content: 05;
Decimal value: 5.00000
Line: 7
Col : 9-9

FIGURE C.4 – Interface de saisie numérique de tables

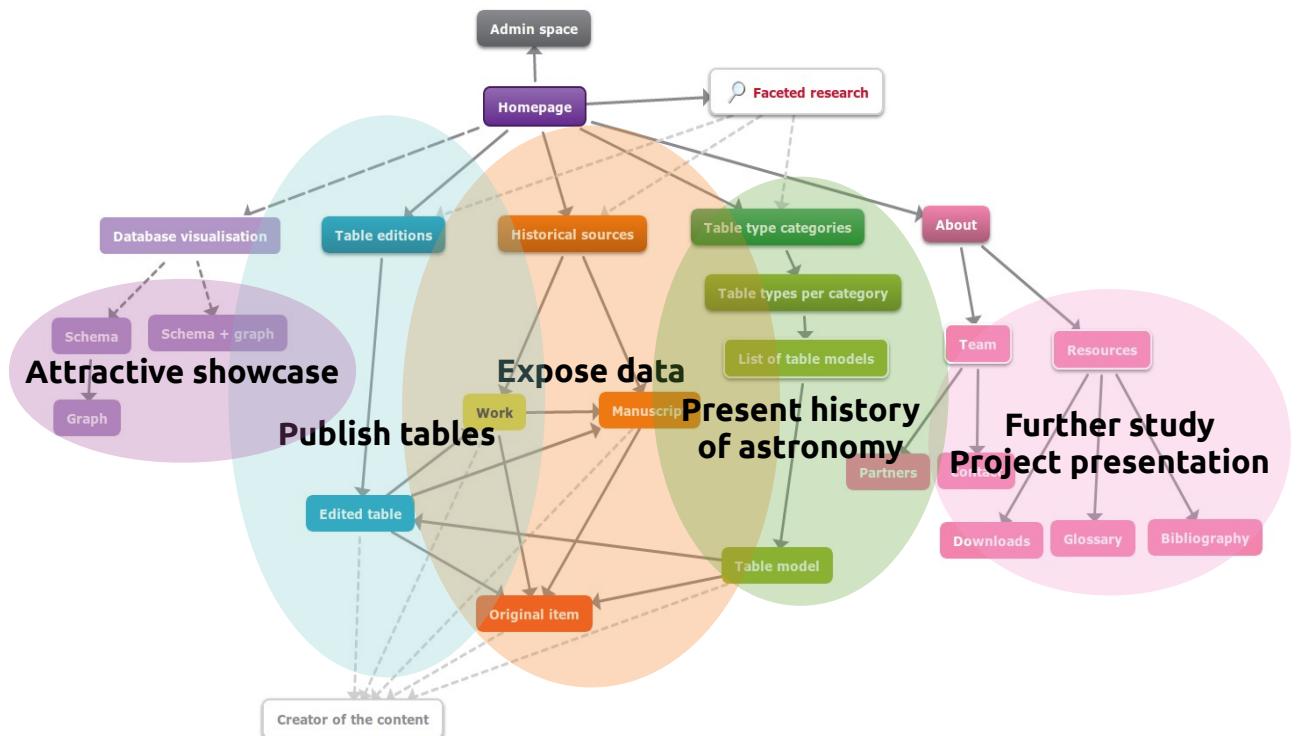
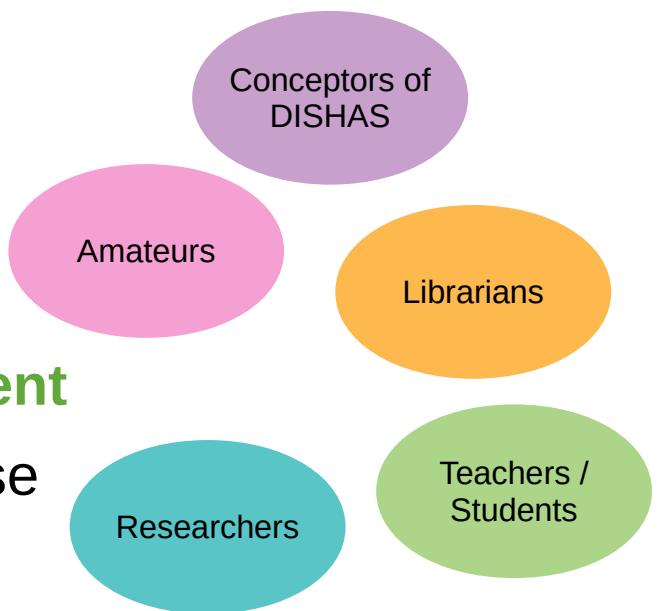
Annexe D

Présentation du *front office*

Document de présentation utilisé lors du séminaire numérique de l'équipe de DIS-HAS du 13 mai 2019, en présence de l'équipe de recherche. Certaines diapositives ont été retirées du document pour éviter la redondance de l'information. Cette présentation visait à exposer les propositions de développements concernant les parties de l'interface utilisateur.

Purpose of the front office

- **Expose** the platform data
- **Publish** tables
- **Present** the project
- Provide **pedagogic content**
- Be an **attractive** showcase



Aim of the visualizations

- Explore DISHAS's **data** in its specificity
- Aggregate a lot of data : **quantitative analysis**
- Illustrate the **context of production** of the tables
- Approach the subject in an **interactive way**
- Discover unexpected **correlations** between data
- Provide **course materials**

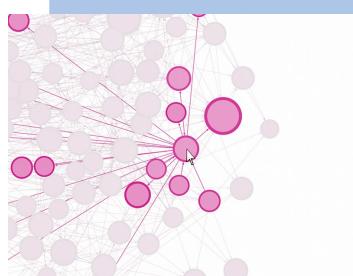
Granularity of visualizations



Project

Global visualization
of all the data

Ex : graph of the
entire database



List

Visualization
of an entity
of the database

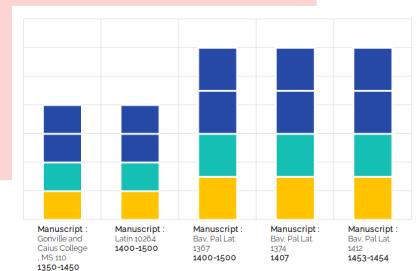
Ex : map of the
places of conceptions
of all the works



Record

Visualization of
a single record

Ex : a column chart to
represent the manuscript
originated from a work



Entire Database



Graph of the database

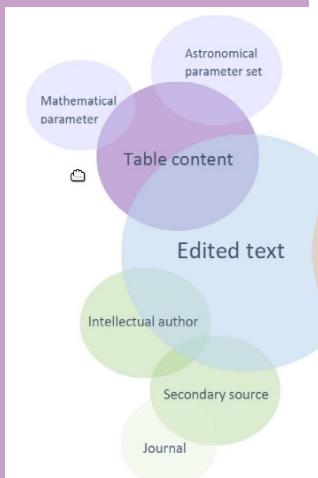
PURPOSES

- Better understanding the conceptual model of the database
- Discover unexpected correlations between records
- Be an attractive showcase for the project

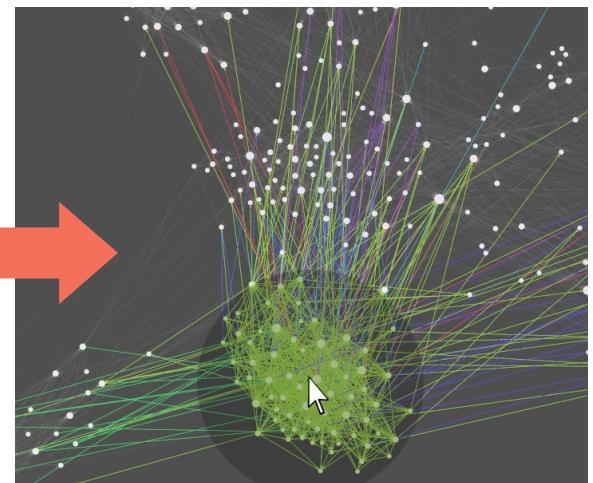
Entire Database



Schema + graph



Example



Tools : [Sigma js](#), [cytoscape](#), [Gephi](#), [Kibana](#)



Astronomical object
Table type
Formula definition

Discover table models

PURPOSES

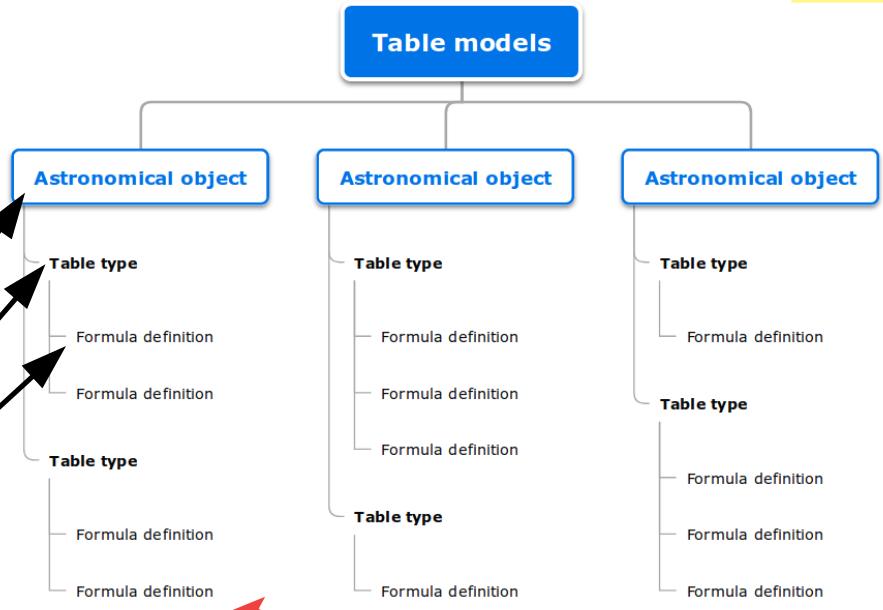
- Provide editorialized scientific content for the platform
- Give access to the tables via an astronomical logic
- First milestones of a virtual exhibition on medieval astronomy

Astronomical object
Table type
Formula definition

Discover table models



Trigonometrical
→ Sine
→ Simple sinus

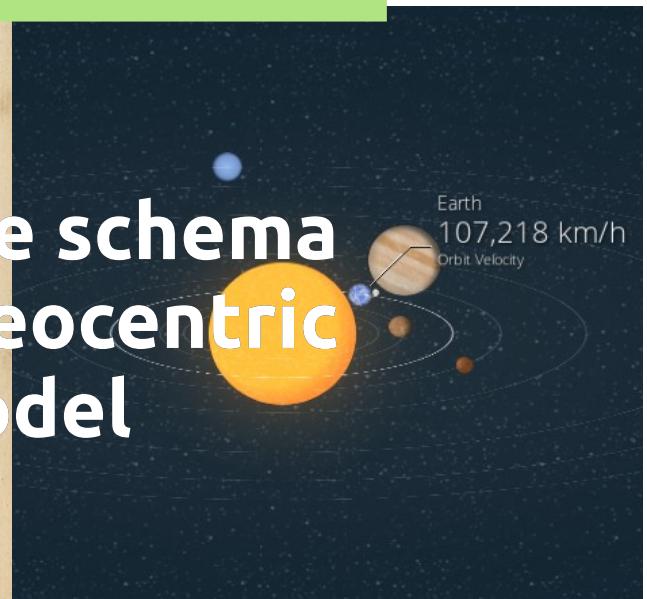


Link to the tables that use this model

Astronomical object



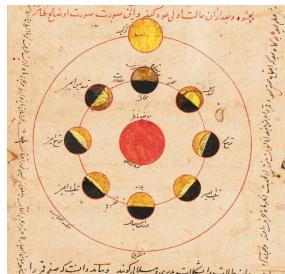
Clickable schema
of the geocentric
model



Example of an animated schema

Astronomical
object

Extensive
iconography

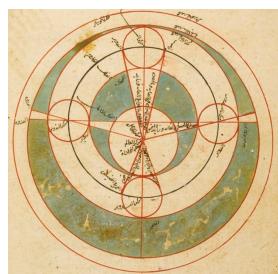


Eclipse

Moon

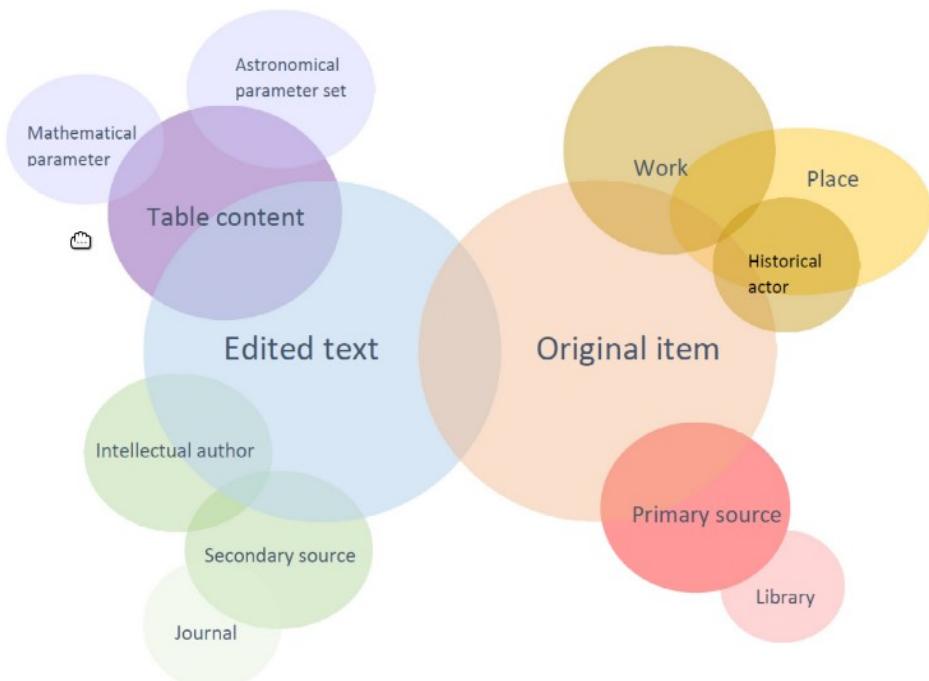
Mars

Trigonometrical Spherical

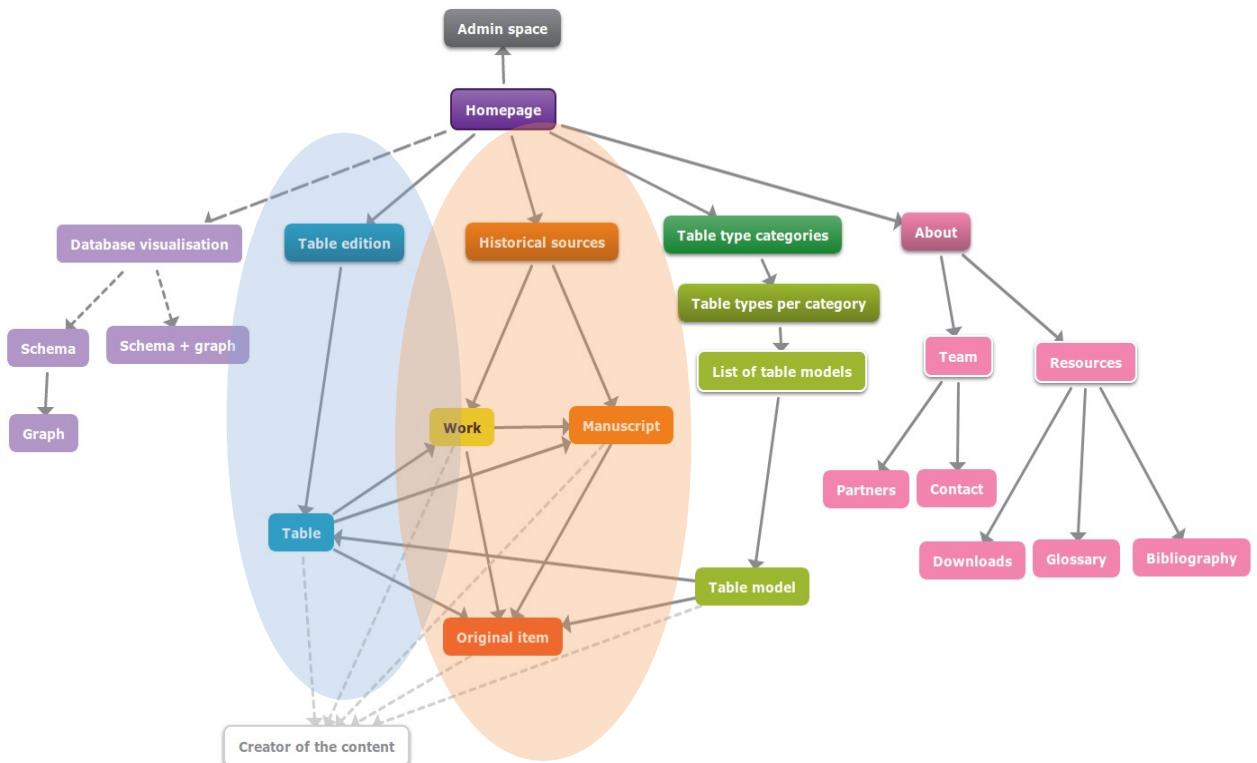


Sun

Table edition



Historical sources



Work
Original
item

Historical sources



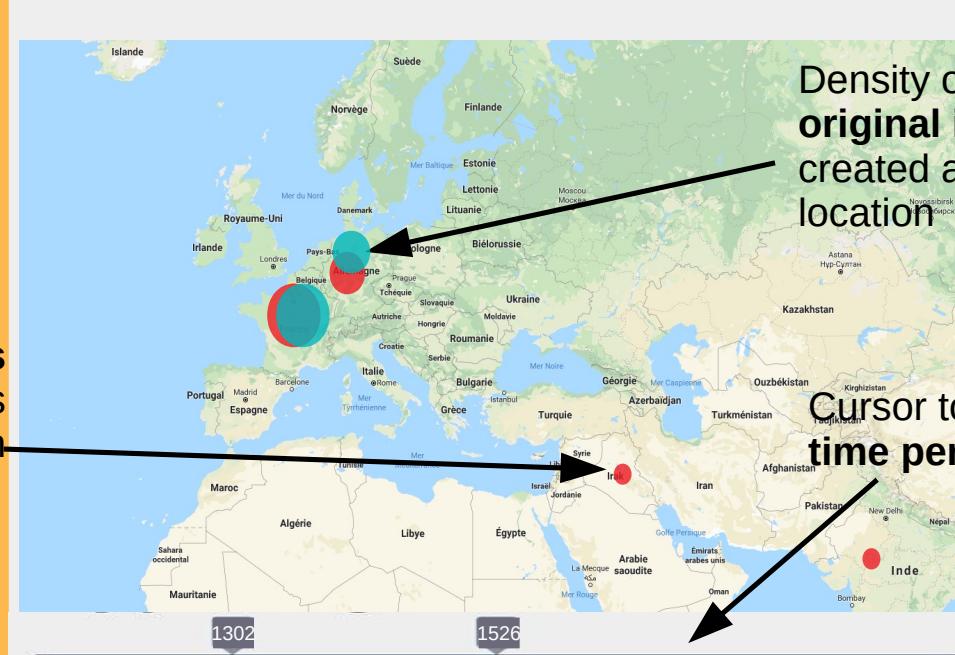
PURPOSES

- Focus on the creation context of the tables
- Illustrate the circulation of knowledge
- Show places and periods of intellectual profusion
- Give access to the records of work and primary source

Work Original item



Density of **works**
created at this
location



Density of
original items
created at this
location

Cursor to set a
time period

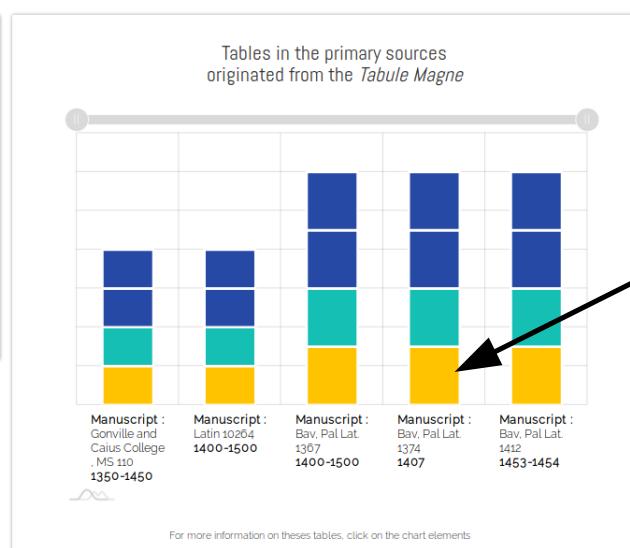
Work



General
informations

GENERAL INFORMATION	
<i>Tabule Magne</i>	
Incipit	Multiplicis philosophie
Creator	John of lignères (1290-1350)
Date of conception	1325
Place of conception	Paris, France

Link to the **list of
works** sharing
the same
characteristics



Visualisation
of the
**primary
sources** and
**original
items**
originated
from the work

Example

Primary source



General informations

► Primary source

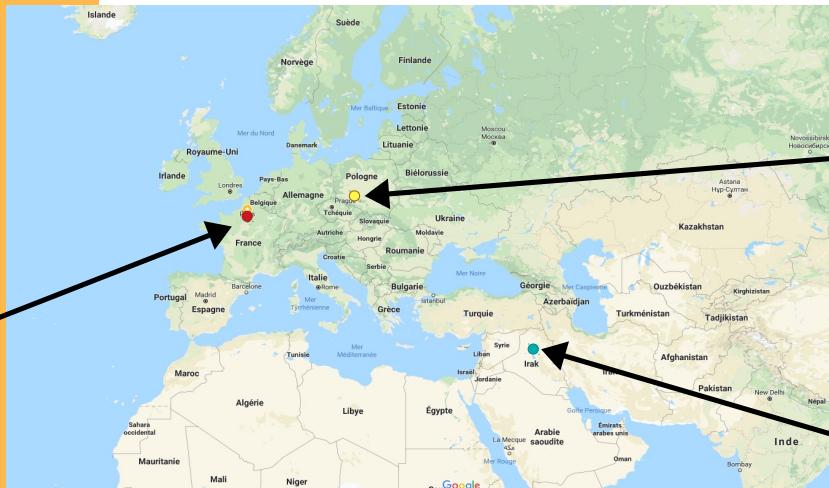
Manuscript — Gonville and Caius College, MS 110

Link to the list of primary sources sharing the same characteristics

Original item



Place of creation of the work of the original item

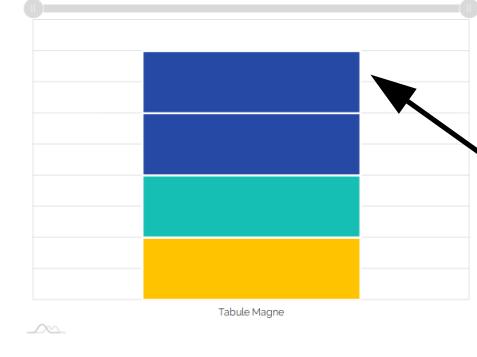


Place of creation of the original item

Place of curation of the manuscript containing the original item

► Primary source

Works present in Cambridge | Gonville and Caius College, MS 110



Example

Visualisation of the **work** and **original items** present in the primary source

Edited text
Table
content



Table editions

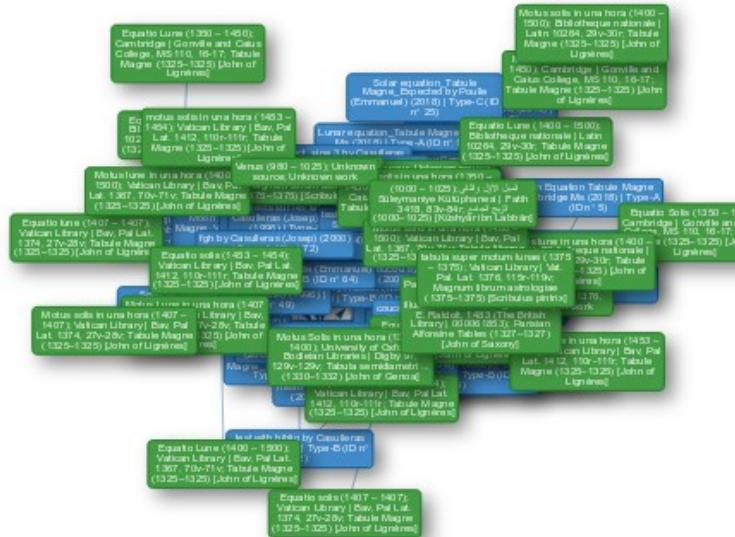
PURPOSES

- Give an overall view of the edited content of the platform
- Show tendencies in type of table / astronomical parameter
- Give access to the records of table edition

Edited text
Table
content



Admin side



Edited text
Table
content

Edited
works

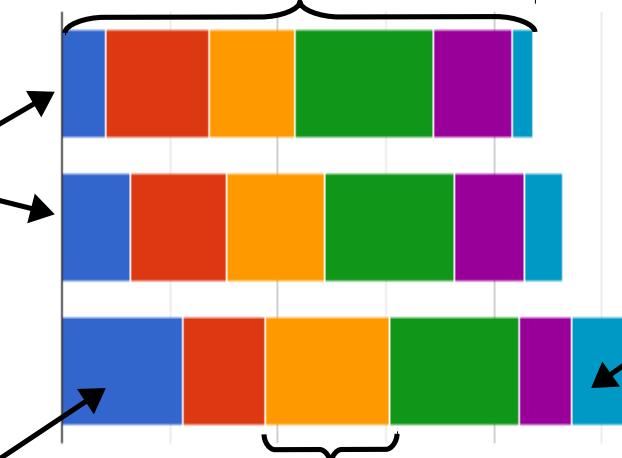
One original
item

Edited text
Table
content

Moon

Stacked bars

Amount of editions of a particular work



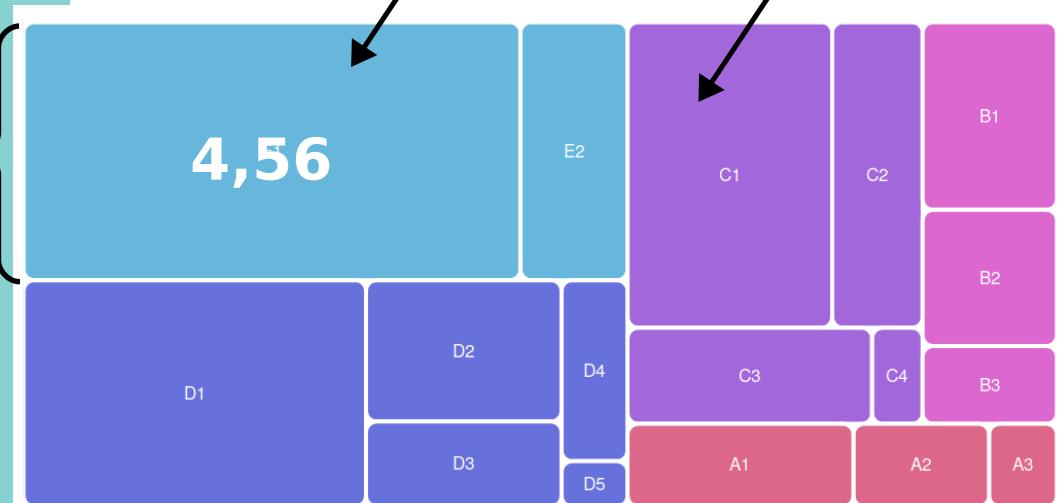
Color to specify a table type

Amount of editions of a particular original item

Treemap

Color for an astronomical object

A cell for a same parameter set



Edited text Table content



Column :
country or
period

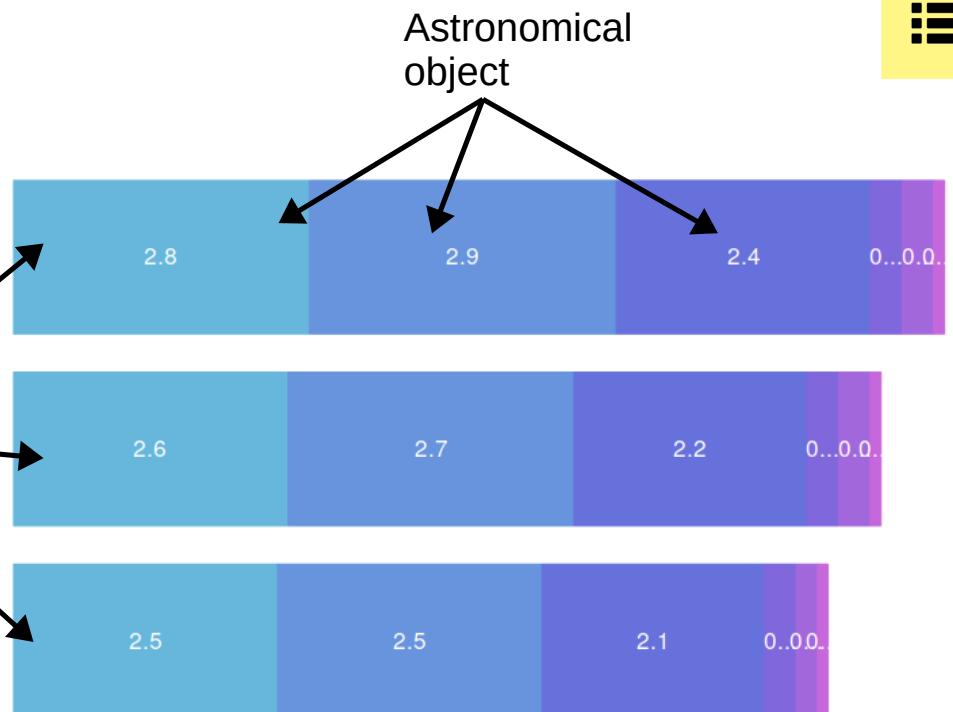


Table edition



Historical Context

(edition type A
and B)

- related work
- related manuscript
- etc.

Mathematical Context

- astronomical parameter set
- table type
- etc.

Editorial Context

- sources of the edition
- intellectual author
- edition type

Table content

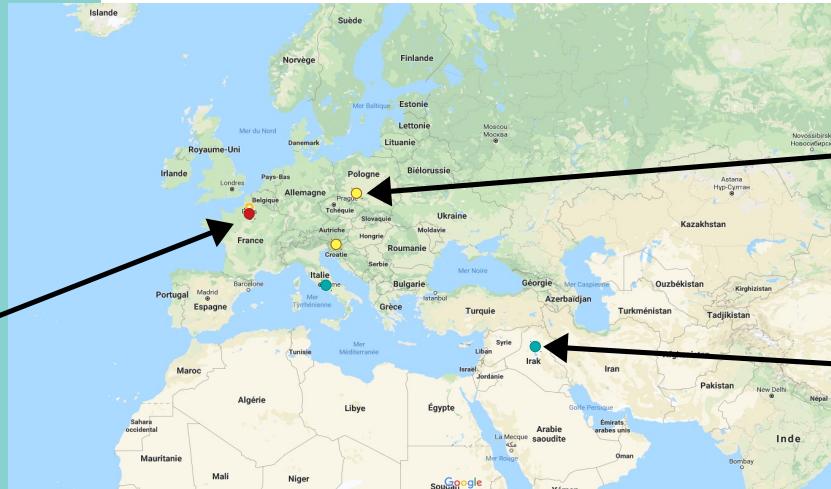
- View of the table
- Critical apparatus
- Comments

Edited text



Historical context

Place of conception of the work



Writing places of the original items

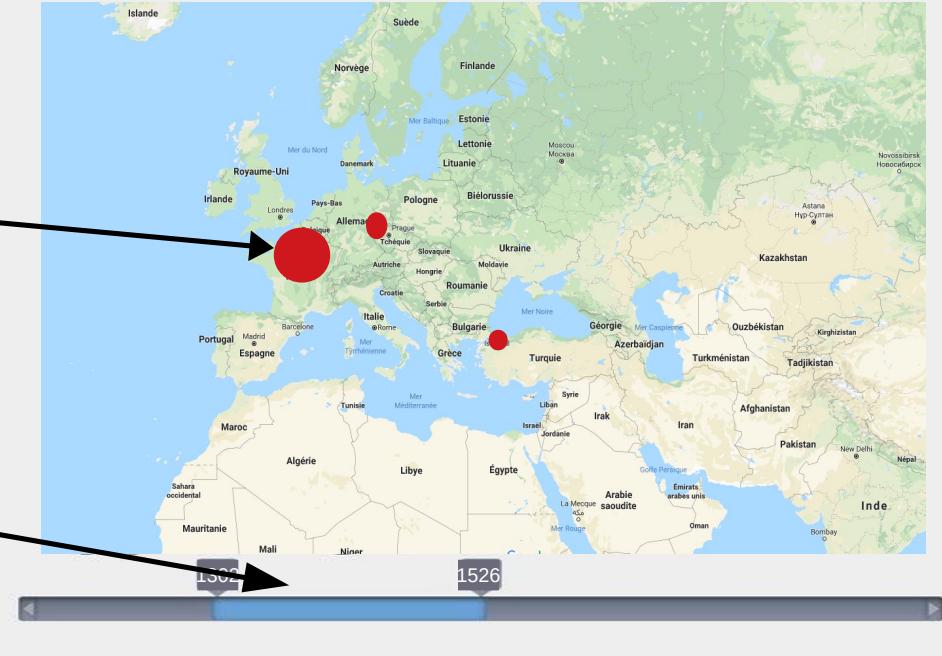
Edited text
Table content



Mathematical context

Density of original item using the same parameter set

Cursor to set a time period



Edited text
Table
content

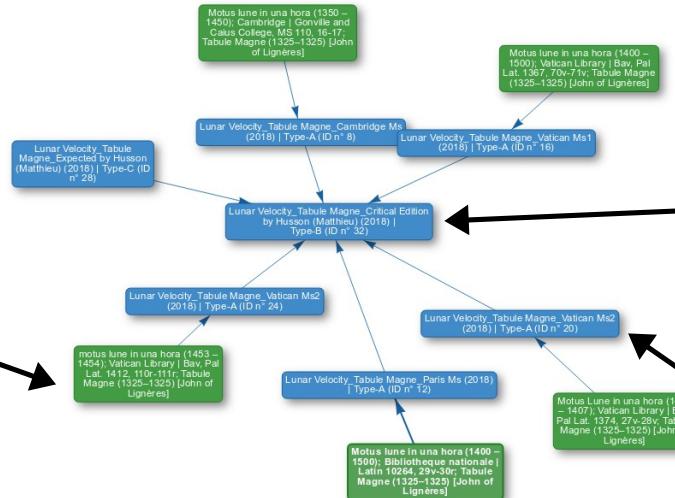


Edited
original
item

Table
content



Editorial context



Current
edition

Edited text
on which
this edition
is based

Table content

Intensity of
color : number
of variants at
the same spot

Outline :
comment

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1														
2														
3														
4				Var										
5														
6														
7							Var							
8								Var						
9														
10														
11								Var						
12														
13														
14														
15														

Metadata of the cell
that was clicked

Location 8 – 7

Variants :

- Ms 1 : 12
- Ms 2 : 13
- Ep 3 : 11

Comment :

« blah blah blah »

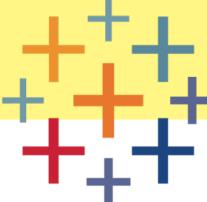
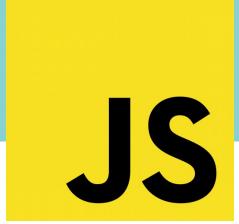
Link to modify the
table in DTI

Exposing the data : technical choices

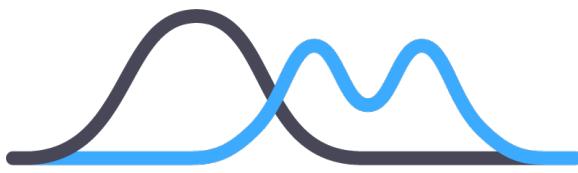
Three way to query the database



Technologies for visualization

<p>Visualization Software</p> <p>+ graphic interface + no loading time - static views - not very customizable</p> 	<p>Javascript libraries</p> <p>+ great diversity available - no control over updates & backward compatibility</p> 	<p>Kibana</p> <p>+ graphic interface + easy installation with elastic search - not designed for public visualizations</p> 
--	--	--

Choice of a javascript library



AMCHARTS

- **Versatile library** : chart, map, animated graph
- A **lot of documentation** available
- Interesting features : zoom, exports, etc.
- Easy implementation and handling
- Pricing : free on given conditions

Annexe E

Cahier des charges fonctionnel des visualisations pour l'interface utilisateur

Ce document est un cahier des charges fonctionnel où sont décrites en détail toutes les interfaces de la plateforme publique ; chaque rangée correspond à une page de la plateforme, chaque colonne concernant quant à elle un aspect précis de cette page :

- une première colonne présente l'interface de manière générale : son objectif, son architecture globale avec la description de ses différents composants ainsi qu'éventuellement des interrogations vis-à-vis de sa conception ;
- une deuxième colonne expose l'ensemble des données qui seront récupérées de la base pour l'affichage, en particulier pour les métadonnées requises dans les pages de notices ;
- une troisième colonne s'attache à expliciter précisément la visualisation de donnée apparaissant sur la page (si tel est le cas), tant dans son aspect visuel que dans ses enjeux intellectuels ;
- une quatrième colonne concerne les données nécessaires à la constitution de cette visualisation ;
- une cinquième colonne dresse la liste de tous les liens de redirection présents à l'intérieur de la page, composant ainsi le réseau dans lequel elle s'inscrit ;
- une dernière colonne détaille les corpus de données qui pourront être établis à partir de cette page. Cette partie concerne essentiellement les pages de notice, et fait référence aux « loupes » disposées sur la barre latérale de métadonnées.

	DESCRIPTION DE LA PAGE	MÉTADONNÉES A AFFICHER	VISUALISATION DE DONNÉES	MÉTADONNÉES DE LA VISUALISATION	LIENS ET REDIRECTION	PAGE RECHERCHE AVEC LISTES D'ENTITÉS
HOMEPAGE	BLOC 1 Eléments : - Bannière : logo + phrase résumant le projet très succinctement (type : plateforme de recherche en astronomie médiévale) - Barre de navigation : Logo (accueil), Découvrir (<i>historical navigation + astronomical navigation</i>), Recherche, À propos, Back office - Court paragraphe d'introduction au projet (lien pour en savoir plus : à propos) dans une bannière pour le mettre en avant - Accès aux parties <i>Historical navigation / Astronomical navigation</i> + paragraphe de présentation des deux parcours de découverte de la donnée + mention du glossaire - Accès à DTI + paragraphe de présentation de l'outil - Accès à la visualisation en <u>graphe de la base de données</u> + paragraphe de présentation - footer : projets partenaires + RGPD Enjeux : - Faire comprendre de quoi le projet retourne - Guider l'utilisateur dans son parcours dans le site - Constituer une entrée en matière attractive				Liens vers autres pages BLOC 2 : <i>Historical navigation</i> BLOC 6 : <i>Astronomical navigation</i> BLOC 12 : Recherche facettée BLOC 13 : Database graph BLOC 15 : About BLOC 16 : DISHAS's Table Interface	
HISTOIRE	BLOC 2 Enjeux : - Point d'accès aux données d'enregistrements - Présenter des données de la base dans leur aspect historique - Représenter de manière simple des corrélations complexes entre les données Description de la page : Point d'entrée aux enregistrements d'œuvres et de sources primaires. Page constituant un portail historique vers la donnée, il s'agit de contextualiser la donnée historique dans son milieu et de présenter les interdépendances et causalités entre les entités de la base de manière claire. Page conçue comme un pendant du BLOC 6, constituant une approche davantage astronomique de la donnée Problème : données de la base qui peuvent tendre à constituer une visualisation biaisée : rappel que les données présentées ne sont pas vouées à être exhaustives et qu'elles sont probablement trompeusement représentatives.	Métadonnées affichées : Selon la facilité d'accès aux pages d'œuvre (BLOC 3) et de source primaire (BLOC 4) grâce à la visualisation, il sera nécessaire ou non d'afficher une liste d'enregistrements en plus à la visualisation. Si tel est le cas (il est possible de penser à une liste se mettant à jour en fonction des données affichées sur la visualisation). On peut imaginer également pouvoir filtrer les données affichées sur la visualisation grâce à un système de case à cocher au sein des listes d'enregistrements)	VISU 2 Enjeux : Mettre en valeur le contexte de production des œuvres et d'enfigurer la dispersion géographique Visualisation de données : Carte des <i>original items</i> et des œuvres conçue comme un point d'entrée aux pages d'enregistrements d'œuvres et de sources primaires : chaque œuvre sera représentée par un point sur la carte correspondant à son lieu de conception. Si de nombreuses œuvres sont placées au même endroit, les points seront regroupés pour constituer des points plus ou moins gros pour signifier la densité : au zoom, les différents enregistrements apparaîtront distingués. À chaque point sera de plus assigné une période de temps : des curseurs pourront permettre de déterminer une période de temps pour laquelle on souhaite configurer l'affichage des enregistrements. La même chose sera faite pour les données <i>original item</i> avec leur lieu de rédaction ou d'édition. Problème : si la visualisation est pensée comme un point d'accès, il faut que chacun des enregistrements représentés soit accessible, or, si deux œuvres sont placées exactement au même endroit, la question se pose de l'affichage différencié des ces données.	Données de la visualisation : WORK (all records) - id + title / small incipit - tpq + tqq - PLACE → placeName + lat + long ORIGINAL TEXT (all records) - id - tpq + tqq - PLACE → placeName + lat + long - PRIM SOURCE → id + shelfmark / primTitle	Liens vers autres pages BLOC 3 : Au clic sur un point représentant une œuvre, redirection vers la page de l'œuvre correspondante. BLOC 4 : Les points représentant un <i>original item</i> redirigent vers la page de la source primaire qui le contient. BLOC 6 : <i>Astronomical navigation</i> BLOC 12 : Lien vers la recherche facettée	Listes Liste de toutes les œuvres Liste de toutes les sources primaires
WORK	BLOC 3 Enjeux : - Mise en avant des métadonnées - Monter les liens avec les entités issues de cette œuvre (<i>original item, primary source et edited text</i>) - Situer le contexte de création de l'œuvre Description de la page : Notice de métadonnées pour l'entité œuvre : il s'agit de présenter et contextualiser le plus complètement, car l'entité <i>work</i> a censément la primauté par rapport à toutes les autres entités de la base : il faut donc souligner cette parenté intellectuelle. De plus, il faut rappeler le contexte de production d'une œuvre en suggérant le milieu qui a entouré la création de cette œuvre et évoquant la tradition dans laquelle elle s'inscrit. Question : Est-ce nécessaire de fournir une visualisation en carte de l'œuvre pour figurer la répartition géographique des témoins et sources primaires ? Possibilité soit de donner un accès à une visualisation en carte sur une page extérieure ou une visualisation plus développée pour la page <i>Original item</i>)	Métadonnées affichées : WORK (single record) - id + title + titleOriginalChar - incipit + incipitOriginalChar - tpq + tqq - translator + creator(s) : HIST ACTOR → id + name + tpq + tqq + viaf - PLACE → id + placeName - createdBy : FOS USER → username ORIGINAL ITEM (array) (<i>originalItem.work = current work</i>) - LANGUAGE (array) → languageName - SCRIPT (array) → scriptName DEFINITION (single record) (<i>id = 26</i>) - longDefinition - userInterfaceColor	VISU 3 Enjeux : Visualisation qui met en avant la complétude de l'œuvre dans les différentes sources primaires et donne un aperçu du contenu scientifique de celle-ci. Description de la visualisation : Visualisation en colonnes proche de la VISU 4 où chaque colonne représente une source primaire contenant l'œuvre en question. Chaque colonne est ensuite subdivisée en cellules représentant les <i>original items</i> issus de l'œuvre contenue dans la présente source primaire. C'est-à-dire que ce ne sont pas affichés les <i>original items</i> dans ces sources qui ne sont pas issus de l'œuvre. La hauteur d'une cellule est déterminée par l'importance en terme de page de chaque item. En outre, la couleur d'une cellule correspond à l'objet astronomique de la table: une légende sur le côté indique la signification de ces différentes couleurs. L'ordre des colonnes est chronologique. Chaque élément de la visualisation est cliquable et donne accès à la page de l'élément concerné. Enfin, il est possible de zoomer dans la visualisation pour accéder à plus de détail. Au dessous de ce graphique, il est possible d'ajouter une frise chronologique où seront figuré les sources primaires ou témoins. (En outre, pour témoigner de la diffusion de l'œuvre dans différentes traditions, une visualisation en diagramme circulaire des langues peut être envisagée : se pose la question en revanche du statut des œuvres qui ne sont qu'une traduction, ce qui fausserait la visualisation).	Données de la visualisation : ORIGINAL ITEM (array) (<i>originalItem.work = current work</i>) - id + originalTextTitle - tpq + tqq - TABLE TYPE → ASTRONOMICAL OBJECT → objectName - pageMin + pageMax (→ nombre de pages) - PRIMARY SOURCE → id - LANGUAGE → languageName PRIMARY SOURCE (array) (<i>originalItem.primarySource = those primary Source</i>) - id + shelfmark - LIBRARY → libraryName - primType - ORIGINAL ITEM (array) → tpq + tqq	Liens vers autres pages BLOC 4 : au clic sur le label des colonnes, redirection vers la notice de la source primaire désignée BLOC 5 : au clic sur les cellules, redirection vers la notice de l' <i>original item</i> concerné BLOC 2 : <i>Historical navigation</i> BLOC 6 : au clic sur un élément de la légende, redirection vers une page de présentation de l'objet astronomique correspondant Sites extérieurs - VIAF : lien vers la notice de la personne en question lorsqu'un identifiant est renseigné	Listes - Œuvres du même auteur - Éditions (publiques) de cette œuvre - Œuvres concues dans la même ère géographique (± 1 long, ± 2 lat) - Œuvres concues dans la même période temporelle (± 50 tpq, ± 50 tqq) Question : Si l'œuvre représentée est une traduction, la date donnée par l'utilisateur est-elle celle de la conception ou de la traduction ? Déterminer une période de temps autour de cette date est-il encore pertinent ?

P R O M A R C E	BLOC 4 Enjeux : <ul style="list-style-type: none">- Vocation à constituer une notice de bibliographique à citer et référencer- Insister sur les aspects de la source qui permettent d'en comprendre l'usage et le contexte de conception Description de la page : <p>Page de présentation des sources primaires contenant les tables astronomiques. Cet élément de la base est à valoriser notamment pour un public de bibliothécaires qui pourraient être intéressé de retrouver des manuscrits en lien avec leurs collections.</p> <p>Mise en avant des référentiels permettant de faire des liens entre plusieurs ontologies. Mise en valeur des métadonnées de catalogage utilisées par la plateforme – notice qui a vocation à être liée à des notices bibliographiques dans les catalogues numériques de bibliothèques.</p> <p>Penser à un URL de la page qui puisse constituer une URI pérenne et facilement citable.</p>	Métadonnées affichées : PRIMARY SOURCE (single record) <ul style="list-style-type: none">- id + shelfmark- primType- primTitle + primEditor + date- tpx + tq (de tous ses original items)- digitalIdentifier- LIBRARY → libraryName + city + country + isni- createdBy : FOS USER → username ORIGINAL ITEM (array) <ul style="list-style-type: none">- originalItem.primarySource = current primary source- PLACE → placeName- HIST ACTOR → actorName + actorNameOrigChar + tpx + tq + viaf- SCRIPT → scriptName + iso15924- LANGUAGE → languageName + iso6392 DEFINITION (single record) <ul style="list-style-type: none">(Id = 20)- longDefinition- userInterfaceColor	VISU 4 Enjeux : <p>Mise en avant d'aspects ayant trait à la bibliographie matérielle, c'est-à-dire l'agencement des folios, la composition de l'ouvrage, tant en tant qu'assemblage de différentes œuvres que en tant qu'objets composites ayant gardé la traces d'usages et de multiples possesseurs.</p> Description de la visualisation : <p>Visualisation en colonnes proche de la VISU 3. Chaque colonne représente une œuvre présente dans la source figurée. Chaque cellule correspond à un <i>original item</i> de cette œuvre dans la source primaire ; sa couleur est liée à l'objet astronomique et sa hauteur, au nombre de pages. Chaque élément de la visualisation est cliquable et donne accès à la page de l'élément concerné. Enfin, il est possible de zoomer dans la visualisation pour accéder à plus de détail.</p> <p>En cochant une case, l'axe des ordonnées correspond à la numérotation des pages : les cellules se repartissent alors en fonction de leur position dans l'ouvrage, révélant ainsi la composition des différentes œuvres et la répartition des tables.</p>	Données de la visualisation : ORIGINAL ITEM (array) <ul style="list-style-type: none">(originalItem.work = current primary source)- id + originalTextTitle- tpx + tq TABLE TYPE → ASTRONOMICAL OBJECT → objectName <ul style="list-style-type: none">- pageMin + pageMax (→ nombre de pages) WORK → id	Liens vers autres pages BLOC 3 : au clic sur le label des colonnes, redirection vers l'œuvre concernée BLOC 5 : au clic sur les cellules, redirection vers la notice de l' <i>original item</i> concerné BLOC 2 : Historical navigation BLOC 6 : au clic sur un élément de la légende, redirection vers une page de présentation de l'objet astronomique correspondant	Listes <ul style="list-style-type: none">- Sources primaires conservées dans la même bibliothèque- Editions (publiques) de cette source primaire- Sources primaires écrites avec la même écriture- Sources primaires rédigées dans la même langue- Sources primaires conçues dans la même ère géographique (± 1 long, ± 2 lat)- Sources primaires conçues dans la même période temporelle (± 50 tpx, ± 50 tq)
	BLOC 5 Enjeux : <ul style="list-style-type: none">- Mise en avant des métadonnées- Expliciter la nature de l'entité <i>original item</i>- Montrer le statut central de cette donnée : connexion avec de nombreuses parties du <i>front office</i> Description de la page : <p>Page de présentation d'un témoin (<i>original item</i>), entité centrale dans la base de donnée mais plus difficile à appréhender dans sa nature pour des non-connaissants.</p> <p>Item à replacer à la fois au sein des témoins de la même source primaire, au sein des témoins issus de la même œuvre mais peut-être aussi des items similaires (items issus de la même œuvre partageant un <i>paramater set</i>). Il faut faire comprendre que la granularité qui est désignée pour cette notice est celle de la table.</p>	Métadonnées affichées : ORIGINAL TEXT (single record) <ul style="list-style-type: none">- id + originalTextTitle- tpx + tq- comment- public (savoir s'il est possible de montrer la notice dans le <i>front office</i>)- pageMin + pageMax (→ nombre de pages) TABLE TYPE → name <ul style="list-style-type: none">- PLACE → placeName- HIST ACTOR → actorName + actorNameOrigChar + tpx + tq + viaf- WORK → title/incipit + tpx + tq- WORK → HIST ACTOR (array) → actorName + actorNameOrigChar + tpx + tq + viaf- WORK → PLACE → placeName- PRIM SOURCE → shelfmark + digitalIdentifier + primType + primTitle- PRIM SOURCE → LIBRARY → libraryName + city + country + isni- EDITED TEXT (array) → editedTextTitle + date + type- EDITED TEXT (array) → HISTORIAN → firstName + lastName- SCRIPT → scriptName + iso15924- LANGUAGE → languageName + iso6392- createdBy : FOS USER → id + username DEFINITION (single record) <ul style="list-style-type: none">(Id = 11)- longDefinition- userInterfaceColor	VISU 5 Enjeux : <ul style="list-style-type: none">- Monter les déplacements géographiques de l'item de sa conception intellectuelle à son ultime lieu de conservation Description de la visualisation : <p>Visualisation en carte où les entités <i>original item</i>, <i>work</i> et <i>primary source</i> (avec la localisation de la bibliothèque) sont représentées par des points de différentes couleurs. Ces différents points sont cliquables et les métadonnées relatives sont affichées seulement au survol.</p> <p>Possibilité éventuelle d'afficher dans la même couleur que l'<i>original item</i> mais d'une teinte moins soutenue, les autres <i>original items</i> contenus dans la même source primaire. Même chose avec les <i>original items</i> issus de la même œuvre. Ces possibilités d'affichage permettent de représenter le témoin dans sa matérialité (même source primaire) et dans son contexte intellectuel (même œuvre). De telles fonctionnalités permettent de proposer une visualisation en carte pour les entités œuvre et source primaire. (carte proche de la VISU 9)</p>	Données de la visualisation : WORK (single record) <ul style="list-style-type: none">(originalItem.work = this work)- id + title / small incipit- tpx + tq- HIST ACTOR → actorName ORIGINAL ITEM (single record) <ul style="list-style-type: none">(current original item)- id + originalTextTitle- tpx + tq TABLE TYPE → ASTRONOMICAL OBJECT → objectName <ul style="list-style-type: none">- PRIM SOURCE → id- PRIM SOURCE → LIBRARY → libraryName + city + country ORIGINAL ITEM (array) <ul style="list-style-type: none">(originalText.primarySource = same primary source)- id + originalTextTitle- tpx + tq ORIGINAL ITEM (array) <ul style="list-style-type: none">(originalText.work = same work)- id + originalTextTitle- tpx + tq TABLE TYPE → ASTRONOMICAL OBJECT → objectName	Liens vers autres pages BLOC 4 : au clic sur le point correspondant, redirection vers la notice de la source primaire désignée BLOC 3 : au clic sur le point correspondant, redirection vers la notice de l'œuvre concernée BLOC 13 : Lien vers la recherche facetée BLOC 2 : Historical navigation	Listes <ul style="list-style-type: none">- liste des <i>original items</i> de la même source primaire qui sont publics- liste des <i>original items</i> issus de la même œuvre qui sont publics- liste des <i>original items</i> issus de la même œuvre et partageant le même set de paramètres qui sont publics- <i>Original items</i> conçus dans la même ère géographique (± 1 long, ± 2 lat)- <i>Original items</i> conçus dans la même période temporelle (± 50 tpx, ± 50 tq)- <i>Original items</i> du même type de table- <i>Original items</i> du même auteur- <i>Original items</i> du même scribe- Éditions de ces items qui sont publiques- <i>Original items</i> du même type d'écriture- <i>Original items</i> dans la même langue

BLOC 6  Enjeux : <ul style="list-style-type: none"> - Proposer un point d'entrée « astronomique » vers les éditions, les modèles de tables et aussi les témoins - Fournir les textes explicatifs d'introduction aux concepts centraux du projet (qu'est-ce qu'une table ? Modèle / caron ?) Idéalement, présentation des différents « objets astronomiques » - Illustrer l'imbrication des concepts d'objet astronomique, de type de table, de modèles - Figurer tous les objets astronomiques de la base - Donner à voir des représentations médiévales astronomiques - Rappeler la diversité des traditions et conceptions de l'astronomie <p>Description de la page : La page est pensée comme pouvant être une introduction à l'astronomie médiévale ainsi qu'un pendant au BLOC 2 décrivant les données dans leur versant historique. Présence de textes d'introduction aux notions centrales du projet élaborés par des chercheurs avec des « disclaimer » pour expliciter la variété des différentes traditions. La page n'affiche au départ que les différentes images d'objets astronomiques provenant de cultures variées : lorsqu'un objet est sélectionné par l'utilisateur, s'affiche en dessous une série de boutons représentant les différents type de table associés à cet objet. En cliquant sur un type de table en particulier, une liste des éditions, sets de paramètres et modèles liés à ce type de table s'affiche en dessous. Afin de pouvoir opérer une redirection vers la page où un objet astronomique est déjà sélectionné (lorsque l'utilisateur clique sur la légende des VISU 3 et 4), il faut prévoir une URL où l'objet astronomique est renseigné en argument de l'URL.</p>	Métadonnées affichées : ASTRONOMICAL OBJECT (all records) - id + objectName TABLE TYPE (all records) - id + tableTypeName + astronomicalObjectId EDITED TEXT (all records) - id + smallEditedTextTitle - type - tableTypePld - HISTORIAN → firstName + lastName - date - public (Pour les types A : ORIGINAL ITEM → WORK → title) TABLE CONTENT (all records) - id + editedTextId PARAMETER SET → id FORMULA DEFINITION (all records) - id + name tableTypePld PARAMETER VALUE (all records) - id + parameterSetId - typeOfNumber - valueOriginalBase - valueFloat Questions : Est-ce qu'il est pertinent d'afficher des données historiques en tant que paramètre dans la liste des éditions (pour filtrer) sachant que les éditions de type B font référence à plusieurs sources primaires voire même plusieurs œuvres, et que les éditions de type C à aucune de ces entités ? Est-il possible d'afficher le modèle associé à un contenu de table ? Sinon, doit-on implémenter cette fonctionnalité ?	VISU 6 Enjeux : <ul style="list-style-type: none"> - Illustrer avec une iconographie variée les différents objets astronomiques tels que représentés dans les cultures et traditions présentes sur la base. <p>Description de la visualisation : Plutôt qu'une série de boutons pour sélectionner un objet astronomique, des images issues de reproductions de manuscrits ou incunables pourraient donner un aperçu de la diversité représentée sur la plateforme.</p>		Liens vers autres pages BLOC 7 : Modèle de tables BLOC 2 : Historical navigation BLOC 8 : Table edition BLOC 13 : Lien vers la recherche facetTEE BLOC 12 : Lien vers la carte des sets de paramètres	Listes - Liste des <i>astronomical objects</i> - Liste des <i>table types</i> - Liste des <i>Edited texts</i> qui sont publics - Liste des <i>Formula definition</i>
BLOC 7  Enjeux : <ul style="list-style-type: none"> - Expliquer les modèles sous-jacents aux calculs modernes des tables - Fournir du contenu scientifique éditorialisé pour la plateforme - Constituer les premières pierres d'une exposition virtuelle <p>Description de la page : La définition des modèles dans la base de données est déjà stylisée et plus ou moins mise en page : il s'agit donc juste de présenter de manière claire et esthétique les différents éléments qui composent la définition du modèle. Affichage des schémas et formules.</p>	Métadonnées affichées : FORMULA DEFINITION (single record) - id + name imageId explanation modernDefinition bibliography parameterExplanation argNumber latexFormula formulaJSON : « \$p_... » pour récupérer l'id de l'enregistrement de la table parameterFormat PARAMETER FORMAT → id + parameterName TABLE TYPE → id + tableTypeName ASTRONOMICAL OBJECT → id + name authorId : HISTORIAN : id + lastName + firstName createdBy : FOS USER → id + username			Liens vers autres pages BLOC 8 : Table content lié BLOC 6 : Astronomical navigation	Listes - Éditions qui reposent sur ce modèle - Modèles liés au même type de table
BLOC 8  TABLE CONTENT Enjeux : <ul style="list-style-type: none"> (tous les onglets) Présenter les différents aspects inhérents à la nature des éditions de tables Constituer une édition de table citable dans un article Proposer des outils pertinents de visualisations d'apparat critique <p>Description de la page : La visualisation de la table éditée prend l'essentiel de la page : un écart sur le côté permet d'afficher les métadonnées de celle-ci, notamment les métadonnées associées à une cellule précisément en cas de clic de l'utilisateur). La visualisation donne accès également à DTI afin de pouvoir modifier la table présentée. Des fonctionnalités d'exports peuvent être proposées, notamment en format LaTeX et JSON (décomposition des exports entre table, commentaires et métadonnées ?). En cas de plusieurs tables éditées (cas des <i>Mean motions</i>), un dropdown pourraient permettre l'affichage des tables une à une.</p> <p>Chacun des autres aspects de la notice sont présent dans des onglets séparés</p>	Métadonnées affichées : TABLE CONTENT (single record) - tableTypePld - argumentName public - editedTextId - valueOriginal jsonOriginal etc. - createdBy : FOS USER → id + username	VISU 8 Enjeux : <ul style="list-style-type: none"> - Figurer l'apparat critique de manière intuitive et claire d'utilisation - Proposer quelques fonctionnalités pour rendre la lecture de la table plus aisée <p>Description de la visualisation : La table est représentée dans un tableau (présence ou non d'une grille extérieure permettant de localiser une case comme pourraient le faire un document excel) où les valeurs et arguments sont représentés : ajout d'un coloration plus ou moins intense (<i>heat map</i>) en fonction du nombre de variantes présentes dans telle ou telle cellule. Un contour de couleur ou autre système de signalisation peut quant à lui représenter la présence d'un commentaire. Un panel sur le côté de la table permet d'afficher à la fois les commentaires et l'apparat critique lié à la case qui a été cliquée. On peut penser à la possibilité d'afficher également l'intégralité des variantes et des commentaires (alors nécessité de localisation des cases grâce à une grille) : les valeurs pour la case sélectionnée pourraient apparaître en surbrillance. La question subsiste de l'affichage d'un graphique pour représenter graphiquement la table : toutefois, cette fonctionnalité serait de toute manière présente dans DTI</p>	Données de la visualisation : TABLE CONTENT (single record) tableTypePld argumentName public editedTextId valueOriginal jsonOriginal etc. - createdBy : FOS USER → id + username	Liens vers autres pages BLOC 6 : Astronomical navigation BLOC 8/11 : Notice de l'édition d'une table BLOC 5 : original items de la présente édition	Listes - Éditions du même type de table

E T D A B T L I E O N	BLOC 9 HISTORICAL CONTEXT	Métadonnées affichées : EDITED TEXT (single record) - id + editedTextTitle ORIGINAL TEXT (array) - id + originalTextTitle - tpq + taq - comment - PRIMARY SOURCE → id + primType + shelfmark + LIBRARY (→ name + city) + primType + primEditor - WORK → id + title/incipit + tpq + taq + HIST ACTOR (- actorName + tpq + tag) + PLACE (- id + placeName) - PLACE → id + placeName - HIST ACTOR → actorName + tpq + taq - SCRIPT → scriptName - LANGUAGE → languageName	VISU 9 Enjeux : - Figurer la provenance historique et géographique de l'édition. Description de la visualisation : Visualisation proche de la VISU 5 : carte des <i>original items</i> édités, de l'œuvre associée et des bibliothèques de conservation de ces items.	Données de la visualisation : EDITED TEXT (single record) - id ORIGINAL TEXT (array) (editedText.itemId = those original items) - id + originalTextTitle - tpq + taq - PLACE → id + placeName + lat + long PRIMAR Y SOURCE (array) (itemId.primarySource = those primary sources) - id + primType + shelfmark - LIBRARY → name + city WORK (single record) (origItem.work = this work) - id + title/incipit + tpq + taq - PLACE → id + placeName + lat + long	Liens vers autres pages BLOC 6 : Astronomical navigation BLOC 8/11 : Notice de l'édition d'une table BLOC 5 : original items édités BLOC 2 : Historical navigation BLOC 3 : œuvre BLOC 4 : source primaire	Listes - Liste des éditions du même manuscrit - Liste des éditions du même work - Liste des éditions du même <i>original item</i> - Listes des éditions issue de l'œuvre du même acteur historique
	BLOC 10 MATHEMATICAL CONTEXT	Métadonnées affichées : TABLE CONTENT (single record) - TABLE TYPE → id + tableName + ASTRO OBJECT → id + name - mathematicalParameterId - entryTypeOfNumberId, argument1TypeOfNumberId, ... : TYPE OF NUMBER → id + typeName + typeDefinition - entryNumberUnitId, argument1NumberUnitId, ... : NUMBER UNIT → id + unit - FORMULA DEFINITION → id + name + latexFormula createdBy : FOS USER → id + username - PARAMETER SET (array) → id MATHEMATICAL PARAMETER (single record) id + argNumber + entryShift - typeOfParameter - argument1DisplacementOriginalBase - entryDisplacementOriginalBase - argument1Shift, argument2Shift, ... PARAMETER VALUE (array) (parameterValue,parameterSetId tableContent.parameterSetId) - id + valueOriginalBase - TYPE OF NUMBER → typeName - PARAMETER FORMAT → id + parameterUnitId + parameterName parameterName CALENDAR (single record) id + name + monthList - ERA → id + name + epoch DEFINITION (single record) (id = 23) - longDefinition - userInterfaceColor	VISU 10 Enjeux : - Donner un schéma récapitulatif des paramètres de la tables Description de la visualisation : Deux types de visualisations peuvent être envisagées : soit une visualisation en carte à la manière de la VISU 12 (ne mettant en scène que le set de paramètres de la table éditée), soit une visualisation synthétisant tous les paramètres avec leur type de nombre et leur métadonnées respectives, ce qui pourrait entraîner des redondances avec les données affichées sur la partie notice. La visualisation pourrait constituer l'encart concernant les paramètres astronomiques : c'est-à-dire à la fois le <i>parameter set</i> , les <i>parameter values</i> associées et leurs <i>parameter formats</i> liés. Il pourrait y avoir autant d'encarts que de sets de paramètres associés à l'édition : leur contenu serait calqué sur la mise en page dans le <i>back office</i> (elle-même reflétant le formulaire pour rentrer un <i>parameter set</i>). Des couleurs et définitions pourraient être ajoutées pour clarifier et rappeler le code couleur pour les objets astronomiques repris dans le <i>front office</i> .	Données de la visualisation : TABLE CONTENT (single record) - PARAMETER SET (array) → id + name + tableTypeId TABLE TYPE (single record) - id + tableName + ASTRONOMICAL OBJECT → id + name PARAMETER VALUE (array) (parameterValue,parameterSetId tableContent.parameterSetId) - id + valueOriginalBase - TYPE OF NUMBER → typeName - PARAMETER FORMAT → id + parameterUnitId + parameterName parameterDefinitionId + formulaDefinitionId	Liens vers autres pages BLOC 6 : Astronomical navigation BLOC 8/11 : Notice de l'édition d'une table	Listes - Éditions partageant le même set de paramètres (redirection vers la VISU 12) - Liste des éditions du même type de table - Liste des éditions utilisant le même paramètre mathématique - Liste d'édition utilisant le même type de nombre
	BLOC 11 EDITORIAL CONTEXT	Métadonnées affichées : EDITED TEXT (single record) - id + date + editedTextTitle + type + created - public - comment + onlineResource + pageRange - secondarySourceId - HISTORIAN → id + firstName + lastName createdBy : FOS USER → id + username SECONDARY SOURCE (single record) secType + secTitle + secIdentifier + secPubDate secPageRange + secPublisher + secVolume - secOnlineIdentifier -journalId + collectiveBookId : JOURNAL → journalTitle DEFINITION (single record) (id = 3) - longDefinition - userInterfaceColor	VISU 11 Enjeux : - Montrer les différentes sources de la présente édition Description de la visualisation : La visualisation reprendrait la le graphe des sources de l'édition disponible dans la partie admin pour les notices des <i>Edited texts</i> . La présente édition se trouve au milieu du graphe, point vers lequel pointent des flèches partant des différentes sources de l'édition : ainsi une édition de type B sera liée à tous les <i>original items</i> et éditions sur lesquelles elle s'appuie, qui elles-mêmes seront liées aux <i>original items</i> desquels sont tirés la transcription. Des différences de coloris pourront indiquer le type d'édition dont il est question, de la même manière, il est possible de rendre les différents éléments cliquables, permettant la redirection vers les notices des items concernés (pose la question de la révélation des liens entre les objets et du déplacement des différents éléments).	Données de la visualisation : EDITED TEXT (single record) - id + smallEditedTextTitle + type + date - HISTORIAN → id + firstName + LastName EDITED TEXT → (même métadonnées) - originalTextId ORIGINAL ITEM (array) (originalText.id = editedText.originalTextId) - id + tpq+tag + smallTextTitle WORK → id + title/incipit + historicalActorId HISTORICAL ACTOR (array) (historicalActor.id originalText.work.historicalActorId) - actorName + tpq + taq	Liens vers autres pages BLOC 8/11 : Notice de l'édition d'une table BLOC 6 : Astronomical navigation BLOC 5 : original items édités	Listes - listes des éditions d'un des mêmes <i>original items</i> - Listes des éditions du même auteur intellectuel - Liste des éditions provenant du même journal

P A R A M E T E R	BLOC 12 Enjeux : <ul style="list-style-type: none">- Fournir une visualisation instructive et claire croisant de nombreuses données de la base- Illustrer concrètement les enjeux relatifs à la donnée de paramètre set- Page de résultats conçue comme étant la page de redirection des listes des différents blocs Description de la page : <p>La visualisation en carte prendrait l'essentiel de la page : en dessous, plusieurs outils pour filtrer les résultats et affiner les enregistrements pourraient être mis à disposition. En outre, une liste des résultats figurant sur la carte pourraient permettre d'accéder aux notices des éléments concernés : idéalement, cette liste serait mise à jour par rapport aux informations présentées dans la visualisation.</p> <p>Cette visualisation serait accessible depuis la page <i>Astronomical navigation</i> mais également comme vue indépendante sur laquelle il serait possible d'afficher des données déjà filtrée selon un certain nombre de paramètres.</p> <p>Question : les éléments affichés sur la carte correspondent-ils à une liste d'<i>edited texts</i> ou bien une liste d'<i>original items</i> sachant l'entité représentée sur la visualisation est en réalité hybride ?</p>	Métadonnées affichées : EDITED TEXT (array) (<i>editedText.tableContent.parameterSet = parameter set affiché sur la carte</i>) - id + date + editedTextTitle + type - public - HISTORIAN → id + firstName + lastName ORIGINAL TEXT (array) (<i>originalText.id = those editedText.originalTextId</i>) - id + originalTextTitle - tpq + taq - WORK → title/incipit - WORK → HIST ACTOR → actorName	VISU 12 Enjeux : <ul style="list-style-type: none">- Illustrer le déplacement des paramètres astronomiques dans le temps et l'espace : figurer des tendances dans la circulation des savoirs astronomiques dans le temps Description de la visualisation : <p>Carte où les sets de paramètres seraient représentés à l'endroit de conception des <i>original items</i> associés aux éditions correspondant aux sets de paramètres sus-mentionnés. Chaque set de paramètre serait symbolisé par un point dont la taille varierait en fonction du nombre d'éditions l'utilisant. Sa localisation serait déterminée donc par l'<i>original item</i> / les <i>original items</i> (ce qui suppose alors plusieurs points) liés à l'édition. Un seul paramètre pourrait être affiché à la fois pour des questions de visibilité et de lourdeur de chargement : l'utilisateur pourrait sélectionner un autre set de paramètres grâce à un dropdown, ou un système de cases à cocher dans la liste au-dessous. Par surcroît, il est possible d'envisager que la sélection d'un objet astronomique ou d'un type de table pourrait circonscire également les données affichées.</p> <p>Enfin, des curseurs placés sur une frise chronologique permettraient de sélectionner une période de temps pour restreindre l'affichage des données sur la carte.</p> Problème : <p>Comment faire figurer les éditions de type C qui ne sont pas associées à aucun <i>original item</i> c'est-à-dire à aucun lieu ? Pas forcément pertinent comme donnée pour illustrer la diffusion des savoirs.</p> Question : <p>Est-ce une visualisation pertinente pour la partie de la navigation plus axée astronomique sachant qu'elle révèle des enjeux davantage historiques ?</p>	Informations à faire figurer : Valeur du paramètre : PARAMETER SET (all records) - id PARAMETER VALUE (all records) - id + parameterSetId - typeOfNumber - valueOriginalBase Nombre d'enregistrement (taille des points) : TABLE CONTENT (array) (<i>table.content.parameterSet = parameterSet.id</i>) - id + editedText EDITED TEXT (array) (<i>editedText.id = tableContent.id</i>) - id + smallEditedTextTitle - tableTypeid (couleur du point) - ORIGINAL TEXT (array) → id Lieu : ORIGINAL TEXT (array) (<i>originalText.id = editedText.originalText.id</i>) - PLACE → placeLat + placeLong Date : ORIGINAL TEXT (array) (<i>originalText.id = editedText.originalText.id</i>) - tpq + taq	Liens vers autres pages BLOC 5 : original items BLOC 6 : Astronomical navigation BLOC 8 : édition de table	Listes <ul style="list-style-type: none">- Liste d'éditions utilisant un même set de paramètres- Liste des <i>original items</i> issus de ces éditions
A R D E V S A C N C E D H	BLOC 13 Enjeux : <ul style="list-style-type: none">- Proposer une grande finesse et souplesse dans les recherches faites à la base de données : possibilité de croiser de nombreux paramètres- Proposer une interface intuitive et facile à appréhender- Page de résultats conçue comme étant la page de redirection des listes des différents blocs Description de la page : <p>Une barrière en haut de la page donnerait accès à une barre de recherche plein texte : mise en avant de cette seule fonctionnalité souvent seule à être employée par les utilisateurs.</p> <p>En dessous, différents filtres peuvent être proposés : la sélection d'un de ces filtres serait rappelée grâce à un signet que l'utilisateur pourrait supprimer pour élargir à nouveau son champ de recherche. Des fonctionnalités d'auto complétion pourraient être mises en place pour faciliter la saisie.</p> <p>La page de recherche pourrait permettre de visualiser toutes les listes de <i>related contents</i> accessibles depuis les pages d'enregistrements</p> Exemples de pages de recherche facetée : <ul style="list-style-type: none">- Biblissima, Isidore, Gallica, sites marchands, europeana, etc. Questions : <ul style="list-style-type: none">- les facettes permettant de filtrer la recherche sont elles à afficher une fois qu'un type d'objet (œuvre, source primaire, item original et édition de table) a été sélectionné ou dès le départ et être davantage standardisées (date / lieu / objet astronomique / auteur ?) ? qu'est-ce qui est le plus intuitif d'utilisation et le moins susceptible d'induire en erreur ? Il serait par exemple utile de filtrer des résultats sans pour autant vouloir sélectionner un type d'objet.- les objets astronomiques sont une manière de trier les types de tables, toutefois, d'autres catégorisations pourraient être utiles à proposer, comme la possibilité de trier par table de latitude. Faut-il créer un filtre correspondant à cela ou ce genre de requête pourrait être mise en place grâce à la recherche plein texte ?- Tri par type d'objet : doit-on envisager de donner des résultats pour les types d'objets « secondaires » (acteur historique / set de paramètre / lieu) ou seules les entités disposant de notices dans le front office peuvent être l'objet d'un résultat de recherche ?	Métadonnées affichées : Possibilité de tri : <ul style="list-style-type: none">- date- lieu (comment déterminer des zones ? La délimitation en terme de pays peut être arbitraire et anachronique)- objet astronomique- auteur (auteur de l'œuvre ? Scribe ? Auteur de l'édition ? → polysémie du terme qui peut être trompeuse)- langue- type de nombre- bibliothèque- date de mise en ligne- manuscrit / imprimé- type de table- journal	VISU 13 On peut imaginer pouvoir trier par journal par exemple et proposer des réponses qui seraient en réalité des entités qui auraient été éditées dans ce journal, par exemple, une œuvre, dont un des <i>original items</i> auraient été éditées et publiées dans ce même journal. Cela en revanche serait assez lourd et créerait sans doute plus de bruit que cela n'aiderait l'utilisateur dans sa recherche. Il est à l'inverse possible de trier sur des champs qui ne seraient spécifiques qu'à certaines entités : la sélection d'un journal en particulier par exemple, indiquerait que les résultats de la recherche ne serait que des éditions. En outre, une manière de mettre en avant les données présentes sur la plateforme peut consister à ajouter sous la barre de recherche, une mention de type « Rechercher parmi les N notices, les N enregistrements, etc. »	Données de la visualisation :	Liens vers autres pages BLOC 2 : Historical navigation BLOC 6 : Astronomical navigation	

 D A T A B A S E	<p>BLOC 14</p> <p>Enjeux :</p> <ul style="list-style-type: none"> - Figurer le modèle conceptuel de la base de donnée - Révéler des corrélations entre les entités de la base - Constituer une vitrine pour le projet <p>Description de la page : La visualisation, pour des questions de lisibilité et de compréhension, doit être accompagnée d'un schéma. Ce schéma peut être pensé comme suffisamment explicite et placé comme image de fond de la visualisation en graphe, soit peut être présenté en prélude à la visualisation et assorti d'un texte explicatif ou d'animations aidant à la compréhension du modèle de donné. Des pictogrammes pourront être réalisés pour apporter un support visuel à la compréhension, notamment au sein du graphe pour figurer les différentes entités.</p>		<p>VISU 14</p> <p>Enjeux :</p> <ul style="list-style-type: none"> - Donner à voir la structure de la base de données - Mieux comprendre le modèle conceptuel sous-jacent prévu pour accueillir les tables <p>Description de la visualisation : La visualisation en graphe est composée de points représentant les enregistrements, et de lignes reliant les enregistrements, symbolisant ainsi les liens entre les données dans la base. Les points peuvent soit correspondre à un seul enregistrement, soit symboliser l'ensemble des enregistrements d'une entité (alors la taille du point sera relative à nombre d'enregistrements). De la même manière, il peut être choisi de ne représenter que les entités principales de la base de données ou alors de tout représenter.</p> <p>Questions : Quelles informations faut-il faire figurer ? Doit-on faire apparaître un label pour chaque enregistrement (visible au survol ?) ? Comment mettre en avant les liens entre entités (en cliquant sur un point pour révéler les lignes de relations ?) ?</p> <p>Outils : Gephi, Kibana, Sigma</p>	<p>Données de la visualisation :</p>	<p>Liens vers autres pages</p> <ul style="list-style-type: none"> - Liens vers les notices d'enregistrements ou les pages de recherche concernant une entité en particulier 	
 A B O U T	<p>BLOC 15</p> <p>Pages :</p> <ul style="list-style-type: none"> - About : texte introductif + liens vers sites de projets pour les actualités - Team → Partenaires + contact - Ressources → Glossaire + bibliographie + téléchargements <p>Enjeux :</p> <ul style="list-style-type: none"> - Fournir toute la documentation parallèle au projet - Accueillir les documents ayant trait au projet sans avoir d'espaces dédiés à leur présentation - Constituer une interface de contact pour les contributeurs du projet 	<p>Métadonnées affichées :</p> <p>GLOSSAIRE DEFINITION (all records) - objectUserName - longDefinition - userInterfaceColor</p> <p>ASTRONOMICAL OBJECT (all records) - id + name => affichage des couleurs</p> <p>DOWNLOADS PDF FILE (all records) - fileName + userFileName</p>				
 D T I	<p>BLOC 16</p> <p>Enjeux :</p> <ul style="list-style-type: none"> - Fournir un outil sur mesure pour la saisie de table - Faciliter la saisie, le calcul, la production automatisée d'appartat critique, la transposition dans différents systèmes numériques 					

Annexe F

Exemples de requêtes avec ElasticSearch

F.1 ElasticSearch query language

ElasticSearch has its own query language, more documentation can be found [here](#)

F.1.1 Anatomy of a query

In most cases, an ElasticSearch query is composed of :

- a **header** defining :
 - which **method** is to be used : `GET`, `POST`, etc.
 - which **index** (i.e. which entity of the database written in snake_case) is going to be queried. Not necessary if all indexes will be queried.
 - what **type** of search is to be made (`_search` most of the time)
- a **body** defining (among other things) :
 - the **fields** that are going to appear in the results (`_source`)
 - the **filters** that are going to narrow the number of results matching those filters (`query`)
 - different properties of the results (size of the results, index from which to begin, etc.)

Tips & tricks The body of a query needs to be a correctly formatted JSON string, even using single quotes instead of double is considered to be an error. The `dev tools` tab in Kibana interface offers help for automatic indentation and autocompletion features that can be very handy.

F.1.2 Get all records

To get all the records of the entire database :

```
1 GET _search
```

To retrieve all records from an index (primary source in this case) :

```
1 GET primary_source/_search
```

Which is equivalent to :

```
1 GET primary_source/_search
2 {
3     "query": {
4         "match_all": {}
5     }
6 }
```

F.1.3 Simple matching query

Exact term match

All the works that contain the string “tabule” in their title.

Note that the case of the letters doesn’t matter, it will match “Tabule” as well. In addition to that, in this configuration, a sub-string will not match the same as the entire string (“tabu” will not match “tabule”) ; the string is treated as a complete word.

```
1 GET work/_search
2 {
3     "query": {
4         "match": {
5             "title": "tabule"
6         }
7     }
8 }
```

All the original items that are associated with a primary source that is kept in the library that have 2 as id.

```
1 GET original_text/_search
2 {
3     "query": {
4         "match": {
5             "primary_source.library.id": "2"
6         }
7     }
8 }
```

```

7     }
8 }
```

Multiple terms match

All original items that have in their title either the word “solis”, either the word “lune”.

In this configuration, each string separated by a space is treated individually and the operator used to connect them is **OR**. In other words, the more you put terms, the more you will match original items.

```

1 GET original_text/_search
2 {
3   "query": {
4     "match": {
5       "original_text_title": "solis lune"
6     }
7   }
8 }
```

To get a response where each terms given are independently going to filter the result (same kind of behavior as a Google query), you need to specifies the operator to be **AND**.

```

1 GET original_text/_search
2 {
3   "query": {
4     "match": {
5       "original_text_title": {
6         "query": "lune solis",
7         "operator": "AND"
8       }
9     }
10   }
11 }
```

F.1.4 Adding some margin of error

Fuzziness The fuzziness allows a certain amount of inaccuracy to be accepted.

All library that approximately have the string “nationale” in their name.

```

1 {
2   "query": {
3     "match": {
4       "library_name": {
5         "query": "nationale",
6         "fuzziness": "auto"
7       }
8     }
9   }
10 }
```

```

7         }
8     }
9 }
10 }
```

You can set the fuzziness to 1 or more but the `auto` settings allows a number of letters that do not match, proportional to the length of the term to be searched.

Full text search on an index

To allow search on every field of an entity, the query has to be set to `multi_match`.

All edited texts that have approximately the string “lune” in one of them fields.

```

1 GET edited_text/_search
2 {
3   "query": {
4     "multi_match": {
5       "query": "lune",
6       "fuzziness": "auto"
7     }
8   }
9 }
```

As is, those kind of requests are deprecated because no fields are specified : ElasticSearch encourages to list the fields you want the query to be executed on. The `fields` allows to reduce noise in the results and to take less time.

All primary sources that match approximately the strings “vatican” and “latin” in the list of fields specified.

```

1 GET primary_source/_search
2 {
3   "query": {
4     "multi_match": {
5       "query": "vatican latin",
6       "fuzziness": "auto",
7       "operator": "and",
8       "fields": [
9         "shelfmark",
10        "digital_identifier",
11        "kibana_name",
12        "tpq.keyword",
13        "taq.keyword",
14        "prim_type",
15        "library.kibana_name",
16        "original_texts.kibana_name",
17        "original_texts.table_type.kibana_name",
18        "original_texts.place.kibana_name",
19        "original_texts.historical_actor.kibana_name",
20        "original_texts.script.script_name",
21        "original_texts.language.language_name"
22      ]
23    }
24  }
```

```

22     ]
23   }
24 }
25 }
```

Notice that on the fields `tpq` and `taq` that are typed as integer, a string query cannot be performed. In order to query those fields as well, you must add `.keyword` after the field name : it corresponds to the field but typed as a string.

Wildcards To find some more documentation for wildcard queries, click [here](#).

F.1.5 Defining the source

If you are not interested in all the metadata (i.e. the content of the fields) associated with the entity you want to query, it is possible to set a list of fields that are going to appear in the response.

Only the shelfmark, and the library name of all primary sources that are a manuscript

```

1 GET primary_source/_search
2 {
3   "_source": [
4     "shelfmark",
5     "library.library_name"
6   ],
7   "query": {
8     "match": {
9       "prim_type": "ms"
10      }
11    }
12 }
```

Note that if a record in the result do not have some information you asked for (let's say, the primary source isn't associated with a library, thus doesn't have a `library.library_name`), the result object will not have the key for this precise field. Instead of looking like that :

```

1 "_source" : {
2   "library" : {
3     "library_name" : "Vatican Library"
4   },
5   "shelfmark" : "Vat. Pal. Lat. 1376"
6 }
```

It will look like :

```

1  "_source" : {
2      "shelfmark" : "Vat. Pal. Lat. 1376"
3  }

```

F.1.6 Special queries

Range queries

Range queries can be made on fields that are numbers (even if the field is typed as a string but contains integer) in the Kibana interface, but does seem to only work on integer/float/date typed field when using ajax.

All the primary source that have an edition date between 1400 and 1500

```

1 GET primary_source/_search
2 {
3     "query": {
4         "range": {
5             "date": {
6                 "gte": 1400, // greater than
7                 "lte": 1500 // less than
8             }
9         }
10    }
11 }

```

If you want to make a range query on a date typed field (fields that end with `_date`, you can use some ElasticSearch tools for date math (those as well, seems to cause problem when used with ajax) :

All original items that have been created between 1000 years before today and 25 years after 1500

```

1 GET original_text/_search
2 {
3     "query": {
4         "bool": {
5             "must": [
6                 {
7                     "range": {
8                         "tpq_date": {
9                             "gte": "now-1000y"
10                        }
11                   }
12                 },
13                 {
14                     "range": {
15                         "taq_date": {
16                             "lte": "1500-01-01||+25y"
17                         }
18                   }
19                 }
20             ]
21         }
22     }
23 }

```

```

17         }
18     }
19   }
20 ]
21 }
22 }
23 }
```

Geo-distance queries

The fields named `location` holds information that is treated as geo point by ElasticSearch : `geo_distance` queries can be executed on them.

All works that have been conceived around 100km from 48 of latitude and 2 of longitude. **NB** : in this syntax, `longitude` comes before `latitude`.

```

1 GET work/_search
2 {
3   "query": {
4     "geo_distance": {
5       "distance": "100km",
6       "place.location": [2,48]
7     }
8   }
9 }
```

The same query can be formulated more explicitly with this syntax :

```

1 GET work/_search
2 {
3   "query": {
4     "geo_distance": {
5       "distance": "100km",
6       "place.location": {
7         "lat" : 48,
8         "lon" : 2
9       }
10     }
11   }
12 }
```

F.1.7 Combining multiple clauses

Every filter you want to combine to build a query can be add with this kind of structure :

```

1 {
2   "query": {
```

```

3     "bool": {
4         "must": [
5             {
6                 // filter one
7             },
8             {
9                 // filter two
10            }
11        ]
12    }
13 }
14 }
```

Putting all together

The shelfmarks of all early printed primary sources that contains an original item that were created near Paris (lat : 48, long : 2).

```

1 GET original_text/_search
2 {
3     "_source": [
4         "primary_source.shelfmark"
5     ],
6     "query": {
7         "bool": {
8             "must": [
9                 {
10                     "geo_distance": {
11                         "distance": "100km",
12                         "place.location": [2,48]
13                     }
14                 },
15                 {
16                     "match": {
17                         "primary_source.prim_type": "ep"
18                     }
19                 }
20             ]
21         }
22     }
23 }
```

Annexe G

Description des développements réalisés pendant le stage

The “Segolene-front-office” branch mostly concerns the historical part of the navigation in the front office. Several modules where developed :

- Data visualisations ;
- Sidebar for metadata ;
- Search interface ;
- Bread crumbs ;
- Related table editions.

NB : the presented functions are here often simplified for demonstration purposes comparing to the way they are integrated into the code.

G.1 Data visualisation

G.1.1 Visualisation description

Three different visualisations were developed for the historical navigation with the AmCharts library :

- **Column chart** : visualisation for the work record page. Each column represents a primary source divided in original items (the cells) which colors are determined by their astronomical object ; the visualisation allows to show completeness and content of the sources derived from a intellectual work.

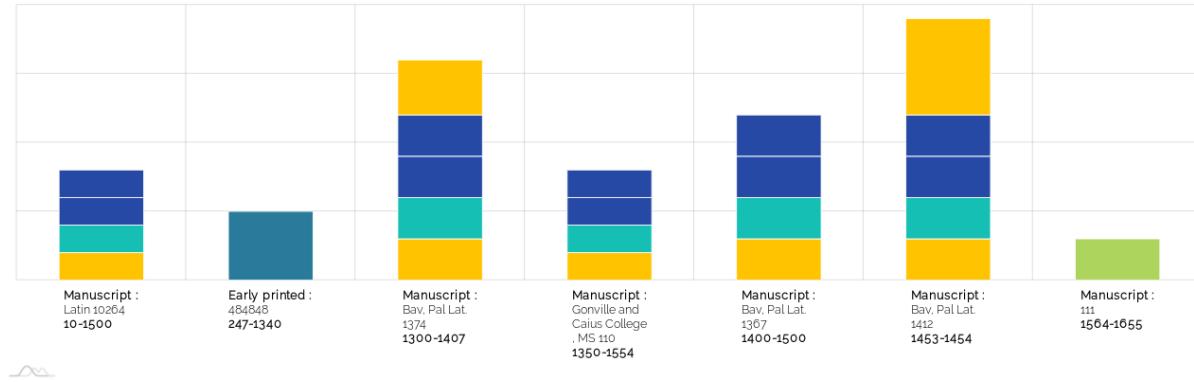


FIGURE G.1 – Column chart for work visualisation

- Bar chart : visualisation for the primary source record page. Each bar represents a work that is contained in the primary source, each bar being divided in original items originated from the work. A select allows two switch between two states of the visualisation : one where all pages of the items are added together, one where they are disposed accordingly to their position in the source. The visualisation thus can show the composite aspect of some primary sources.

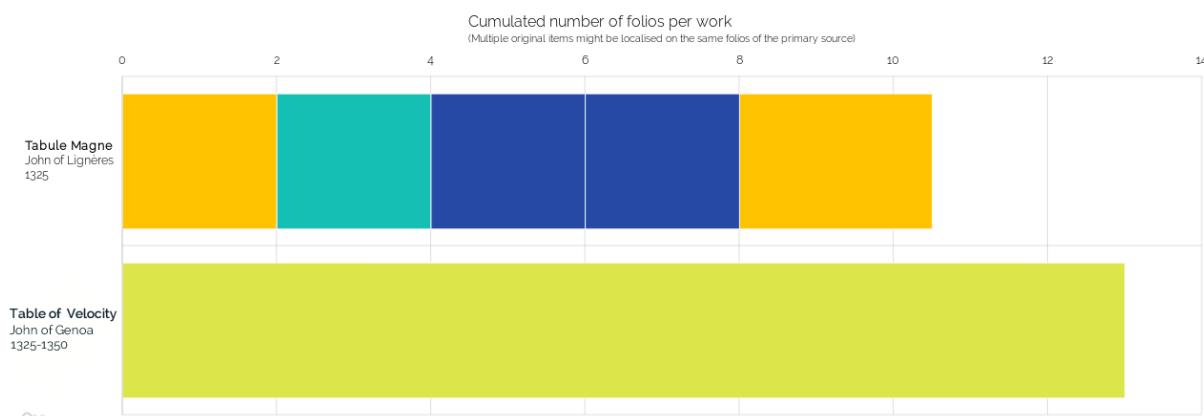


FIGURE G.2 – Bar chart for primary source visualisation : original items stacked

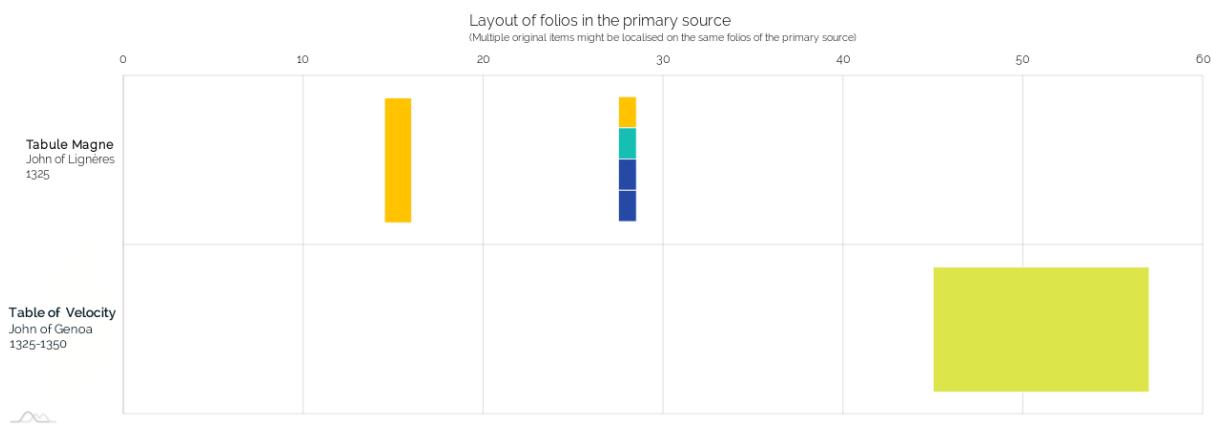


FIGURE G.3 – Bar chart for primary source visualisation : original items spread out

- **Historical map** : visualisation for the original item records, the historical navigation and more to come. The map displays points that represent the place of creation of works (in yellow), primary sources/original items (in red) and combination of both (in orange). A heatmap underneath represents the time axis and show what period were more prolific : the scrollbar allows the user to select a time period so that only the items created during the timerange are shown.

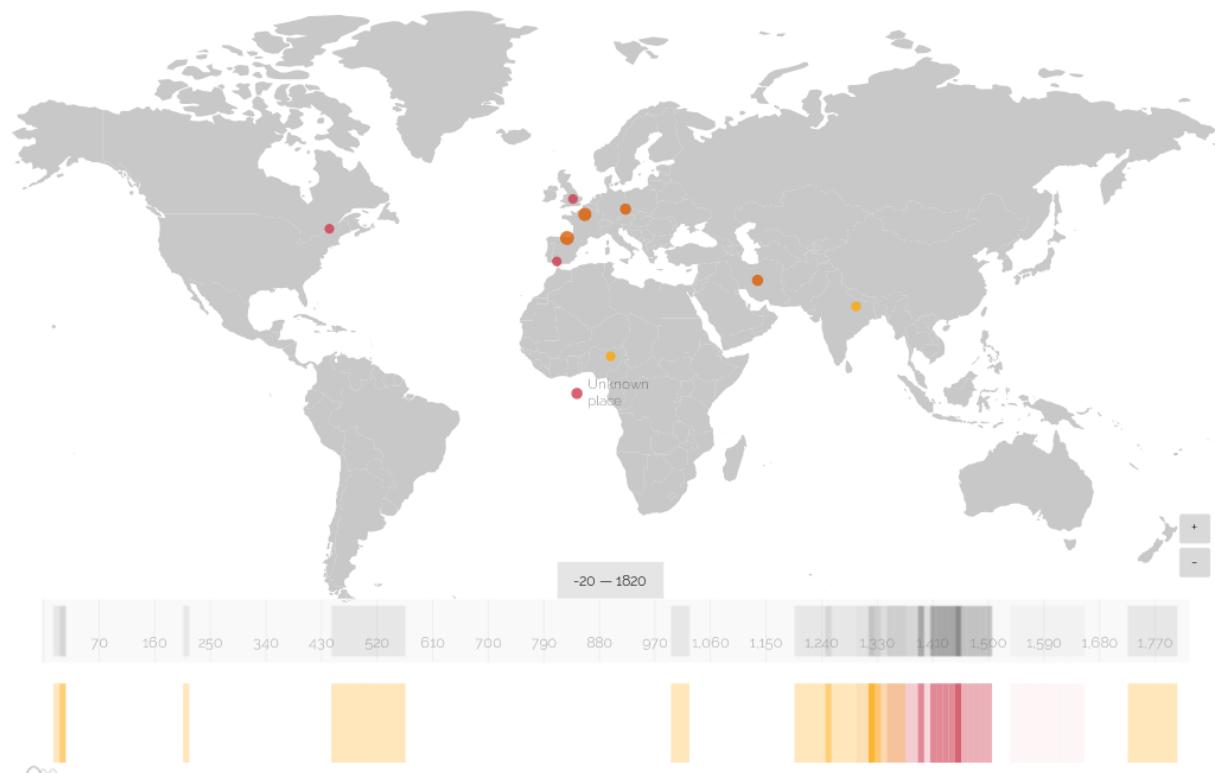


FIGURE G.4 – Map for historical navigation

G.1.2 Function explanation

Historical map

This function is triggered when the user change the range selected by the scrollbar. What it does is that each time the start and the end grips of the scrollbar are both outside (by being under or above it) of the `tpq` and `tap` dates of an item, the item is not shown on the map. When the grips are not both outside, the size, color, tooltip of the map point are set accordingly to the items that were created in the timerange selected.

```

1  scrollbarX.events.on("rangechanged", function () {
2      let cursorMin = scrollbarX.range.start;
3      let cursorMax = scrollbarX.range.end;
4
5      // Conversion of the min and max range value into date
6      let dateMinRange = parseInt((cursorMin / yearRange) + startBar);
7      let dateMaxRange = parseInt((cursorMax / yearRange) + startBar);
8      // Show the timerange selected
9      scrollbarX.startGrip.tooltipText = `${dateMinRange}`;
10     scrollbarX.endGrip.tooltipText = `${dateMaxRange}`;
11     timeframeLabel.text = `${dateMinRange} - ${dateMaxRange}`;
12
13     i = 0;
14     // for each place in mapData
15     for (let index in mapData) {
16         // defining appearance of the circle localised on the place
17         let place = mapData[index];
18         let opacity = 0; // if no item is to be displayed, opacity to 0
19         let countW = 0; // count of the number of works to be displayed : color and radius of
the circle
20         let countPS = 0; // count of the number of primary sources to be displayed : color and
radius
21         let number = 0;
22
23         for (let j = place.items.length - 1; j >= 0; j--) {
24             let dateMin = (place.items[j].from - startBar) * yearRange;
25             let dateMax = (place.items[j].to - startBar) * yearRange;
26             if (!(cursorMin > dateMax && cursorMax > dateMax) || (cursorMin < dateMin &&
cursorMax < dateMin))
27                 // when the two cursors are not both outside of the time frame selected
28                 // either by being under or above it
29                 {
30                     number++;
31                     opacity = 0.8; // set the opacity in order to not be transparent
32                     if (place.items[j].entity === "work") {
33                         countW += 1; // add one to the count of works
34                     } else {
35                         countPS += 1;
36                     }
37                 }
38             }
39
40             let Wtooltip = countW !== 0 ? "[bold]Work[/]\n" + countW + " record(s)" : "";
41             let PStooltip = countPS !== 0 ? "[bold]Primary source[/]\n" + countPS + " record(s)"
: "";
42

```

```

43     let tooltip = "[font-size:18px]" + place.place + "[/]\n" + Wtooltip + PStooltip;
44     number = number !== 0 ? number * 100 : 100;
45
46     placeMarker[i].fillOpacity = opacity;
47     placeMarker[i].radius = Math.log(number);
48     placeMarker[i].tooltipText = tooltip;
49     i++;
50 }
51 })();

```

Bar chart

This function is triggered when the user selects one or another option. If the option is “page disposition”, the height and displacement of the cell are determined in relation to the number of items in the same bound as theirs. In other words, the more items there are on the same pages, the thinner and the more deported from the center will be the cell representing one item.

```

1  function selectView(select) {
2      if (select === "spread") {
3          chart.data = spreadOutSet; // make the chart use the "spread out" data set
4          let numberOfWorks = Object.keys(worksBounds).length;
5          let heightWork = 292/numberOfWorks; // height of the horizontal part dedicated to one
6          work
7          let heightBar = 225/numberOfWorks; // height of the bar in itself
8
8          for (let work of Object.keys(worksBounds)){
9              bounds = worksBounds[work];
10             for (key of Object.keys(bounds)){
11                 i = 0;
12                 let bound = bounds[key];
13                 let numberOrigItems = bound.origItems.length;
14                 for (let id of bound.origItems) {
15                     let heightItem = heightWork/numberOrigItems;
16                     // setting the height of the cell
17                     series[id].columns.template.height = heightItem;
18                     // configuring the displacement of the cell comparing to the center
19                     if (heightItem !== heightWork){
20                         series[id].columns.template.dy = -(heightBar/2) + (heightItem * i);
21                     }
22                     i++;
23                 }
24             }
25         }
26     } else if (select === "stack") {
27         chart.data = stackedSet; // make the chart use the "stacked" data set
28         for (id of ids){
29             series[id].columns.template.dy = 0;
30             series[id].columns.template.height = am4core.percent(80);
31         }
32     }
33 }

```

After creating an object detailing how the different items are disposed in the primary source (`workBounds = {"workTitle1" : [{bound : [min, max], origItems : [id1, id2]}, {bound : [min, max], origItems : [id3]}], "workTitle2" : [...]}`), some items being on the same pages, each series of the chart (i.e. each cell) is individually styled according to how the items are disposed.

G.2 Sidebar of metadata

The sidebar is a template that is integrated to the record pages of work, primary source and original item. The template is filled with an associative array filled with metadata of the record that is being displayed : this associative array is build thanks to the `getMetadataTable()` method located in each repository of the concerned entities.

General information		
<i>Tabule Magne</i>		
Incipit	Multiplicis philosophie	
Creator	John of Lignères (1290–1350)	
Date of conception	1325	
Place of conception	Paris, France	
Author of the record	Matthieu Husson	

FIGURE G.5 – Metadata display in the sidebar

G.2.1 Data structure

The data produced by the `getMetadataTable()` method is always structured the same way in order to correctly generate the sidebar using the template `sideMetadata.html.twig`.

```
1 $metadata = ["entity" => $entityName // index name that will be used to construct an
   Elasticsearch query
```

```

2   "title" => $title, // text to display in the upper part of the sidebar
3   ("subtitle" => $subtitle), // facultative
4   "field name" => ["values" => ["text to display"],
5     "search" => ["json" => ["query to send to ElasticSearch"],
6       "hover" => "text to display when hovering the magnifier"],
7       "title" => ["title of the query"]],
8   "other field" => ["values" => [[["html" => "text to display", // when text must be a
link
9     "id" => "id of the entity for the link",
10    "path" => "path name to construct link"]],
11   "search" => ["json" => [], // no magnifier when this array is empty
12     "hover" => ""],
13     "title" => []]
14 ];

```

Each key of the array (except for the key `entity`) corresponds to a field that is going to appear in the sidebar. The upper part of the sidebar is filled with the keys `"title"` and `"subtitle"` (not required), and for the part in the white insert, each key constitutes the text that is going to appear as field name.

To each of these keys is attached an associative array defining how the metadata related to these fields are going to be displayed. Each key refers to something specific :

- **values** : the text in `html` that is going to appear next to the field name. In the array associated to this key, can be added an array if the text is going to be a link :
 - **html** : text to display
 - **id** : id of the entity in order to build the link
 - **path** : path of the page to redirect to
- **search** : data used for generating the magnifier (i.e. redirecting to the search page for more related items). To this key is attached an other associative array :
 - **json** : query in the DSL language of ElasticSearch. Not all the query is needed, only the `query` part. The `source`, `size`, etc are generated after that.
 - **hover** : text that is going to appear when hovering the magnifier, which is a description of the query
 - **title** : title of the query that will appear in the result page

The keys `values`, `json` and `title` are associated to an array because some fields can have several values to display, for instance is a work has multiple creators. For example, if a primary source is written in two different scripts, the metadata associated to the “script field” will look like this :

```

1  ["scripts" =>
2   ["values" =>
3     ["<span class='mainContent'>Latin</span>",
4      "<span class='mainContent'>Gothic</span>"],
5   "search" =>

```

```

6      ["json" =>
7          ['[{"bool": ["must" => [[[{"match" => ["original_texts.script.id" => 85}]]]}],',
8              [{"bool": ["must" => [[[{"match" => ["original_texts.script.id" => 46}]]]}],}],
9              "hover" => "Find all primary sources written in the same script",
10             "title" =>
11                 ["All primary sources written in latin",
12                  "All primary sources written in gothic"]
13         ],
14     ];

```

String formatting

The text that is going to be displayed is formatted thanks to the `toPublicString` and `toPublicTitle` methods :

- `toPublicString` is used when the information will be displayed as metadata (i.e. in the record page of an original item, the name of the work of this item will be formatted with this method) ;
- `toPublicTitle` is used when the information will be displayed as main content (i.e. in the record page of a work, the name of the work will be formatted with this method).

The `getTpaq()` methods returns for each entity associated with a tpq and a taq date, a correctly formatted date, for example :

- if `tpq` = 1245, and `taq` = 1325 ==> “1245—1325”
- if `tpq` = 1325, and `taq` = 1325 ==> “1325”

Managing missing information

If an information is missing, (for example, if the current entity doesn't have place associated with it), several layout options are available :

- *textbfNot showing the field at all* : the key for this field must not appear in the array of metadata, or the array associated to the `values` key must be lefted empty.

```

1 // the field translator must appear only if there is one
2 if ($work->getTranslator()){
3     $metadata["translator"] = $metadataTable;
4     $metadata["translator"]["values"][] = $work->getTranslator()->toPublicString();
5 }

```

- *Showing the field with a string indicating the missing information* : in the array associated to the `values` key, a string can be set when there is no information available for the current field. It must be encapsulated in a `` tag.

```

1 // if there is no incipit, display "Missing incipit"
2 $incipit = $work->getIncipit() ? $work->getIncipit() : "<span class='noInfo'>Missing incipit</span>";

```

- ***Keeping the field empty of information*** : when the array of the `values` key is lefted with an empty string in it, the template of the sidebar will automatically display `No information provided`

```

1 // if there is no historical actors, set the values to be []
2 $values = count($actors) != 0 ? [] : [""];

```

Pie chart

A pie chart can be displayed at the bottom of the sidebar when creating a key named `visualisation` in the metadata array, filled with an associative array defining a title and a dataset for the chart.

```

1 $metadata["visualisation"] = [
2     "title" => "Title of the visualisation"
3     "data" => [[{"category" => "cat1", "value" => 7}, {"category" => "cat2", "value" => 4}
4     ], [...]
5 ];

```



FIGURE G.6 – Pie chart at the bottom of the sidebar

G.2.2 Sidebar template

The sidebar is opened with a click on a toolbar icon, and closed when clicked outside the sidebar. The javascript allows to switch between two states that are defined in the public CSS stylesheet : when opened, the classes `open-toolbar` and `open-sidebar` are added to the element of the sidebar in order to appear on the page.

```

1 $(document).ready(function(){
2     // open the sidebar on a toolbar item
3     $('.toolbarItem>span').click(function(){
4         let id = $(this).attr('id');
5
6         $("#sidebar").attr("isOpen", "true");
7         switchSidebar();
8         let sideContent = "side-"+id;
9         $('.backgroundSidebar').hide();
10        $('#'+sideContent).show();
11    });
12
13    // close the sidebar when clicked outside the side bar
14    $(document).click(function (e) {
15        let thatTarget = $(event.target);
16        // if the clicked target is not a child of sidebar AND sidebar is open
17        if (!thatTarget.closest('#sidebar').length && $("#sidebar").attr('isOpen') === "true")
18        {
19            $("#sidebar").attr("isOpen", "false");
20            switchSidebar();
21        }
22    });
23
24    function switchSidebar() {
25        if ($("#sidebar").attr("isOpen") === "false") {
26            $("#sideBorder, #toolbar").removeClass("open-toolbar");
27            $("#innerSidebar").removeClass("open-sidebar");
28        } else {
29            $("#sideBorder, #toolbar").addClass("open-toolbar");
30            $("#innerSidebar").addClass("open-sidebar");
31        }
32    };

```

G.3 Search interface

The search page serves for now two purposes :

- allowing the user to perform full-text queries, only on the entities **work**, **primary source** and **original item**,
- displaying the related items when clicking on a magnifier in the sidebar of a entity record.

The queries are performed on the ElasticSearch database through an AJAX call : TODO when launching in production, the requested server (from `localhost` to `dis has.obspm`) will change. The variable to be modified to make this change is located in the `myElasticSearch.js` :

```
1 function generateUrl(query="", index="", from = 0, size = 10000, server="localhost:9200", http
= "http") { ... }
```

The retrieved results from this call are then formatted to be displayed in a DataTable thanks to the `TAMASListTableTemplate` objects. Those templated are defined with the `getPublicObjectList()` in the entity repositories.

G.3.1 Data structure : TAMASListTableTemplate

The objects of the class `TAMASListTableTemplate` defines for each field (i.e. column) of an entity, how the information is going to be presented inside the DataTable. For instance, the template of a work title will defined that the text is going to be in *italic*, redirect to the page of the concerned work and that the string “Unknown title” will be displayed if the information is missing.

The `getPublicObjectList()` is defined as follows :

```
1 public function getPublicObjectList()
2 {
3     $fieldList = [
4         new TAMASListTableTemplate(
5             'name of the key where the information in the array of result will be located'
6             ,
7             'Column title for this field',
8             ['class'=>'array of CSS classes to style the text'],
9             'path' => 'defined if a the text is a link',
10            'id' => 'name of the key where the id in the results is located'],
11            'unknown' => 'text to display if the information is missing',
12            'fields needed to display this field, correspond to the `'_sources` in a
13            ElasticSearch query'
14        );
15
16        return $fieldList;
17    }
18}
```

Each column needs its template, each entity combining between 5 and 10 templates. A template object can be defined with a certain number of properties :

- **Name** : provides the location in an array where DataTable will find the information to display
- **Title** : corresponds to the column label associated with the information
- **Properties** : defines how the cell content will be formatted :

- **class** (will surround the text in a ``) :
- **list** : to indicate that multiple values can be displayed in the same cell
- **number** : in order to align the text to the right
- **title-italic** : to style the text of the cell in italic
- **path + id** :
- **path** : routing path to generate a link
- **id** : location of the id in the result object
- **Source** : defines which fields will appear in the elasticsearch response corresponding to the fields that are added to the array associated with the key “source” in the query (multiple fields can be added for a single column with “+”)

```

1 // Field "Work" of an original item defining how the work titles will appear
2 new TAMASListTableTemplate(
3   'work',
4   'Work',
5   ['class' => ['title-italic']], 'path' => 'tamas_astro_viewWork', 'id' => 'work.id'],
6   '', // ifOnly is not used in the public interface
7   'work.default_title+work.id'
8 )

```

G.3.2 Specific functions

The functions and variables used to create the search interface are located in several files :

- `publicResult.html.twig` : twig template for the result page
- `myElasticSearch.js` : functions to build queries, to retrieve data from response or generate URL
- `myDataTable.js` : functions related to DataTables
- `mainJavascript` : utility functions

The main functions help :

- to **retrieve results** from ElasticSearch

```

1 /**
2 * This function takes a response from an ajax call to elasticsearch
3 * and returns an array containing all the results of this query
4 *
5 * @param queryResponse : object
6 * @return {Array}
7 */
8 function retrieveResults(queryResponse) {
9   if (typeof queryResponse.hits.hits !== 'undefined') {

```

```

10     let data = queryResponse.hits.hits; // data is an array of results, but only the value
11     of the key "_source" is needed
12     let results = [];
13
14     $.each(data, function (key, value) {
15         results.push(value._source);
16     });
17
18     return results;
19 }

```

- to **format those results** (using the properties defined in the `TAMASListTableItem` `plate`)

```

1 /**
2 * to either return the information associated with a field
3 * or return a tag indicating missing information
4 * The properties will define how the cell content is formatted.
5 *
6 * @param text : result values to display inside an array (even if there is only 1 value)
7 * @param properties : object defined in the template object list (getPublicObjectList)
8 * @param result : object containing the results send by Elasticsearch
9 * @returns string
10 */
11 function textDisplay (text, properties=[], result={}){
12     let classes = "";
13     if (isDefined(properties.class)){ // putting all classes in the same string
14         for (let i = 0; i < properties.class.length; i++){
15             let className = properties.class[i];
16             classes = classes + className;
17             classes = i < properties.class.length ? classes + " " : "";
18         }
19     }
20
21     if (isDefined(text)){
22         if (isDefined(properties.path)){
23             let id = [selectNodeOfObject(properties.id, result)];
24             // selectNodeOfObject will select the id in the result if there is only one id to
25             // find,
26             // unlike when the cell is supposed to contain a list
27             if (isDefined(properties.class) && properties.class.includes("list")){
28                 // if the cell is supposed to contain a list of links
29                 // the "text" array looks like : ["text", id, "text", id, ...]
30                 let texts = [];
31                 id = [];
32                 for (let j = 0; j < text.length; j += 2){
33                     texts.push(text[j]);
34                     id.push(text[j+1]);
35                 }
36                 text = texts; // now it looks like ["text", "text", ...]
37             }
38             let linkText = [];
39             for (i = 0; i < text.length; i++){
40                 linkText.push(`<a href="${generateRoute(properties.path, id[i])}">${text[i]}</
41                 a>`);
42             }
43         }
44     }
45 }

```

```

40         }
41         text = linkText;
42     }
43
44     let cellContent = "";
45     let cellSummary = "";
46     for (i = 0; i < text.length; i++){
47         if (i === 0){
48             cellContent = classes !== "" ? `<span class="${classes}">${text[i]}</span>` :
49             `${text[i]}`;
50         } else {
51             cellSummary = cellSummary + `<span class="${classes}">${text[i]}</span>`;
52             cellSummary = i < text.length ? cellSummary + "<br/>" : "";
53         }
54     if (text.length > 1){ // if there is multiple data to display, put it in a summary
55         let Nrecord = text.length > 2 ? "s" : "";
56         let summary = `<summary class="mainContent">
57                         ${text.length-1} <span style="font-variant: small-caps
58                         ">more record${Nrecord}</span>
59                         <span style="font-size: 10px"></span>
60                         </summary>`;
61         cellContent = `${cellContent}<details>${summary}${cellSummary}</details>`;
62     }
63     return cellContent;
64 } else { // if the key doesn't exist in the array of results => missing information
65     if (isDefined(properties.unknown)){
66         return `<span class='noInfo'>${properties.unknown}</span>`;
67     }
68     return "<span class='noInfo'>No information provided</span>";
69 }

```

— to display those results in a DataTable

```

1 /**
2 * This function generates an entire Datatable of results retrieved from a query to
3     elasticsearch
4 *
5 * on the given index for the filters given as parameter query
6 *
7 * @param index string : label of an entity (ex : "primary_source")
8 * @param query object : the filters of the query (ex: {"match":{"id":"5"}})
9 *                     => the value associated with the key "query" in the query language of
10    elasticsearch
11 * @param queryTitle string : title detailing the query in natural language (EX => "All
12     original items kept in the British Library")
13 * @param queryTerm string : terms used to filter the results
14 * @param from int : the number of results from which the query starts
15 * @param size int : number of results wanted in the response
16 * @param generateHeader bool : defines if the header is generated or not
17 * @return url : url of the query
18 */
19 function generateResultTable(index, query, queryTitle="", queryTerm="", from=0, size=10000,
20     generateHeader=true) {

```

```

18     let fieldList = fieldLists[index];
19     let dataStructure = generateDataStructure(fieldList);
20     let sourceFields = generateSources(fieldList, true);
21
22     // generates columns labels
23     generateColumnHeader("results", fieldList)
24
25     // generates URL in order to make an ajax call
26     let url = generateUrl(query, index, from, size);
27
28     // defining the function formatting the response of an elasticsearch ajax call
29     // into an array of results ready to be displayed by DataTable
30     let formatData = function (queryResponse){
31         let results = retrieveResults(queryResponse);
32         for (let result of results){
33             let i = 0;
34             for (let field of fieldList){
35                 // field.name is the key where DataTable is going to search for the value to
36                 // fill the cells of one specific column
37                 // textDisplay() returns the value that is going to be displayed in the cell :
38                 // - the text correctly formatted (according to the field.properties)
39                 // - OR no information provided (if there is no value in the result array
40             )
41                 // selectNodeOfObject() returns the value of the node in the array of result
42                 // corresponding to string given as first argument
43                 result[field.name] = textDisplay(selectNodeOfObject(sourceFields[i], result,
44                     field.properties), field.properties, result);
45                 i++;
46             }
47         }
48         return results;
49     };
50
51     if (generateHeader){
52         // generates a header for the current query
53         generateSearchHeader(index, queryTitle, queryTerm);
54     }
55
56     let table = $('#results').DataTable({
57         "ajax": {
58             "url": url, // url on which the query is done
59             "cache": true, // mandatory in order to send the query without wildly added
60             "parameters": {
61                 "dataSrc": formatData // variable containing the function formatting the results
62             }
63         },
64         "columns": dataStructure, // specify where DataTable will find information associated
65         "data": []
66     });
67     return url;
68 }
```

G.4 Bread crumbs and table editions

G.4.1 Bread crumbs

The breadcrumbs allows the user to find his way around the front office, it indicates at which level of the site hierarchy they are located. The javascript code to generate them takes as argument the name of the current node, i.e. the name of the page, and it reconstructs the page tree structure to get there.

G.4.2 Related table editions

The relation of the edited texts and the original texts are modelled in a graph structure in order to limit aberration and loops (preventing an edition to be based on an edition that is itself based upon this edition), and to display graph visualisations for the editions. With the aim of gathering all the editions that were based on one particular original item, including the *indirect* ones (type B editions), it is necessary to use the graph structure.

DISHAS/src/TAMAS/AstroBundle/Repository/OriginalTextRepository.php

```

1  /**
2  * getAllEditedTexts
3  *
4  * this method returns an array of the editions of an original item, including indirect
5  * editions (type B)
6  *
7  * @param \TAMAS\AstroBundle\Entity\OriginalText $originalText
8  * @return array
9  */
10 public function getAllEditedTexts(\TAMAS\AstroBundle\Entity\OriginalText $originalText)
11 {
12     if (! $originalText)
13         return [];
14
15     $editedTextRepo = $this->getEntityManager()
16         ->getRepository(\TAMAS\AstroBundle\Entity\EditedText::class);
17
18     $dependanceTree = $editedTextRepo->getDependanceTree();
19     $graph = new \TAMAS\AstroBundle\DISHASToolbox\Graph\TAMASGraph();
20     $graph->loadJSONTree($dependanceTree);
21
22     $origItemId = $originalText->getId();
23
24     $originalId = 'o' . $origItemId; // the id of the original item is preceded by "o" in the
25     // graph tree
26     $original = $graph->getNode($originalId); // find the node corresponding
27     $editedTexts = $original->getAncestors(); // find the edited texts related to this node
28
29     $editions = [];
30     foreach ($editedTexts as $editedText){
31         // fill the array with all the edited texts corresponding to the id
32         $editions[] = $editedTextRepo->findBy(['id' => str_replace("e", "", $editedText->
33             getLabel())])[0];
34     }
35
36     return $editions;
37 }
```


Annexe H

Introduction à la bibliothèque *AmCharts*

This page is a general introduction to the charts javascript library AmCharts in its version V4. An extensive documentaion and API can be found here, demos of all the charts that can be made can be found here and tips to customize chart can be found there.

H.1 Instantiating a chart

H.1.1 Modules

AmCharts allows to create several kind of charts. Depending on the one you wish to create, you will need to import several modules from the library. In all cases, you will need the main module `am4core` (which should always comes first) and an other module corresponding to the type of chart : either the `am4charts` for all the *regular* charts, or `am4maps` for map visualisations.

Some kind of charts requires additionnal plugins like force-directed graphs (`forceDirected`) or maps that need to import some geodata to display a background map (`worldLow`).

On top of that, you can import modules for setting assets like translations or themes that will define behaviors and color set of the charts.

```
1 <script src="https://www.amcharts.com/lib/4/core.js"></script>
2 <script src="https://www.amcharts.com/lib/4/charts.js"></script>
3 <script src="https://www.amcharts.com/lib/4/maps.js"></script>
4 <script src="https://www.amcharts.com/lib/4/geodata/worldLow.js"></script>
5 <script src="https://www.amcharts.com/lib/4/themes/animated.js"></script>
```

H.1.2 Creating a chart

The `create()` function is always first used to add new instances to your HTML page. It takes two arguments :

- The **tag id** of the `html` element in which is going to contain the chart (the element must already exist in your DOM, the function will not create it) ;
- The **type of chart**, *i.e.* the class reference.

```
1 var chart = am4core.create("chartdiv", am4charts.XYChart);
```

Combining multiple charts

Is it possible to put several charts in the same `html` element. To do that you only need to create first a container in which you will create charts instances :

```
1 var container = am4core.create("chartdiv", am4core.Container);
2 var mapChart = container.createChild(am4maps.MapChart);
3 var XYChart = mapChart.chartContainer.createChild(am4charts.XYChart);
```

Configuration : object/JSON approach

There are two way of setting up the parameters and data for the charts : either a **JSON-based** approach, or an **object-based** approach. Object approach is more readable and less prone to bug whereas JSON approach, being fundamentally a string, can be stored in a database or load dynamically with AJAX.

The rest of this introduction will only give examples with the object approach but you can find more information here on the JSON way. Object-based only means that the chart is treated as an object whose properties we're going to change/call.

H.2 Anatomy of a chart

A chart is mainly composed of two elements : the “*frame*” and the “*shapes*” representing the data :

- The *frame* is, in most cases, composed of two `axes` defining the structure of the chart.
- The data can be displayed in different *shapes* depending of the type of chart : columns, bubbles, points, line, etc. Whatever its shape, data is an instance of one or several `series`.

H.2.1 Axes

Axes are the structure of the chart ; they are specially fundamental to XYChart and some other type of charts. The axes determines how the data is going to be spread out through out the chart.

The axes can be of mainly two types :

- **Category axis** : used to position data items on text-based categories. For example, for a dataset detailling the number of folio per manuscript, the category axis would represent all the manuscripts.
- **Value axis** : used to position data items on a numeric scale. In our previous example, it would represent the total number of folios.
- Other types of axes :
 - **Duration axis**
 - **Date axis**

And, in the case of XYCharts, axes can be horizontal – or circular in case of some charts like radar – (`xAxes`) or vertical (`yAxes`) ; any of them can be a category axis or value axis, and both of them can be of the same type.

Creating axes

To associate axis to a chart, you need to push them to your existing chart like so :

```
1 var valueAxis = chart.yAxes.push(new am4charts.ValueAxis());
2 var categoryAxis = chart.xAxes.push(new am4charts.CategoryAxis());
```

Configuring axes

Most of the appearance parameters and settings can be configured directly on the axis variable or with the `renderer` property :

```
1 // define a title for the value axis
2 valueAxis.title.text = "Number of folios";
3 // display the axis one the opposite side than the default one
4 categoryAxis.renderer.opposite = true;
5 // to set the background grid of the value axis to be transparent
6 valueAxis.renderer.grid.template.strokeOpacity = 0;
7 // set the font size of the label from the category axis to 20
8 categoryAxis.renderer.labels.template.fontSize = 20;
```

H.2.2 Series

A series is a collection of similar, logically grouped data points, it gathers the configuration defining how the data is going to be displayed.

One series aggregates data you want to have the same properties. Sometimes one series will be enough to display gigantic datasets, but sometimes you will need to provide each data its series in order to specify finely the settings for each one in particular. A series, in a column chart for instance, can be either — depending on what you want to show — a group of columns, a single column, a group of cells through out the columns or a single cell.

Each type of chart allows only certain kind of series : a `pieSeries` on a `mapChart` will never be displayed.

Series have two main purposes : - Setting appearance and behavior of a collection of chart items ; - Binding individual chart items to source data.

Creating series

To create a new series for your chart, you must proceed as for the axes :

```
1 var series = chart.series.push(new am4charts.ColumnSeries());
```

Configuring series

You can set configuration for a series either directly on the series variable or through its `template` property (more on templates later on) :

```
1 // setting width of the stroke
2 series.strokeWidth = 0;
3 // setting text appearing when hovering on columns
4 series.columns.template.tooltipText = "Number of folios in manuscript";
```

H.2.3 Additionnal components

Inside a chart, you can add several type components like **SVG shapes**, **labels**, etc. To create them, you need to push them in your chart container, then you can configure them as you like :

```
1 var rectangle = chart.chartContainer.createChild(am4core.Container);
2 var rectLabel = rectangle.createChild(am4core.Label);
```

H.3 Data

H.3.1 Data structure

All data, in order to be used by your chart, need it to be formatted as an **array of objects** :

```

1 var msData = [
2   {
3     "ms": "Latin 10264",
4     "folios": 3
5   },
6   {
7     "ms": "Bav. Lat. 1374",
8     "folios": 14
9   }
]

```

H.3.2 Binding data

In order to make your chart use your data, you first need to associate your array to the data property of your chart :

```
1 chart.data = msData;
```

With this line, when we will set data properties to our chart series, it will know automatically where to find the information it needs. You can also bind a dataset to a specific series for complex layouts.

H.3.3 Using data fields

Now, the axes and the series that have been earlier create needs to be associated with the data from our array. The **dataFields** property is a simple object that binds specific field in series, to a data field in data objects : using **dataField** the series, as well as the category axis, will know what key in the dataset they are binded to.

```

1 series.dataFields.categoryX = "ms";
2 series.dataFields.valueY = "folios";
3 categoryAxis.dataFields.category = "ms";

```

What those lines means is : for each object in the array of data, the numerical value for the series will be found to the “folios” key and the string-based value will be found to the “ms” key. The value axis will generate itself according to the maximal value associated with the series to display, but the category axis needs to know which labels to show.

NB : for each kind of chart, the syntax might vary. You will find more details on that in the numerous examples in the demos.

H.3.4 Using properties fields

The `dataFields` property only holds information that is needed for the chart to be displayed, additionnal information can be binded to the chart to customize it further : `propertyFields` are the way to tie chart element's properties to data.

Let's say some fields were added to your dataset like that :

```

1 var msData = [
2   {
3     "ms": "Latin 10264",
4     "folios": 3,
5     "color": "#fcba03",
6     "url": "https://archivesetmanuscrits.bnf.fr/ark:/12148/cc71999k"
7   },
8   {
9     "ms": "Bav. Lat. 1374",
10    "folios": 14,
11    "color": "#fc3103",
12    "url": "https://digi.ub.uni-heidelberg.de/diglit/bav_pal_lat_1374"
13  }
14];

```

Pretty much any part of your chart can be customize using information from your dataset with the `propertiesField` property :

```

1 // make column labels a link
2 categoryAxis.renderer.labels.template.propertyFields.url = "url";
3 // colour the columns
4 series.columns.template.propertyFields.fill = "color";

```

Using dummy data

You can bind any data of your choice to a chart item using its `dummyData` property : it is a sort of storage container to put anything you want in it, without affecting the overall behavior of your chart.

It can be very useful when you will need to customize your chart creating tailored function using event listeners or adapters : it can become handy if you are dynamically generating data you want to access later on, for instance to bind behavior for two distinct events.

H.4 Appearance configuration

H.4.1 Templates

Practically all of the charts components have a `template` property : it is some kind of pattern for the layout and behavior of specific parts of your chart. Using templates, you can, for example, set that all filling colors of columns to be half transparent or for all axis labels to become green when clicked on :

```

1 // Half transparency of the columns
2 series.columns.template.fillOpacity = 0.5;
3 // Green labels
4 valueAxis.renderer.labels.template.events.on("hit", function(ev) {
5   valueAxis.renderer.labels.template.fill = "green";
6 });

```

Templates can be modified dynamically using DOM events, which is not the case with data values (associated with `dataFields`) : to change the values your chart is using, you will need either to switch datasets or use incremental loading).

AmCharts is an incredibly customisable library : if you would like to change any aspects of your chart, by browsing through the properties of the element you want to modify in the API, you will surely find something to suit your desires.

H.4.2 Colors, percentage, etc.

To unlock all methods applicable to colors and percentages, you will need to use special Amcharts functions. The `color()` method takes only one argument of color and transform it into an object that can be later on modified. The same is true for `percent()` method. To define the color or the percentage of a property :

```

1 rectangle.width = am4core.percent(50);
2 rectangle.background.fill = am4core.color("orange");

```

Heat rules

Heat rules defines a property of a chart item that will be bind to a numeric value. Any property is usable, as long as you can define a min and max state like so :

```

1 // The higher the value, the wider the column
2 series.heatRules.push({
3   target: series.columns.template,
4   property: "width",
5   min: 80,
6   max: 160

```

```
7  } );
```

H.4.3 States

Any part of the chart can be associated with a `state` object that holds properties defining how the element will behave when on this particular state. States define alternative properties that will override “*normal state*” properties on certain occasions.

As many other objects in AmCharts, defining state is a two steps process : first **creating** the state for a specific element, then **configuring** it :

```
1 var hoverState = series.columns.template.states.create("hover");
2 hoverState.properties.fillOpacity = 0.9;
```

H.4.4 Formatting text in labels and tooltips

Font styling

Text styling options are always contained in `[square brackets]`. It works like HTML tag so each opening bracket should be closed with a closing bracket (`[/]`), but an other opening bracket will act as a closing one for one opened earlier.

Some styling have an already implemented syntax in AmCharts, but you can use any CSS property :

```
1 rectLabel.text = "[bold]Work diffusion[/] :\n the example of the [font-style: italic]Tabule
Magne [/]";
```

Accessing data values

Is it possible to display data in labels with the `{curly brackets}` notation. Depending on what you write in your brackets, the parser (or the “formatter” to use AmCharts taxonomy) will try to find matching data in your dataset. You can either use the name of a key in your dataset, or the class reference of the data :

```
1 series.columns.template.tooltipText = "Numbers of [{color}]folios[/] in {ms} : \n {valueY}";
```

You can even perform some data computing on your numeric values with the formatter using modifiers such as `sum` or `count`.

Using HTML

You can also display any HTML elements in your tooltips and labels using the `html` or `tooltipHTML` property :

```
1 rectangle.tooltipHTML = '<span style="font-variant: small-caps;">Regarding folio quantity in  
manuscript</span>';
```

H.5 Additional features

Satellite elements can be instantiated by creating a new object of their class, they will behave accordingly to their normal use but you can completely divert their utility

H.5.1 Cursors

Cursors are only available for XY charts and radar charts : they allow to add interactivity to charts as well as providing zoom modalities.

```
1 chart.cursor = new am4charts.XYCursor();
```

By default, cursors, when moving on the chart, display two crossing lines ending with tooltips to better pinpoint and understand the data, but all of this can be disabled or modified. You can as well, add some cursor behaviors to allow zooming and panning :

```
1 chart.cursor.behavior = "selectX";
```

H.5.2 Scrollbar

A scrollbar is associated to an axis of the chart and allows to zoom in it. Is it possible to push series in it to display a preview of the value of your chart, even to treat it as a chart in itself.

```
1 var scrollbarX = new am4charts.XYChartScrollbar();  
2 scrollbarX.series.push(series);  
3 chart.scrollbarX = scrollbarX;
```

H.5.3 Legends

A legend, when created, will automatically display the series that are associated to your chart. When clicking of the items of the legend, corresponding series disappear, giving possibilities to vary the visualization to the user.

In the case of our manuscript chart, adding a legend is not very insightful, because our data is contained in an unique series representing all the columns, which colors were assigned individually. If we wanted to display a matching legend, we would need to build a custom legend.

```
1 chart.legend = new am4charts.Legend();
```

H.5.4 Events listeners

Event listeners allows to bind custom functions to any action perform on a chart element. Each part of the chart has a list of events tied to variables that can be accessed and modified when the event is triggered.

Two kind of functions can be associated to the events :

- `on()` : any time the event happens ;
- `once()` : the first time the event happens.s

```
1 chart.cursor.events.on("selectended", function(ev) {
2   var range = ev.target.xRange;
3   var axis = ev.target.chart.xAxes.getIndex(0);
4   var from = axis.getPositionLabel(axis.toAxisPosition(range.start));
5   var to = axis.getPositionLabel(axis.toAxisPosition(range.end));
6   alert("Selected from " + from + " to " + to);
7 });
```

The syntax for setting custom functions is always the same :

```
{chartName}.{elementName}.events.[on/once]("{eventName}", function(args){...});
```

Some events are universal to any type of element, as `hit`, when the user is clicking or tapping on a touch screen.

H.6 Demos

H.6.1 Our manuscript chart

Here is the chart that was made during this introduction : Number of folios in manuscript.

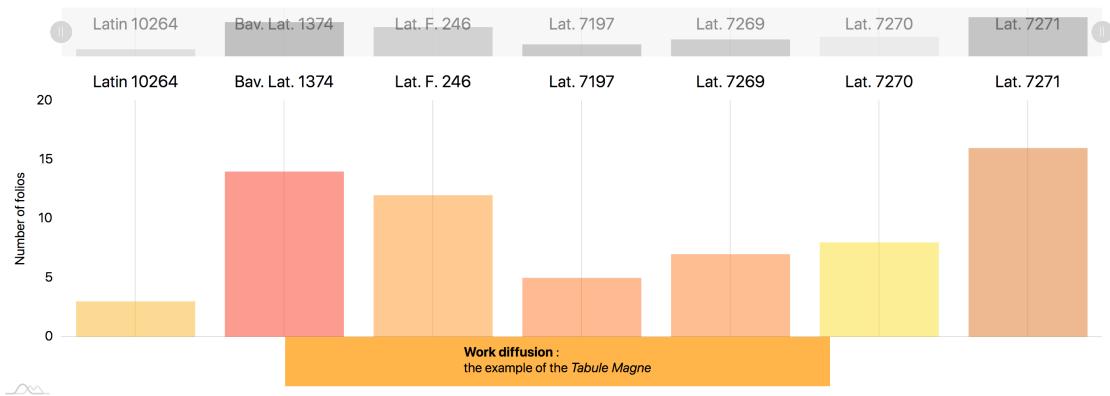


FIGURE H.1 – Chart made during the AmCharts tutorial

H.6.2 Examples in DISHAS

- Primary source : dynamic bar chart
- Historical sources : historical map

Liste des acronymes

- ALFA *Alfonsine Astronomy.* 5–9, 43
- API *Application Programming Interface.* 7, 42, 43, 77, 90, 98
- Ajax *Asynchronous JavaScript and XML.* 97, 101
- CHAMA *Commission for the History of Ancient and Medieval Astronomy.* 3
- CINES Centre Informatique National de l’Enseignement Supérieur. 75, 76
- CMS *Content Management System.* 30
- CRUD *Create Read Update Delete.* 77
- DH *Digital Humanities.* 87, 89, 90, 107
- DISHAS *Digital Information System for the History of Astral Sciences.* xix, xx, 3–10, 13, 17–19, 21–23, 25, 26, 28–32, 34, 35, 40–47, 49, 53–56, 58–60, 62, 63, 66, 68, 70, 74–79, 81, 82, 85, 87, 89, 90, 93–96, 98, 99, 102, 103, 105, 106, 110, 111, 113, 115, 116, 199
- DOM *Document Object Model.* 106
- DTI *DISHAS Table Interface.* 35, 66, 67, 90
- ERC *European Research Council.* 7, 8, 43
- FAIR *Findable Accessible Interoperable Reusable.* xviii, xx, 8, 43, 44, 75, 116
- FRBR *Functional Requirements for Bibliographic Records.* 22
- HAMSI *History of Astronomical Sciences in India.* 5, 6, 9
- IIIF *International Image Interoperability Framework.* 8
- ISNI *International Standard Name Identifier.* 33, 42
- JSON *JavaScript Object Notation.* 31, 35, 42, 77, 78, 98, 100, 104
- MVC *Model-View-Controller.* 93–95
- NOSQL *Not Only SQL.* 98
- ORM *Object-Relational Mapping.* 95, 96, 99, 101, 102
- PAL *Ptolemaeus Arabus et Latinus.* 5, 9
- RDF *Resource Description Framework.* 7, 10, 76, 98

- RGPD Règlement Général sur la Protection des Données. 30, 44
- SEO *Search Engine Optimization.* 44
- SPARQL *SPARQL Protocol and RDF Query Language.* 76
- SQL *Structured Query Language.* 32, 102
- SVG *Scalable Vector Graphics.* 106
- SYRTE Systèmes de Référence Temps-Espace. 7
- UX *User eXperience.* 44, 47, 48, 199
- VIAF *Virtual International Authority File.* 33, 42
- XML *eXtensible Markup Language.* 31, 87, 98, 104

Table des figures

2.1	BnF, Département des manuscrits – Latin 14481, f. 81v – Équation de la Lune	18
3.1	Modèle conceptuel de la base de données de DISHAS	22
3.2	Visualisation de l'architecture des données de paramètres astronomiques dans la base	26
3.3	Exemple de structuration de modèles mathématiques dans la structure de données	28
4.1	Schéma inspiré du diagramme de l'UX <i>design</i> conçu par J.J. Garrett	48
4.2	Schéma inspiré du livre <i>The Elements of User Experience</i>	49
4.3	Diagramme d'architecture de l'interface publique	50
5.1	Organigramme réalisé avec <i>Google charts</i> lors de phase d'élaboration des interfaces	57
5.2	Diagramme en colonnes conçu avec <i>AmCharts</i> pour les notices d'œuvres . .	57
5.3	Pictogrammes élaborés pour les principales entités de la plateforme de DISHAS	59
5.4	Exemple de page de notice pour l'œuvre de la <i>Tabule Magne</i>	60
6.1	Barre latérale des métadonnées de l'œuvre de la <i>Tabule Magne</i>	65
6.2	Visualisation du nombre de folios accumulé dans une source primaire . . .	69
6.3	Visualisation de la disposition des folios dans une source primaire	70
6.4	Carte temporelle de la page <i>Historical navigation</i>	73
6.5	Sélection d'une période sur la carte temporelle	74
7.1	Interaction entre la base de données de DISHAS, l'API et ElasticSearch . . .	77
7.2	Différence de syntaxe entre Javascript simple et jQuery	79
7.3	Exemple de commentaires de code intralinéaires	80
9.1	Fonctionnement simplifié de l'architecture logicielle MVC	93
9.2	Étapes pour l'affichage d'une donnée de la base au sein d'une page Web . .	95
9.3	Formulation dans le langage de requête d'ElasticSearch	100
9.4	Formulation dans le langage de requête du client PHP Elastica	100

9.5 Processus de requête asynchrone à ElasticSearch pour la page de recherche	101
10.1 Logo de la bibliothèque JavaScript AmCharts	106
10.2 Évolution de la visualisation pour la page de notice de sources primaires	108
10.3 Conséquences visuelles du changement d'échelle d'une visualisation	112
A.1 Schéma de la base de données réalisé avec <i>MySQL Workbench</i>	120
C.1 Page d'accueil de l'interface administrateur	129
C.2 Page de liste des enregistrements des éditions de tables astronomiques	130
C.3 Page de liste des enregistrements des <i>original items</i>	130
C.4 Interface de saisie numérique de tables	131
G.1 Column chart for work visualisation	168
G.2 Bar chart for primary source visualisation : original items stacked	168
G.3 Bar chart for primary source visualisation : original items spread out	169
G.4 Map for historical navigation	169
G.5 Metadata display in the sidebar	172
G.6 Pie chart at the bottom of the sidebar	175
H.1 Chart made during the AmCharts tutorial	195

Table des matières

Résumé	iii
Remerciements	v
Introduction	xvii
I Production, nature et méthode d'acquisition de la donnée	1
1 Le projet DISHAS	3
1.1 Présentation du projet	3
1.1.1 Objectifs du projet	3
1.1.2 Collaboration au sein du projet	5
1.2 Projets de recherche en histoire de l'astronomie	7
1.2.1 Le projet ALFA	7
1.2.2 Autres projets : description des ressources en ligne	9
2 Corpus de recherche	13
2.1 Bornes historiques du corpus	13
2.1.1 Ptolémée et l' <i>Almageste</i>	13
2.1.2 Propagation du modèle ptoléméen	14
2.2 Nature documentaire du corpus	16
2.2.1 Les sources primaires astronomiques	16
2.2.2 Les paramètres	18
3 Numérisation et accueil de la donnée	21
3.1 Modélisation de la donnée	21
3.1.1 Les sources historiques	23
3.1.2 Structuration des données d'édition	24
3.1.3 Développements futurs de la base de donnée	28
3.2 Données de gestion et encodage de l'information	29
3.2.1 Données structurelles	29

3.2.2	Encodage et typologie de la donnée	30
3.3	Saisie de la donnée	32
3.3.1	Interface administrateur	32
3.3.2	Saisie assistée de tables astronomiques : présentation de DTI	34
II	Modalités d'exposition de la donnée	37
4	Analyse des besoins	39
4.1	Définition des publics	39
4.1.1	Chercheurs	39
4.1.2	Non-spécialistes	41
4.1.3	Ingénieurs numériques	42
4.2	Enjeux liés à la donnée	43
4.2.1	Contraintes qualitatives	43
4.2.2	Spécificités des données de DISHAS	44
4.3	Constitution d'une interface publique	46
4.3.1	Objectifs d'une interface publique	46
4.3.2	Notions d'expérience utilisateur	47
4.3.3	Constitution d'un parcours de navigation	48
5	Médiation pour les publics non-spécialistes	53
5.1	Garantir l'ouverture à des publics non-spécialistes	53
5.1.1	Organisation de l'information comme vecteur de compréhension . .	53
5.1.2	Nécessité du contenu éditorialisé	55
5.2	Attractivité des contenus	58
5.2.1	Cohérence des interfaces	58
5.2.2	Intégration de modules interactifs	60
6	Participation au travail de recherche	63
6.1	Décloisonnement des corpus spécialisés	64
6.1.1	Enrichir les données par l'interconnexion	64
6.1.2	Constitution d'une interface spécialisée : le cas des éditions de tables	66
6.2	Analyse par la visualisation	68
6.2.1	Des manuscrits composites : montrer l'agencement d'un source primaire	68
6.2.2	Représenter un milieu intellectuel : le cas de la carte historique . .	70
7	Assurer la réutilisation du corpus numérique	75
7.1	Accessibilité des données brutes	75
7.1.1	Modalités d'accès à la donnée	75

TABLE DES MATIÈRES	203
7.1.2 Fonctionnalités implémentées dans l’interface publique	78
7.2 Exposer les méthodes : documentation et maintien de code	78
7.2.1 Lisibilité et documentation	79
7.2.2 Gestion de <i>bugs</i>	80
 III Traitement technique de la donnée et management des interactions	83
 8 Gestion de projet	85
8.1 Dialogue entre recherche et ingénierie	85
8.1.1 Coopération dans le domaine des humanités numériques	85
8.1.2 Processus de conception des interfaces	87
8.2 Organisation du travail intra-équipe	89
8.2.1 Instruments de gestion du travail	89
8.2.2 Méthodes de développement	91
 9 Environnement technique	93
9.1 Architecture de l’application Web	93
9.1.1 Modèle-Vue-Contrôleur	93
9.1.2 Le paradigme client/serveur	96
9.2 Implémentation d’ElasticSearch	98
9.2.1 Présentation d’ElasticSearch	98
9.2.2 Utilisation dans le <i>front office</i>	99
 10 De la conception à l’implémentation	103
10.1 Technologies de visualisation de données	103
10.1.1 Choix du moyen de réalisation	103
10.1.2 Choix d’une bibliothèque JavaScript	105
10.2 Constitution d’une visualisation	107
10.2.1 Dialogue entre conception et réalisation	107
10.2.2 Anticiper les obstacles à la visualisation	108
10.2.3 Limites de la visualisation	111
 Conclusion	115
 Annexes	119
 A Base de données du projet	119

B Glossaire	121
B.1 Astronomical object	121
B.2 Astronomical parameter set	121
B.3 Edited text	121
B.4 Historical actor	122
B.5 Intellectual author	122
B.6 Journal	123
B.7 Library	123
B.8 Mathematical parameter	123
B.9 Original text	123
B.10 Place	124
B.11 Primary source	124
B.12 Secondary source	124
B.13 Table content	125
B.14 Table type	125
B.15 Type of edition	125
B.16 Work	127
C Captures d'écrans de l'interface administrateur	129
D Présentation du <i>front office</i>	133
E Cahier des charges fonctionnel des visualisations pour l'interface utilisateur	151
F Exemples de requêtes avec ElasticSearch	159
F.1 ElasticSearch query language	159
F.1.1 Anatomy of a query	159
F.1.2 Get all records	160
F.1.3 Simple matching query	160
F.1.4 Adding some margin of error	161
F.1.5 Defining the source	163
F.1.6 Special queries	164
F.1.7 Combining multiple clauses	165
G Description des développements réalisés pendant le stage	167
G.1 Data visualisation	167
G.1.1 Visualisation description	167
G.1.2 Function explanation	170
G.2 Sidebar of metadata	172
G.2.1 Data struture	172

G.2.2 Sidebar template	176
G.3 Search interface	176
G.3.1 Data structure : TAMASListTableTemplate	177
G.3.2 Specific functions	178
G.4 Bread crumbs and table editions	182
G.4.1 Bread crumbs	182
G.4.2 Related table editions	183
H Introduction à la bibliothèque <i>AmCharts</i>	185
H.1 Instantiating a chart	185
H.1.1 Modules	185
H.1.2 Creating a chart	186
H.2 Anatomy of a chart	186
H.2.1 Axes	187
H.2.2 Series	188
H.2.3 Additionnal components	188
H.3 Data	189
H.3.1 Data structure	189
H.3.2 Binding data	189
H.3.3 Using data fields	189
H.3.4 Using properties fields	190
H.4 Appearance configuration	191
H.4.1 Templates	191
H.4.2 Colors, percentage, etc.	191
H.4.3 States	192
H.4.4 Formatting text in labels and tooltips	192
H.5 Additionnal features	193
H.5.1 Cursors	193
H.5.2 Scrollbar	193
H.5.3 Legends	194
H.5.4 Events listeners	194
H.6 Demos	194
H.6.1 Our manuscript chart	194
H.6.2 Examples in DISHAS	195
Acronymes	197
Table des figures	199
Table des matières	201