

ÉCOLE NATIONALE DES CHARTES

Pauline Breton-Chauvet

docteure ès lettres

Du faisceau au catalogue

**La patrimonialisation des données de New
AGLAE au prisme de leurs usages**
(projet Euphrosyne)

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

2021

Résumé

Ce mémoire a été rédigé à la suite d'un stage de quatre mois à l'Accélérateur Grand Louvre d'Analyse Élémentaire (New AGLAE), instrument d'analyse par faisceau d'ions exclusivement dédié à l'étude des objets du patrimoine. Installé dans les sous-sols du Louvre et intégré au laboratoire du Centre de Recherche et de Restauration des Musées de France (C2RMF), il permet de dévoiler la composition chimique d'un matériau avec un haut degré de précision et de fiabilité. En avril 2021, New AGLAE obtient pour Euphrosyne le statut de start-up d'État du Ministère de la Culture. Prévue en plusieurs phases de déploiement, cette interface baptisée du nom de l'une des Trois Grâces doit d'abord offrir un accès à distance aux données expérimentales et aux outils nécessaires à leur traitement. Elle aura ensuite vocation à permettre la diffusion et la réutilisation des données définitives. Ce stage a permis de contribuer aux fondations conceptuelles et fonctionnelles d'Euphrosyne, concomitamment à l'analyse de l'état FAIR des données produites par l'accélérateur. Dans un contexte de labilité des frontières entre le patrimoine et les informations qui en sont issues, nous avons profité des missions confiées pour envisager le processus de patrimonialisation des données de New AGLAE au prisme de leurs usages. Après une réflexion introductory sur la nature et la valeur des données au regard de l'objet patrimonial, les chapitres de ce mémoire abordent ainsi dans l'ordre naturel les « gestes » constitutifs de la transformation en bien patrimonial que sont : l'identification et la reconnaissance par la communauté d'AGLAE ; la conservation et l'archivage pérennes en vue d'une transmission aux générations futures ; la valorisation par l'exposition et la diffusion, dans le cadre actuel de l'*Open Science*.

Mots-clés : archivage numérique ; catalogue ; donnée ; Euphrosyne ; métadonnée ; IBA ; FAIR ; New AGLAE ; patrimonialisation ; pérennisation ; Open Science ; référentiel ; sciences du patrimoine ; sémiologie ; web sémantique.

Informations bibliographiques : Pauline Breton-Chauvet, *Du faisceau au catalogue : la patrimonialisation des données de New AGLAE au prisme de leurs usages (projet Euphrosyne)*, mémoire de master « Technologies numériques appliquées à l'histoire », dir. Gautier Poupeau, École nationale des chartes, 2021.

Remerciements

Je tiens tout d'abord à remercier mon directeur de mémoire, Gautier Poupeau, pour avoir accepté d'encadrer mes travaux tout en leur inspirant des pistes de réflexion fécondes, avec disponibilité et pédagogie. Je m'en estime sincèrement privilégiée.

Toute ma gratitude va également à ma tutrice de stage, Claire Pacheco, cheffe du groupe New AGLAE, qui m'a accompagnée avec confiance et dévouement dans le monde de l'analyse par faisceau d'ions appliquée au patrimoine. C'est tout naturellement que mes remerciements s'adressent désormais à Marine Zelverte, cheffe du Service Archives et Documentation du C2RMF, qui a co-encadré mon stage avec un engagement et une attention dont je lui suis particulièrement reconnaissante.

Je remercie bien entendu chaleureusement chaque membre de l'équipe de l'accélérateur : Quentin Lemasson, qui m'a transmis beaucoup de l'« âme » d'AGLAE en se mettant toujours à la portée de la néophyte que je suis ; Laurent Pichon, qui a été un véritable binôme et sans lequel je n'aurais pu percer le mystère technologique des données ; Brice Moignard, pour sa prévenance et sa gentillesse ; Jean-Paul Berthet, pour son écoute et nos échanges toujours enrichissants ; et Eric Laval, pour son accueil et son humour.

Merci au comité scientifique de la 8e rencontre IBAF de m'avoir permis de présenter le projet Euphrosyne et ses enjeux patrimoniaux auprès de la communauté élargie de l'analyse par faisceau d'ions. J'adresse aussi mes remerciements aux membres des institutions partenaires de SSHOC et de DIGILAB, en particulier Joseph Padfield, de la National Gallery, ainsi qu'à Orla Delaney qui ont exposé leurs expérimentations sémantiques avec générosité.

Je souhaite de même exprimer ma reconnaissance à mes camarades et amis de Master : Maxime Challon, Caroline Corbières, Mathilde Daugas, Arsène Georges, Morgane Rousselot et Hugo Scheithauer. Ils m'ont apporté une aide précieuse, et ont formé une chaîne de solidarité indispensable dans la situation particulière qui était la mienne.

Merci également à mon frère ainsi qu'à mes amis, pour leur soutien et leur compréhension. Enfin, je ne saurais suffisamment remercier Steeve, sans qui cette reprise d'études et sa charge de travail auraient été inenvisageables. Je lui dédie ce mémoire, ainsi qu'à Louise et à Rose.

Bibliographie

Diffusion, exposition

- BERMÈS (Emmanuelle), *Vers de nouveaux catalogues*, Paris, 2016 (Bibliothèques).
- BERMÈS (Emmanuelle), ISAAC (Antoine) et POUPEAU (Gautier), « Convergence et interopérabilité : vers le Web de données », *Bibliotheques* (, 2013), p. 29-46, URL : <https://www.cairn.info/le-web-semantique-en-bibliotheque--9782765414179-page-29.htm> (visité le 05/06/2021).
- GUILLAUD (Hubert), *Un monde de données*, Montpellier, 2011 (Washing machine).
- HEATH (Tom) et BIZER (Christian), *Web de données : méthodes et outils pour les données liées*, Montreuil, 2012.
- INFORMATIQUE (Séminaire IST et), *Le document numérique à l'heure du Web de données*, dir. Institut national de recherche en informatique et en automatique, 1 t., Paris, 2012 (Sciences et techniques de l'information).
- ISAAC (Antoine), « Les référentiels : typologie et interopérabilité », dans *Le Document numérique à l'heure du web de données*, Paris, 2012, p. 94-100.
- JEANNERET (Yves), *Critique de la trivialité : les médiations de la communication, enjeu de pouvoir*, 1 t., Paris, 2014 (Collection SIC, 04).
- SOWA (J.), « Ontology, Metadata, and Semiotics », dans B. Ganter, *Conceptual Structures : Logical, Linguistic and Computational Issues*, Springer, Berlin, 2000, p. 55-81.

Open Science

- Loi pour une République numérique*, 7 oct. 2016, URL : <https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000033202746/> (visité le 05/07/2021).
- MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR (de la Recherche et de l'Innovation), *Discours Science Ouverte à LIBER de Madame la Ministre Frédérique Vidal*, 4 juil. 2018.
- « Qui a peur de l'Open Access ? », *Le Monde* (, 15 mars 2013), URL : https://www.lemonde.fr/sciences/article/2013/03/15/qui-a-peur-de-l-open-acces_1848930_1650684.html (visité le 28/05/2021).

Recommandation de la Commission européenne relative à l'accès aux informations scientifiques et à leur conservation, 17 juil. 2012, URL : <https://eur-lex.europa.eu/legal-content/FR/TXT/PDF/?uri=CELEX:32012H0417&from=FR> (visité le 09/07/2021).

Recommandation de la Commission européenne relative à l'accès aux informations scientifiques et à leur conservation, 25 avr. 2018, URL : <https://eur-lex.europa.eu/legal-content/FR/TXT/PDF/?uri=CELEX:32018H0790&from=GA> (visité le 09/07/2021).

Open access and digital libraries : social science libraries in action, dir. Lynne M. Rudasill, Maria Elena Dorta-Duque et Fédération internationale des associations de bibliothécaires et des bibliothèques, 1 t., Berlin Boston, 2013 (IFLA publications).

Open access and the future of scholarly communication : policy and infrastructure, dir. Kevin Lindsay Smith et Katherine A. Dickson, Lanham, Md, 2016 (Creating the 21st-century academic library).

Patrimoine culturel, patrimoine numérique

Émotions patrimoniales, dir. Annick Arnaud et Laboratoire d'anthropologie et d'histoire de l'institution de la culture, avec la coll. de Daniel Fabre, 1 t., Paris, 2013 (Ethnologie de la France).

Pour une histoire des politiques du patrimoine, dir. Centre d'histoire culturelle des sociétés contemporaines et Institut d'histoire contemporaine, avec la coll. de Philippe Poirrier, Loïc Vadelorge et France, Paris, 2003 (Travaux et documents).

DAVALLON (Jean), *Le don du patrimoine : une approche communicationnelle de la patrimonialisation*, Paris, 2006 (Collection Communication, médiation et construits sociaux).

DESPRÈS-LONNET (Marie-Pierre), « L'écriture numérique du patrimoine, de l'inventaire à l'exposition », *Culture et musées*–14 (2009), p. 19-38.

GAUTHIER (Christophe), *Patrimoine et patrimonialisation du cinéma*, Paris, 2020 (Études et rencontres de l'École des chartes).

HEINICH (Nathalie), *La fabrique du patrimoine : de la cathédrale à la petite cuillère*, 1 t., Paris, 2009 (Ethnologie de la France).

Le tournant patrimonial : mutations contemporaines des métiers du patrimoine, avec la coll. de Christian Hottin et Claudie Voisenat, 1 t., Paris, 2016 (Ethnologie de la France).

MALRAUX (André), *Le musée imaginaire : [épreuves d'imprimerie]*, Genève, 1947, URL : <http://gallica.bnf.fr/ark:/12148/bpt6k1312096> (visité le 29/08/2021).

— *Psychologie de l'art*, 3 t., Genève, 1947.

- Numérisation du patrimoine : quelles médiations ? quels accès ? quelles cultures ?, avec la coll. de Bernadette Nadia Saou-Dufrêne, Madjid Ihadjadene et Denis Bruckmann, Paris, 2013 (Cultures numériques).*
- PATRIMOINE (Entretiens du), *Patrimoine et passions identitaires*, dir. Jacques Le Goff, Paris, 1998 (Collection des actes des Entretiens du patrimoine).
- Patrimoine culturel immatériel et numérique : transmission, participation, enjeux*, avec la coll. de Marta Severo et Séverine Cachat, 1 t., Paris, 2016 (Humanités numériques).
- POULOT (Dominique), *Une histoire du patrimoine en Occident, XVIIIe-XXIe siècle : du monument aux valeurs*, Paris, 2006 (Le noeud gordien).
- TRELEANI (Matteo), « Le patrimoine en ligne a-t-il un sens ? », *e-Dossiers de l'audiovisuel, L'extension des usages de l'archive audiovisuelle* (, 2014), URL : <http://www.ina-expert.com/e-dossiers-de-l-audiovisuel/le-patrimoine-en-ligne-a-t-il-un-sens.html> (visité le 23/07/2021).
- *Qu'est-ce que le patrimoine numérique ? : une sémiologie de la circulation des archives*, Lormont, 2017 (Collection UDPN).
- UNESCO, *Convention concernant la protection du patrimoine mondial, culturel et naturel*, 1972, URL : http://portal.unesco.org/fr/ev.php?URL_ID=13055&URL_DO=DO_TOPIC&URL_SECTION=201.html (visité le 18/07/2021).

Pérennisation, conservation, archivage

- ARNAUD (Michel) et MERZEAU (Louise), « Introduction », *Hermès*–53 (2009), p. 9-12.
- B2C, *D'un standard national d'échange de données pour l'archivage à un projet de norme ISO*, Modernisation et archives, URL : <https://siaf.hypotheses.org/806> (visité le 09/07/2021).
- BACHIMONT (Bruno), *Patrimoine et numérique : technique et politique de la mémoire*, Bry-sur-Marne, 2017 (Médias et humanités).
- BAUMACH (Donna) et MILLER (Linda), *Less is more, a practical guide to weeding school library collections*, USA, 2006.
- BOURDELOIE (Hélène) et CHEVRET-CASTELLANI (Christine), *L'impossible patrimoine numérique ? : mémoire et traces*, 1 t., Lormont, 2019 (Collection UDPN).
- CAMUS (Albert), *Le mythe de Sisyphe : [essai sur l'absurde]*, Paris, 1942 (Les Essais, N° 12).
- CHABIN (Marie-Anne), *Le management de l'archive*, Paris, 2000.
- FARGE (Arlette), *Le goût de l'archive*, Paris, 1997 (Points, 233).
- GALINON-MÉLÉNEC (Béatrice), « À la recherche de la trace », *Communication et organisation*–47 (2015), p. 31-50.
- GAMBA (Fiorenza), *Mémoire et immortalité aux temps du numérique : l'enjeu de nouveaux rituels de commémoration*, Paris, 2016.

INFORMATIQUE (Séminaire IST et), *Pérenniser le document numérique*, avec la coll. de Bernard Hidoine, Lisette Calderan et Jacques Millet, Paris, 2006 (Sciences et techniques de l'information).

JEANNERET (Yves), « Les Harmoniques du Web : espaces d'inscription et mémoire des pratiques », dans M. Lavigne et N. Pignier, *Mémoires et Internet, MEI*, Paris, 2010. *Vers un nouvel archiviste numérique*, dir. Institut national de l'audiovisuel et Université de Paris VII, avec la coll. de Valentine Frey et Matteo Treleani, Paris [Bry-sur-Marne], 2013 (Les médias en actes).

MERZEAU (Louise), « Du signe à la trace », *Médium*–18 (2009), p. 21-36.

Mémoire et médias, dir. Louise Merzeau et Thomas Weber, Paris, 2001.

RICOEUR (Paul), *La mémoire, l'histoire, l'oubli*, 1 t., Paris, 2003 (Points, 494).

ROUSSEAU (Jean-Yves) et COUTURE (Carol), *Les fondements de la discipline archivistique*, Sainte-Foy, 1994 (Gestion de l'information).

SIBILLE (Claire) et NICHELE (Baptiste), « Le Standard d'échange de données pour l'archivage (SEDA), un outil structurant pour l'archivage », *Gazette des archives*, 240–4 (2015), Publisher : Persée - Portail des revues scientifiques en SHS, p. 153-164, DOI : 10.3406/gazar.2015.5291.

Sciences du patrimoine

CALLIGARO (Thomas) et PACHECO (Claire), « Un accélérateur de particules fait parler les œuvres d'art et les objets archéologiques », *Reflets de la physique*–63 (1^{er} oct. 2019), Number : 63 Publisher : Société Française de Physique, p. 14-20, DOI : 10.1051/refdp/201963014.

CHALMIN (Émilie) et SALOMON (Hélène), « Constitution d'une « pigmentothèque » : un outil pour comprendre l'approvisionnement en matériaux colorants durant la Préhistoire », *ADLF. Archéologie de la France - Informations. une revue Gallia* (, 1^{er} mars 2021), Publisher : ministère de la Culture, URL : <https://journals.openedition.org/adlfi/65664> (visité le 13/06/2021).

La chimie et l'art : le génie au service de l'homme, dir. Maison de la chimie, 1 t., Les Ulis, 2010 (L'actualité chimique).

GUIAVARC'H (Mikaël), QUERRÉ (Guirec) et LEMASSON (Quentin), *Analyses de la fibrolite par PIXE et XRF pour l'étude de la provenance de haches polies néolithiques de l'Ouest de France*, Published : GMPCA 2017 - XXIème colloque international du Groupe des Méthodes Pluridisciplinaires Contribuant à l'Archéologie, avr. 2017, URL : <https://hal-univ-rennes1.archives-ouvertes.fr/hal-01982864> (visité le 12/06/2021).

LEROI-GOURHAN (André), *Évolution et techniques*, 2 t., Paris, 1943.

- *Origine et diffusion de la connaissance scientifique... [Conférence donnée à la Maison des sciences, à Paris, le 5 mars 1952.]* Paris, S.E.T. [Besançon, Impr. franccomtoise], 1953.
- *Le geste et la parole*, Paris, 1989 (Sciences d'aujourd'hui).
- *L'homme et la matière*, 1 t., Paris, 1992 (Sciences d'aujourd'hui , 50; Évolution et techniques , 1), URL : <http://gallica.bnf.fr/ark:/12148/bpt6k21430n> (visité le 29/08/2021).

MEEK (Andrew), BOUQUILLON (Anne), LEHUÉDÉ (Patrice), MASSON (Aurélia), VILLING (Alexandra), PIERRAT-BONNEFOIS (Geneviève) et WEBB (Virginia), « Discerning differences : Ion beam analysis of ancient faience from Naukratis and Rhodes », *Technè. La science au service de l'histoire de l'art et de la préservation des biens culturels*–43 (1^{er} août 2016), ISBN : 9782711863389 Number : 43 Publisher : Centre de recherche et de restauration des musées de France, p. 94-101, DOI : 10.4000/techne.804.

PACHECO (Claire), *Étude des films d'or sur matière vitreuse : application à la céramique glaçurée islamique médiévale : Asie Centrale, XIV^e - XV^e siècles – Iran XII^e – XIII^e siècles*, thèse de doct., Bordeaux, Université de Bordeaux III, 2007.

- « L'accès transnational FIXLAB : l'opportunité pour les scientifiques du patrimoine culturel d'accéder aux grands instruments », *Technè. La science au service de l'histoire de l'art et de la préservation des biens culturels*–43 (1^{er} août 2016), ISBN : 9782711863389 Number : 43 Publisher : Centre de recherche et de restauration des musées de France, p. 26-31, DOI : 10.4000/techne.589.

PACHECO (Claire), LEMASSON (Quentin), MOIGNARD (Brice), PICHON (Laurent), RADEPONT (Marie) et GOURIER (Didier), « D'AGLAE à New AGLAE », *Technè. La science au service de l'histoire de l'art et de la préservation des biens culturels*–43 (1^{er} août 2016), ISBN : 9782711863389 Number : 43 Publisher : Centre de recherche et de restauration des musées de France, p. 63-69, DOI : 10.4000/techne.689.

SCHMIDT (Bernd) et WETZIG (Klaus), *Ion beams in materials processing and analysis*, Wien New York, 2013.

X-RAY ARCHAEOOMETRY (International Symposium on), *X-rays for archaeology*, avec la coll. de Guy Demortier, Izumi Nakai et Masayuki Uda, Dordrecht, 2005.

ZWEIG (Stefan), *Le mystère de la création artistique*, trad. par Dominique Tassel, Tessere (Suisse), 2017 (Ciel vague).

Sémiologie

CARTIER (Centre Jacques), *Penser l'esprit : des sciences de la cognition à une philosophie cognitive*, avec la coll. de Vincent Rialle, Denis Fisette et Daniel Payette, Grenoble, 1996 (Sciences et technologies de la connaissance).

- Eco (Umberto), *Le signe : histoire et analyse d'un concept*, avec la coll. de Jean-Marie Klinkenberg, Bruxelles, 1990 (Collection Média).
- *La production des signes*, Paris, 1992 (Le livre de poche, 4152).
 - *Kant et l'ornithorynque*, trad. par Julien Gayrard, Paris, 1999.
 - *De l'arbre au labyrinthe : [études historiques sur le signe et l'interprétation]*, trad. par Hélène Sauvage, Paris, 2010.
- KLEIN (Robert), *La Forme et l'intelligible : écrits sur la Renaissance et l'art moderne*, dir. André Chastel, Paris, 1983 (Collection Tel, 83).
- MAILLET (Fernand), *La forme sous le regard : des origines au sens, l'immuable essai*, Paris, 2019.

Web sémantique

- ALLEMANG (Dean), *Semantic Web for the working ontologist : effective modeling in RDFS and OWL*, avec la coll. de James A. Hendler, 2nd ed, Waltham, MA, 2011.
- BERMÈS (Emmanuelle), *Le web sémantique en bibliothèque*, avec la coll. d'Antoine Isaac et Gautier Poupeau, Paris, 2013 (Collection Bibliothèques).
- *Le web sémantique en bibliothèque*, avec la coll. d'Antoine Isaac et Gautier Poupeau, 1 t., Country : FR ill., tabl., fig. 24 cm. En appendice, dictionnaire des sigles. Bibliogr. p. 165-166. Webliogr. p. 166. Notes webliogr., Paris, 2013 (Collection Bibliothèques).
- DELESTRE (Nicolas) et MALANDAIN (Nicolas), *Du web des documents au web sémantique*, Bois-Guillaume, 2017 (En pratique).
- GANDON (Fabien L.), FARON-ZUCKER (Catherine) et CORBY (Olivier), *Le Web sémantique : comment lier les données et les schémas sur le Web ?*, 1 t., Paris, 2012 (Info-Pro).
- KASSEL (Gilles), « Une alternative à la distinction 'continuant' vs 'occurrent' », dans *29e Journées Francophones d'Ingénierie des connaissances*, Nancy, 2018, p. 147-162.
- POUPEAU (Gautier), *Au-delà des limites, que reste-t-il concrètement du web sémantique ?*, Les Petites Cases, 6 oct. 2018, URL : <http://www.lespetitescases.net/au-delà-des-limites-que-reste-t-il-concrètement-du-web-sémantique> (visité le 01/08/2021).

Liste des acronymes

- AGLAE** Accélérateur Grand Louvre d'Analyse Élémentaire
- CEA** Commissariat à l'énergie atomique
- CIAF** Comité interministériel aux Archives de France
- CINES** Centre informatique national de l'Enseignement supérieur
- CNRS** Centre national de la recherche scientifique
- CRC** Centre de Recherche sur la Conservation
- C2RMF** Centre de Recherche et de Restauration des Musées de France
- EDF** ESRF Data Format
- EPA** Établissement Public à caractère administratif national
- E-RIHS** European Research Infrastructure for Heritage Science
- EROS** European Research Open System
- ESPADON** En Sciences du Patrimoine, l'Analyse Dynamique des Objets anciens et Numériques
- DINUM** Direction interministérielle du numérique
- FAIR** Findable, Accessible, Interoperable, Reusable
- IBIL** Ion Beam Induced Luminescence
- IN2P3** Institut national de physique nucléaire et de physique des particules
- LRMH** Laboratoire de recherche des monuments historiques
- NRA** Nuclear Reaction Analysis
- PARCOURS** PAtrimoine culturel et Restauration-Conservation : Ontologie pour l'Usage d'un Référentiel commun aux différentes Sources de données
- PIGE** Particle-Induced Gamma-ray Emission
- PIXE** Particle-Induced X-ray Emission
- POP** Plateforme ouverte du patrimoine
- RBS** Rutherford backscattering spectrometry
- SAED** Service Archives et Documentation

SIAF Service interministériel des Archives de France

TGIR Très grande infrastructure de recherche

VITAM Valeurs immatérielles transmises aux archives pour mémoire

Introduction

Ainsi espérons-nous n'avoir pas manqué de respect en essayant ici d'approcher le secret de la création artistique, cet instant indescriptible où finit la limitation terrestre imposée aux mortels que nous sommes, et où commence l'éternité.¹

Avec une facture traversée par un élan hagiographique envers les auteurs d'œuvres d'art, Stefan Zweig exprime la dimension métaphysique intrinsèque à la révélation du « secret du faire d'un artiste »², cette saisie de la transformation d'une image onirique intérieure en vision extérieure traduite par des formes et des matériaux. Il reprend ainsi à son compte la pensée de Goethe, lequel affirme plus d'un siècle plus tôt : « On ne connaît pas les œuvres d'art, quand on les voit uniquement achevées, on doit aussi les avoir connues dans leur devenir »³. Ces deux grands penseurs n'envisagent alors certainement pas que cette percée du mystère de la création puisse être portée, partiellement du moins, par un alliage complexe de technologie, de science, d'intuition et d'empirisme tel que celui qui sous-tend l'Accélérateur Grand Louvre d'Analyse Élémentaire (AGLAE). Devenu New AGLAE en 2017 avec l'amélioration et l'automatisation de la ligne de faisceau, il est l'unique accélérateur de particules au monde dédié exclusivement à l'étude d'objets patrimoniaux. Instrument de physique nucléaire, son utilisation dans les sous-sols du Musée du Louvre est ainsi singulière et exigeante. Par un procédé complexe de traitement et de projection de particules sur une cible, AGLAE permet une analyse non invasive d'un large panel de matériaux, des métaux aux céramiques en passant par certains minéraux, à un niveau de précision et de fiabilité qui en fait sa renommée. Les expériences menées à l'aide de l'accélérateur sur le patrimoine génèrent depuis presque quarante ans de nombreux flux de données, qui soulèvent désormais de façon urgente la question de leur conservation, de leur stockage, de leur accessibilité, de leur diffusion et de leur partage. En effet, l'absence de politique de gestion de données au-delà de leur destination applicative et de leur temporalité courante a abouti à une dégradation ainsi qu'à une perte d'informations extrêmement délétères et souvent irréversibles pour la compréhension et le rejet intellectuel

1. Stefan Zweig, *Le Mystère de la création artistique*, trad. de l'allemand par Dominique Tassel, Tessere (Suisse), Pagine d'Arte, 2017, p. 44.

2. *Ibid.*, p. 20.

3. Goethe, « Dialogue sur la vérité et la vraisemblance dans les œuvres d'art », dans Friedrich Schelling, *Écrits philosophiques et morceaux propres à donner une idée générale de son système*, trad. de l'allemand par Charles Bénard, Paris, Joubert, 1847, p. 418.

a posteriori de l'expérience.

À l'heure où les données, les méthodes et les outils de la recherche scientifique sont happés avec une force et une pression croissantes par les courants de la Science Ouverte – ou *Open Science* – , il est nécessaire de les réenvisager à travers un paradigme renouvelé, actuellement promu par le quadriptyque FAIR : *findable, accessible, interoperable, reusable*. En réalité, ces quatre volets reprennent des préoccupations déjà relativement anciennes et inhérentes à la gouvernance des données⁴, particulièrement prégnantes dans les institutions culturelles et patrimoniales, chargées de pérenniser des ressources et de transmettre sur le long terme les connaissances qui en sont issues aux différentes strates de la société. Selon une logique d'alimentation réciproque, cette perspective éminemment politique et stratégique de « gouvernance » est étroitement entrelacée, dans le cas d'AGLAE, à la dynamique exponentielle d'inflation patrimoniale actuellement en cours, et qui conduit au débordement de la notion initiale de patrimoine pour inclure certains éléments qui y étaient associés jusque-là sans se confondre à lui. Défini par l'Unesco en 1972 comme un héritage du passé que nous devons transmettre aux générations futures⁵, les frontières labiles du patrimoine autorisent de nouvelles formes de symbioses conceptuelles et culturelles entre la matérialité patrimoniale à proprement parler et l'arsenal sémiologique et cognitif qui vient lui conférer sa charge et son épaisseur testimoniales à l'égard de l'histoire de l'Humanité.

Sans nier la réalité et l'importance des enjeux économiques à l'œuvre dans ce nouvel effort de rationalisation et d'optimisation des données produites au moyen de fonds publics, la stratégie de pilotage définie conjointement par AGLAE et le Centre de Recherche et de Restauration des Musées de France (C2RMF) dont il dépend démontre avant tout une modification profonde du prisme à travers lequel les données expérimentales sont perçues, traitées et valorisées. Grâce au flou sémantique et typologique qui imprègne la notion de « patrimoine » depuis cinquante ans, un jeu fécond de syncedoque permet de réévaluer le périmètre et la portée des données générées à l'aide de l'accélérateur. Il est ainsi une interrogation qui surplombe et détermine désormais la prise en charge des données d'AGLAE dans les différentes étapes de leur cycle de vie : celle de leur nature, ou plutôt de leur valeur au regard d'une construction socio-historique à transmettre. Issues d'analyses pratiquées sur le patrimoine, peuvent-elles, par capillarité et en raison de leur contexte singulier de production, devenir elles-mêmes constitutives du patrimoine ? C'est

4. La gouvernance des données renvoie à une stratégie mise en place au sein d'une entreprise ou d'une institution afin d'offrir un cadre et des procédures destinés à structurer les flux de données, de leur collecte à leur utilisation finale. Elle recouvre les enjeux, les processus et les outils : juridiques, de sécurité, de pilotage, de modélisation et de référencement. Elle environne et détermine les modalités de traitement et de stockage des données. [Voir le support de cours « Visite guidée au pays de la donnée » de Gautier Poupeau, accessible en ligne : <https://docs.google.com/presentation/d/1PkV6QS0mRRrGZJT2Ib0jcdQV8T9AKwEP/edit#slide=id.p1>]

5. *Convention pour la protection du patrimoine mondial, culturel et naturel*, Unesco, 16 novembre 1972. Accessible en ligne : <https://whc.unesco.org/fr/conventiontexte/>

en tout cas le postulat qui fonde la réflexion actuellement en cours au C2RMF autour de la conservation pérenne et de l'ouverture des données en vue de leur diffusion. Il irrigue également les sources d' Euphrosyne, portée par la direction et par l'équipe d'AGLAE, lauréate du premier appel à projet de l'Atelier numérique, incubateur de start-ups d'État du Ministère de la Culture. Totalement pensée pour les utilisateurs d'AGLAE, elle a vocation à offrir dans un premier temps aux chercheurs et aux spécialistes de l'accélérateur, la « communauté-cible » prioritaire, un accès distant aux données expérimentales et aux outils nécessaires à leur traitement. À moyen et long termes, Euphrosyne doit ensuite permettre l'ouverture, l'interrogation et le partage des données dans le respect des principes FAIR précédemment évoqués, en conformité avec les normes et référentiels bâties collectivement au sein des infrastructures et des programmes – européens et français – auxquels AGLAE est associé pour l'identification, la préservation et l'exposition des données des sciences du patrimoine. Ces deux phases de déploiement ont respectivement été baptisées « Euphrosyne-manip » et « Euphrosyne-data ».

Par sa dépendance au C2RMF, membre pilote d' E-RIHS (*European Research Infrastructure for Heritage Science*), consortium européen d'excellence et d'intérêt global, AGLAE est introduit dans le programme IPERION-HS et dans la construction actuelle de la plateforme partagée DIGILAB, dédiée à l'accès et à la diffusion des données quantitatives produites par les différents instruments d'analyse du patrimoine. Innervée par les principes cardinaux de l'*Open Science* soutenus par la Commission Européenne depuis 2012, elle doit permettre l'interopérabilité de différents gisements nationaux, et optimiser leur ouverture et leur réutilisation par les chercheurs. Par ailleurs, AGLAE est intégré à ESPADON, projet d'équipement d'excellence (EquipEx) lauréat de l'appel à manifestation d'intérêt (AMI) Équipements structurants pour la Recherche, soutenu par la Fondation des Sciences du patrimoine et qui prévoit une gestion mutualisée des données afin de créer un « objet patrimonial augmenté »⁶.

C'est dans ce contexte d'émulation multiscalaire en faveur d'une politique transversale et holistique de gestion de données que s'est inscrit le stage effectué du 29 mars au 31 juillet 2021 à AGLAE. D'abord consacré à l'inventaire de l'existant et à l'analyse de l'état FAIR des données et des métadonnées de l'accélérateur, il devait ensuite être dédié à l'approfondissement de chacun des axes afin de proposer une méthodologie de « FAIRisation » exhaustive et valable pour l'intégralité du cycle de vie des données, tant dans leur état courant qu'intermédiaire et définitif. La simultanéité chronologique du début du stage avec la première phase de construction d'Euphrosyne en a remodelé partiellement les contours en orientant les missions du stage sur l'analyse interdépendante des flux de données avec les pratiques des utilisateurs afin de traduire le credo FAIR pour des

6. « ESPADON-PATRIMEX+, Lauréat de l'A.M.I. Équipements structurants pour la Recherche », *sciences-patrimoine*, <http://www.sciences-patrimoine.org/2020/12/selection-espadon/>, visité le 05 mai 2021.

usages-métiers spécifiques et préalablement définis. Selon une méthode itérative inspirée des projets AGILE, nous avons construit en collaboration avec l’Atelier Numérique⁷ l’architecture fonctionnelle des deux volets d’Euphrosyne, en cartographiant simultanément l’existant et en esquissant les modifications à apporter en matière de format, de structure, d’organisation et de renseignement des (méta) données. Parce qu’elles lient étroitement les données à leur justification sociale, les modalités d’édification génétique d’Euphrosyne ont constitué un prisme de choix pour l’observation et l’étude du processus de patrimonialisation des données d’AGLAE à travers les usages qui en sont faits.

Aussi, par quels processus et selon quels enjeux les données produites à partir du patrimoine peuvent-elles devenir biens patrimoniaux elles-mêmes ? En quelle mesure la FAIRisation encouragée politiquement et culturellement par l’Open Science favorise-t-elle cette transformation ?

Le postulat initial de données patrimoniales du simple fait de leur filiation heuristique à l’objet a été remis en question à l’aune des différents « gestes » constitutifs du patrimoine, adaptés à la nature des données d’AGLAE et qui sont : la reconnaissance par une communauté légitimante ; l’identification et la classification en fonction de référentiels communs ; la préservation au long cours à travers la conservation et l’archivage pérenne ; et enfin l’exposition, pas seulement réservée à une niche de spécialistes mais d’envergure universelle. Après un effeuillage conceptuel de la dialectique qui lie, voire confronte, la donnée et l’objet patrimonial afin de déterminer la nature et les apports de l’une au regard de l’autre, nous verrons en quelle mesure et selon quelles modalités la communauté d’AGLAE est devenue matrice de patrimonialisation à travers la genèse d’Euphrosyne. Conformément au tracé naturel des différentes étapes du processus, nous démontrerons ensuite en quoi la conservation et la pérennisation des (méta) données expérimentales, à travers la stratégie d’archivage et de désherbage en particulier, s’apparentent au mythe de Sisyphe, soit un effort imposé par son enjeu mais voué à l’inachèvement perpétuel. Enfin, nous aborderons les développements d’Euphrosyne-data en vue de l’exposition et de la dissémination des données d’AGLAE dans les différents circuits de l’Open Science. Nous démêlerons notamment l’écheveau méthodologique et contextuel qui a conduit à l’esquisse d’un catalogue, à l’aide d’une classification pour l’analyse par faisceau d’ions appliquée au patrimoine. Cette dernière étape sera l’occasion de clarifier les apports réels des outils et d’une forme de déontologie de l’*Open Science* pour l’universalité d’accès et d’exposition patrimonialisante.

7. Lancé en septembre 2020, l’Atelier Numérique est l’incubateur du ministère de la Culture et a pour mission d’accompagner les intrapreneurs de l’administration centrale, des directions régionales des affaires culturelles (DRAC) et des services à compétence nationale dans la conception de services publics numériques centrés utilisateurs.

Première partie

De la donnée à la recomposition de
l'objet patrimonial

Ce que nous disent la Ronde de nuit, les derniers Titien et la Montagne Sainte-Victoire, le Penseur, le tympan de Moissac, la statue du prince Goudéa et celle du pharaon Djéser, l'Ancêtre africain, ne peut être dit que par des formes, de même que ce que nous disent le Kyrie de Palestrina, Orféo, Don Juan ou la Neuvième Symphonie, ne peut être dit que par des notes.⁸

Dans *La Fabrique du patrimoine*, Nathalie Heinich rappelle à juste titre le caractère récent du mot « patrimoine ». Il y a encore quarante ans, ce terme n'apparaissait que dans quelques circulaires administratives et n'avait pas encore de sens pour le grand public⁹. La notion de patrimoine est intronisée par l'UNESCO au début des années 1970 avec le vote d'une Convention pour la protection du patrimoine culturel mondial. Trois ans après que 1975 ait été désignée comme « année européenne du patrimoine », une nouvelle « direction du Patrimoine » est créée au sein du Ministère de la Culture, suivie de « l'année du patrimoine » en 1980. Sans approfondir excessivement la dimension polysémique de cette notion, il est nécessaire de comprendre la dynamique d'inflation qui nous conduit aujourd'hui à nous interroger sur la patrimonialisation de certaines données, bien éloignées – par leur nature – des monuments de pierre englobés dans la définition initiale du patrimoine. Cette labilité du terme est à la fois la finalité et la cause de la « réinvention perpétuelle du patrimoine »¹⁰ qui profite désormais à une partie des productions numériques. Malgré ses anamorphoses, la définition du patrimoine porte intrinsèquement une visée inaliénable et consensuelle : son héritage et sa nécessaire transmission. Ainsi, la Convention du patrimoine mondial de l'Unesco de 1972 reconnaît comme patrimoine « l'héritage du passé dont nous profitons aujourd'hui et que nous transmettons aux générations à venir »¹¹. Cette définition bien plus fonctionnelle que typologique engage implicitement une architecture sémiologique indispensable, laquelle offre une corporéité sémantique à la subjectivité humaine qui sous-tend la reconnaissance patrimoniale.

La donnée pose en ce sens un défi conceptuel majeur, puisqu'elle est un pur produit de la sémiologie perceptive et peut, par ailleurs, relever à la fois du signifiant et du signifié. Sa porosité est telle qu'elle vient tout autant nourrir un « continuum amorphe »¹², parce qu'elle circule et existe indépendamment de toute interprétation, qu'une construction de l'esprit qui lui confère son sens et son épaisseur cognitive. Dans cette perspective, nous pouvons nous interroger sur la valeur et la nature des données produites par AGLAE : constituent-elles une augmentation de l'objet patrimonial dont elles seraient pleinement constitutives, sorte de versant immatériel d'une entité concrète, comme un jumeau numé-

8. André Malraux, *Le Musée imaginaire*, Paris, Gallimard, 1996, p. 263.

9. Nathalie Heinich, *La Fabrique du patrimoine : de la cathédrale à la petite cuillère*, Paris, Éd. de la Maison des sciences de l'Homme, 2009, p. 13.

10. Loïc Vadelorge et Philippe Poirier (dir.), *Pour une histoire des politiques du patrimoine*, Paris, Fondation Maison des sciences de l'homme, 2003, p. 14.

11. *Convention pour la protection du patrimoine mondial, culturel et naturel...*

12. Umberto Eco, *Kant et l'ornithorynque*, trad. de l'italien par Julien Gaynard, Paris, le Grand livre du mois, 1999, p. 350.

rique composé par les données IBA (*Ion Beam Analysis*) ? Ou bien s'inscrivent-elles dans sa périphérie directe, comme un prolongement à la fois interdépendant et bien distinct de l'objet patrimonial ?

Chapitre 1

Le « jeu de données » : augmentation ou prolongement de l'objet patrimonial ?

1.1 Nature et périmètre d'un « jeu de données »

Avant de nous attarder sur la nature d'un jeu de données au regard de l'objet patrimonial auquel il se rapporte, il convient d'abord d'en déterminer le périmètre. En effet, lorsque nous évoquons la *donnée*, nous abusons en réalité de la généricté du terme qui recouvre beaucoup d'interprétations et d'usages différents. La donnée se dérobant à toute tentative d'universalité, nous allons présenter ce que la communauté AGLAE – en particulier les experts en science du patrimoine mais également les ingénieurs, les archivistes et chargés de médiation – entend par ce terme, lequel doit toujours être abordé au prisme de ses usages.

Dans le cadre de la construction de la première phase de déploiement d'Euphrosyne, exclusivement dédiée au premier cercle d'utilisateurs d'AGLAE, l'équipe pluridisciplinaire du projet a dû circonscrire les limites d'un jeu de données de façon à l'encapsuler dans une unité sémiologique cohérente, lourde d'enjeux pour la structure numérique qui la sous-tend. Cette étape génétique d'Euphrosyne a été appréhendée de façon collective et itérative, chaque proposition étant soumise au débat et aux éventuelles reprises des membres de l'équipe. Le fil d'Ariane de cette réflexion a été tissé dans une subjectivité prospective : toute personne souhaitant « rejouer » l'expérience pratiquée sur un objet patrimonial dans un futur proche comme lointain doit pouvoir le faire à partir de l'ensemble de données et d'informations mis à sa disposition. A priori concrète et peu sujette à controverse, cette injonction a mis en lumière la variété et la complexité des interprétations possibles, même parmi ceux exerçant des fonctions et des activités similaires. La définition du niveau de granularité est sans conteste l'un des points de divergence les

plus importants. Cela est notamment dû à la particularité temporelle et l’arborescence des expériences réalisées à AGLAE, marquées par une forte hétérogénéité et des ruptures de linéarité avec des régularisations correctives fréquentes. Par ailleurs, une même expérience peut être pratiquée sur un seul ou plusieurs objets patrimoniaux, brouillant toute recherche d’unicité par l’intermédiaire de ces derniers. Ainsi, une expérience à AGLAE, appelée communément *run*, peut être initiée à la suite d’une demande officielle de temps de faisceau ou pas (cas des expériences en interne), dans le cadre d’une demande de service ou d’un projet de recherche, national ou européen (Voir Annexes : B.1, B.3), porter sur un ou plusieurs objets patrimoniaux, s’étendre sur quelques heures à plusieurs jours, et nécessite le plus souvent l’association de plusieurs techniques d’analyse, l’éventail de ces dernières allant de la méthode PIXE (*Particle-Induced X-ray Emission*) à la méthode PIGE (*Particle-Induced Gamma-ray Emission*), en passant par la ionoluminescence et la spectroscopie de rétrodiffusion de Rutherford (RBS).

Par conséquent, pour définir par capillarité le périmètre d’un jeu de données, nous nous sommes d’abord accordés sur le cadre sémantique d’une expérience (*run*). Cette dernière est ainsi d’abord caractérisée par une stabilité du contexte expérimental avec des paramètres et des standards définis en début de manipulation¹, la moindre modification contrevenant à la cohérence et à la compréhension de l’expérience par ceux qui en sont extérieurs. Une expérience peut être multitechnique, les données PIGE, PIXE et RBS étant générées automatiquement et ne demandant pas de sélection initiale. Elle donne lieu à la création d’un dossier dont le nommage est composé actuellement du nom de l’analyste, de la référence de l’objet et de la date². Un fichier Excel est généré automatiquement pour toute l’expérience et contenant le numéro de dossier, la référence de l’objet, le nom du projet, la Dose/seconde, la durée de l’analyse, les détecteurs, les filtres, les standards, la taille de la cartographie, l’énergie du faisceau et le type de particule. Dans le contexte d’une expérience réalisée par AGLAE, un jeu de données renvoie donc à l’ensemble des données brutes – fichiers LST et fichiers spectres primaires – et des données extraites via la suite logicielle ad hoc pour un run, avec le fichier Excel intégrant les métadonnées du contexte expérimental, les données traitées mettant en évidence le phénomène physique voulu, les éléments du cahier de laboratoire, qu’il soit imprimé et/ou dématérialisé, les photographies inhérentes aux zones d’intérêt, les standards, le tout contenu dans un dossier d’utilisateur.

Ce périmètre du jeu de données défini, nous pouvons désormais en interroger la nature au regard du processus de patrimonialisation. Si l’on reprend la « volonté d’art » et le « sentiment de métamorphose »³ qui président à la création du *Musée imaginaire* pensé par André Malraux, nous devons questionner la valeur même de ces données : qu’expriment-elles de notre histoire et de notre culture pour mériter l’intronsation pa-

1. Voir Annexe : B.5

2. Voir Annexe : A.2

3. U. Eco, *Le Musée imaginaire...*, p. 246.

trrimoniale ? Dans le cas d'AGLAE, et plus précisément des expériences menées à l'accélérateur sur des objets patrimoniaux⁴, les données générées par l'analyse par faisceau d'ions nous renseignent sur la composition chimique d'un objet et éventuellement sa structure physique⁵. Ces informations permettent d'identifier la provenance et les techniques de fabrication d'un objet, améliorant ainsi notre compréhension des sociétés et de leurs interactions dans un contexte géochronologique bien défini. De plus, les analyses produites servent à l'authentification de certaines pièces à l'origine controversée ainsi qu'à la connaissance des mécanismes de dégradation. L'analyse de la surface des objets en révèle les signes d'altération, éclairant les processus de corrosion et de dégradation pour stopper leur progression et recommander des conditions de conservation adaptées.

Les données produites au sein d'AGLAE nourrissent de fait aussi bien les savoirs nécessaires aux protocoles de restauration du patrimoine que l'histoire des faits culturels et sociaux depuis la Préhistoire jusqu'à l'époque contemporaine. Tandis que l'objet patrimonial du « monde réel » donne à voir une certaine esthétique dans une proximité avec le public riche d'enjeux culturels et philosophiques, les données en offrent une dimension cognitive indiscutable. Cette valeur ajoutée justifie t-elle à elle seule la reconnaissance patrimoniale ? Rien n'est moins certain si l'on se réfère à l'analogie du catalogue d'art et des œuvres qu'il présente. Le catalogue renseigne lui aussi sur l'histoire et la provenance des œuvres et n'est pas pour autant – sauf exception – considéré comme constitutif de notre patrimoine. Or, par leur singularité, leur rareté, leur contexte de production et leur objet, les données d'AGLAE échappent à toute catégorisation péremptoire, et s'inscrivent dans une frontière conceptuelle ambiguë et poreuse.

1.2 Les données d'AGLAE : enrichissement périphérique ou intrinsèque à l'objet patrimonial ?

La spécificité sémiotique des données d'AGLAE bouleverse jusqu'à leur sens même. Elles peuvent être à la fois l'expression et le contenu d'informations issues directement des objets patrimoniaux analysés. Nous ne pouvons donc en l'occurrence aborder la patrimonialisation des données de l'accélérateur de façon univoque pour chacune des étapes de leur cycle de vie. Par ailleurs, la nature patrimoniale des données doit nécessairement être envisagée à travers leur médiation et le caractère tronqué que cette dernière peut induire. Nous pouvons être tentés de considérer que les données brutes, en tant qu'émanation non-interprétée de l'objet soumis au faisceau, peuvent être appréhendées comme

4. L'Accélérateur Grand Louvre d'Analyse Élémentaire participe également ponctuellement à des recherches expérimentales inter-laboratoires en physique nucléaire (CIL) et à des analyses sur des matériaux non patrimoniaux en vue d'offrir des « marqueurs » et des éléments de référence aux expériences pratiquées sur les objets patrimoniaux.

5. Nous faisons ici référence à la méthode RBS et à l'analyse stratigraphique par couches minces. Voir Annexe : C.4

éléments numériques intrinsèques à l'objet patrimonial qu'elles « augmentent » et révèlent sous d'autres aspects. Il faut cependant noter que l'extraction des données par les outils logiciels d'AGLAE – qu'il s'agisse de GUPIXWin ou de PlotAndRoi pour les méthodes PIXE et PIGE – exige déjà une sélection des paramètres expérimentaux⁶, des éléments recherchés et du mode de calcul, impliquant inévitablement une interprétation et une troncature déterminées par la subjectivité et l'état des connaissances de l'analyste. Dès lors, les données obtenues par intervention de l'intelligence humaine dévoilent une coupe orientée de l'objet, comme une sorte d'état parcellaire et sélectionné. Malgré la médiation qui préside à leur génération, ces données sont malgré tout issues de l'œuvre elle-même et ne résultent pas d'une forme quelconque d'invention. D'une certaine façon, l'objet étudié « montre » ce qu'on lui demande de présenter de lui, un certain profil à l'orientation déjà définie pour lui. En ce sens, les données d'AGLAE ne constituent ni un prolongement ni une amplification intrinsèque de l'objet patrimonial, qui n'acquiert pas *stricto sensu* de nouveaux caractères ou éléments. L'accroissement de cet objet est d'ordre cognitif puisque les données récoltées en révèlent des aspects scientifiques jusqu'ici inconnus et non explorés. On peut donc plutôt évoquer une « mise en profondeur » de l'objet patrimonial, qui se trouve ainsi effeuillé par l'association féconde de la technique scientifique et de l'imaginaire.

Quant aux données traitées, fruit du travail intellectuel de sélection et d'interprétation de l'analyste chargé de l'expérience, elles posent de façon plus complexe la question de leur rapport à l'objet auxquelles elles se réfèrent. Totalement interdépendantes des données brutes puis extraites, elles ont cependant subi un processus d'intervention et de traitement qui altère l'« authenticité brute » de leur lien de filiation avec l'objet. Dès lors, nous devons interroger plus en profondeur la frontière conceptuelle et perceptive entre l'objet et les données en sciences du patrimoine qui s'y rapportent. Doit-on nécessairement considérer l'apport de l'analyste sur les données produites comme une altération à l'égard de la filiation à l'objet ? Ne pourrait-on pas plutôt percevoir ces ajouts comme un enrichissement intellectuel de l'objet analysé en considérant que, après tout, le patrimoine repose tout entier sur le principe de médiation et d'appropriation ?

6. Voir Annexes : C.1,C.2

Chapitre 2

Les données en sciences du patrimoine : lentilles de contemplation de l'inaccessible ?

Bien que les sciences du patrimoine aient le vent en poupe depuis plusieurs dizaines d'années, il est périlleux d'associer les termes de patrimoine et de science sans en approfondir la dialectique. L'émotion et la visée ontologique voire existentialiste qui irriguent les sources primitives du patrimoine ne paraissent pas forcément compatibles, de prime abord, avec la rationalité rugueuse des sciences appliquées. Le langage et les concepts de la physique et de la chimie n'ont ainsi que peu à voir avec le système sémantique déployé par les historiens de l'art. Aux premiers, l'austérité cartésienne, aux seconds le rêve et l'insaisissable, serions-nous tentés de conclure. Pourtant, avant l'émotion et l'appropriation sentimentale, l'objet patrimonial est avant tout forme et matière. Les céramiques médiévales, les peintures rupestres, le masque d'or de Toutankhamon ne sont d'abord rien d'autre que des minéraux et des métaux façonnés, des pigments et composés naturels. Dès lors, la physique et la chimie des matériaux s'articulent au patrimoine dans un mouvement de symbiose, où, tel un Ouroboros, nous ne pouvons plus exactement distinguer où débute l'expertise du chimiste et où commence l'esthétisme de l'œuvre :

Les autres (chimistes) se font enquêteurs et démasquent derrière les fards égyptiens le souci de la prophylaxie des yeux, restaurent au-delà de leur aspect actuel la réalité des bronzes antiques, admirent les performances de l'empirisme des artistes de l'Antiquité lorsqu'ils osent des transformations thermiques qui conduisent au verre, à la faïence. Et puis il y a la couleur ! Merveille de la perception humaine qui est aussi un grand acquis de la science moderne et particulièrement de la chimie qui en comprend la gamme des variations que leur apporte le vieillissement.¹

1. Christian Amatore, Anne Bouquillon, *et al.*, *La Chimie et l'Art : le génie au service de l'Homme*,

C'est ainsi que les chimistes en viennent à seconder les « transformeurs de matière » que sont les artistes par leurs analyses, la puissance de leurs prédictions sur l'avenir et l'évolution des matériaux. Les activités d'AGLAE représentent en cela un cas pratique incontestable de l'apport des techniques chimiques et de l'utilisation de la physique pour la mise en relief des processus de création artistique, qu'ils relèvent du passé le plus lointain ou du présent le plus actuel.

Parce qu'elles éclairent la matérialité du patrimoine sous un jour qui demeurerait inaccessible sans leur intercession, les données obtenues grâce à AGLAE peuvent constituer un viatique singulier de contemplation de ce qui est invisible à l'œil nu. En effet, nul regard humain, aussi aguerri soit-il à l'expertise visuelle des matériaux et des objets patrimoniaux ne peut en déduire la composition chimique précise. Les aspérités et la présentation chromatique de l'objet pourront éventuellement livrer de premiers indices qui définiront le cadre initial du contexte expérimental. Cependant, la répartition des taux entre les éléments chimiques et la représentation de la structure physique interne de l'objet ne peuvent être exprimées que sous la forme de données après l'intervention du faisceau.

2.1 Les données d'AGLAE lèvent-elles le voile de l'invisible ?

Les données d'AGLAE nous donnent accès à l'infiniment petit de l'objet patrimonial, à travers deux approches possibles et souvent combinées : la cartographie et l'analyse ponctuelle. Tandis que la première correspond à plusieurs points de mesure répartis sur une zone donnée – normalement rectangulaire avec un espacement régulier des points -, la seconde renvoie à un point de mesure². Avec les améliorations portées par New AGLAE en 2017, les analyses qui s'effectuaient jusqu'alors point par point sur le matériau étudié se sont ainsi vues enrichies par l'accessibilité – grâce à un système d'acquisition multi-paramétrique fonctionnant en List Mode – d'une imagerie chimique systématique³. Les

Les Ulis, EDP sciences, 2010, p. 11.

2. Un point de mesure peut correspondre à la taille du faisceau (typiquement quelques dizaines de μm de diamètre) ou à l'enregistrement global des différents événements sur une zone également rectangulaire balayée par le faisceau sur quelques centaines de μm^2 à quelques mm^2 .

3. Claire Pacheco, Quentin Lemasson, Brice Moignard, Laurent Pichon, Marie Radepond et Didier Gourier, « D'AGLAE à New AGLAE », *Technè* [En ligne], 43, 2016, p. 63-69, <http://journals.openedition.org/techne/689> (visité le 12 juin 2021). DOI : <https://doi.org/10.4000/techne.689> Dans le cas de l'imagerie, une surface maximale de 200 x 200 mm^2 est parcourue en combinant un balayage vertical du faisceau au moyen de bobines électromagnétiques et le déplacement mécanique de l'objet. Le mode d'acquisition sélectionné permet de corrélérer chaque rayonnement détecté avec la position (X, Y) du faisceau sur la cible. Ce couplage permet ensuite de reconstruire la carte de répartition spatiale des données IBA. [Thomas Calligaro et Claire Pacheco, « Un accélérateur de particules fait parler les œuvres d'art et les objets archéologiques », *Reflets de la physique*, 63, octobre 2019, p. 14-20, <https://www.refletsdelaphysique.fr/articles/refdp/pdf/2019/03/refdp201963p14.pdf> (visité le 25 juin 2021). DOI :<https://doi.org/10.1051/refdp/201963014>]

émissions secondaires produites par l'irradiation de la cible⁴, qu'ils s'agissent de rayons X, gamma ou de lumière, permettent l'identification, y compris à l'état de traces, des éléments chimiques présents dans les couches superficielles de l'objet. Par la saisie de ses facettes physico-chimiques, nous pénétrons dans sa chaîne opératoire⁵, à travers les matérialités successives qui ont précédé son état définitif et les différentes phases d'intervention humaine qui lui ont conféré sa valeur ainsi que sa rareté. Ce faisant, les données d'AGLAE élucident, par une technique d'analyse ou l'association de plusieurs, des énigmes de provenance – historique et géographique –, d'authenticité ou de procédé de fabrication de la mineure à la plus déterminante pour l'évolution des sciences humaines, en particulier pour l'histoire culturelle, économique et politique des sociétés depuis l'âge de pierre jusqu'à nos jours. Leur singularité et leur dimension patrimoniales « natives » s'en trouvent alors légitimées si l'on considère le patrimoine comme une construction perpétuelle vouée à conserver et à contextualiser un héritage pour le transmettre aux générations à venir. Parce qu'elles portent intrinsèquement une certaine vérité de l'objet patrimonial, nous pourrions considérer que les données brutes et extraites méritent d'être conservées au même titre que ce dernier. Néanmoins, plusieurs exemples distincts passés sous le faisceau d'AGLAE démontrent que cette perspective connaît des variations interdépendantes au contexte expérimental et à l'objet de l'analyse.

Commençons par deux exemples d'expériences réalisées au cours de la période du stage, marqués par une dimension expérimentale commune. Dans le premier cas, une équipe italienne de chercheurs, venue dans le cadre d'un projet de recherche européen, a consacré une semaine de temps de faisceau à l'étude d'échantillons-tests de toile recouverte de peinture, elle-même enduite d'une couche de titane à différents degrés d'épaisseur. Les expériences menées visaient à établir le comportement du titane pulvérisé en surface, afin de déterminer s'il pouvait jouer un rôle efficace et non altérant de protection, dans un objectif de préservation des œuvres picturales. Les méthodes d'analyse mobilisées ont été PIXE et RBS afin de visualiser la répartition des éléments par couche et de déterminer précisément la profondeur d'absorption du titane. Dans le second cas, une équipe française de chercheurs en sciences archéologiques a souhaité analyser des pièces ainsi que des échantillons minéraux en fibrolite⁶ afin de rechercher d'éventuels « marqueurs » géolo-

4. L'irradiation peut être réalisée avec l'accélération de protons ou de particules alpha, projetés ensuite sur la cible. Plus rarement, des ions plus lourds appelés « deutons » peuvent aussi être tirés, mais plutôt à des fins expérimentales en physique nucléaire.

5. La chaîne opératoire renvoie à un acte technique mobilisant des matériaux, des outils, des gestes et des savoirs. Pour approfondir cette notion : A. Leroi-Gourhan, *L'homme et la matière*, Paris, Albin Michel, 1943. Du même auteur : *Milieu et techniques*, Paris, Albin Michel, 1945.

6. Voir Annexe : B.2. La fibrolite est une variété massive et compacte de la sillimanite, particulièrement présente en Bretagne. C'est un matériau utilisé notamment pour l'outillage poli de la période néolithique dans l'Ouest de la France. Voir : Mikaël Guiavarc'H, Guirec Querré, Quentin Lemasson, « Analyse de la fibrolite par PIXE et XRF pour l'étude de la provenance de haches polies néolithiques de l'Ouest de la France », *Journées du "CReAAH", Archéologie, Archéosciences, Histoire, Rennes, 12-13 mai 2017. Résumés des communications*, 2017, p. 44, <https://hal-univ-rennes1.archives-ouvertes.fr/hal-01982864>(visité le 14 juillet 2021).

giques qui permettraient de caractériser certaines zones de fouille. La méthode PIXE a été largement favorisée tout en étant complétée ponctuellement par les données issues de la ionoluminescence (IBIL), laquelle permet de détecter les irrégularités et les impuretés à l'intérieur des matériaux par la détection des photons. D'objectifs bien distincts, ces deux projets partagent cependant une visée expérimentale commune, en lien indirect avec le patrimoine d'ordinaire analysé par AGLAE. Il s'agit pour l'un de nourrir la recherche nécessaire à la préservation des œuvres ; pour l'autre d'améliorer le degré de certitude des connaissances obtenues par les sciences archéologiques. Les données ainsi récoltées, croisées puis traitées nous offriront des informations utiles pour la connaissance et la protection du patrimoine, mais elles ne nous disent rien directement d'un héritage à conserver et à transmettre, et s'inscrivent donc plutôt dans sa périphérie.

Poursuivons par les données d'imagerie récoltées sur des éléments de vitraux de la cathédrale de Chartres, dont la filiation patrimoniale semble au contraire relativement évidente, tant la méconnaissance artisanale, la fragilité et l'inaccessibilité spatiale de ces objets d'étude imposent d'emblée la valeur des coupes numériques et scientifiques qui en sont faites. Ensemble exceptionnel de 93 baies datant essentiellement du XIII^e siècle, les vitraux de la cathédrale de Chartres ont fait l'objet d'un programme de restauration continu de 2009 à 2016⁷. Ce chantier a offert une occasion unique d'observer les vitraux et de les soumettre à une analyse scientifique multitechnique et inter-laboratoires, dont celui du C2RMF avec AGLAE. Les expériences menées à l'accélérateur découlent toujours d'une question posée, explicitement formulée lors de la demande administrative de temps de faisceau⁸. En l'occurrence, celle qui a guidé l'analyse des vitraux concernait la recette du lavis utilisée pour les carnations et les cheveux : s'agissait-il de grisaille – mélange à base d'oxydes de cuivre et/ou de fer dans du vin ou de l'urine - diluée ou avait-il été réalisé à partir de silicate de plomb ? En effet, si la technique de réalisation du décor du vitrail a été consignée au XII^e siècle, elle supposait deux recettes de lavis possibles. La méthode PIXE, par rayons X, a permis de distinguer nettement les deux recettes en représentant les répartitions spatiales du calcium, du fer et du plomb⁹. La complémentarité des données recueillies simultanément par PIXE et RBS (figure c) a été le seul moyen de déterminer la profondeur à laquelle se trouve le plomb par rapport à la surface. Ce profilage élémentaire en profondeur livre des indices extrêmement précieux sur les procédés de fabrication des vitraux, impossibles à obtenir autrement que par l'ingénierie scientifique.

Si nous pouvons envisager la reconnaissance de la valeur voire de la nature patrimoniale de ces données, en raison de leur contexte de production et des informations uniques qu'elles livrent, ce n'est que sous réserve de prendre en compte leur nécessaire entrecroise-

7. Rapport sur la campagne de restauration menée sur la cathédrale Notre-Dame de Chartres (Eure-et-Loir), 2009-2016, Ministère de la Culture et de la Communication, accessible en ligne : <https://www.culture.gouv.fr>

8. Voir Annexe : B.1

9. Voir figure 2.1

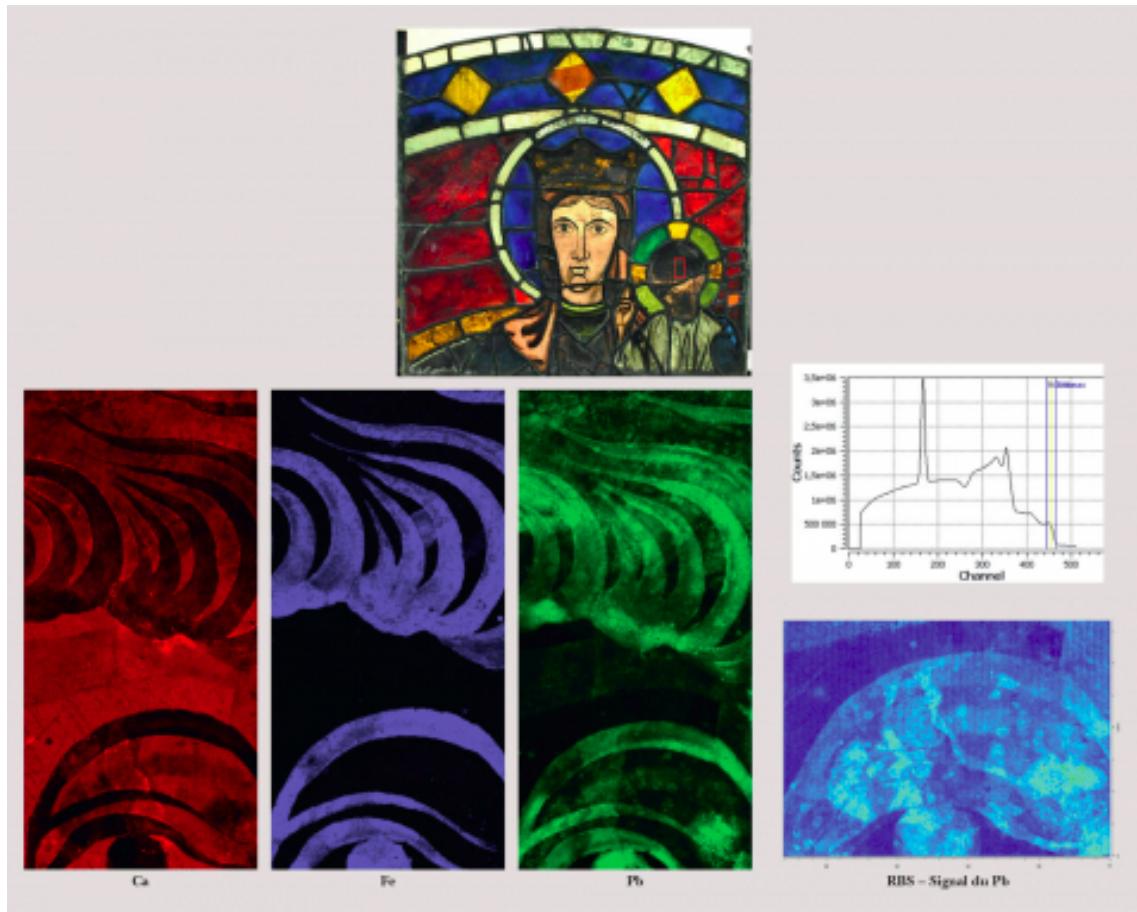


FIGURE 2.1 – Cartographie de 6 x 3 cm² réalisée sur un vitrail de la cathédrale de Chartres du XII-XIII^e siècles. ©Claire Pacheco, Quentin Lemasson, *et al.*, « D'AGLAE à New AGLAE... »

- a) Le cadre rouge localise la zone cartographiée sur le vitrail.b) Cartographie PIXE de cette zone, donnant la répartition des éléments calcium, fer et plomb. c) Spectre RBS global du vitrail. En bleu, la zone correspondant à la bande de rétrodiffusion du plomb. d) Cartographie RBS de la répartition spatiale du plomb dans la partie inférieure de la figure 3b.

ment avec les données obtenues simultanément par les autres méthodes d'analyse. Ainsi, les données PIXE obtenues constituent la base nécessaire à partir de laquelle les données livrées grâce aux détecteurs des autres méthodes peuvent être rendues intelligibles. Par conséquent, les données PIGE consacrées à l'étude des éléments plus légers d'un matériau, et les données RBS dédiées à l'analyse de l'évolution des concentrations sur le trajet du faisceau, n'ont aucun intérêt si elles ne sont pas traitées à la lumière des données PIXE. Quant à ces dernières, elles permettent la quantification de tous les éléments de Na (sodium) à U (uranium) mais considèrent le matériau comme homogène en général, et ne permettent donc pas de connaître en profondeur la structure de l'objet analysé. Si la patrimonialisation des données d'AGLAE peut être legitimée par les informations qu'elles nous livrent de façon singulière et unique, nous devons les envisager obligatoirement sous l'angle de leur association complémentaire.

2.2 L'ingénierie informatique au service du nécessaire entrecroisement des techniques d'analyse

L'ingénierie informatique et numérique, en particulier par l'intermédiaire de la conception et de l'utilisation des logiciels, joue un rôle déterminant dans la viabilité et le degré de valeur de cette association. Nous devons en effet garder à l'esprit que les sciences, et plus spécifiquement celles du patrimoine en raison de leur charge morale, sont tenues à un impératif de rigueur et de réduction maximale de la marge d'erreur. Tout comme la contrefaçon – ignorée ou connue par l'institution - contrevient gravement à la réputation d'un musée, des données erronées dévaluent la crédibilité et la valeur de l'intégralité des résultats d'une expérience menée. En cela, les logiciels de traitement et de conversion d'AGLAE ne sont plus seulement des outils mais également des creusets de patrimonialisation des données, auxquelles ils confèrent corps et dimension cognitive. D'une certaine façon, ce sont donc moins les données qui « lèvent le voile de l'invisible » que les logiciels nécessaires à leur traitement. AGLAE dispose de plusieurs outils quantitatifs de simulation et de traitement des spectres¹⁰, mettant à profit l'amélioration collective et continue des données fondamentales nécessaires au calcul¹¹. Des codes de calcul performants ont été développés et utilisés à AGLAE à travers les programmes GUPIXWIN pour PIXE, SIMNRA pour RBS, PIGE et NRA (*Nuclear Reaction Analysis*), ainsi que DATAFURNACE pour RBS et NRA. Ils permettent d'obtenir les concentrations élémentaires et les profils en profondeur à partir des spectres IBA avec une précision de quelques pourcents :

Leur principe consiste généralement à calculer le spectre produit par l'échantillon et à optimiser à l'aide de méthodes non linéaires les paramètres de la simulation, notamment la composition et la structure de la cible, jusqu'à ce que le spectre simulé coïncide avec le spectre expérimental.¹²

Chaque groupe de données (PIGE, PIXE, RBS ou IBIL) est pris en charge par un outil développé par l'équipe d'AGLAE pour ses usages particuliers et faisant appel à un moteur de calcul spécifique. Les données PIXE sont ainsi traitées par le logiciel « maison » TRAUPIXE qui fait appel au moteur de calcul GUPIXWin. Dans le cadre de l'extraction et du traitement quantitatif, les données converties par les logiciels ad hoc sont ensuite extraites en format EDF (*ESRF Data Format*) permettant leur manipulation par des outils en accès libre tels que PyMCA ou des logiciels « maison » de traitement et/ou de visualisation des données. À titre d'exemple, le logiciel AGLAEMap permet de sélectionner et de visualiser rapidement les données enregistrées par chaque détecteur sur une même zone, qu'il s'agisse de la méthode PIXE, PIGE, RBS ou IBIL.

10. Voir Annexe : A.3

11. Parmi ces données fondamentales de calcul, nous pouvons notamment relever la perte d'énergie des ions dans la matière, les coefficients d'atténuation des rayons X et Y, les sections efficaces d'interaction, etc. [Thomas Calligaro et Claire Pacheco, « Un accélérateur de particules... », p. 18.]

12. *Ibid.*

2.2. L'INGÉNIERIE INFORMATIQUE AU SERVICE DU NÉCESSAIRE ENTRECROISEMENT DES

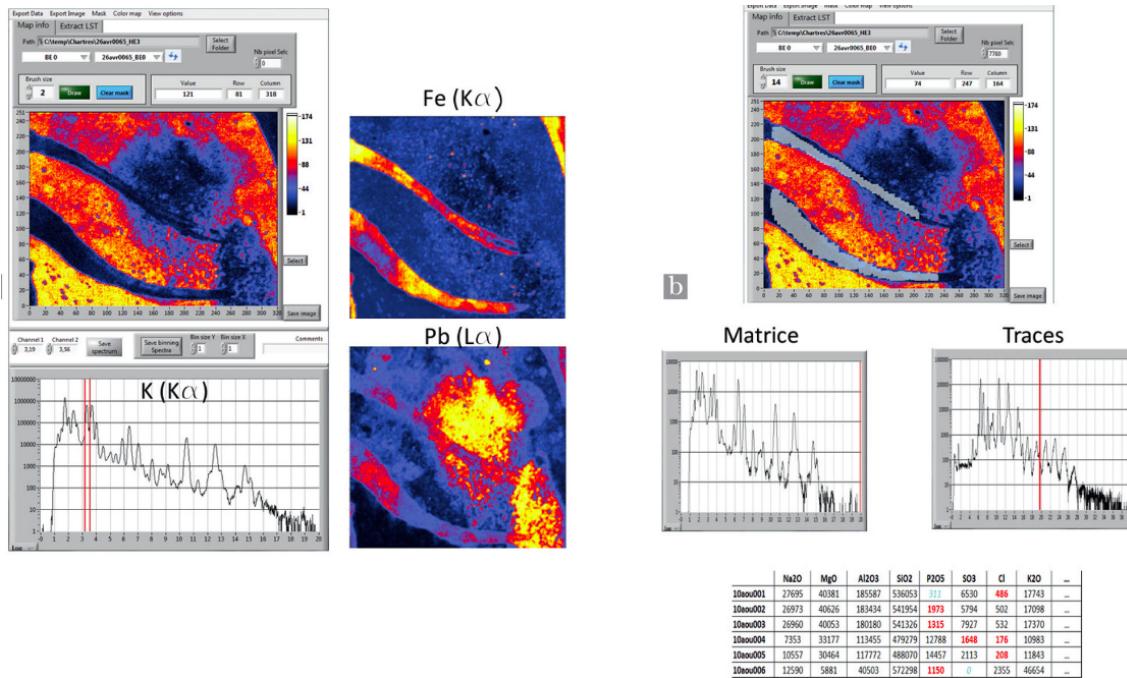


FIGURE 2.2 – Exemple de traitement de données réalisable par le logiciel **AGLAEMap**. @Claire Pacheco, Quentin Lemasson, *et al.*, « D'AGLAE à New AGLAE... »

a) Sur le spectre PIXE de la zone analysée, un pic est sélectionné. La cartographie se met alors à jour et révèle la répartition spatiale de l'élément choisi. Ici, trois éléments ont été choisis l'un après l'autre (K, Fe et Pb). b) Sur la cartographie, une zone d'intérêt est sélectionnée (en gris sur l'image). Les données enregistrées par chaque détecteur sur cette zone-là uniquement peuvent être traitées comme des analyses ponctuelles.

Un spectre somme inhérent à cette zone s'affiche, permettant à l'utilisateur de sélectionner un pic du spectre, la répartition spatiale de l'élément correspondant se mettant automatiquement à jour. Lorsqu'une zone spécifique fait l'objet de l'étude, l'utilisateur peut sélectionner les pixels de cette zone et enregistrer les données de chaque détecteur pour cette dernière. Le traitement des données s'effectue alors comme pour une analyse ponctuelle¹³. Précisons que dans le cas de l'imagerie, l'enjeu de révélation des caractères invisibles de l'objet patrimonial se traduit par une dimension visuelle particulièrement marquée et significative. À l'instar de l'archéologue qui exhume progressivement les contours de sa trouvaille, l'analyste d'AGLAE fait œuvre de découverte en observant les résultats cartographiés obtenus par une conversion et un traitement croisé rigoureux des données. La cartographie PIXE a ainsi révélé des traces de gravure volontairement effacées sur une plaque en argent du XVI^e siècle, insérée sous un centre de table réalisé par l'orfèvre Wenzel Jamnitzer.

Jusqu'alors, aucune autre technique mise en œuvre – qu'il s'agisse de la radiographie ou de la photographie en incidence rasante – n'avait permis de lire le texte. Non seulement les analyses d'AGLAE ont offert la possibilité d'identifier des lettres invisibles à l'œil nu

13. Claire Pacheco, Quentin Lemasson, Brice Moignard, Laurent Pichon, Marie Radepont et Didier Gourier, « D'AGLAE à New AGLAE... », p. 65.

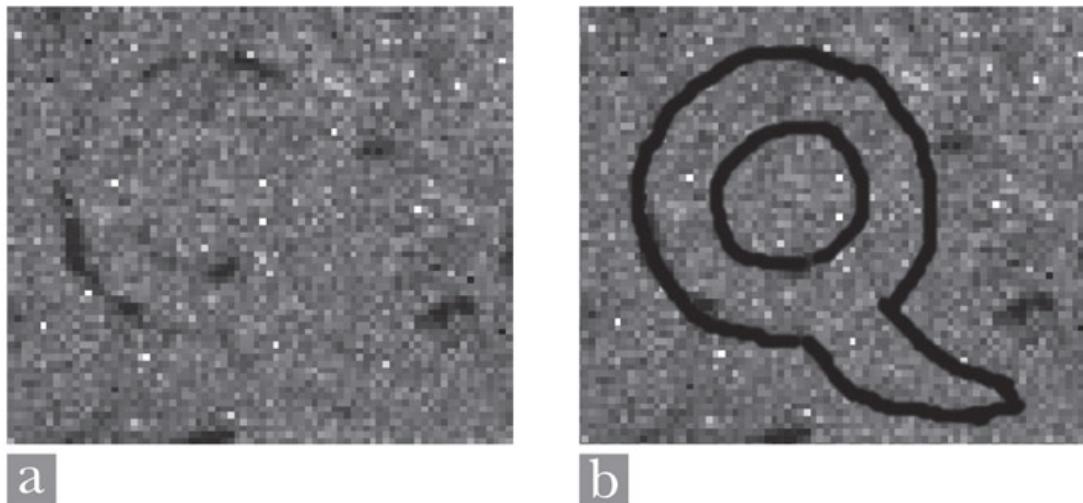


FIGURE 2.3 – Cartographie de la raie Ka du cuivre obtenue sur le spectre somme PIXE de haute énergie. Le contour d'une lettre se distingue et révèle la lettre Q. @Claire Pacheco, Quentin Lemasson, *et al.*, « D'AGLAE à New AGLAE...

et irréductibles à plusieurs techniques d'analyse de pointe, mais elles ont pu également valider plusieurs hypothèses quant à certains procédés de création tels que le mode de gravure à l'eau-forte et la technique utilisée pour supprimer l'inscription. De la même façon, l'association des méthodes RBS et PIXE a révélé le secret du lustre des céramiques médiévales. Les données RBS traduites par le logiciel SIMNRA et associées aux données PIXE ont en effet servi à déterminer la distribution de nanoparticules d'argent au sein de la glaçure¹⁴.

Capables d'exprimer la nature et le comportement de la matière des matériaux utilisés pour les objets patrimoniaux, de façon non invasive, les données d'AGLAE ne se contentent pas de donner accès à des dimensions jusque-là invisibles de notre patrimoine. Parce que ce dernier est irremplaçable et fragile, il est voué à être préservé de toute forme de prélèvement qui altérerait son intégrité et son état de conservation. Ainsi, les données issues des techniques d'analyse non-invasives et non-destructives sont seules à pouvoir nous transmettre des informations normalement condamnées à demeurer dans les profondeurs inaccessibles de la matière. En cela, elles contribuent bien à la découverte et à la contemplation inhérentes au patrimoine, mais également à sa transmission multidimensionnelle aux générations futures. Par conséquent, ce que nous disons des données d'AGLAE au regard du patrimoine est également valable pour d'autres techniques analytiques émergentes. Dans la chaîne des techniques et des instruments mis au service du patrimoine, les méthodes légères et mobiles telles que la fluorescence de rayons X portable (p-XRF), la spectroscopie Raman et la spectrométrie d'émission atomique induite par laser (LIBS) livrent ainsi d'une autre façon des informations uniques et précieuses,

14. Thomas Calligaro et Claire Pacheco, « Un accélérateur de particules... », p. 18.

2.2. L'INGÉNIERIE INFORMATIQUE AU SERVICE DU NÉCESSAIRE ENTRECROISEMENT DES

en offrant de surcroît une possibilité d'analyse *in situ* dans les musées ou sur les sites de fouille. Dans ce dernier cas de figure, elles permettent une quasi-simultanéité entre la trouvaille archéologique et la découverte de ses propriétés physico-chimiques.

Chapitre 3

D'ESPADON à IPERION-HS : un cadre d' institutionnalisation multiscalaire

Le questionnement de la nature patrimoniale intrinsèque des données d'AGLAE mène inévitablement à la question de leur institutionnalisation. Ce dernier terme renvoie au fait que des données générées au sein d'une institution à vocation patrimoniale pourraient devenir elles-mêmes constitutives du patrimoine en raison de leur cadre de production. Or, dans notre cas, le C2RMF dont dépend AGLAE n'est pas à proprement parler une institution chargée de conserver du patrimoine, même si l'établissement assure des missions fondamentales de soutien à la conservation et à la restauration des biens patrimoniaux. Les différentes techniques d'analyse regroupées par son laboratoire s'appliquent au patrimoine, et produisent des données qui lui sont directement liées. Mais cette capillarité suffirait-elle à conférer une valeur patrimoniale à ces dernières ? Tout ce qui émane d'un bien patrimonial et l'environne est-il voué à en partager la reconnaissance et la préservation ? Où se situe désormais la limite entre le patrimoine et son environnement sémantique ? Sans apporter de réponse univoque à cette interrogation conceptuellement et culturellement complexe, nous pouvons approfondir l'exploration de la frontière qu'elle induit entre la donnée et son statut. Cette étape de réflexion sera l'occasion d'élargir notre définition de l'« institutionnalisation » en l'enrichissant de la notion d'identification collective.

3.1 ESPADON : un EquipEX en quête de l'« objet patrimonial augmenté »

Dans sa réflexion sur la dynamique de patrimonialisation du cinéma, Christophe Gauthier retient l'institutionnalisation comme l'une des phases indispensables du proces-

sus de transformation d'un objet – au sens générique du terme – en bien patrimonial¹. Si l'on considère l'action d'institutionnaliser comme un processus de placement d'une entité dans une organisation patrimoniale, nous pouvons en déduire que la patrimonialisation relève avant tout d'un rite collectif de reconnaissance et d'« estampillage » symbolique par une collectivité patrimoniale légitimante. La clarté de ce processus tend cependant à se dissoudre lorsque cette collectivité se diffracte dans une forte imbrication d'échelles et se divise en de multiples institutions et regroupements extrêmement hétérogènes par leur nature, leurs statuts et leurs visées. Ainsi, les données générées par AGLAE font l'objet d'une institutionnalisation multiscalaire selon une dynamique ascendante, depuis le Centre de Restauration des Musées de France (C2RMF) jusqu'à IPERION-HS et E-RIHS en passant par ESPADON.

La dénomination recouverte par ce dernier sigle mérite une attention particulière : « En Sciences du Patrimoine, l'Analyse Dynamique des Objets anciens et Numériques ». Sous couvert d'une mise en relief de l'événement que constitue la technique expérimentale appliquée au patrimoine, l'ambition de ce projet d'Équipement d'Excellence (EquipEx) est en réalité moins technologique que documentaire et sémiologique. Porté par la Fondation des Sciences du patrimoine et sélectionné par le Programme d'Investissements d'Avenir (PIA 3), ESPADON doit créer « l'objet patrimonial augmenté »² selon une double sémantique : scientifique d'une part, avec le perfectionnement des techniques d'analyse par l'amélioration des instruments existants ; numérique d'autre part, avec la mutualisation et l'enrichissement des données produites. Les fonds alloués financent actuellement la mise en place d'un réseau d'appareils travaillant à différentes échelles et à différents niveaux de mobilité pour transcender les difficultés logistiques d'acheminement des œuvres. À terme, ESPADON doit mettre à la disposition des communautés de chercheurs, des restaurateurs, des acteurs du patrimoine et des scientifiques de nouveaux moyens instrumentaux de tomographies 3D multi-échelles et d'imagerie multi-physics 2D, mais également des ressources et savoir-faire numériques pour le traitement, la gestion et le stockage de données massives. Par ailleurs, la mise en place d'un système optimisé d'archivage pour les données brutes et les données semi-intermédiaires a été appréhendée comme support de restitution de cet « objet patrimonial augmenté ». La structuration effective d'un réseau de chercheurs et de laboratoires à travers le territoire français et, plus largement, européen constituera le maillage communautaire indispensable à cette transmutation patrimoniale. Outre le noyau partenarial naturel formé par la réunion des laboratoires du Centre de Recherche sur la Conservation (CRC), du C2RMF et du Laboratoire de recherche des monuments historiques (LRMH), de nombreux laboratoires, centres de recherche et d'enseignement supérieur français se sont associés au projet, de façon plus ou moins étroite.

1. Christophe Gauthier, *Patrimoine et patrimonialisation du cinéma*, Paris, École nationale des Chartes, 2020, <https://hal-enc.archives-ouvertes.fr/hal-02193020>

2. « ESPADON-PATRIMEX+... », *op. cit.*

Notons pour leur charge symbolique en matière patrimoniale la présence du Muséum National d'Histoire Naturelle et celle de l'Institut national du patrimoine parmi les différents partenaires officiellement recensés.

Partie organique du C2RMF, AGLAE doit *de facto* suivre le sillage tracé par ESPADON, profiter des opportunités qu'il ouvre et participer à la résolution des défis logistiques, heuristiques et sémiologiques qu'il soulève au sein de la chaîne patrimoniale. Le caractère imposé et non concerté de cette filiation n'est pas sans effet sur la fluidité et l'aisance des changements de pratiques et de paradigmes à opérer pour une mise en conformité avec les lignes d'envergure de l'EquipEx. La création récente du projet PARCOURS³ et de sa suite SoCoRe⁴ a ainsi révélé la complexité de réalisation d'une architecture organisationnelle, documentaire et sémantique apte à englober toute la diversité, les singularités et les contraintes respectives des différentes techniques d'analyse mobilisées autour du patrimoine. Nous aborderons plus spécifiquement les enjeux patrimoniaux et les modalités de la structuration sémantique des données d'AGLAE dans un chapitre ultérieur. Toutefois, sans nous attarder prématûrément sur les choix de référentiels⁵ et d'architecture des projets portés par ESPADON autour des données, il nous semble nécessaire d'appréhender leurs implications dans le processus de reconnaissance et d'identification patrimonialisantes.

Mené sous l'égide du laboratoire d'excellence PATRIMA de la Fondation des sciences du patrimoine, le projet PARCOURS s'inscrit dans le contexte de la mise en place d'un système holistique d'information pour les sciences du patrimoine, orienté vers les données de conservation-restauration et les données d'analyse. Il consiste en une ontologie⁶ d'application « dont l'objectif faciliterait le partage d'information entre les différents acteurs de la conservation-restauration »⁷. En d'autres termes, il s'agit de créer des référentiels partagés aptes à sous-tendre l'interopérabilité sémantique. Mais plutôt que d'opter en amont pour une dynamique horizontale de réflexion collective transdisciplinaire et inter-laboratoires nécessairement chronophage et incertaine, ESPADON propose, à travers PARCOURS, une ontologie construite selon une logique verticale par les spécialistes en nouvelles technologies numériques. Baptisée CIDOC-CRM Cr - les deux dernières lettres renvoyant à « conservation-restauration »-, elle a vocation à structurer une base de données RDF distribuée, laquelle impliquera une exportation des bases de données locales vers le modèle commun en RDF. Les bases EROS du C2RMF et CASTOR du LRMH sont directement

3. PAtrimoine culturel et Restauration-Conservation : Ontologie pour l'Usage d'un Référentiel commun aux différentes Sources de données

4. SoCoRe est un projet financé par l'Agence nationale de la Recherche (ANR)

5. De façon générique, les référentiels peuvent être considérés comme des vocabulaires et des listes de références assurant la cohérence des valeurs de données.

6. Une ontologie doit être comprise comme un ensemble structuré des termes et concepts représentant le sens d'un champ d'informations.

7. Présentation du projet PARCOURS, accessible en ligne : <http://projet-parcours.eu/2018/02/01/bonjour-tout-le-monde/#more-1> (visité le 18 avril 2021).

concernées, impliquant par capillarité les données achevées d'AGLAE, prochainement mutualisées et versées dans un serveur commun à l'ensemble des techniques d'analyse du C2RMF.

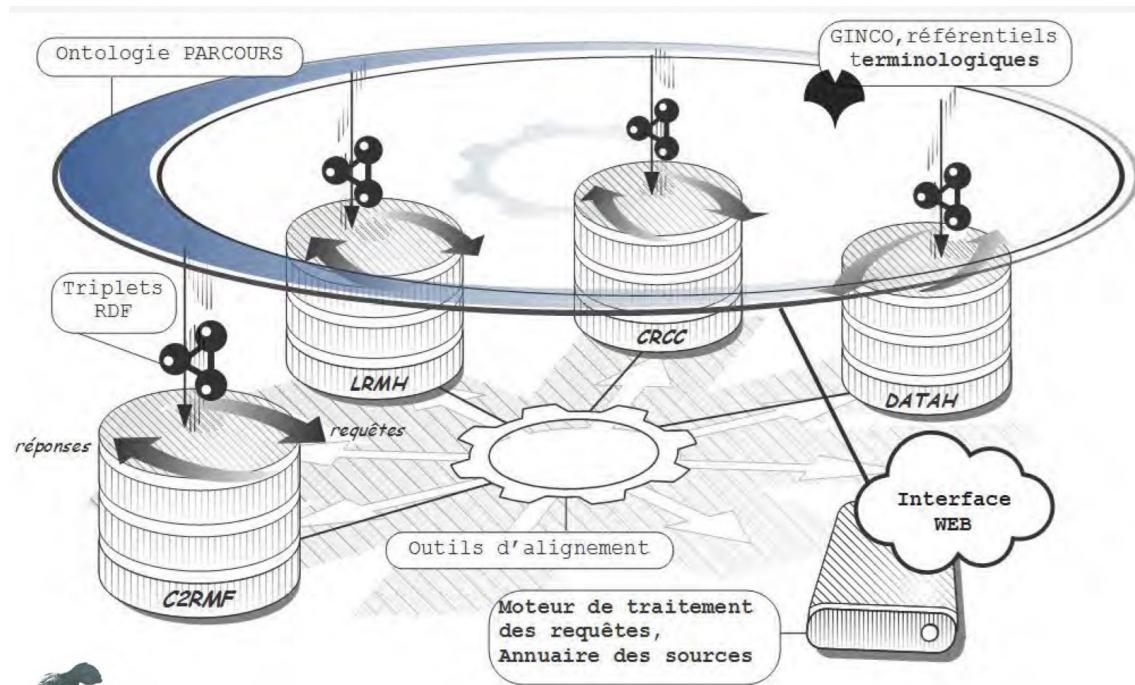


FIGURE 3.1 – Capture d'écran d'un extrait de la présentation du projet PARCOURS, avril 2021. @Luc Bouiller-C2RMF

Par son intention et sa dynamique, PARCOURS induit ainsi l'apprentissage d'un langage commun aux acteurs de la conservation-restauration, lesquels sont tenus à une forme d'acculturation nouvelle aux technologies du web sémantique et de l'information documentaire, parfois fort éloignée des usages et des prismes conceptuels ancrés de longue date dans les microsociétés scientifiques. C'est particulièrement le cas d'AGLAE qui, à l'instar de la plupart des autres accélérateurs de particules – français comme européens –, s'en est longtemps tenue à une gestion exclusivement utilitaire et restrictive des données, sans intégrer les contraintes de partage, de réutilisation et d'optimisation d'accès dans les différentes étapes de leur cycle de vie. Aussi, ce n'est pas sans une certaine réserve que sont appréhendés les changements inhérents à l'introduction dans l'« Open Science » par l'intermédiaire des dispositifs tels que PARCOURS. L'une des causes de cette résistance tient sans doute au fait que les utilisateurs d'AGLAE sont encore en voie de transition d'un écosystème informationnel centré sur le document à un écosystème orienté sur la donnée. Ils pensent ainsi davantage en termes de « dossier » et de « fichier » qu'en termes de données à proprement parler. Par ailleurs, la crainte du pillage, particulièrement marquée chez les chercheurs en sciences du patrimoine, explique pour beaucoup la défiance à l'égard du libre-accès et du partage des données. Nous pouvons relever un paradoxe inhérent à cette approche : c'est en raison d'une conscience et d'une connaissance pro-

fondes de la valeur patrimoniale des données qu'ils traitent que les chercheurs tendent à les cloisonner tels des biens précieux à protéger et préserver. Mais ce faisant, en obstruant les mécanismes d'identification et d'accès, ils entravent le processus d'institutionnalisation et de reconnaissance collective nécessaire à la mise en œuvre de la patrimonialisation de ces données rares et singulières. Nous devons malgré tout préciser que 80% des chercheurs sondés lors de la phase d'investigation d'Euphrosyne se sont déclarés prêts à modifier ces pratiques afin d'ouvrir leurs données à d'autres spécialistes.

De force ou de gré, les institutions, organismes et laboratoires intégrés dans PAR-COURS devront progressivement alimenter un courant puissant d'identification et de valorisation de cet « objet patrimonial augmenté » hybride et multifacette, à la fois matériel et immatériel, présent dans la réalité physique tout comme dans l'abstraction numérique du web de données. Il en va de la conformité déontologique avec le Plan national pour la science ouverte (juillet 2018)⁸, juridique avec la Loi pour une République numérique (octobre 2016)⁹, et européenne avec les cadres de l'*Open Science* soutenus par la Commission Européenne.

3.2 DIGILAB : une infrastructure numérique européenne à construire

Si ESPADON constitue une première échelle d'institutionnalisation singulière pour la patrimonialisation des données d'AGLAE, du fait de la dépendance fonctionnelle du laboratoire avec l'EquipEX et ses instances muséales, DIGILAB en forme la seconde selon une dynamique centrifuge. Organe d'E-RIHS, infrastructure d'excellence et d'intérêt global pour l'étude des matériaux du patrimoine, DIGILAB doit permettre l'accès à une infrastructure numérique pour le traitement des données quantitatives produites par les différents instruments, selon une politique respectueuse des principes FAIR. Il s'intègre dans un éventail de structures complémentaires au périmètre d'action bien défini : FIX-LAB, qui offre l'accès à des équipements fixes permettant des analyses physico-chimiques et structurelles sur les matériaux ; MOLAB, qui renvoie aux équipements mobiles destinés à étudier les matériaux *in situ*¹⁰ ; ARCHLAB, qui optimise l'accès au réseau d'archives scientifiques de musées et d'institutions culturelles ; EXPERTLAB, qui regroupe des panels de spécialistes pour l'initiation de projets d'étude de biens patrimoniaux intégrant une expertise pluridisciplinaire. Irriguée par la philosophie de l'*Open Science*, la volonté stratégique qui a présidé à la création de ces structures est approximativement la même

8. Le Plan national pour la Science Ouverte est accessible en ligne : <https://www.ouvrirla.science.fr/plan-national-pour-la-science-ouverte/>

9. Loi n°2016-1321 du 7 octobre 2016 pour une République numérique, accessible en ligne : <https://www.legifrance.gouv.fr/dossierlegislatif/JORFDOLE000031589829/>

10. Voir Annexe : B.3

que celle qui a conduit à l’impulsion d’ESPADON. Rien d’étonnant donc à ce que l’on retrouve – de façon plus implicite toutefois – en filigrane des objectifs cette idée d’objet patrimonial d’un genre nouveau, né de l’entrecroisement des techniques de pointe appliquées à sa matérialité et d’un traitement numérique mutualisé des données qui en sont issues.

Membre de FIXLAB et de DIGILAB, AGLAE en partage les préoccupations et les expérimentations tout en s’affirmant comme fer de lance de ces dernières. En effet, aucune équipe d’accélérateur de particules œuvrant – même partiellement – pour les sciences du patrimoine n’a pour l’heure entrepris la mue de ses données de façon à assurer leur ouverture et leur partage. ATOMKI, l’Institut de recherche nucléaire de l’Académie hongroise des sciences, produit essentiellement des fichiers ASCII non accessibles à distance. Quant à l’équipe pluridisciplinaire de SOLEIL¹¹, synchrotron de troisième génération, elle utilise le format Nexus (HDF5) et une arborescence de métadonnées adaptée, mais on constate une forte hétérogénéité de remplissage selon les lignes de faisceau. Par ailleurs, les données sont stockées sur un réseau à part et dédiées exclusivement aux utilisateurs. Il n’existe donc pas d’accès « tous publics » aux données historiques, la plupart étant conservées sur les disques durs personnels des utilisateurs. En ce sens, l’identification et la reconnaissance des données issues des analyses patrimoniales en France et à l’échelle élargie de l’Europe en est encore à sa genèse et Euphrosyne a été conçu comme un modèle à essaimer par la communauté au sens large de l’analyse par faisceau d’ions. Notons qu’il n’est pas tant question d’*Open Science* que de favoriser et d’optimiser des pratiques au long cours. Les échanges intra-européens autour des techniques d’analyse pré-existent bien à l’ouverture des données que DIGILAB s’efforce d’encourager. Par conséquent, la circulation et la reconnaissance nommée des données parachèvent un mouvement de fédération des acteurs patrimoniaux et para-patrimoniaux autour de la valorisation des analyses pratiquées sur le patrimoine.

11. Acronyme de « Source Optimisée de Lumière d’énergie Intermédiaire du LURE ».

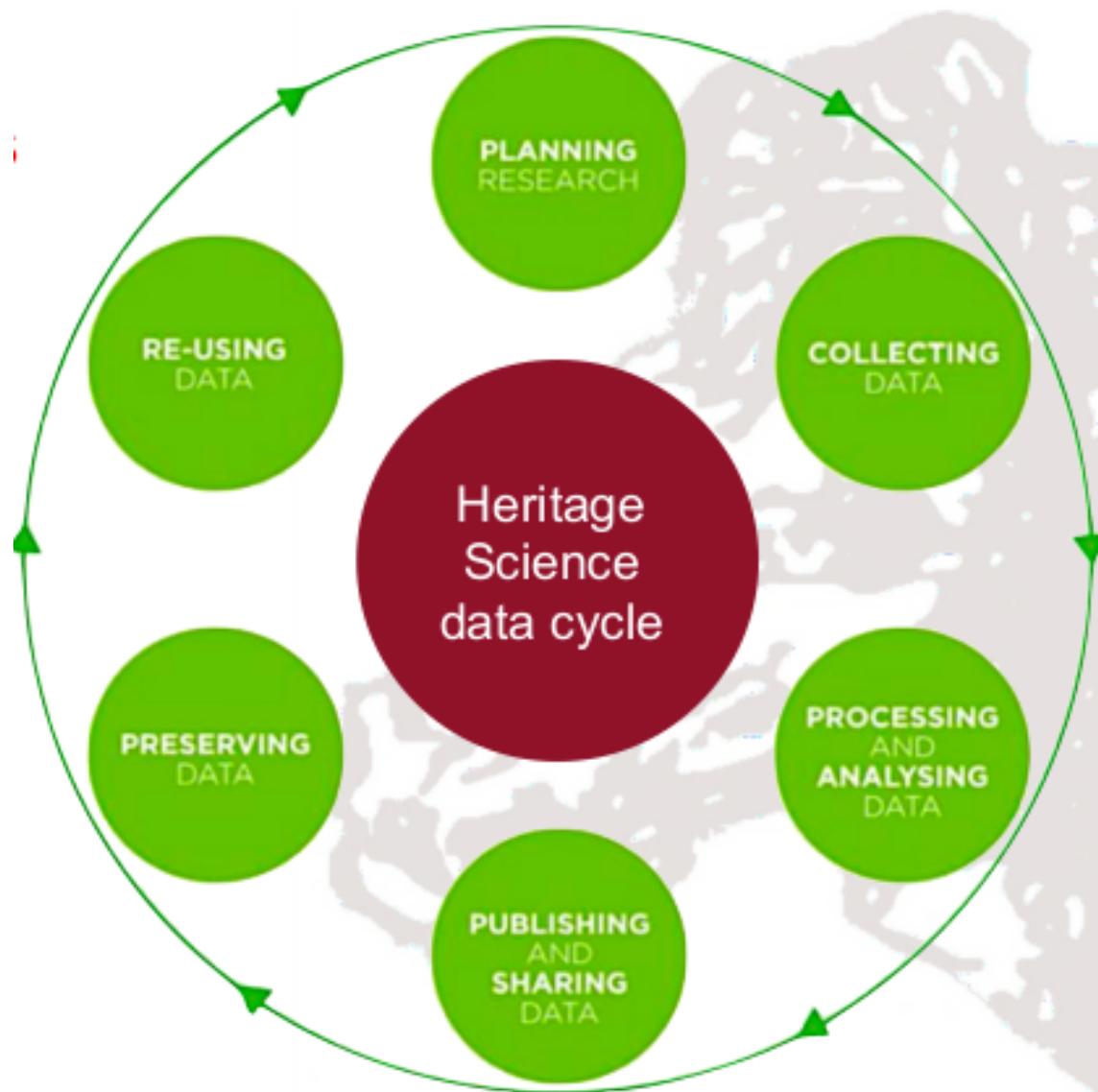


FIGURE 3.2 – Capture d'écran de la présentation introductory « DIGILAB and JRA joint meeting », 21 avril 2021. @Sorin Hermon/Vincent Detalle/Victor Etgens

Deuxième partie

**Euphrosyne : une communauté
d'utilisateurs devenue matrice de
patrimonialisation**

À la genèse du projet Euphrosyne, on trouve l'un des axes cardinaux du processus de patrimonialisation : la constitution d'une communauté qui s'agrège et trouve sa légitimité dans le sens et la valeur qu'elle reconnaît aux objets qu'elle manie, en l'occurrence ses données. Selon Emmanuelle Bermès et plusieurs théoriciens du patrimoine, c'est bien par l'existence de l'objet patrimonial que la communauté existe, et réciproquement, il n'est pas de patrimoine sans communauté pour le sous-tendre et lui faire sens¹². Par *communauté*, nous entendons bien un groupe social partageant des normes, une culture et des valeurs semblables. Elle se distingue en cela de la *collectivité* qui regroupe des individus ayant des intérêts ou des traits communs, sans nécessairement de dimension intellectuelle, affective ou socio-culturelle. La communauté implique un degré supérieur d'homogénéité des pratiques et des systèmes de pensée. Dans le cas d'AGLAE, la reconnaissance testimoniale et/ou patrimoniale de certaines données produites par les différentes techniques d'analyse est précisément ce qui a permis la transformation d'une *collectivité* de chercheurs et de spécialistes en *communauté* aux contours bien dessinés. Son existence trouve sa pleine expression dans la validation officielle et la construction du projet Euphrosyne, à travers ses phases de déploiement envisagées. Parce qu'elle induit des enjeux sémiologiques et structurels déterminants, tout en étant pensée et conçue pour les utilisateurs auxquels elle se destine, cette interface contribue à faire de la communauté AGLAE une véritable matrice de patrimonialisation. S'interroger sur les modalités de transformation de la collectivité AGLAE en communauté, notamment à travers la genèse d'Euphrosyne équivaut ainsi à remonter la source du processus de patrimonialisation de certaines données produites avec l'accélérateur.

12. Emmanuelle Bermès, *Le numérique en bibliothèque : naissance d'un patrimoine : l'exemple de la Bibliothèque Nationale de France (1997-2019)*, Paris, École nationale des chartes, 2020, p. 66.

Chapitre 4

De la *collectivité* à la *communauté* AGLAE

Au principe de l'éclosion du groupe initial d'AGLAE constitué de l'équipe et des utilisateurs, il y a d'abord une volonté politique et stratégique mûrie dans les cercles de décision gouvernementaux¹. Il est essentiel de restituer précisément ce contexte génétique avant de nous attarder sur les développements récents qui ont conduit à la création d'une start-up d'État dédiée au financement et à la conception d'une interface exclusivement consacrée – dans un premier temps - à la communauté d'AGLAE.

4.1 Le Laboratoire de recherche des musées de France, incubateur d'une collectivité singulière

Au début des années 1980, AGLAE doit sa création à un pari tant scientifique que culturel sur les apports des techniques d'analyse par faisceaux d'ions pour l'amélioration des connaissances du Patrimoine, en répondant aux besoins des conservateurs, historiens de l'art et archéologues :

L'enjeu était de réaliser de 3000 à 5000 analyses par an sur des objets, des prélèvements mais également des aérosols pour des études environnementales dans les musées. L'avantage des méthodes d'analyses par faisceaux d'ions était leur facilité d'utilisation pour des analyses panoramiques non destructives, en série, quelles que soient les natures des matériaux constitutifs des œuvres d'art et les objets archéologiques.²

1. C'est en Comité interministériel de la recherche scientifique et technique, fin 1982, qu'une première enveloppe budgétaire a été décidée, puis complétée en 1983 par le Ministère de la Recherche, et les années suivantes par le Ministère de la Culture. Laurent Fabius, Hubert Curien et Jack Lang ont joué, dans leurs ministères respectifs, un rôle décisif car ils étaient convaincus de l'intérêt de l'analyse par faisceau d'ions pour l'étude et la conservation scientifiques des biens patrimoniaux.

2. Christian Amatore, Anne Bouquillon, *et al.*, *La Chimie et l'Art...*, p. 15.

Les directeurs des musées de France successifs ont également été convaincus des émules potentiels d'un tel système d'analyse pour un laboratoire de musées. C'est ainsi qu'un projet, nommé PIXE du nom de la méthode du même nom, est lancé avec le soutien de la communauté scientifique concernée, nourri par des contacts étroits différents laboratoires du Centre national de la recherche scientifique (CNRS), de plusieurs universités, du Commissariat à l'énergie atomique (CEA) et de l'Institut national de physique nucléaire et de physique des particules (IN2P3). La conception de la machine et le choix de l'accélérateur sont réalisés en relation avec le Groupe de Physique des Solides de l'Université Paris VII et du CNRS, sous la direction du physicien Georges Amsel. Un groupe d'experts est simultanément constitué pour préciser le cahier des charges. Georges Amsel et Michel Menu, chargés de la direction scientifique et technique du projet, constituent rapidement une équipe locale dédiée en installant les premiers équipements d'AGLAE sur l'accélérateur du campus de Jussieu. Nous pouvons souligner que si l'embryogenèse d'AGLAE est de nature politique, sa genèse appartient bien d'emblée aux physiciens et aux spécialistes des sciences du patrimoine.

En 1985, une étape de consolidation est ensuite franchie grâce au rapport du scientifique Roger Bird, spécialisé dans la physique des neutrons et le développement des méthodes IBA (*Ion Beam analysis*). Son étude exhaustive sur les applications de ces dernières, suivie d'un colloque international sur le même thème à l'initiative du LRMF, a confirmé le caractère incontournable des analyses par faisceau d'ions pour l'étude non-invasive de trésors patrimoniaux tels que la Bible de Gutenberg. À partir de 1987, des travaux sont menés sous l'autorité de l'établissement public du Grand Louvre, afin de créer les locaux définitifs d'AGLAE dans les souterrains du LRMF, en proximité étroite avec les espaces muséaux. Cet environnement privilégié a largement favorisé l'interdisciplinarité scientifique et la possibilité de disposer de différentes méthodes de caractérisation physico-chimique de la matière. Initialement, l'accélérateur a été conçu pour permettre à terme une diversification des techniques, et devait comprendre en plus de son microfaisceau un spectromètre de masse pour la datation radiocarbone. Ce second dispositif est finalement installé à Saclay en 2003, au Laboratoire de mesure du carbone 14.

Livrés en 1988, les éléments de l'accélérateur sont ensuite assemblés et montés l'année suivante, ouvrant une période de tests et de réglages de deux ans. Durant cette période, la « collectivité » AGLAE n'existe pas encore. Nous pouvons tout au plus évoquer un « groupe » qui ne renvoie alors qu'à l'équipe composée de physiciens nucléaires, d'ingénieurs d'étude et de recherche spécialisés dans les développements mécaniques, l'électronique et la détection. Toutefois, cette première phase de mise en place, marquée par un constant dialogue entre les techniciens, les informaticiens, les physiciens, les chimistes et les spécialistes de la conservation et de la restauration patrimoniales, érige les fondements et impulse la dynamique originelle de la collectivité singulière qui se constituera par la suite autour d'AGLAE. Des objets du Musée du Louvre inaugurent l'activité de la

ligne de faisceau, afin de caractériser les matières utilisées ainsi que leurs origines. Sont ainsi analysées des pièces issues des collections de l'Égypte ancienne et des Antiquités orientales, pour ne citer qu'elles³. Dès les débuts d'AGLAE, la complexité matérielle et multidimensionnelle des matériaux du patrimoine impose un dialogue constant entre les scientifiques, les ingénieurs et les spécialistes du patrimoine, les incitant à trouver un système commun de références et de normes pour analyser les objets et en exploiter les résultats avec pertinence. En effet, la composition des objets est souvent hétérogène puisqu'elle peut être en grains, en couches ou composite avec une association de matière organique et inorganique. De plus, leur surface, leur dimension et leur forme accroissent le défi technique et scientifique de leur analyse. Par un phénomène d'acculturation réciproque, les scientifiques et techniciens d'AGLAE se sont ainsi appropriés la terminologie et les contraintes du champ patrimonial ; tandis que les conservateurs, les archéologues et les historiens de l'art ont acquis les connaissances élémentaires nécessaires à la définition du contexte expérimental souhaité. De l'efficience de cette acculturation dépend le succès de l'expérience. Bien sûr, l'expertise et la connaissance fine des paramètres des techniques d'analyse demeurent l'apanage des analystes d'AGLAE et des spécialistes en sciences des matériaux (chimistes, physiciens, etc.), induisant une hétérogénéité de maîtrise relativement forte parmi les utilisateurs. Plusieurs cercles distincts se sont ainsi progressivement dessinés entre ceux, totalement autonomes, qui ne sollicitent l'appui de l'équipe d'AGLAE qu'en cas de besoin ou difficulté ; les chercheurs partiellement autonomes qui ne maîtrisent qu'une partie des étapes de la chaîne expérimentale, et ceux totalement extérieurs au processus d'acquisition et de traitement des résultats.

4.2 De New AGLAE à Euphrosyne : enrichissement et consolidation identitaire

À partir de 1995, des améliorations successives sont apportées à l'accélérateur, jusqu'à l'inauguration de New AGLAE en 2017. Du ressort de la technique et de l'ingénierie, ces modifications influent étroitement sur le périmètre de la collectivité AGLAE, dans un mouvement de co-construction que nous devons éclairer.

Les ingénieurs et les chercheurs améliorent d'abord le transport du faisceau jusqu'à l'objet à analyser, simultanément à l'optimisation de la géométrie d'extraction, grâce à la conception d'un nouveau système plus compact en laiton. D'autres obstacles aux performances d'AGLAE pour l'analyse du patrimoine émergent rapidement dans la réalité quotidienne des usages. L'absence d'automatisation de l'accélérateur et le manque de stabilité du faisceau sont sans doute parmi les plus problématiques puisqu'ils limitent son temps d'utilisation à la journée et ne permettent pas l'analyse en toute sécurité d'œuvres fragiles.

3. *Ibid.*, p. 25.

Les utilisateurs se heurtent alors également à la difficulté de réaliser des cartographies chimiques systématiques, ainsi qu'à la nécessité de renforcer les conditions de radioprotection lors des tirs de faisceau de noyaux de deutérium. La candidature d'AGLAE à un appel d'offre en Équipement d'Excellence (EquipEX), remporté en 2012, s'inscrit directement dans ce nœud de contraintes à résoudre et pose le premier jalon de ce qui deviendra New AGLAE. Financés par le programme d'investissements d'avenir de l'Agence nationale de la recherche (ANR-10-EQPX-22), le Ministère de la Culture, le CNRS, Rennes métropole et la Mairie de Paris, les travaux de stabilisation du faisceau et d'automatisation de la ligne sont réalisés de 2016 à 2017 en collaboration avec la société Thalès. Ils entraînent l'arrêt forcé d'AGLAE et son inaccessibilité aux équipes de recherche pendant plusieurs mois. En contrepartie, leur achèvement et l'inauguration de New AGLAE en 2017 transforment et facilitent significativement le cadre d'expérience en permettant un accès 24h/24, 5 jours sur 7. Pour des contraintes liées à la sécurité et aux polices d'assurance, la journée reste consacrée aux œuvres tandis que la nuit est dédiée à l'analyse automatisée de séries d'échantillons ou à des cartographies induisant des temps d'acquisition longs. Ce mode de fonctionnement est introduit avec succès et perdure jusqu'à aujourd'hui. Par ailleurs, un nouveau système de balayage du faisceau autorise l'étude d'objets du patrimoine dont les zones d'intérêt peuvent atteindre quelques centimètres carrés. Enfin, les processus d'acquisition des données d'imagerie ont été largement optimisés, de même que l'association des différentes techniques d'analyse et le développement de la ionoluminescence (IBIL).

Tous ces développements instrumentaux et méthodologiques ont toujours eu pour objectif l'amélioration continue des performances pour l'étude des objets du patrimoine, en prenant en compte leurs spécificités et leurs contraintes particulières telles qu'une forte incurvation, ou l'impossibilité éventuelle d'acheminement jusqu'à la ligne de faisceau. Les scientifiques et les techniciens se sont ainsi pleinement adaptés aux enjeux parfois lourds et contraignants de la préservation patrimoniale. L'esprit d'invention et d'innovation scientifique s'en est trouvé sensiblement affûté, mais toujours dans une perspective de mise au service du patrimoine et de ses acteurs. Cette vocation est d'ailleurs explicitement énoncée et régulièrement réaffirmée par l'équipe d'AGLAE qui déclare concevoir et développer « l'instrument et les méthodes d'analyse par faisceau d'ions dans le but d'en repousser les limites actuelles pour les applications patrimoniales »⁴. L'approfondissement scientifique du couplage des méthodes PIXE et PIGE lors des développements de New AGLAE⁵ a ainsi nourri des interactions particulièrement fécondes entre les spécialistes de l'analyse par faisceau d'ions et les « connasseurs » du patrimoine, toutes disciplines confondues, qui ont vu dans ces expérimentations la promesse de nouvelles voies d'exploration, utiles tant pour l'appréhension cognitive du patrimoine que pour sa préservation matérielle.

4. Claire Pacheco, Quentin Lemasson, *et al.*, « D'AGLAE à New AGLAE », p. 68.

5. Cet approfondissement est toujours en cours avec le projet QuEL@AGLAE, dédié à la mise en place de nouveaux détecteurs gamma et financé par la région Ile-de-France, par l'intermédiaire de la DIM-MAP (Domaine d'Intérêt Majeur-Matériaux anciens et patrimoniaux).

Les améliorations et les perfectionnements apportés à AGLAE entraînent dans leur sillage l’élargissement symbiotique des cercles d’utilisateurs et de collaborateurs de l’accélérateur, guidés par une communauté croissante d’objectifs et de valeurs.

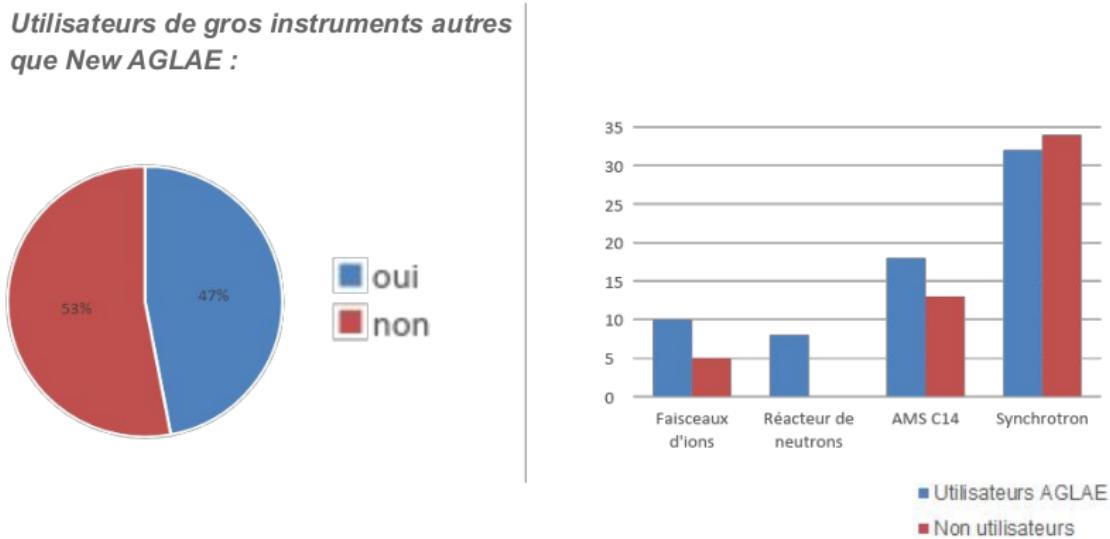


FIGURE 4.1 – Illustration extraite de la présentation d’audience d’Euphrosyne pour l’obtention du statut de start-up d’État du Ministère de la Culture, 8 avril 2021. @Ned Baldessin/Jonathan Bourguignon/Claire Pacheco

En permettant à de nouveaux projets de recherche français et européens de voir le jour, New AGLAE facilite la conjonction d’idées, de pratiques et de normes entre les équipes et les utilisateurs des autres accélérateurs travaillant - au moins partiellement - à l’étude du patrimoine. Ainsi, 47% des utilisateurs sondés lors de la phase de prospection préparatoire du projet Euphrosyne ont indiqué l’usage de gros instruments autres que New AGLAE, avec une vaste prédominance des expérimentateurs Synchrotron (voir figure ci-dessus). Dans le cadre de la grande infrastructure de recherche⁶ IPERION-CH (devenue IPERION-HS), AGLAE s’engage dans des activités de recherche au côté d’IPANEMA et de BNC-Wigner afin d’optimiser, lors d’une cartographie, la présentation et le positionnement d’une cible devant le faisceau. Dans une démarche de réduction des risques, AGLAE collabore également avec ATOMKI pour l’étude des modifications des matériaux induites par faisceau d’ions, en vue de permettre l’étude de matériaux fragiles jusqu’alors potentiellement altérés par l’irradiation. D’expérience en expérimentation, une communauté sans la lettre affleure ainsi parmi celle, plus large, des physiciens nucléaires, tout en agrégant les flux d’idées et de pratiques des disciplines dédiées à la conservation et à la connaissance patrimoniales. C’est ainsi que des spécialistes du patrimoine que l’on considère issus des « sciences molles » s’insèrent, à géométrie variable, dans les cercles hautement

6. Selon la Commission européenne, les infrastructures de recherche sont des installations, des ressources et des services qui sont utilisés par les communautés scientifiques pour mener leurs recherches et favoriser l’innovation : grands instruments scientifiques, ressources cognitives, infrastructures électriques et informatiques.

technico-scientifiques d'AGLAE. Notons que cette dynamique n'est pas vraiment propre à AGLAE et qu'elle touche l'ensemble des sciences appliquées au patrimoine, interdisciplinaires par essence, et par extension, vouées à transcender les cloisonnements conceptuels et professionnels entre les disciplines mobilisées. Toutefois, la singularité d'AGLAE, seul accélérateur de particules au monde exclusivement dédié à l'analyse du patrimoine, renforce la spécificité identitaire de sa communauté. Cette distinction s'exprime d'abord par une terminologie construite et partagée au fil des usages et des échanges, limon fertile de référentiels toujours en cours de construction. Lors des expériences menées à AGLAE au cours de la période du stage, il était ainsi particulièrement difficile de deviner le profil des chercheurs présents, tant le dialogue avec l'équipe d'AGLAE se nourrit de notions et de termes communs à tous les utilisateurs, puisés dans une terminologie hybride où les mots de l'analyse par faisceau d'ions côtoient ceux de la description et de l'identification patrimoniales. Les éléments chimiques recherchés jouent un rôle pivot dans cette acculturation, car ils constituent une base d'alignement et de jonction sémantique relativement accessible. Dans une complémentarité féconde, les uns – spécialistes du patrimoine – indiquent quoi rechercher, les autres – ingénieurs et/ou spécialistes en sciences des matériaux – , savent comment y parvenir. Marqueur identitaire solide, ce langage commun n'est pas le seul vecteur communautaire d'AGLAE.

L'absence de cloisonnement pyramidal entre les ressources d'AGLAE – équipe et équipements inclus – et les utilisateurs explique sans doute pour beaucoup leur appropriation souvent familière. Les responsables d'expérience autonomes ou semi-autonomes bénéficient en effet d'un accès non restreint à la machine et conservent la maîtrise de l'installation de l'objet-cible devant le faisceau. Les statistiques réalisées pour Euphrosyne ont révélé l'importante proportion de ces expérimentateurs qui se substituent aisément aux analystes de l'équipe d'AGLAE. Sur les 75 répondants, presque 70% d'entre eux indiquent être en capacité de mener l'expérience seuls. Précisons que ce panel se décompose ainsi : 80% sont chercheurs et/ou ingénieurs spécialistes en sciences dures ; 20% sont archéologues, un tiers et 90% sont issus de laboratoires, d'instituts de recherche ou d'universités⁷. Ces résultats doivent cependant être interprétés avec prudence et approfondis. La catégorie « spécialistes en sciences dures » recouvre implicitement un certain nombre de profils interdisciplinaires à la croisée de l'histoire de l'art ou de l'histoire et des sciences appliquées. Claire Pacheco, cheffe du groupe New AGLAE, est ainsi elle-même physicienne spécialisée dans les archéomatériaux⁸. Malgré cette interdisciplinarité représentée et valorisée au sein d'AGLAE, on constate que les historiens de l'art et les conservateurs issus des formations en sciences humaines ne se sont pas appropriés les usages d'AGLAE

7. Ned Baldessin, Jonathan Bourguignon, Claire Pacheco, *Présentation d'audience d'Euphrosyne pour l'obtention du statut de start-up d'État du Ministère de la Culture*, 8 avril 2021.

8. Claire Pacheco est l'auteure d'une thèse intitulée « Étude des films d'or sur matière vitreuse : application à la céramique glaçurée islamique médiévale : Asie centrale XIVe - XVe siècles – Iran XIIe – XIIIe siècles », sous la direction de Rémy Chapoulie, soutenue en 2007 à l'Université de Bordeaux III.

avec la même aisance que les autres utilisateurs. Cette maîtrise moindre ne les exclut pas pour autant de l'expérience, l'étape d'interprétation des résultats relevant plus particulièrement de leur champ de compétences. La chaîne d'acquisition, de traitement et d'interprétation des données expérimentales constitue en cela un creuset d'harmonisation conceptuelle et sémiologique, puisque la lecture historique et scientifique des résultats est interdépendante de la compréhension des choix de sélection de l'analyste, lequel n'est pas toujours le responsable du projet de recherche ou de la demande de service.

Unis autour d'AGLAE par la force de cohésion de l'expérience, les utilisateurs directs ou indirects de l'accélérateur en partagent bien une forme de culture commune, mêlant pratiques et terminologie spécifiques, dans la poursuite d'objectifs communs centrés sur la connaissance, la préservation et la transmission patrimoniales. Pour autant, c'est à travers l'appropriation et la reconnaissance de la valeur patrimoniale d'une partie des données qu'ils traitent et interprètent que naît véritablement la *communauté* d'AGLAE, auréolée d'une légitimité qui déborde du cadre strictement scientifique.

Chapitre 5

Reconnaissance et appropriation : fondement de la transformation en bien patrimonial

La singularité patrimoniale des données d'AGLAE tient notamment au fait que la reconnaissance de leur valeur est à la fois venue des utilisateurs au cours d'un processus d'appropriation complexe et progressif; et des circuits de l'information et de la documentation qui en ont fait des objets sémantiques singuliers au regard de l'étude et de la signalisation des ressources patrimoniales, notamment grâce à l'émergence de politiques holistiques de gestion de données.

5.1 Un processus d'appropriation interdépendant du cycle de vie des données

Une difficulté majeure d'appréhension de la patrimonialisation des données d'AGLAE tient en l'extrême hétérogénéité de ces dernières, dont la nature, le format et l'intérêt fluctuent en fonction des étapes de leur cycle de vie. Rappelons que notre propos vise à démontrer en quoi les usages, la culture et l'attitude des utilisateurs d'AGLAE à l'égard de certaines données innervent le processus de patrimonialisation. Cet argumentaire implique d'aborder en profondeur les différences de regard et d'appropriation induites par les états mouvants et successifs des données récoltées.

Les enquêtes menées au cours du stage auprès des utilisateurs dans le cadre des manipulations et des phases de traitement des données ont abouti à une évidence : l'appropriation intellectuelle et personnelle n'intervient pas au même moment pour tous les utilisateurs. De plus, elle s'inscrit toujours dans une dynamique rétrospective, après vérification et confrontation rigoureuse des résultats obtenus avec ceux espérés ou avec des données fiables de comparaison. Ainsi, il n'est pas rare de voir certains jeux de données

brutes et/ou intermédiaires être supprimés en cours ou en fin d'expérience, car jugés erronés ou peu satisfaisants quant aux informations obtenues. Même si AGLAE est un instrument dédié à l'étude patrimoniale, il n'en demeure pas moins un outil au service de l'expérience et de l'expérimentation, dont les résultats sont incertains par nature et souvent perfectibles. Dans le cadre temporel du *run*, il n'y a donc pas d'embrée une quelconque forme de sacralisation des données au début de la chaîne d'acquisition. Juste après le tir de la ligne de faisceau sur la cible, des fichiers LST et des fichiers spectres primaires sont générés simultanément et constituent les données brutes de l'imagerie et de l'analyse ponctuelle. Brutes par opposition aux données ensuite converties, sélectionnées puis traitées, ces données sont cependant déjà le fruit de paramètres préalablement définis par l'analyste et le(s) chercheur(s) en charge du projet – ces deux rôles se confondant souvent – tels que les choix de standards, de filtres, la DOSE par seconde ou encore l'énergie du faisceau. Dès cette première étape peuvent apparaître des anomalies de détection, des erreurs de paramétrages ou une mauvaise installation de la cible devant le faisceau. Dès lors, les premières données obtenues n'ont aucune valeur et seule l'historique de l'expérience, s'il a été correctement consigné, en restituera une trace.

Selon la technique (PIXE, PIGE, RBS, IBIL) et le mode d'analyse (ponctuelle ou imagerie) voulus, les données sont ensuite converties à l'aide des logiciels dédiés, lesquels génèrent à leur tour des fichiers voués à subir une nouvelle phase de traitement. Si nous prenons l'exemple du flux de données PIXE, les fichiers spectres nommés *.x* suivis d'un numéro auto-incrémenté sont convertis par GUPIXWin en fonction d'un mode de calcul¹ sélectionné par l'utilisateur. Des fichiers *.par*, pour « paramètre », sont alors produits et reliés par leur nommage au nom du détecteur concerné². Destinés à être traités par TRAUIXE, ces fichiers n'ont pas de valeur intrinsèque, si ce n'est comme base de réajustement lorsque les données des fichiers TRAUIXE sont insatisfaisantes. On remarque en effet de fréquents va-et-vient avec les fichiers *.par* pour la modification des temps de réglage des paramètres. Les fichiers intermédiaires qui précèdent ces correctifs sont le plus souvent supprimés. Il en va de même pour les fichiers *.png* produits simultanément par TRAUIXE. Les paramètres et les détecteurs sélectionnés lors de la requête réalisée dans TRAUIXE sont déterminants pour la compréhension de ces spectres définitifs, qui ne sont pas conservés en cas d'erreur ou de non-utilité. Ce rapport extrêmement rationnel et méthodique aux données est renforcé dans le cas de leur croisement avec celles obtenues par les autres techniques d'analyse telles que PIGE ou RBS. Ce cadre de traitement et d'interprétation tend à renforcer la valeur des données définitives en réduisant celle des données brutes et intermédiaires. Au cours de l'expérience, les données contenues dans les fichiers générés et convertis ne sont encore qu'une glaise malléable à pétrir et remodeler

1. GUPIXWin permet de choisir entre un mode de calcul itératif prenant en compte la matrice ou un mode de calcul utilisant la trace.

2. Voir Annexe : C.3

jusqu'à l'obtention de la forme souhaitée.

Du point de vue de la plupart des utilisateurs, l'appropriation à travers l'attribution silencieuse d'une propriété intellectuelle au sens moral commence malgré tout dès la définition et la sélection de paramètres nécessaires à la découverte et à la révélation inhérentes à l'objet étudié. Cela explique la fréquence des pratiques de rétention des données par les utilisateurs d'AGLAE, craignant de s'en voir dérober la paternité. À l'instar de l'archéologue revendiquant la responsabilité de la découverte d'un objet enfoui depuis des millénaires, le chercheur est attaché de façon presque affective aux informations qu'il a exhumées au moyen d'un complexe outillage physique, technique et numérique, d'une expertise et d'une formation intellectuelle exigeantes, et d'un protocole d'acheminement des œuvres souvent lourd et contraignant. La singularité de cette chaîne, qu'il est souvent impossible de répéter tant elle est coûteuse et complexe, suffit à justifier le sentiment de la « trouvaille » et la perception de sa préciosité, une fois l'expérience terminée. Tant que cette dernière se poursuit, les données PIXE à haute valeur ajoutée sont celles contenues dans les fichiers TRAUPIXE en format Excel ainsi que les spectres définitifs. Dès lors qu'elle est achevée, le paradigme change et l'ensemble du jeu de données acquiert une nouvelle dimension, y compris les données brutes qui, à l'aide des logiciels de traitement adéquats, peuvent donner lieu à de nouveaux traitements et à un véritable rejeu de l'expérience. C'est pourquoi les utilisateurs envisagent et reconnaissent la valeur de leurs différentes données selon une double temporalité, présente et future ; et selon une double vocation, scientifique et patrimoniale. Cet enchevêtrement cognitif et perceptif n'est pas sans incidence sur la distinction univoque de ce qui relève des données précieuses et de ce qui appartient au brouillon expérimental, tant la capillarité est forte et les ramifications interdépendantes d'un bout à l'autre du flux de données.

Le mouvement d'ouverture et de partage des données induit par l'« Open Science » a cependant constraint les utilisateurs d'AGLAE à réenvisager plus précisément la frontière entre les données brutes à vocation universelle et les données interprétées à verrouiller jusqu'à leur publicité éditoriale officielle. Les contours de cette délimitation fluctuent d'un utilisateur à l'autre, certains considérant que les fichiers LST et les spectres primaires peuvent être ouverts sans délai ; d'autres arguant que ces mêmes fichiers devraient demeurer fermés – au moins pendant la durée qui précède la publication - en raison de la subjectivité des paramètres du contexte expérimental dont ils sont issus. Rappelons au passage que cette pratique de l'embargo est de l'ordre des usages en vigueur dans le domaine des sciences du patrimoine, et ne repose sur aucun socle légal³. Tous s'accordent sur la nécessité de fermeture temporaire des données intermédiaires, y compris celles is-

3. Entretien avec Laurent Romary, directeur de recherche à l'INRIA et président de l'ISO/TC 37, comité technique de l'Organisation internationale de normalisation, 3 juin 2021. Étaient présents : les membres de l'Atelier Numérique, ceux de l'équipe d'AGLAE, Marine Zelverte, cheffe du SAED, et moi-même.

sues des logiciels de la première étape de conversion, et des données traitées⁴. Ce faisant, ils reconnaissent implicitement le caractère gradué et inégal de la valeur scientifique de leurs différents états de données, tout en laissant en suspens la question de leur nature et de leur devenir patrimoniaux. Les positions varient en effet entre ceux qui désirent ouvrir certaines données en raison de leur intérêt patrimonial, et les partisans de leur rétention pour ce même motif et la notion de rareté qui le caractérise. Il est de plus particulièrement difficile de démêler la reconnaissance de la valeur scientifique des données par les utilisateurs d'AGLAE de la reconnaissance patrimoniale intrinsèque, les deux s'interpénétrant plus ou moins.

5.2 Euphrosyne : creuset d'une symbiose entre intérêt scientifique et valeur patrimoniale des données

Par sa genèse, sa construction et sa destination, le projet Euphrosyne achève la confusion de l'intérêt scientifique et de la valeur patrimoniale d'une partie des données d'AGLAE en leur offrant le cadre d'une symbiose. En effet, cette interface a vocation à être à la fois un lieu de mutualisation et d'exposition presque muséale des données produites, et un outil d'accès et de traitement scientifique à part entière. Les enquêtes menées auprès des utilisateurs et des spécialistes des sciences du patrimoine au cours de la phase d'investigation d'Euphrosyne ont ainsi mis en lumière des attentes variées. Tandis que certains voient dans cette interface l'opportunité de nourrir des bases de données dédiées à des archéomatériaux bien spécifiques⁵, d'autres saluent d'abord la praticité d'un accès distant et par conséquent mobile aux ressources d'AGLAE, qu'ils s'agissent des données expérimentales, de l'expertise de l'équipe de l'accélérateur et/ou des outils informatiques de visualisation et de traitement⁶. Les réponses recueillies démontrent que l'étincelle primitive d'Euphrosyne est née d'un rapport pluriel aux données, dans un double contexte d'exploitation scientifique et de valorisation patrimoniale, avec des usages évolutifs et

4. Conclusions issues des enquêtes menées sur place auprès des utilisateurs lors d'expériences (*runs*) menées tant dans le cadre d'un projet de recherche que d'une demande de service.

5. Émilie Chalmin, maître de conférence à l'Université Savoie-Mont-Blanc déclare ainsi : « Un outil comme Euphrosyne pourrait servir d'exemple à la pigmentothèque pour la préhistoire que nous construisons. Ces deux bases doivent être connectées. » Nous pouvons également citer Lore Troalen, responsable du laboratoire du Musée national d'Écosse, qui considère : « Savoir que des données acquises sur des matériaux similaires par d'autres utilisateurs me permettrait d'enrichir ma base de données sur les ors égyptiens. » [Ned Baldessin, Jonathan Bourguignon, Claire Pacheco, *Présentation d'audience d'Euphrosyne...*]

6. Nous pouvons notamment citer Anne Bouquillon, ingénierie de recherche spécialisée sur les céramiques au C2RMF : « Pouvoir traiter les données à distance me changerait la vie et me ferait gagner énormément de temps. » Dans le même esprit, Lucile Beck, directrice du LMC14 affirme : « Un outil me permettant d'interagir plus facilement et rapidement avec l'équipe d'AGLAE m'aiderait pour le traitement des cartographies. » [*Ibid.*]

partiellement satisfaits.

Le besoin d'interopérabilité avec d'autres ressources numériques - patrimoniales, scientifiques et documentaires – constitue sans conteste l'une des attentes les plus fortes à l'égard de la future interface. Les personnes interrogées ont ainsi exprimé la volonté d'accéder, aussi bien pendant la manipulation que lors du traitement des données, à des sources photographiques ainsi qu'à d'autres types de données, en particulier celles issues des autres techniques d'analyse du C2RMF et versées dans la base EROS (*European Research Open System*). Il faut en effet garder à l'esprit que l'expérience ou la série d'expériences menée(s) à AGLAE sur un objet ou un lot d'objets – au sens générique du terme – n'est souvent qu'une étape dans un circuit d'étude et d'analyse plus large mobilisant plusieurs techniques voire laboratoires différents. Lors de son arrivée au C2RMF, un objet ou un lot est photographié de façon rigoureuse avant d'être ensuite soumis à l'imagerie et/ou aux techniques d'analyse élémentaire et moléculaire.

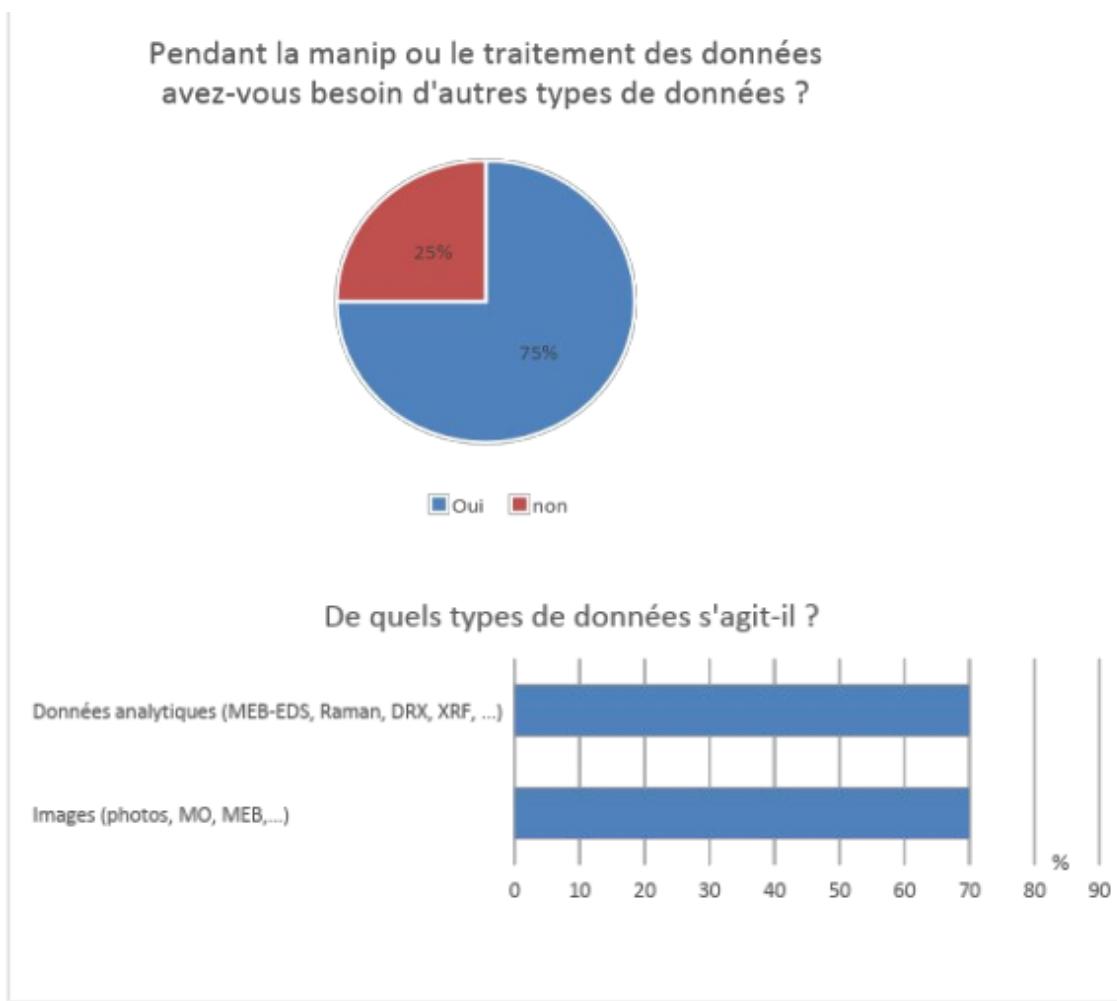


FIGURE 5.1 – Illustration extraite de la présentation d'audience d'Euphrosyne pour l'obtention du statut de start-up d'État du Ministère de la Culture, 8 avril 2021. @Ned Baldessin/Jonathan Bourguignon/Claire Pacheco

et de la conservation-restauration des collections de musées, EROS recense cinq catégories d'informations :

1. Les métadonnées des œuvres.
2. Les métadonnées des images obtenues par des numérisations de clichés photographiques et radiographiques, ainsi que des images électroniques et des images numériques par capture directe de l'objet.
3. Les métadonnées et les résultats des analyses de composition élémentaire, structurale, isotopique et moléculaire.
4. Les métadonnées des rapports d'étude et de restauration, des procès-verbaux des commissions de restauration, des constats d'état.
5. Les métadonnées de la régie des œuvres.⁷

Une œuvre peut donc être associée à plusieurs images, plusieurs analyses ou plusieurs documents, et un document peut être lié à plusieurs œuvres. Par l'interopérabilité future d'Euphrosyne avec EROS, les utilisateurs d'AGLAE souhaiteraient ainsi à juste titre définir leur expérience à la lumière des résultats déjà obtenus et mettre à profit les images numériques réalisées en les reliant aux zones d'intérêt passées sous le faisceau. À travers cette mutualisation, la valeur patrimoniale et l'apport scientifique des données d'AGLAE s'en trouveraient sensiblement enrichis et élargis.

Puissant vecteur d'appropriation et de valorisation des données, l'interopérabilité envisagée d'Euphrosyne en facilite le processus de patrimonialisation. Cela est d'autant plus vrai que ce dernier repose sur la consolidation culturelle et identitaire d'une communauté, elle-même renforcée et étendue grâce à la proximité induite par l'interface entre les ressources d'AGLAE et les utilisateurs ou simples chercheurs intéressés par les résultats d'expériences passées. Euphrosyne tend ainsi à réduire les frontières entre les utilisateurs autonomes, semi-autonomes et les autres, en mettant à disposition les ressources et les compétences d'AGLAE auprès de ceux qui en formuleraient le besoin par l'intermédiaire de l'interface. Par ailleurs, en donnant la possibilité d'accéder à distance aux logiciels de conversion et de traitement, Euphrosyne ne peut qu'accroître le sentiment d'appropriation à l'égard des données obtenues, non seulement chez le responsable de projet, mais également chez ses collaborateurs grâce à la fonctionnalité multi-utilisateurs en cours de construction.

7. Christian Lahanier, Geneviève Aitken, Ruven Pillay, « EROS : European Research Open System », *Cultural institutions and digital technology*, École du Louvre, 8-12 septembre 2003.

Chapitre 6

Construire une sémiologie préventive contre la perte

Comme nous l'avons vu au cours du premier chapitre, les données issues des analyses effectuées sur le patrimoine se trouvent, par un jeu complexe de synecdoque et de filiation, investies d'une charge presque thaumaturgique à l'égard du risque d'altération, de destruction et plus largement de perte. Comme si la rémanence de l'expérience pouvait entraîner la recomposition d'un patrimoine devenu inaccessible, les données sont préservées comme des traces ou des vestiges précieux car garants de la restitution permanente d'une matérialité éphémère. Dès lors, la structure et la construction sémiologique des données deviennent déterminantes car seules à permettre l'intelligibilité pérenne des contenus à travers le temps et les évolutions – cognitives, numériques et techniques - qu'il induit. Nous avons éclairé précédemment l'importance de déterminer la nature et le périmètre d'un jeu de données au regard de l'objet ou du lot patrimonial auquel il se réfère. Cette étape ayant été définie, il est ensuite nécessaire d'en approfondir l'arborescence sémantique et les différents niveaux de granularité, à travers une réflexion collective sur les métadonnées et l'optimisation de leur structuration. Dans le contexte d'AGLAE et de la création d'Euphrosyne, s'est adjointe à ce débat la prise en compte de la migration prochaine des données vers le format HDF5.

6.1 De l'arborescence numérique à la subsidiarité sémantique

Après avoir déterminé qu'un jeu de données général¹ pour AGLAE équivaut à une expérience (*run*) définie par une stabilité des paramètres du contexte expérimental, il nous faut aller au plus près des flux de données pour leur faire sens dans une arborescence

1. Il s'agit donc là du haut niveau d'arborescence, un jeu de données expérimentales en contenant lui-même plusieurs, par techniques d'analyse.

adaptée et suffisamment souple pour s'adapter à la variété des cas d'usage. De plus, l'hétérogénéité et la discontinuité éventuelle des données imposent le recours à la subsidiarité sémantique, comprise comme l'ajustement des métadonnées au niveau de granularité le plus proche des contenus décrits. Néanmoins, ce choix ne doit pas se faire aux dépens de la liberté d'interprétation des chercheurs et des analystes, et encore moins de l'aisance de requête dans la future base contenue dans Euphrosyne. Afin d'évoluer dans une voie de consensus, il est apparu judicieux de s'inspirer des différents usages des données, en cherchant à restituer les différentes étapes cognitives de traitement, de sélection et d'interprétation, dans la pluralité des techniques d'analyses par faisceau d'ions potentiellement mobilisables. L'intention de notre démarche reste identique à celle de la définition générale du périmètre d'un jeu de données : permettre la compréhension voire le rejet de l'expérience par une personne extérieure dans une temporalité future, parfois lointaine, sans perte d'intelligibilité. Cette perspective nécessite l'identification explicite des phases de l'expérience dans une logique qui n'apparaît le plus souvent que de façon rétrospective.

Notons que le statut de start-up d'État du Ministère de la Culture obtenu par Euphrosyne en avril 2021 inscrit le projet dans le cadre des méthodes AGILE, lesquelles impliquent une gestion adaptative ainsi qu'une livraison précoce appelée familièrement *quick and dirty*. Il signifie pour Euphrosyne le rendu d'un premier prototype, ou *a minima* l'esquisse de son *back-office* au début du mois de juillet, nous laissant donc trois mois pour réaliser collectivement un modèle de données et trancher les enjeux tant conceptuels qu'informatiques qui en découlent. Dans un contexte de télétravail prédominant et de réunions à distance bi-hebdomadaires, cette tâche s'est parfois embourbée alternativement dans un excès de technicité ou au contraire, de conceptualisation. Nous sommes néanmoins parvenus à combiner ces deux dimensions grâce à des réajustements constants ainsi qu'à l'apprentissage d'un dialogue transdisciplinaire étroit.

Ce calendrier sous échéance a impliqué l'intégration rapide des contraintes d'implémentation dès l'étape de modélisation, avec des effets sur la définition des entités, de leurs attributs et sur l'architecture des associations, chargées d'être les moins complexes et les plus synthétiques possibles. Nous avons d'abord procédé à la réalisation du modèle conceptuel, qui a ensuite servi de base au modèle relationnel de données², ces deux propositions ayant l'entité/la classe « expérience » pour épicentre.

Dans un jeu de données inhérent à une expérience, un niveau supérieur d'arborescence correspondant au contexte expérimental général s'est d'abord imposé de façon évidente, avant de poser la question épingleuse de la granularité des informations contenues dans ce champ. Dans un premier temps, en plus de l'énergie du faisceau et du type de particule projetée, nous avons considéré la DOSE par seconde et la taille du pinceau du faisceau comme des attributs du contexte expérimental. Les remarques et corrections ap-

2. Les modèles relationnel et conceptuel évoqués sont accessibles dans le dépôt suivant : https://github.com/PaulineChauvet/Livrables_techniques_AGLAE

portées par l'équipe d'AGLAE ont permis d'affiner la compréhension des métadonnées et d'en rectifier la portée parfois exagérément étendue ou trop restreinte. Dans le cas présent, la DOSE par seconde et la taille du pinceau du faisceau sont en réalité des attributs affiliés à un « point de mesure »- lui-même lié à une zone d'intérêt -, générant la création d'une entité du même nom. Cet ajout a entraîné celui d'une entité « spectre » associée aux entités « point de mesure » et « fichier ». En effet, il est apparu que la première proposition de modèle conceptuel, s'arrêtant au contenant et non au contenu des données, souffrait d'un excès de synthétisation, et ne reflétait pas la variabilité scalaire d'une expérience menée à AGLAE. Au cours d'une analyse ponctuelle, les données recueillies correspondent bien à un point précis de l'objet-cible et non à l'intégralité de ce dernier. Ce raffinement conceptuel est absolument déterminant pour la compréhension des choix expérimentaux et de ses résultats. Par ailleurs, c'est justement par la comparaison entre les spectres générés à partir de points de mesure distincts que l'analyste peut confirmer ou infirmer solidement les hypothèses initiales de provenance, de composition et de fabrication. Enfin, ce procédé d'analyse est incontournable dans le cas d'objets composites ou ayant subi des modifications/ajouts au cours du temps. Pour gagner en intelligibilité rétrospective à l'égard du jeu de données, il a donc fallu s'écartier de la zone de confort de l'implémentation et descendre à l'échelle presque microscopique du point de mesure, en augmentant le défi de la masse à structurer et de la scalabilité future. Ce choix de modélisation cristallise la difficulté majeure inhérente à la multiplicité des usages prévus d'Euphrosyne, avec la prise en compte des différentes étapes du cycle de vie des données. En effet, la structure et l'arborescence indispensables à l'utilisateur responsable d'une expérience peuvent être lourdes à implémenter et insatisfaisantes dans une perspective d'accès historique aux données froides, trop massives.

La complexité et l'absence de linéarité du workflow concret des utilisateurs ont par ailleurs largement corsé l'effort d'universalité d'application de notre modèle à l'ensemble des cas expérimentaux envisagés. Le traitement conceptuel des standards de l'expérience est en cela un véritable cas d'école qui mérite une attention particulière. Essentiels à la fiabilité, à la rigueur et à la compréhension d'une expérience, les standards constituent un maillon particulier et central de la chaîne d'acquisition des données. Dans le monde physique, ils consistent en de modestes pastilles échantillonnées et placées en rang devant la ligne de faisceau. De composition parfaitement connue et référencée, ils sont sélectionnés par l'expérimentateur en fonction du matériau étudié et de la technique d'analyse utilisée. La pertinence de leur utilisation détermine celle de l'expérience elle-même, puisque les données obtenues sont interprétées à l'aide de celles acquises sur les standards. Passés en début et en fin d'expérience, parfois également au milieu de cette dernière, ils participent pleinement de la définition du contexte expérimental. C'est pourquoi, dans un premier temps, une entité « standard » a été créée et associée au contexte expérimental. Les analystes de l'équipe d'AGLAE ont alors soulevé l'impossibilité de limiter un standard

à un nom et à une référence alphanumérique, car il acquiert son sens dans un usage dynamique et une savante manipulation qui doit s'inscrire dans un circuit conceptuel. De surcroît, les standards s'inscrivent dans une double temporalité : d'abord celle de leur analyse par la ligne de faisceau, puis celle du traitement croisé des données obtenues avec celles de l'expérience réalisée sur un lot ou un objet-cible. Les standards ont donc une portée à la fois synchronique et diachronique au regard de l'expérience. Ils sont tout autant à sa périphérie que dans son flux interne en générant des fichiers spectres qui serviront à l'interprétation d'autres fichiers spectres. La possibilité de faire figurer un standard en attribut de type booléen d'une entité « spectre » a été évoquée et finalement écartée. En vue de la traduction du modèle conceptuel en modèle relationnel de données, il fallait envisager le respect des formes normales et ne pas définir d'attribut lui-même porteur d'autres attributs qui ne dépendraient pas directement de l'entité. De plus, nous ne souhaitions pas perdre la possibilité de mise en exergue de la spécificité conceptuelle d'un standard, qui ne renvoie pas seulement à un spectre lui-même lié à un point de mesure, mais également à un échantillon doté d'une matérialité à part entière. Par conséquent, nous avons finalement créé une entité « standard » avec ses attributs d'identification matérielle, associé à une entité « point de mesure » reliée à la fois à un spectre et à une technique d'analyse. Adapté à une modélisation respectueuse des flux réels de données, ce choix est bien plus complexe à traduire dans une arborescence de métadonnées, en particulier avec l'incertitude de leur format et du conteneur à venir du jeu de données. Une migration vers un format conteneur HDF5 est en effet envisagée et toujours en cours d'étude à ce jour.

Déjà utilisé par le synchrotron SOLEIL et d'autres grands instruments d'optique et d'imagerie, HDF5 pour *Hierarchical Data Format* est un format permettant de sauvegarder et de structurer des fichiers contenant des données massives. Un fichier HDF5 est ainsi un conteneur de fichiers aux formats divers, qui offre déjà une arborescence par sa structure et sa composition en groupes, datasets et l'inclusion d'attributs équivalant à des métadonnées. À l'origine, sa première version exploitable sert aux données de la NASA, qui a besoin d'effectuer des calculs complexes sur de grandes masses de données, grâce à une indexation efficace. AGLAE ne produit pas de données massives, même dans le cas de l'imagerie, qui ne pèse guère plus que quelques teraoctets. Mais le recours à un format conteneur adapté aux données scientifiques hétérogènes, apte à supporter des fichiers d'imagerie mais également des fichiers d'analyse avec des métadonnées associées aux niveaux d'arborescence adéquats a ouvert une piste de mise en cohérence avec le standard dominant, en vue de l'ouverture et du partage souhaités des données d'AGLAE dans le contexte de l'*Open Science*³. Nous sommes donc dans le cas d'un usage du format HDF5 quelque peu détourné de sa vocation initiale et utilisé davantage pour sa capacité à associer le stockage et la description de contenus que pour l'interprétation et le calcul de données massives. Lors de la première phase de création d'Euphrosyne, il est

3. Ned Baldessin, Jonathan Bourguignon, Claire Pacheco, *Présentation d'audience d'Euphrosyne...*

rapidement apparu que si la migration des données d'imagerie telles que celles générées par AGLAEMap vers HDF5 se ferait sans difficulté, il n'en irait pas de même pour les données de l'analyse ponctuelle, dont les formats de fichiers convertis ne sont pour l'heure pas supportés par HDF5. À ce jour, nous ignorons donc encore si ce format sera utilisé pour l'ensemble des données d'AGLAE ou seulement une partie d'entre elles, et s'il sera implémenté dans le cadre de la seconde phase de déploiement – autrement dit la création d'Euphrosyne-data – ou non. Cette incertitude nous a incités à réfléchir simultanément à deux types d'arborescence éventuels : l'un dans le cadre d'une intégration des données dans un format conteneur HDF5 ; l'autre, inspiré de l'organisation actuelle des données et des fichiers qui les renferment. Dans le premier cas, le principe de subsidiarité sémantique est poussé à son maximum avec la possibilité de renseigner des métadonnées à un niveau de granularité extrêmement fin. Dans un format xml, des informations de contenu et de description seraient associées à chaque fichier ou groupe de fichiers selon les cas, préparant ainsi le travail de pérennisation et d'archivage. L'arborescence serait fragmentée et ne serait pas aisément compréhensible dans une perspective d'ensemble, mais elle aurait le mérite d'être conçue sur mesure et de s'adapter à la forte plasticité des données d'AGLAE. Dans le second cas, un document xml valable pour l'intégralité d'un jeu de données serait produit et renseigné au moyen d'une extraction automatique des métadonnées incluses dans le corps des fichiers, et de leur attribution dans des champs préalablement balisés. Ce procédé impliquerait certes, une certaine rigidité structurelle et sémantique, mais aussi une cohérence harmonieuse pour tous les jeux de données qui y seraient soumis. Un niveau d'arborescence général relatif au contexte expérimental serait lui-même subdivisé par techniques d'analyse (PIXE, PIGE, RBS, IBIL, etc.), découpées en types d'analyse (imagerie, analyse ponctuelle) contenant eux-mêmes un niveau inférieur correspondant à chaque point de mesure. Dans le cas des métadonnées relatives aux données issues de l'association de plusieurs techniques d'analyse – ce qui représente une large proportion des expériences –, elles bénéficieraient d'un niveau d'arborescence à part entière, contenu dans celui du contexte expérimental, et profiteraient de l'utilisation de pointeurs vers les techniques et les fichiers initiaux concernés. Nous nous sommes inscrits dans la perspective d'un utilisateur ou d'un individu extérieur susceptible de vouloir accéder uniquement aux données traitées définitives ou d'exploiter et croiser des données primaires et intermédiaires peu ou pas utilisées lors de l'expérience. À l'instar de la description archivistique numérique, les métadonnées seraient renseignées au niveau le plus approprié et ne seraient donc pas répétées aux éventuels niveaux inférieurs.

6.2 Métadonnées et données : modalités d'une circulation sémantique fluctuante

Arbre sémiologique par lequel la signification et le sens des données circulent et jaillissent, l'arborescence est déterminante pour garantir la viabilité de la circulation sémantique et documentaire entre les métadonnées et les données. La plasticité des données d'AGLAE, dont l'usage et l'interprétation sont souvent pluriels, a rendu le choix et la définition des métadonnées particulièrement complexes, avant même d'envisager les contraintes technologiques d'implémentation. Si nous prenons l'exemple d'une notice bibliographique d'œuvre littéraire, les métadonnées qui la composent renvoient de façon explicite et univoque à des contenus clairement catégorisés et aisément identifiables, ne serait-ce que par leur forme. Ainsi, un code visuel implicite et commun à la plupart des monographies permet au catalogueur de repérer sans doute possible ce qui relève du nom de l'auteur, du titre, du sous-titre, de l'éditeur, du lieu d'édition et de la date de publication. Ces informations sont renseignées dès la phase d'édition de l'œuvre selon des règles de normalisation en adéquation avec les codes de description bibliographiques. Il en va autrement pour les données d'AGLAE qui ne bénéficient pas encore, à ce jour, d'une stratégie d'indexation et de renseignement rationalisée et valable pour un jeu de données expérimentales. En l'état, il existe bien des métadonnées, générées le plus souvent automatiquement, disséminées au sein même des données et au travers du nommage des fichiers¹. Il a été nécessaire de procéder à la cartographie précise de cet existant⁴ avant d'envisager sa redéfinition à l'aune du nouveau périmètre tracé pour un jeu de données, désormais correspondant à une expérience (*run*).

Cette étape de travail a révélé un processus de transition informationnelle déjà en cours, avec de nettes inégalités de renseignement entre les différents types de fichier et, dans une moindre mesure, entre les différentes techniques d'analyse, celles particulièrement utilisées comme PIXE et PIGE ayant bénéficié d'un soin particulier. Par ailleurs, nous avons constaté des redondances dues à la variété des supports de métadonnées. Ainsi, les informations renseignées dans le cahier de laboratoire imprimé reprennent certains éléments figurant sur la demande de temps de faisceau tels que le nom et la responsabilité de l'analyste, la technique d'analyse souhaitée, l'objet ou le lot ciblé par l'expérience. Y figurent également les détecteurs choisis, les filtres et les standards sélectionnés, autant d'informations réitérées dans le fichier Excel généré automatiquement et valable pour toute l'expérience⁵. Toutefois, ce cahier détient également des informations uniques et non dématérialisées sur certaines anomalies ou corrections expérimentales, sans lesquelles il est impossible de comprendre la manipulation et le traitement des données qui ont conduit

4. La cartographie des flux de données PIXE d'AGLAE ainsi que celle des métadonnées sont accessibles dans le dépôt suivant : https://github.com/PaulineChauvet/Livrables_techniques_AGLAE

5. Voir Annexe : B.4

à leur état définitif. Les informations du contexte expérimental existent donc mais ne sont pas organisées de façon à offrir une intelligibilité pérenne au niveau de granularité approprié. Par ailleurs, certains utilisateurs renseignent uniquement les champs référencés du cahier de laboratoire, sans utiliser l'espace libre de rédaction, et réservent les commentaires inhérents à la chaîne expérimentale à des supports personnels qu'ils conservent et sur lesquels l'équipe d'AGLAE n'a donc pas la main. Il n'est ainsi pas rare de devoir faire appel à la mémoire des analystes et des chercheurs présents lors d'une expérience passée lorsqu'il y a nécessité d'en comprendre les résultats. De plus, il n'existe pas encore de jeu de données tel que nous l'avons récemment défini mais un regroupement de fichiers par nom d'utilisateur et date de manipulation. L'accès à des données expérimentales précises s'en trouve par conséquent fortement entravé si le cahier de laboratoire n'offre pas d'informations contextuelles complètes et satisfaisantes. La compréhension et la possibilité du rejet de certaines expériences sont donc, à l'heure actuelle, l'apanage exclusif de l'expérimentateur et de l'éventuel groupe de chercheurs qui l'accompagne. Nous constatons ici un revers possible de l'autonomie de certains utilisateurs, tentés de s'émanciper d'AGLAE pour la gestion et le stockage des données et des métadonnées expérimentales, tant pour des raisons pratiques d'accès qu'en vertu d'une appropriation intellectuelle. En plus de ces informations textuelles et chiffrées, de nombreux utilisateurs prennent également des photographies des zones d'intérêt analysées au cours de l'expérience, qui constituent des métadonnées singulières mais précieuses, échappant à la conservation et au stockage d'AGLAE.

Parallèlement aux métadonnées renseignées manuellement par les utilisateurs, il y a celles récupérées et versées automatiquement dans l'en-tête et dans le nommage de la plupart des fichiers à l'aide d'un programme développé sur mesure à AGLAE, en vue de la compréhension rétrospective des données. Il s'agit alors moins d'une volonté de pérennisation dans une perspective d'ouverture que d'améliorer l'intelligibilité des données par les utilisateurs et l'équipe d'AGLAE. Ces deux paradigmes se rejoignent toutefois dans les évolutions de pratique et de structure informationnelles induites. Que l'on soit dans le cadre de la méthode d'analyse PIXE ou PIGE, nous retrouvons des champs de métadonnées similaires par type de fichier. Les fichiers spectres de l'analyse ponctuelle en format ASCII contiennent tous un en-tête de deux lignes : la première indique le nombre de canaux du spectre, la seconde répète certaines informations du contexte expérimental figurant à la fois dans le fichier Excel général et dans le cahier de laboratoire. S'y ajoutent des métadonnées à la granularité plus fine - bien que partiellement appropriée - telles que la somme du spectre ou le temps d'acquisition. Les en-têtes des fichiers LST de l'imagerie se décomposent également en deux lignes : la première relative à la dimension ainsi qu'à la résolution de la cartographie ; la seconde réitérant les informations générales du contexte expérimental. Cette redondance est également présente dans les fichiers EDF résultant de la conversion des fichiers LST. Elle témoigne d'une capillarité contextuelle nécessaire

lorsqu'un jeu de données est organisé en fonction d'une arborescence d'usage et non de hiérarchisation sémantique de contenus.

Certaines limites technologiques placent certaines données hors de leur environnement informationnel et contextuel. En effet, les métadonnées d'un fichier LST ne peuvent être lues par le convertisseur et ne sont donc pas incorporées automatiquement au fichier EDF de sortie. Le logiciel de traitement de certaines données RBS ne parvient pas non plus à traiter de fichiers comportant un en-tête. Pour ces cas spécifiques, les seules métadonnées sont celles réinjectées *a posteriori* dans les fichiers issus de la conversion, et/ou celles intégrées dans le nommage des fichiers, ce dernier n'étant pas normalisé dans le cas de la méthode RBS. Qu'il soit chargé de pallier à certains écueils technologiques ou à améliorer l'aisance d'accès et de recherche pour les utilisateurs, le nommage a acquis une dimension prédominante dans la signalisation et l'organisation des données d'AGLAE. Les logiciels de conversion et de traitement dépendent d'ailleurs de lui pour la sélection de paramètres et d'éléments spécifiques. À titre d'exemple, les fichiers .par générés par GUPIXWin puis traités par TRAUPIXE doivent forcément se terminer, dans leur nommage, par « nom du détecteur ». Le lien de chaque fichier .par au détecteur adéquat est indispensable pour TRAUPIXE qui le traite nécessairement en fonction du/des détecteur(s) sélectionné(s)⁶. Ce système d'intégration de métadonnées au nommage, déployé faute d'un espace approprié d'indexation et de renseignement, entraîne l'élaboration de noms de fichiers parfois excessivement longs et parfois délétères pour leur accessibilité. Le nom des fichiers .png – ou .dat si le contenu est en format ASCII - pour la fin de flux de l'imagerie PIXE ou PIGE est ainsi composé : de la date, du numéro d'analyse, de la référence du projet, du nom du détecteur et d'un éventuel commentaire de l'utilisateur. Quant aux fichiers spectres primaires, leur nom résulte de la combinaison de la date, du numéro d'analyse, de la référence de l'objet, du nom du projet suivi de « IBA.x ». Cette structuration composite peut être problématique selon les points d'accès recherchés, variables selon les utilisateurs, lesquels souhaitent parfois effectuer une recherche sur le dernier élément renseigné dans le nom. Les noms des fichiers TRAUPIXE en offrent un exemple éloquent. Ils se composent de la date, de la matrice avec le détecteur associé à la matrice ou à la trace et se terminent par l'indication de l'élément pivot. Ce dernier correspond à un élément chimique choisi par l'utilisateur, sous réserve d'être présent dans les différents détecteurs PIXE et en quantité suffisante. Extrêmement important pour le traitement et l'interprétation des données PIXE, il peut faire l'objet d'une recherche ciblée dans des données historiques. Son apparition en fin de nommage constitue en cela un obstacle car seul le début du nom de fichier est directement visualisable. En l'état, les métadonnées associées aux différents fichiers contenus dans un répertoire utilisateur ne fournissent qu'une intelligibilité partielle des données dans les étapes successives de leur

6. En plus des détecteurs réels d'AGLAE, il existe des détecteurs qualifiés de virtuels qui correspondent à la combinaison de plusieurs détecteurs réels.

cycle de vie. Une réflexion collective a été menée dans le cadre de la création d'Euphrosyne afin de déterminer les modalités de révision de leur organisation, de leur format et de leur structure afin d'optimiser la pertinence de leur charge sémantique et documentaire. Cette démarche s'est inscrite en étroite corrélation avec les différentes instances du C2RMF afin d'aligner les pratiques de gestion et de pérennisation des données d'AGLAE sur celles du laboratoire⁷. La mission Nouvelles Technologies de l'Information (NTI) et le Service Archives et Documentation (SAED) ont ainsi participé, selon leurs domaines de connaissance et d'application respectifs, au chantier de révision des métadonnées d'AGLAE, en l'inscrivant dans la problématique plus large de l'ouverture et de l'interopérabilité des données des sciences du patrimoine. La coïncidence chronologique de cette réflexion avec les prospections menées dans le contexte d'ESPADON et dans celui d'IPERION-HS, par l'intermédiaire des groupes de travail DIGILAB autour des données, a été réellement féconde et a favorisé la résolution rapide et concertée de certains défis méthodologiques et conceptuels. Certaines institutions particulièrement avancées en *data management*, telles que la National Gallery, sont en effet parvenues à mettre en place un système d'information, de traitement et de partage des données extrêmement efficace dont nous avons pu partiellement nous inspirer. Multiscalaire, notre approche a cependant d'abord favorisé les propositions de l'échelle immédiatement supérieure – à savoir le C2RMF –, afin de respecter la graduation intrinsèque des dynamiques de mutualisation, non sans soulever de réels enjeux politiques et stratégiques.

Afin de prendre en compte la nécessaire transition des pratiques, la première étape d'amélioration de la signalisation des données consiste en l'harmonisation des règles de nommage par type de fichier, en prenant en compte les recommandations du SAED en la matière. La première d'entre elle suggère une mise en cohérence des noms de fichiers qui permettent leur distinction et leur accessibilité logique. Au cours des usages antérieurs à cette recomposition, il n'existe pas de règle commune ou même de bonne pratique visant à harmoniser les dénominations, laissées à la libre appréciation des utilisateurs. Nous sommes désormais dans une transition hybride où la gestion des fichiers répond à la fois aux besoins pratiques des analystes et à un effort documentaire récent en vue de l'accès et de l'ouverture des données historiques. Par leur organisation, les données sont ainsi toujours reliées de façon prépondérante aux utilisateurs. En revanche, la dénomination actuelle des fichiers traduit une volonté de restitution du flux de données dans ses composantes technologiques et scientifiques, sans présomption des usages qui en seront faits. À titre d'exemple, les fichiers générés par TRAUIPIXE, qui ne faisaient pas initialement l'objet d'un nommage standardisé, ont ensuite été préfixés par GUPIX, suscitant une ambiguïté avec les fichiers issus de GUPIXWin. TRAUIPIXE s'est ainsi finalement substitué à ce choix, mais sans entraîner de modifications rétrospectives sur les noms de

7. Le C2RMF contient un laboratoire qui regroupe différentes techniques d'analyse scientifique pratiquées sur le patrimoine, et qui peuvent être, par abus de langage, qualifiées elles-mêmes de laboratoires.

fichiers antérieurs à cette syntaxe. Les fichiers 2D-EDF ont également connu plusieurs versions de nommage, et sont désormais regroupés par répertoire utilisateur comprenant des répertoires par année, eux-mêmes conteneurs de répertoires par run. Leurs noms se terminent maintenant systématiquement par « IBA - nom du détecteur_EDF ». Il faut souligner le fait que chacune de ces modifications a ensuite dû être appropriée par des analystes souvent soucieux de conserver des repères familiers lors de l'accès et du traitement des données. Par ailleurs, bien qu'incités à respecter les bonnes pratiques de nommage mises en place à AGLAE, les utilisateurs demeurent libres d'influer sur la sémiologie des données qu'ils contribuent à produire.

De réels efforts pédagogiques sont déployés auprès des chercheurs par l'équipe d'AGLAE pour les accompagner dans l'assimilation de ce qui représente un véritable changement de culture. En effet, les utilisateurs de l'accélérateur n'ont jamais, pour la plupart d'entre eux, envisagé que la matière première de leurs publications puisse être rendue accessible à un public élargi, ou même entièrement ouverte. Nommer et classer leurs données de façon à les rendre intelligibles par le plus grand nombre est, dans une certaine mesure, à contresens des pratiques courantes des spécialistes en sciences du patrimoine, parmi lesquels les utilisateurs d'AGLAE. Chacun d'eux se considère comme responsable de son expérience et donc de ses données, auxquelles il s'attache dans une forme de propriété intellectuelle et morale. Il appose son estampille par une architecture informationnelle et documentaire qui lui est propre, notamment à travers l'ajout de commentaires et de signes destinés à être d'abord compris de lui, ainsi que par des modifications de noms de fichiers et l'usage de métadonnées spécifiques à son champ de valeurs et de références. Le renoncement partiel à ce mode d'appropriation au bénéfice d'une libre circulation des données scientifiques, inégalement comprise et acceptée, suscite des résistances chez certains utilisateurs. Dès lors, la démarche itérative inhérente à la création d'Euphrosyne prend tout son sens. Plutôt que de proposer une interface conçue selon une dynamique descendante, en prenant le risque d'un rejet ou d'une insatisfaction des utilisateurs, il a semblé indispensable d'articuler en amont les besoins de ces derniers avec les changements induits par la mise en conformité au regard des principes FAIR d'accès, d'identification, d'interopérabilité et de réutilisation. Cette articulation, qui s'est apparentée à une réflexion à la fois prospective et rétrospective, a mis en évidence la variété d'usages, d'intérêts et de temporalités des métadonnées. Ainsi, tandis que certaines telles que la DOSE ou les filtres sont indispensables au travail de traitement et d'interprétation de l'analyste, d'autres informations relatives au contexte expérimental ou au contenu des fichiers n'ont de valeur que pour la compréhension des données historiques, *a posteriori* donc, et pour leur pérennisation. La dimension et la résolution des spectres seront de cette façon plus utiles dans le cadre de l'archivage numérique que pour l'intelligibilité synchronique de l'expérience.

Notre inventaire de l'existant a mis en relief une abondance de métadonnées, dont certaines sont susceptibles d'être réorganisées sans affecter les utilisateurs, puisque in-

usités. Insistons sur le fait qu'il s'agit bien plutôt d'un réajustement que d'une somme d'ajouts et de suppressions, car les métadonnées renseignées actuellement ont toutes leur utilité lors d'une ou plusieurs étape(s) du cycle de vie des données. Elles peuvent participer à la lecture diachronique de l'expérience et/ou à sa viabilité elle-même, tout en offrant chacune une pierre à l'édifice de la pérennisation. Un questionnaire élaboré et diffusé en début de stage⁸ a révélé la nécessité d'un seul ajout aux métadonnées déjà existantes : les photographies des objets analysés, réalisées en début de circuit lors de leur arrivée au C2RMF, ainsi que celles des zones d'intérêt étudiées, souvent faites et conservées par les chercheurs extérieurs au laboratoire. Par une interopérabilité interne toujours en cours d'étude, les clichés numériques versés par le groupe de photographie du C2RMF dans EROS ont ainsi vocation à être prochainement liés à un jeu de données par objet ou lot, mutualisant les résultats des différentes techniques scientifiques du laboratoire tout en permettant à chacune d'elles d'agréger les photographies et toute autre donnée utile à son propre gisement. L'optimisation des métadonnées d'AGLAE passerait également par l'instauration d'une version dématérialisée du cahier de laboratoire, facilitant l'interrogation et l'exploitation des informations qu'il contient. Il serait également intéressant d'envisager un lien numérique entre un jeu de données expérimentales et la demande de temps de faisceau, laquelle contient des renseignements détaillés sur la genèse institutionnelle, scientifique et patrimoniale de l'expérience.

Outre ces quelques nouveautés suggérées, l'essentiel de la restructuration des métadonnées a tenu en un recalibrage conceptuel et relationnel des entités/classes auxquelles elles se sont parfois trouvées directement reliées de façon inappropriée au regard de la subsidiarité sémantique. Le premier prototype d'Euphrosyne étant toujours en cours de réalisation à ce jour, nous n'avons pas encore eu de retours d'utilisateurs sur nos propositions relatives à la réorganisation ainsi qu'à l'architecture conceptuelle et numérique des (méta)données. Par conséquent, nous n'avons pas de hauteur de vue et demeurons des « prospecteurs des abîmes »⁹, progressant à vue.

8. Voir Annexe : A.1 : Questionnaire réalisé en début de stage, destiné à l'équipe d'AGLAE ainsi qu'à Marine Zelverte, afin d'initier l'analyse de l'existant.

9. Georges Duhamel, *Le Temps de la recherche*, Paris, P. Hartmann, 1947, p. 13.

Troisième partie

**La conservation et la pérennisation
des données ou le mythe de Sisyphe**

Tout au long de ce long effort mesuré par l'espace sans ciel et le temps sans profondeur, le but est atteint. Sisyphe regarde alors la pierre dévaler en quelques instants vers ce monde inférieur d'où il faudra la remonter vers les sommets. Il redescend dans la plaine. C'est pendant ce retour, cette pause, que Sisyphe m'intéresse [...] Cette heure qui est comme une respiration et qui revient aussi sûrement que son malheur, cette heure est celle de la conscience. À chacun de ces instants, où il quitte les sommets et s'enfonce peu à peu vers les tanières des dieux, il est supérieur à son destin. Il est plus fort que son rocher. Si ce mythe est tragique, c'est que son héros est conscient. Où serait en effet sa peine, si à chaque pas l'espoir de réussir le soutenait ?¹⁰

On ne saurait offrir meilleure allégorie du travail de pérennisation que cette peinture existentialiste du mythe de Sisyphe, mise en profondeur par la facture aux accents nietzschéens d'Albert Camus. En effet, en quoi consiste le travail de pérennisation si ce n'est en un perpétuel mouvement de progression contre l'emprise du temps et sa menace de destruction ? Tel Sisyphe répétant inexorablement son ascension, les bâtisseurs et les sentinelles de la conservation du patrimoine numérique doivent veiller à maintenir les contenus et leur intelligibilité à travers un mouvement de métamorphose régulier, tant des supports et des cadres technologiques que des référentiels et de la toile sémantique. Sitôt achevé, tout l'édifice est déjà à reconstruire, dans une folle échappée à l'égard de l'obsolescence. Cette course contre le temps est intrinsèquement patrimonialisante, car elle a bien vocation à garantir le leg d'un héritage aux générations futures et mêmes lointaines. Ce faisant, elle pose implicitement la question sociale de la pérennisation : quel est le sens du déploiement continu de tant de moyens financiers, humains et technologiques voué à ne jamais finir ? Pourquoi conserver et transmettre certains contenus plutôt que d'autres à ceux qui nous succéderont ?

L'inflation à l'œuvre avec l'extension toujours plus large de l'axiologie patrimoniale et du périmètre du patrimoine numérique entraîne une hausse exponentielle du coût et de la charge de la pérennisation. En conséquence, la question de sa légitimité se fait de plus en plus pressante et exige d'être saisie et mise en perspective. Bien entendu, il ne s'agit pas d'offrir de réponse ferme et univoque mais de réfléchir à l'adéquation de nos modèles – puisqu'il y en a bien plusieurs – de conservation et d'archivage à l'égard des objectifs socio-culturels, politiques, éthiques voire métaphysiques poursuivis. Dans le cas des données des sciences du patrimoine, parmi lesquelles celles d'AGLAE, la question du pourquoi s'articule étroitement à celle du comment, d'autant plus à l'aune de la création d'Euphrosyne, dont la seconde et la troisième phase de déploiement sont corrélées à l'archivage et à l'accès aux données définitives et historiques. Elle réactive également l'importance de l'échelle de réflexion pour la viabilité organisationnelle, logistique et conceptuelle d'un système d'archivage et de diffusion des données stockées. Notons que cette échelle peut

10. Albert Camus, *Le Mythe de Sisyphe*, Paris, Gallimard, 1942, p. 165.

varier d'une étape à l'autre du processus, ainsi qu'en raison de son évolution dans le temps.

Chapitre 7

L’archivage OAIS des données d’AGLAE : un défi organisationnel ?

Après avoir procédé à la reconnaissance, à l’identification, à l’organisation et au renseignement des données à valeur scientifique et/ou patrimoniale d’AGLAE, il est indispensable d’envisager la logique fonctionnelle, conceptuelle et organisationnelle apte à sous-tendre leur pérennisation, sous-entendue comme une action dédiée à la durabilité et à la permanence d’un objet. Guidés par le tropisme des utilisateurs et de leurs besoins, nous nous sommes naturellement orientés vers le modèle de référence OAIS pour réfléchir à l’archivage des données d’AGLAE dans le cadre élargi du C2RMF. Nous allons voir en quoi cette perspective soulève un vaste éventail d’enjeux, d’un bout à l’autre de la chaîne de traitement.

7.1 De l’émergence d’un besoin au choix d’un modèle

Avant d’aborder les implications protéiformes pour les données d’AGLAE et pour Euphrosyne du recours au modèle OAIS, il est nécessaire d’en comprendre la genèse politique, stratégique et culturelle, laquelle déborde du cadre du C2RMF pour rejoindre une réflexion plus globale sur la gestion pérenne des données en sciences du patrimoine. Nous verrons ensuite quelles en sont les conséquences technologiques, documentaires et organisationnelles pour les données d’AGLAE, dans le contexte préparatoire de migration et de conception d’Euphrosyne axée sur des usages métiers.

À l’origine des premières démarches du C2RMF en faveur de l’archivage des données produites au sein de son laboratoire, on retrouve une contrainte technique et matérielle élémentaire : la nécessité d’accroître l’espace de stockage des données numériques¹. L’absence de stratégie de désherbage et de sélection a en effet entraîné une hausse exponentielle de la masse stockée sur les différents serveurs du laboratoire. Porte-voix des préoccupa-

1. Entretien avec Marine Zelverte, cheffe du Service Archives et Documentation du C2RMF, mené le lundi 10 mai 2021.

tions du C2RMF en la matière, le SAED a donc émis le souhait de bénéficier d'un serveur commun au laboratoire, favorisant à la fois la mutualisation de données jusqu'ici cloisonnées, et la gestion centralisée de leur conservation. Cette demande, toujours en attente, s'est accompagnée d'une réflexion sur l'élaboration d'une politique d'archivage adaptée à l'hétérogénéité, ainsi qu'à la vocation scientifique et patrimoniale des données conservées. Cette prospection s'est notamment appuyée sur les développements et les expérimentations en cours dans les circuits parallèles d'E-RIHS et d'Huma-Num. Rappelons que cette dernière correspond à une très grande infrastructure de recherche (TGIR) visant à faciliter le tournant numérique de la recherche en sciences humaines et sociales. Parmi ses missions, elle propose un dispositif technologique unique permettant le traitement, la conservation, l'accès et l'interopérabilité des données de la recherche². Ce dispositif donne notamment accès à une plateforme d'accès unifié (ISIDORE) ainsi qu'à une procédure d'archivage à long terme³. Précisons que la collaboration avec E-RIHS est cependant plus approfondie en raison de l'intégration effective du C2RMF dans l'infrastructure⁴, contrairement à Huma-Num qui demeure à ce jour un interlocuteur privilégié mais non une organisation partenaire.

E-RIHS cherche à répondre aux problèmes expérimentaux que posent la connaissance et la conservation des matériaux du patrimoine. Dans cette perspective, elle met en place un accès transnational à des technologies de pointe – parmi lesquelles l'analyse par faisceau d'ions – et à des archives scientifiques⁵. Les données nativement numériques récoltées grâce à ces technologies posent de façon cruciale la question de leur pérennisation et d'un archivage à long terme compatible avec les besoins de souplesse d'accès et de stockage des différentes communautés concernées. Les organes de débat ainsi que les groupes de travail transnationaux organisés à fréquence régulière depuis la phase préparatoire d'E-RIHS lancée en 2017 ont permis d'infuser une réflexion partagée ainsi qu'une préoccupation plus forte à l'égard du devenir des données définitives et historiques. La question de l'archivage au sens strict demeure toutefois quelque peu subsidiaire au regard des problématiques d'interopérabilité, de structuration et d'indexation. Les rencontres organisées par DIGILAB le 21 avril dernier l'ont en effet traité de façon transverse, sous

2. Portée par l'Unité Mixte de Services 3598 associant le CNRS, l'Université d'Aix-Marseille et le Campus Condorcet, la TGIR Huma-Num est bâtie sur une organisation originale consistant à mettre en œuvre un dispositif humain (concertation collective) et technologique (services numériques pérennes) à l'échelle nationale et européenne en s'appuyant sur un important réseau de partenaires et d'opérateurs. Elle porte notamment la participation de la France dans le projet DARIAH en coordonnant les contributions nationales.[Stéphane Pouyllau, « Huma-Num, la TGIR des humanités numériques », UpTV (visionnée le 24 juillet 2021)]

3. Par « long terme », nous comprenons une durée suffisamment longue pour être soumise à l'impact des évolutions technologiques.

4. Isabelle Pallot-Frossard, directrice du C2RMF, est également coordinatrice nationale pour E-RIHS France.

5. Site Internet d'E-RIHS (*European Research Infrastructure for Heritage Science*) : <http://www.erihs.fr/>

l'angle de l'optimisation d'un accès documenté aux données à travers le temps⁶.

Sans offrir de méthodologie d'archivage sur mesure pour les données scientifiques du patrimoine, ces flux d'échanges irriguent une véritable culture de la pérennisation adaptée à leur hétérogénéité, leur « matérialité numérique » singulière ainsi qu'aux enjeux actuels de l'*Open Science*. Jusqu'à aujourd'hui, le C2RMF s'est surtout concentré sur un vaste chantier de numérisation des archives photographiques, radiographiques et textuelles afin de permettre l'accès virtuel et l'enrichissement numérique de documents historiques tels que les procès-verbaux des commissions de restauration, les rapports, etc. Ce travail répond aux besoins des utilisateurs, contraints de venir sur place pour accéder aux documents non-numérisés, et par conséquent désireux de bénéficier d'un accès à distance aux ressources les plus élargies possibles. Cependant, le poids croissant des ressources numériques, l'évolution des technologies inhérentes aux données ainsi que l'urgence croissante de trouver une solution de stockage au long cours imposent désormais l'établissement d'une politique d'archivage à la fois favorable aux utilisateurs et aux contenus archivés. Il en va de la justification sociale de la pérennisation des ressources du C2RMF.

Soucieux d'opter pour un système orienté vers les utilisateurs et compatible avec la masse hétérogène des données produites au sein du laboratoire, le SAED s'est rapidement intéressé aux services proposés par le CINES (Centre informatique National de l'Enseignement Supérieur). Établissement Public à caractère administratif national (EPA) placé sous la tutelle du Ministère de l'Enseignement Supérieur, de la Recherche et de l'Innovation, le CINES propose trois volets de moyens aux différentes communautés scientifiques : le calcul numérique intensif, l'hébergement de plates-formes informatiques d'envergure nationale, et l'archivage pérenne de données électroniques. Ce dernier consiste à conserver le document dans son aspect physique comme intellectuel sur le moyen et long terme, de façon à ce qu'il soit en permanence accessible et compréhensible. Cela suppose le respect de trois critères majeurs que sont l'intégrité, la lisibilité et l'intelligibilité. Les contenus pris en charge par le CINES sont : les données scientifiques telles que les données d'observation, d'expérimentation et les résultats de calcul ; les archives publiques courantes et intermédiaires ; ainsi que les ressources des bibliothèques et des centres de documentation. Il a ainsi l'avantage de gérer aussi bien les photothèques, les documents numérisés que les données produites par des instruments scientifiques tels qu'un accélérateur de particules ou un séquenceur de génomes. Il est par ailleurs particulièrement compétent dans l'archivage de contenus soumis à des contraintes de taille de données, d'exigence d'identification unique, de variété des formats et de pluralité de besoins de conservation. Il permet aux services versants de bénéficier de solutions d'archivage pérenne et/ou intermédiaire, tout en offrant la possibilité de diffuser à des communautés distribuées. En 2010, le CINES obtient l'agrément des archives de France en tant qu'infrastructure d'archivage intermédiaire.

6. Sorin Hermon, Vincent Detalle, Victor Etgens, Présentation introductory « DIGILAB and JRA joint meeting », 21 avril 2021.

Par conséquent, le CINES semble *a priori* compétent pour répondre aux besoins d'archivage spécifiques du C2RMF, notamment des données produites par son laboratoire de recherche mosaïque. Rappelons que ce dernier regroupe : des instruments d'imagerie en ultraviolet et infrarouge, la photographie directe en lumière visible, la radiographie, les techniques 3D de surface tels que la microtopographie ou la captation de surface 3D laser, l'imagerie spectrale et l'analyse par faisceau d'ions. Ces techniques génèrent des données extrêmement diverses par leur taille, leur format, leur nature, leur traitement et leur exploitation. Les données d'AGLAE ne constituent qu'une fraction de cet ensemble et leur archivage pérenne ne se conçoit que dans cet environnement composite. Néanmoins, la création d'Euphrosyne, et plus particulièrement d'Euphrosyne-data, a entraîné la demande de solutions particulières tant pour l'archivage intermédiaire que pour l'archivage pérenne. En tant que première technique du C2RMF à porter un projet d'accès et de conservation pérenne de ses données, AGLAE a vocation à ouvrir la voie pour l'ensemble du laboratoire. Au cours du stage, l'équipe d'AGLAE et moi-même avons dès lors initié, en collaboration avec le SAED, les prospections ainsi que les premiers échanges avec le CINES au nom du C2RMF.

Notre interlocuteur, M. Olivier Rouchon, chef du département archivage et diffusion, nous a permis d'ancrer notre demande dans le contexte actuel de transition du CINES. Ce dernier vient en effet d'obtenir l'agrément du SIAF (Service interministériel des Archives de France) pour la conservation d'archives publiques courantes et intermédiaires au sein de son nouveau système d'archivage reposant sur la solution logicielle VITAM (Valeurs immatérielles transmises aux Archives pour mémoire), laquelle a donné son nom au programme interministériel d'archivage numérique du même nom⁷ :

La solution logicielle VITAM doit permettre aux services producteurs d'archives numériques de gérer les étapes du cycle de vie (archivage courant, intermédiaire, définitif) et l'accès à leurs données. Elle doit également permettre aux services d'archives de remplir leurs missions de collecte, de conservation, de gestion et de communication des archives numériques et de répondre aux besoins de pérennisation.⁸

Porté conjointement par les ministères de l'Europe et des Affaires étrangères, de la Culture et des Armées, piloté par la DINSIC (Direction interministérielle du numérique et des systèmes d'information et de communication de l'État), devenue DINUM (Direction interministérielle du numérique)⁹, et le CIAF (Comité interministériel aux Archives de France), financé dans le cadre du Programme d'investissements d'avenir, VITAM vise d'abord à dé-

7. Ministère de la Culture, *Arrêté du 8 décembre 2020 portant agrément pour la conservation d'archives publiques courantes et intermédiaires*.

8. Site web du programme VITAM : https://www.programmevitam.fr/pages/presentation/pres_archivistes/ (visité le 12 juin 2021).

9. C'est par le décret du 25 octobre 2019 qu'elle a pris la suite de la direction interministérielle du numérique et du système d'information et de communication de l'État (DINSIC).

velopper un socle d'archivage électronique réutilisable par l'ensemble des administrations françaises pour classer, conserver et sécuriser les documents numériques. Cette extension

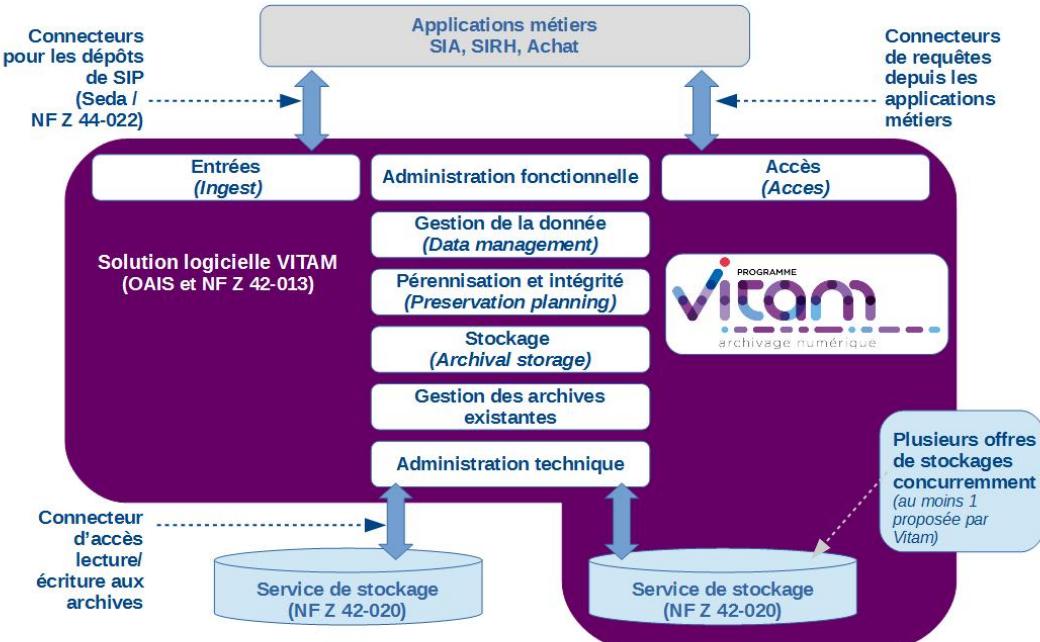


FIGURE 7.1 – Présentation fonctionnelle du programme VITAM. @CINES

du périmètre d'archivage du CINES au regard du cycle de vie a des incidences sur la plateforme PAC (Plateforme d'Archivage au CINES)¹⁰, dont l'infrastructure informatique et matérielle est en cours de migration vers VITAM.

Ce contexte impacte notre projet d'archivage puisqu'il allonge les délais de prise en charge et d'accompagnement des services versants dans leur processus de versement. Cette période traditionnellement comprise entre 6 mois et un an est ainsi rallongée sans échéance précise, augmentant le temps dévolu à chaque étape.

À l'issue de notre entretien, nous avons confirmé la compatibilité des solutions d'archivage du CINES avec les besoins du C2RMF, et notamment ceux d'AGLAE pour l'archivage « à froid » de leurs ressources numériques. L'une de nos réserves majeures était celle de la capacité du Centre à prendre en charge des données contenues dans un format HDF5, ce dernier étant relativement spécifique et nécessitant des outils de lecture adaptés. En effet, bien que non utilisé encore à ce jour par AGLAE, la volonté de recourir à l'HDF5 dans un futur proche oblige à penser dès maintenant sa compatibilité avec le protocole archivistique envisagé. Si aucune donnée n'est aujourd'hui archivée par le CINES dans ce format, le Centre dispose cependant des outils de validation nécessaires. HDF5 est réfé-

10. Voir Figure 7.2.

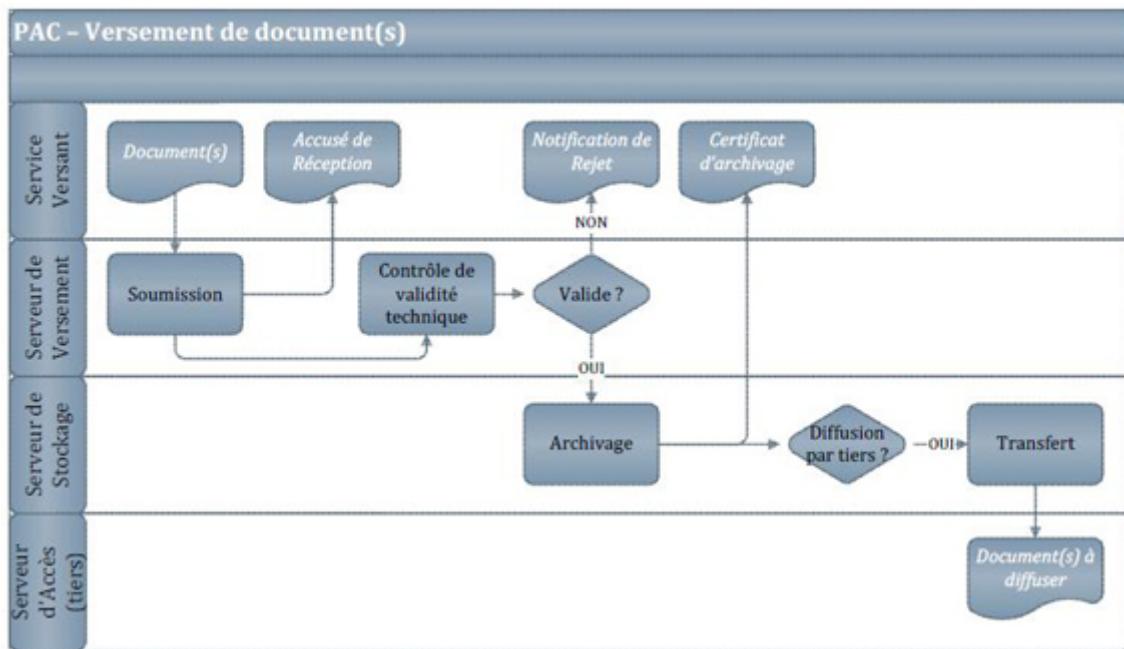


FIGURE 7.2 – Présentation du processus de versement au sein de la PAC (Plateforme d'archivage au CINES) @CINES

rencé et son utilisation est prévue sur le long terme. Par conséquent, le CINES s'engage à l'implémenter si nous en faisons la demande. Cette réserve soulevée et l'obstacle de la durée d'accompagnement accepté, les points particulièrement attractifs du CINES ont été renforcés. Parmi eux, la conformité avec le modèle de référence OAIS (*Reference Model for an Open Archival Information System*) est sans conteste l'argument le plus persuasif et déterminant.

7.2 Pertinence et tropisme du modèle de référence OAIS

Les procédures d'archivage du CINES sont en effet respectueuses du modèle conceptuel et généraliste OAIS, surtout utilisé pour penser l'organisation des systèmes d'archivage électroniques (SAE), bien que ne faisant jamais référence explicitement à des archives. Conçu originellement pour l'archivage des données spatiales¹¹, l'OAIS propose une organisation générique constituée d'une équipe et de systèmes, dont la responsabilité est de pérenniser des informations et de les rendre accessibles à une communauté d'utili-

11. La rédaction de l'OAIS a été pilotée par le *Consultative Committee for Space Data Systems* (Comité consultatif pour les systèmes de données spatiales) mais, de par son caractère abstrait, il est conçu pour s'appliquer à toute autre organisation conservant de l'information. La version 2.0 a été publiée en juin 2012. Elle a fait l'objet d'une traduction française (en accès libre) en octobre 2017. L'OAIS est également enregistré comme norme ISO sous la référence ISO 14721 :2012. Il fait partie des recommandations du Référentiel général d'interopérabilité pour les systèmes d'archivage électronique. Sa documentation est accessible en ligne : <https://www.iso.org/fr/standard/57284.html>

sateurs cibles. Il définit : un vocabulaire commun, des acteurs, des processus, un modèle fonctionnel générique apte à s'appliquer au numérique comme aux contenus imprimés. Il a été conçu pour être indépendant des évolutions technologiques afin de pouvoir inclure tous les standards et normes. Sa dimension largement organisationnelle en fait un outil d'aide à la planification de la pérennisation, au prisme de la gestion de risque. L' OAIS nous offre ainsi l'opportunité d'une jonction entre notre volonté de préservation pérenne, notre ignorance relative des risques technologiques, et notre besoin d'un cadre conceptuel et organisationnel pour bâtir un système d'archivage conforme aux besoins des utilisateurs présents et futurs.



FIGURE 7.3 – Présentation des six fonctions principales du modèle OAIS, orienté vers les besoins utilisateurs. @CINES

Nous avons vu précédemment l'importance de la reconnaissance, de l'identification et de l'appropriation par la communauté d'AGLAE dans le processus de patrimonialisation de certaines données au périmètre sémantique, à la classification et à l'organisation bien définis. Il s'agit ensuite naturellement de les conserver comme les biens rares et précieux qu'elles semblent être, comme on exposerait un objet archéologique dans un musée après s'être assuré de sa valeur culturelle, de son authenticité et de son identification. Or, contrairement au vase exhumé des couches sableuses d'un site lointain, le patrimoine nativement numérique induit une autre forme de vulnérabilité, plus complexe à prévenir que la fragilité matérielle des objets physiques. La ressource numérique est difficile à conserver sur le long terme – et parfois à court et moyen terme – car elle est codée, ce qui pose des obstacles dans le cas de certains formats propriétaires. De plus certains outils logiciels nécessaires à la lecture et à la compréhension des données sont insuffisamment ou non documentés. La vulnérabilité du contenu numérique est multiple car il dépend d'une

technologie intrinsèquement éphémère et de ses variations tant au niveau du support, que du matériel et des logiciels. Le support physique de stockage de l'information numérique est en effet soumise – en fonction du choix qui a été fait - à des risques magnétiques, chimiques, biologiques ou encore thermiques. En cas de dégradation, la perte d'informations peut être brutale et massive. Songeons à la gravité de la perte scientifique et patrimoniale en cas de destruction des données d'AGLAE récoltées lors de l'analyse sur les vitraux de la cathédrale de Chartres, qui ne pourra sans doute jamais être répétée avant plusieurs générations. Nous prenons dès lors toute la mesure de la charge morale et culturelle qui pèse sur l'efficience de la procédure de pérennisation, et l'importance de la méthode qui la gouverne.

Nous avons donc d'abord réfléchi à cette dernière, conformément à l'organisation et aux processus de l' OAIS, en fonction des procédures du CINES et selon la double échelle du C2RMF et d'AGLAE. Cette réflexion multiscalaire n'a pas été aisée et n'est pas encore achevée, en raison notamment de l'actualité des prospections spécifiques au stockage ainsi qu'à la préservation des données produites dans le cadre d'Euphrosyne. De plus, en raison de leurs missions et de leur champ d'action respectifs, les besoins du C2RMF – via le SAED – et ceux d'AGLAE ne sont pas complètement convergents. Ainsi, les utilisateurs présents et envisagés pour le futur du C2RMF désirent consulter des ressources extrêmement hétérogènes par leur nature, leur format de stockage et leur contexte de production. Toutes cependant se caractérisent par leur caractère achevé, souvent même ancien voire historique. Les ressources numériques ne représentent qu'une partie – la plus récente – des fonds conservés depuis plus de 80 ans. Il en va très différemment des utilisateurs d'AGLAE qui souhaitent accéder à des informations nativement numériques et qui sont habitués à évoluer dans le cycle intermédiaire des données. Contrairement au SAED pour l'ensemble du C2RMF, la préoccupation de la pérennisation est récente et ne nous permet pas de définir avec précision les besoins futurs d'accès aux données historiques, faute de précédent en la matière. Pour la dernière étape du cycle de vie, nous sommes plutôt dans l'anticipation d'un besoin futur en raison de la transition culturelle et informationnelle en cours que dans la réponse à une nécessité actuelle. Par ailleurs, le cloisonnement des silos de données répartis par technique scientifique au sein du laboratoire constraint à concevoir au préalable la gestion et l'organisation mutualisées des ressources. En somme, le C2RMF est encore une entité-mosaïque dont l'hétérogénéité intrinsèque constitue un ralentisseur pour la mise en œuvre d'une politique d'ensemble, en l'occurrence celle de l'archivage pérenne.

Aussi, il nous a semblé plus judicieux de procéder par « échantillonnage » en envisageant d'abord l'archivage OAIS des données d'AGLAE, afin de tester la viabilité du processus défini en amont, avant de penser le versement conforme de toutes les ressources à archiver du C2RMF. Notre réflexion a essentiellement porté sur l'organisation et la répartition des différentes tâches inhérentes à la préparation et au traitement archivistiques

internes, préalables au transfert vers une plateforme agréée telle que le CINES¹². Nous avons également abordé collectivement, au sein de l'équipe d'AGLAE et du SAED, les implications structurelles, organisationnelles, cognitives et documentaires des informations de représentation, de contenu et de provenance.

À cette fin, nous avons d'abord dû appréhender les données d'AGLAE sous un angle renouvelé, en nous familiarisant avec la distinction portée par la norme OAIS entre l'information et la donnée. La première fait référence à la connaissance que l'on peut échanger, tandis que la seconde renvoie à la représentation formalisée de cette connaissance. Lors du traitement préparatoire pour l'archivage, il est donc nécessaire de renseigner et de documenter le fond et la forme, cette dernière étant soumise à des technologies fluctuantes dans le temps. Toute une terminologie propre à l'archivage numérique a également dû être apprivoisée, en particulier le vocabulaire relatif aux « paquets d'information », propre au modèle et au circuit OAIS.



FIGURE 7.4 – Présentation des acteurs principaux du modèle OAIS, de la production à la diffusion des « paquets d'information ». @CINES

Il a fallu déterminer les lignes des SIP (*submission information package*) au regard des jeux de données d'AGLAE, et qui correspondent aux objets numériques sur lesquels travaillent les producteurs avant archivage. Afin de conserver la cohérence contextuelle, scientifique et informationnelle, les SIP seraient délimités par le cadre d'un projet de recherche, d'une demande de service, ou d'une expérience réalisée en interne. Tout objet archivé selon l'OAIS est étroitement lié à son information de représentation, laquelle peut être aussi bien une information de structure expliquant l'organisation des informations, qu'une information sémantique fournissant une information complémentaire sur la signification particulière à associer à chacune des informations de structure. *A priori* relativement simples à identifier et à renseigner, ces informations nécessitent une distinction cognitive entre ce qui relève de la forme et ce qui est de l'ordre du contenu, la frontière entre ces deux abstractions intellectuelles étant parfois poreuse. Pour être utile, une infor-

12. Voir Annexe : D.1

mation de représentation doit de plus être adaptée à une base de connaissances partagées par les utilisateurs du contenu archivé, aussi bien actuels que futurs. Il a donc été préconisé à l'équipe d'AGLAE de renseigner l'organisation, l'éventuel standard et le format des contenus sous forme de métadonnées structurées incluses dans le futur fichier HDF5 ou dans un document xml accompagnant le jeu de données inclus dans un fichier zip.

En plus des informations de représentation, le contenu d'information archivé doit également être associé à un certain nombre d'informations de pérennisation, indispensables à son intelligibilité dans le temps. Ces informations en regroupent plusieurs : les informations de provenance qui documentent l'historique du contenu d'information ; des informations d'identification, qui permettent d'identifier sans équivoque le contenu d'information ; des informations de contexte nécessaires à l'établissement des liens entre le contenu et son environnement ; et enfin des informations d'intégrité qui décrivent les mécanismes garantissant le contrôle d'intégrité des informations archivées. La description des paquets d'informations, transmise à l'entité en charge de la gestion des données, est essentielle aux fonctions de recherche, de commande et de récupération des informations contenues dans le système d'archivage. Actuellement, ces informations n'existent que de façon extrêmement incomplète et disséminée, en étant produites à la fois par les humains et par les machines selon un programme de renseignement automatique. Elles constituent sans conteste le volet le plus complexe à mettre en œuvre au sein du déploiement du système d'archivage souhaité. Leur renseignement induit non seulement des ressources humaines, financières et logistiques, mais également une assimilation des logiques et de la culture archivistiques, extrêmement éloignées de l'habitus des scientifiques et des ingénieurs d'AGLAE. Le SAED joue en cela un rôle pivot indispensable pour initier et former les producteurs et utilisateurs de données du laboratoire aux pratiques, ainsi qu'aux réflexes intellectuels relatifs à la pérennisation et à l'archivage. Cette phase d'assimilation et d'élargissement professionnels de longue haleine réclame un temps impossible à quantifier précisément mais incontournable. Elle dépend complètement de la disposition d'esprit des personnes concernées, et de leurs propres rapports aux données et à leur devenir sur le long terme. Certains sont particulièrement sensibles et réceptifs à cette problématique, d'autres considèrent les données d'AGLAE comme inutilement archivables du fait de la caducité rapide liée aux évolutions des connaissances et des techniques scientifiques. De plus, l'absence de précédent rend la légitimité du projet d'archivage encore fragile et à consolider auprès de certains membres de l'équipe qui y voient une charge de travail supplémentaire¹³.

L'organisation et la répartition des responsabilités dans la préparation et le renseignement des paquets d'information ont été au cœur de notre travail d'étude et de prospection, avec pour fil d'Ariane les différents types de contenus descriptifs à produire. À partir de l'organigramme-type du modèle OAIS, nous avons décomposé les tâches spé-

13. Entretiens menés avec les ingénieurs et scientifiques de l'équipe d'AGLAE de mai à juillet 2021.

cifiques de chaque entité et déterminé si elles relevaient plutôt des compétences du SAED ou de celles de l'équipe d'AGLAE. Rappelons qu'un système d'archivage conforme à l'OAIS est organiquement composé de plusieurs entités fonctionnelles : la planification de la préservation, les entrées, l'accès, le stockage, la gestion et l'administration des données. L'entité « entrées » reçoit, contrôle et valide les objets à archiver avant leur transmission à l'entité « stockage ». Les informations nécessaires à leur description et à leur management sont transmises quant à elles à l'entité « gestion des données ». Cette étape est étroitement dépendante du mécanisme de dépôt et de contrôle des données dans la première phase de leur cycle de vie. Procédure à la fois scientifique, informatique et hiérarchique, la vérification de la conformité de la structure et du contenu des métadonnées (données) relève des compétences de l'équipe d'AGLAE, seule à pouvoir déterminer leur intégrité et leur viabilité au regard des objectifs expérimentaux, scientifiques et technologiques. Nous avons considéré les modalités de cette procédure de vérification dans le cadre d'Euphrosyne puisque toutes les données d'AGLAE auront vocation à être collectées et traitées par l'intermédiaire de son interface. Afin de ne pas alourdir excessivement la charge de travail des analystes d'AGLAE, nous avons réfléchi à un protocole automatisé de vérification d'intégrité, inclus dans le *back-office* d'Euphrosyne, et qui devra être développé par l'ingénieur-informaticien de l'équipe et qui nécessitera une phase de test en situation expérimentale. Cette ambition devra résoudre les obstacles liés à la complexité de l'automatisation d'un processus cognitif extrêmement ingénieux et difficilement reproductible par une machine. Pour l'heure, elle prendra la forme d'une vérification de l'effectivité de renseignement des champs obligatoires – comme certaines informations d'organisation – et de conformité numérique par rapport au format et à la structure de données attendus. Avant le transfert à l'entité « stockage », les informations de contrôle et de validation auront donc été renseignées, du moins automatiquement, lors de leur collecte à AGLAE. Notons dès à présent que cette dynamique de transformation des SIP en AIP impliquera une structuration des informations de représentation quelque peu différente selon que le jeu de données à archiver soit contenu dans un fichier HDF5 ou dans un dossier compressé. Dans un fichier HDF5, il serait intéressant d'exploiter le potentiel fort de granularité en diffinant les informations de représentation dans des niveaux d'arborescence adaptés à l'hétérogénéité des données et de leurs formats.

Si les responsabilités de l'entité « entrées » incombent exclusivement à l'équipe d'AGLAE, elles sont en revanche partagées avec le SAED dans le cadre de la gestion des données. Cette entité fonctionnelle prévoit la fourniture des informations de description des données archivées, ainsi que la mise à jour des informations internes nécessaires au système d'archivage. Elle implique notamment l'information de l'évolution des formats utilisés ainsi que de toute modification éventuelle de structure et de choix de métadonnées. Le renseignement des informations de description présentées précédemment de façon générique prend la forme de métadonnées structurées en *Dublin Core* ou en SEDA (Stan-

dard d'échange de données pour l'archivage), le CINES – par l'intermédiaire de VITAM – favorisant plutôt cette deuxième option, plus riche et informative. Structuré en *xml*, le SEDA, standard tant archivistique que technique, est directement inspiré des normes et des structures de description – notamment l'EAD – des archives imprimées comme numériques. Son rôle est de fournir un cadre pour l'échange des archives et de leurs métadonnées, en garantissant la qualité des informations indispensables à leur conservation¹⁴. Il prévoit le renseignement de métadonnées sémantiques selon des balises préétablies. Les métadonnées descriptives du SEDA sont ainsi :

1. L'unité d'archives (*ArchiveUnit*), potentiellement subdivisable en sous-unités intellectuelles.
2. Le renseignement du contenu d'archives avec une description libre (*Description*), des dates (*StartDate*, *EndDate*, etc.), la langue du contenu d'information (*Language*), l'historique de la conservation (*CustodialHistory*), des références à des objets liés (*DataObjectReference*), la position du contenu d'information dans le plan de classement du service producteur (*FilePlanPosition*), ainsi que d'autres données descriptives libres¹⁵.

Ce travail de renseignement nécessite une capacité d'analyse, de repérage et de discrimination qualitatifs des informations. Il ne peut être assumé que par des individus ayant une connaissance fine des données, de leur contexte de production, de leur signification dans le cadre d'usages déterminés, et de leur intégration dans le plan de conservation et de classement si ce dernier existe. On devine aisément que cette dernière information relève plutôt des compétences du SAED, en charge de la politique générale de conservation des ressources du C2RMF. En revanche, les autres métadonnées descriptives sont du ressort des spécialistes et gestionnaires à proximité immédiate des données d'AGLAE. Toute évolution dans le choix des formats et des métadonnées s'inscrit également dans le périmètre professionnel direct de l'accélérateur. Contrairement aux informations inhérentes aux entrées, il n'est pas possible – du moins à l'heure des développements technologiques actuels – de substituer l'effort humain par la prouesse de l'intelligence artificielle et de l'automatisation. Le renseignement de toutes ces métadonnées descriptives va donc s'ajouter aux missions déjà conséquentes des membres de l'équipe, soulevant une fois de plus la difficile question des ressources humaines allouées à cette tâche. Par ailleurs, seulement trois membres d'AGLAE sur les cinq permanents de l'équipe sont véritablement en mesure de renseigner les informations relatives aux contenus et au contexte de production en

14. Le SEDA spécifie six transactions : demande de transfert, transfert, élimination, modification, communication, restitution. Elles ont lieu entre six acteurs différents : service producteur, service versant, service d'archives, opérateur de versement, demandeur d'archives et service de contrôle. [Mintika, « Le SEDA dans VITAM », 7 mars 2019, <https://hub.mintika.fr/2019/03/07/le-seda-dans-vitam/> (visité le 17 juin 2021)]

15. Ministère de la Culture et de la Communication, Service interministériel des Archives de France, *Standard d'échange de données pour l'archivage. Transfert-communication-élimination-restitution-modification*, version 2.1, juin 2018.

raison de leur suivi des projets de recherche, des demandes de service et des expériences menées en interne. Les deux autres assurent en effet l'expertise mécanique, technique et métrologique de l'accélérateur. Là encore, une formation à la description archivistique des contenus semble au préalable nécessaire, ne serait-ce que pour assimiler le langage de balisage xml, ainsi que les logiques sous-jacentes au SEDA et à son arborescence sémantique. Le passage de relais au SAED s'effectue quant à lui lors de l'indication des informations propres à la gestion technique et archivistique nécessaire à l'accès. Cette responsabilité s'insère de façon cohérente dans le cadre de missions du SAED en l'enrichissant de procédures nouvelles spécifiques à l'archivage en OAIS des ressources numériques, et conformément au SEDA.

L'achèvement de ce travail de renseignement archivistique sur le fond et la forme des contenus archivés donne ensuite lieu à l'intervention de l'entité « stockage et sauvegarde des données ». Cette dernière assure la conservation physique des objets archivés et les tient à disposition de l'entité « accès ». En conformité avec les normes établies par l'entité « administration », elle prend en charge la réalisation des copies multiples et le renouvellement des supports anciens. Dans notre cas, la fonction de stockage et de sauvegarde serait supportée par le CINES et éventuellement par Huma-Num pour les archives intermédiaires. Le CINES assure une triple réplication des données : au CINES à deux emplacements géographiques différents, et une fois à Lyon à l'IN2P3. Ainsi, en cas de dommages causés à l'infrastructure matérielle, la sécurité et l'intégrité des contenus versés sont garanties. Le CINES prend en charge la « transformation » des supports technologiques, ce qui recouvre la migration de formats afin de prévenir l'obsolescence. Partenaire étroit d'Huma-Num pour la conservation à long terme des données de la recherche en SHS¹⁶, le CINES manifeste une volonté politique d'inclusion de l'accès et de la transformation des données en proximité directe avec leur lieu de stockage¹⁷. En cela, le stockage et la sauvegarde des données d'Euphrosyne constituerait un « projet-cobaye » intéressant car nécessitant la prise en compte de plusieurs étapes du cycle de vie des données, qui débordent du cadre de l'archivage « à froid ». Une étroite association de l'offre d'Huma-Num et de celle du CINES permettrait de recouvrir avec fluidité la temporalité spécifique des contenus produits par l'intermédiaire d'Euphrosyne, sans rupture marquée entre les infrastructures de stockage/d'accès des données intermédiaires et définitives. Plusieurs options sont possibles, et sont aujourd'hui en cours d'étude mais soulève des enjeux financiers et technologiques. Si nous versions nos données intermédiaires à Huma-Num, ce serait dans l'objectif d'offrir aux utilisateurs un accès souple et fréquent, ce que le CINES n'est pas en mesure de proposer. Par ailleurs, le recours à Huma-Num nous affranchirait des coûts de stockage au CINES des données définitives puisque la TGIR les prend en charge.

16. La plateforme PAC du CINES a ainsi été adaptée aux spécificités des SHS et à leurs métadonnées descriptives, et les ressources de stockage ont été mutualisées avec le centre de calcul de l'IN2P3.

17. Entretien avec Olivier Rouchon, chef du département archivage et diffusion du CINES, 4 juin 2021. Étaient présents : Claire Pacheco, Olivier Rouchon, Marine Zelverte et moi-même.

Toutefois, cet argument financier attractif s'est trouvé nuancé par les réserves émises au sujet de la volumétrie de nos données. Supérieure à 10 To, nous devrions envisager le co-financement pour que le support de nos contenus ne soit pas disproportionné par rapport aux autres projets. De plus, l'offre de Nakala proposée par Huma-Num est mal-adaptée à la taille et à l'état de nos données. Elle est en effet prévue pour des données proches de leur publication et non pour des données toujours en cours d'évolution. Si le versement de nos contenus d'information au CINES ne fait guère de doute, celui de l'accès à nos données intermédiaires à travers les offres d'Huma-Num est toujours en cours d'étude et doit être validé tant par la TGIR que par le C2RMF. Dans le cas des archives courantes d'AGLAE, le stockage et la sauvegarde seraient réalisés par le SAED, éventuellement en fonction d'un plan de gestion des données qui, s'il a déjà été élaboré, n'a jamais été mis en œuvre.

Une fois archivés, les SIP deviennent des AIP (*archival information package*), puis des DIP (*dissemination information package*) lors de leur mise à disposition auprès des utilisateurs. Précisons que l'OAIS identifie une classe particulière d'utilisateurs cible, considérée comme la population prioritairement bénéficiaire du service d'archivage. Dans notre cas, il s'agirait des chercheurs spécialisés en sciences des matériaux patrimoniaux, et d'analystes ayant besoin d'accéder à des données issues d'expériences passées. Selon le modèle OAIS, l'entité « Accès » regroupe tous les services qui permettent l'interface directe avec les utilisateurs. En plus des fonctions de contrôle d'accès, il s'agit principalement de permettre aux utilisateurs de rechercher dans un catalogue des objets archivés et de leur fournir le résultat de leur commande. Afin d'inscrire Euphrosyne à sa juste place dans le futur système d'archivage des données d'AGLAE, l'accès a été modélisé dans le cadre d'Euphrosyne-data, destinée à fournir un catalogue bâti sur les référentiels de l'analyse par faisceau d'ions, ainsi qu'un contrôle des accès utilisateurs pour les archives semi-active/intermédiaires. Euphrosyne-data conserverait également l'historique des requêtes afin de suivre les besoins utilisateurs. Notre modèle ne présente donc pas l'organisation de l'accès pour l'ensemble des données d'AGLAE archivées mais seulement une partie d'entre elles. Cette fragilité conceptuelle tient à l'usage quelque peu détourné d'une partie du modèle au profit de la construction d'Euphrosyne et de la définition de ses fonctionnalités. Par ailleurs, en vue de nos échanges avec le CINES et Huma-Num, nous tenions à présenter le rôle souhaité d'Euphrosyne-data dans le système d'archivage futur, et justifier notre besoin d'une infrastructure « tout-en-un » capable d'héberger Euphrosyne-data tout en assurant la pérennisation de ses ressources dans la quasi-intégralité de leur cycle de vie.

Conformément à l'OAIS, nous avons enfin porté notre réflexion sur l'entité « Administration des données », chargée d'assurer la coordination générale du système et d'en établir les règles internes. Elle veille notamment à la qualité générale du service rendu aux utilisateurs et à son amélioration, tout en rendant compte au management. L'équipe

d'AGLAE est la plus légitime à assurer ces dernières responsabilités. En revanche, la coordination générale du système d'archivage relève plutôt des compétences du SAED.

Bien plus organisationnel et conceptuel que pratique, ce travail de modélisation ne propose qu'un cadre général pour la répartition des rôles et des responsabilités, tout en dessinant l'ossature générale du système d'archivage. À l'aune des prospections menées auprès du CINES et d'Huma-Num, il est apparu que notre demande en recouvrail plusieurs, qui gagneraient à être instruites et englobées dans une réflexion holistique : l'hébergement des outils et des données d'Euphrosyne ; le stockage et l'accès régulier aux données intermédiaires d'AGLAE ; la conservation pérenne et l'accès plutôt ponctuel aux archives définitives sur le très long terme des contenus numériques du C2RMF. Cette approche de l'archivage n'est pas dissidente et trouve écho dans les partenariats actés et en cours entre les infrastructures d'archivage « à froid » et celles dédiées aux contenus intermédiaires et courants.

Chapitre 8

Sélection et désherbage des données : ciseler pour conserver

La notion de patrimonialisation envisage le patrimoine comme une activité dynamique dont les objets ne sont que les supports matériels¹. Dès lors intervient la notion de dépôt, voire de stock² pour la masse numérique archivée. Comme nous l'avons vu précédemment, cette conservation implique la préservation, chargée d'éviter pertes et dommages dus à la non-pérennité des supports. Il s'agit de préserver les objets – là encore, au sens générique du terme – afin de les exposer et de les valoriser. Néanmoins, l'aporie du stock inhérente à sa masse est telle qu'il n'est pas question de tout conserver mais, au contraire, d'opérer une sélection rigoureuse en amont du système d'archivage en fonction de la valeur que l'on souhaite attribuer et diffuser à certains contenus numériques. Les fonds patrimoniaux ne se désherbant pas, il convient de réaliser les coupes éventuelles avant la constitution des paquets d'information à archiver, lors du contrôle et de l'interprétation des données produites par un ou plusieurs analystes. Dans une relation d'interdépendance déterminante, l'élimination et la sélection constituent l'autre versant de la pérennisation appliquée au patrimoine numérique. C'est bien en vertu des suppressions délibérément réalisées que ce qui subsiste acquiert sa valeur. Indispensable, l'acte de suppression porte néanmoins de lourds enjeux et impose une stratégie définie de façon concertée par l'instance en charge de l'archivage et par les producteurs de données. Après avoir restitué les pratiques et les perspectives du SAED en la matière, nous resserrerons la focale sur les implications spécifiques qu'elles soulèvent pour les données d'AGLAE.

1. Matteo Treleani, *Qu'est-ce que le patrimoine numérique ? Une sémiologie de la circulation des archives*, Lormont, Éd. Le Bord de l'eau, 2017, p. 44.

2. Matteo Treleani, « Le patrimoine en ligne a-t-il un sens ? », *e-Dossiers de l'audiovisuel, L'extension des usages de l'archive audiovisuelle*, Ina, 2014, <http://www.ina-expert.com/e-dossiers-de-l-audiovisuel/le-patrimoine-en-ligne-a-t-il-un-sens.html> (visité le 27 juillet 2021)

8.1 L'esquisse d'un plan de gestion des données pour le C2RMF : le projet Lavau

Impliquant les notions de valeur de l'information, de mémoire et de patrimoine, le désherbage appartient initialement à la terminologie des bibliothèques. Complexé, son sens est plus général que celui de l'élimination. Il regroupe à la fois les opérations d'évaluation, de tri et de retrait des documents ainsi que leur traitement ultérieur. Bertrand Calenge le définit comme un « acte volontariste qui associe le récolement régulier d'une collection et le retrait de certains documents d'un fonds particulier »³. Cette interprétation communément admise, valable pour les fonds physiques et non patrimoniaux, s'applique extrêmement mal aux ressources numériques potentiellement patrimoniales. Nous préférerons ainsi la proposition de l'*American Library Association* (ALA) :

Simply put, weeding is selection in reverse. It is deselection. Weeding is the act of reevaluating items in the collection and removing any that are inaccurate, out of date, misleading, inappropriate, unused, in poor condition, or otherwise harmful to students [...] It is a professional responsibility that cannot be taken lightly.⁴

Bien que prévue pour les collections en bibliothèques, cette définition peut également s'appliquer aux contenus numériques qui nous concernent. Le désherbage est avant tout une opération intellectuelle vouée à discriminer ce qui a de la valeur de ce qui n'en a pas, et par conséquent d'anticiper sur les représentations et les usages futurs. Cette évaluation s'insère dans une chaîne complexe de connaissances, de technique et d'expertise.

Traduit au numérique, le désherbage est dissout dans la planification de gestion de données. L'objectif demeure similaire mais il est adapté aux spécificités sémiologiques et structurelles des ressources numériques. Au C2RMF, une première esquisse de plan a été conçue dans le cadre du projet Lavau. Baptisé du nom d'un complexe funéraire monumental de l'Aube, ce dernier s'est inscrit dans un Programme commun de recherche (PCR) entre le C2RMF et l'Institut national de recherches archéologiques préventives (Inrap). Le C2RMF a eu la responsabilité de fournir les moyens d'analyse et les compétences scientifiques nécessaires à l'étude et à la restauration d'une partie des biens archéologiques issus du site de Lavau. Les données produites au cours du programme ont nécessité la réalisation d'un plan de gestion lors de la phase 2 du déroulement du projet, de janvier à mai 2020⁵. Extrêmement brève, cette période a surtout permis de dresser un inventaire des différents types de données et l'ébauche d'une proposition de sélection pour chaque technique scientifique du C2RMF. Sous forme d'un tableau Excel subdivisé par groupe

3. Bertrand Calenge, *Les Politiques d'acquisition*, Paris, Éditions du Cercle de la librairie, 1994, p. 337.

4. Donna J. Baumach, Linda L. Miller, *Less is More, a practical guide to weeding school library collections*, USA, American Library Association, 2006, p. 3.

5. Marine Touret, *Rapport interne du plan de gestion des données du projet Lavau*, C2RMF, 2020.

scientifique, chaque type de donnée – les spectres primaires de la méthode PIXE à titre d'exemple – se voit attribuer une évaluation selon plusieurs critères : l'intérêt à conserver, le risque à supprimer, la préconisation et la durée d'utilité administrative⁶. Ce recensement a été conduit conjointement avec les producteurs des données en fonction de la valeur patrimoniale, pédagogique et historique des données. Lorsque l'élimination est préconisée, les données sont préservées jusqu'à la fin de la durée d'utilité administrative (DUA) puis détruites. Dans le cas d'une conservation recommandée, elles sont destinées à l'archivage à long terme par les Archives Nationales.

8.2 Le tri des données d'AGLAE : un dilemme à la Corneille

Les données d'AGLAE relatives au projet sont incluses dans le plan de gestion de données du projet Lavau et ont ainsi été évaluées, en concertation avec le groupe New AGLAE. Nous disposons ainsi d'une typologie sélective des données, réalisée pour le projet Lavau en 2020 dans le cadre d'une réflexion autour du système d'archivage encore balbutiante. Débattre et mettre en perspective les préconisations du projet Lavau en matière de tri et de sélection est nécessaire pour se saisir de cette étape aussi difficile qu'indispensable à l'entreprise de pérennisation et, plus largement, au processus de patrimonialisation. S'agissant des données d'AGLAE, l'opération est particulièrement hasardeuse voire corrélique, tant les intérêts et les usages s'entrechoquent parfois, notamment en raison de leur plasticité temporelle.

Le plan de gestion propose tout d'abord l'élimination des fichiers LST et des spectres primaires en format ASCII des méthodes PIXE, PIGE, RBS et IBIL, considérés comme non pertinents en raison de leur caractère binaire. L'auteure du plan justifie cette suppression par l'inutilité des données brutes une fois traitées :

Le traitement de la donnée elle-même permet de voir comment a été réalisé le dépouillement et de s'assurer de sa qualité, d'où l'inutilité de conserver sur le long terme les données brutes. Il ne peut pas y avoir de doute quant au résultat, donc pas de nécessité de dépouiller à nouveau la donnée.⁷

Ce postulat démontre que la possibilité du rejet intellectuel de l'expérience n'est pas envisagée. La suggestion de ne conserver que les fichiers déjà convertis pose ainsi la question du respect de l'intelligibilité des données dans la totalité de leur cycle de vie. En cas d'obsolescence ou d'inaccessibilité de l'outil logiciel nécessaire à la conversion, il n'est plus possible d'accéder à l'état primitif des données. Si cette recommandation est totalement valable d'un point de vue scientifique et culturel, elle pose davantage de difficulté pour

6. Le temps de la durée d'utilité administrative (DUA) est de cinq ans à l'issue du projet.

7. Marine Touret, *Rapport interne....*, p. 18.

l’archivage pérenne. Ce dernier exige une compréhension stable des contenus et des informations de représentation qui y sont attachés au-delà des évolutions de formats et de supports. L’élimination de certaines données primaires peut contrevenir à cette exigence et doit être réétudiée en concertation avec l’infrastructure de stockage à long terme – certainement le CINES – qui recevra les versements.

Les propositions de sélection, quant à elles, consacrent largement les données traitées, considérées comme indispensables : les spectres extraits grâce aux différents logiciels en fonction des techniques d’analyse souhaitées ; les données cartographiques en format PNG ; et les fichiers JPEG relatifs aux cartographies traitées. Ces choix reflètent la prépondérance accordée à l’utilisation scientifique des données sur leur pérennisation et leur archivage dans une dynamique patrimonialisante. Cela s’explique aisément par le fait que le plan de gestion de données du projet Lavau a été prévu pour répondre aux besoins des chercheurs et des spécialistes du projet Lavau et n’a pas de vocation archivistique directe. Les évaluations et recommandations réalisées ont prioritairement pour vocation de réduire la charge des données sur les différents supports de stockage du C2RMF, dont les limites risquent d’être éprouvées à court et moyen terme. Elles dévoilent la difficulté de conciliation des enjeux de la pérennisation avec les usages immédiats ou à moyen terme des utilisateurs. Aussi, planifier la gestion de données équivaut à prioriser une temporalité sur les autres, au prix parfois de la transmission aux générations futures, perçues comme une abstraction prospective trop insaisissable. Fruit d’une science et d’une technologie en perpétuelle évolution, les jeux de données expérimentaux d’AGLAE induisent un entrecroisement aisément litigieux de leur exploitation directe et de l’impératif patrimonialisant de leur conservation au long cours. On peut évoquer un dilemme cornélien entre l’espoir d’une transmission incorruptible par le temps et l’appropriation nécessaire mais altérante des utilisateurs actuels. Ces derniers tendent à considérer les données brutes comme une matière première peu utile dès lors qu’elle a servi à l’extraction des informations dans le format recherché. Par ailleurs, les utilisateurs, qui effectuent de réguliers va-et-vient entre les différents états de données et procèdent à des réitérations correctives, éliminent parfois les fichiers intermédiaires des versions jugées erronées, nous interdisant la restitution du cheminement intellectuel et scientifique qui a abouti à l’état définitif des données. À titre de comparaison, cela équivaut à effacer les ratures et les corrections des versions intermédiaires manuscrites d’une œuvre littéraire pour ne conserver que la forme finale. Nous en perdons alors l’épaisseur génétique et l’accès à la construction évolutive de la pensée. Si cette perte n’influe pas négativement sur les utilisateurs présents d’AGLAE, elle est fortement préjudiciable pour les historiens des sciences patrimoniales à venir tout comme pour les scientifiques. Les échanges menés régulièrement avec Marine Zelverte, cheffe du SAED, ont abouti à une posture de prudence et d’humilité à l’égard de l’état actuel de nos connaissances en matière de pérennisation numérique et des modalités d’évaluation des contenus qui pourraient en être écartés. En effet, l’évolution de paradigme qui

a conduit à l'extension exponentielle du patrimoine est encore relativement récente. Il est ainsi particulièrement périlleux de réaliser des coupes sur des contenus en cours de patrimonialisation qui n'ont encore jamais fait l'objet d'une consultation *a posteriori* par un individu extérieur à l'expérience. Si nous pouvons déterminer ce qui est nécessaire au rejet de l'expérience de ce qui est superflu, il est plus délicat de discriminer ce qui sera indispensable à l'intelligibilité de la forme et de la structure des données à travers le temps. Cela exige des compétences spécifiques, détenues par les spécialistes de l'archivage numérique pérenne.

Par conséquent, à ce jour, nous souhaitons recommander aux utilisateurs de ne pas initier de suppressions au sein d'un jeu de données expérimentales, et de soigner particulièrement les noms de fichiers selon les dernières recommandations émises par le SAED afin de restituer une forme de traçabilité depuis les fichiers primaires jusqu'aux fichiers définitifs. Ce *statu quo* sera maintenu jusqu'à l'accompagnement et la formation des équipes du SAED et d'AGLAE par l'infrastructure d'archivage pérenne retenue.

Chapitre 9

Pérenniser : nécessité et utilité d'un inachèvement perpétuel

Si nous abordons en surplomb ce système voué à offrir une permanence à une immatérialité à l'épaisseur culturelle et historique avérée, nous devons affronter la question lancinante de la nécessité et de l'utilité d'un inachèvement perpétuel. En quoi cet état, qui contrevient a priori à l'exigence de rentabilité et de justification des coûts relatifs à un processus aussi lourd et complexe, est-il finalement à la fois nécessaire et utile ?

9.1 Du signe à la trace : une sémiologie nourrie du mouvement

Réfléchir au sens et à la nécessité de l'inachèvement intrinsèque à la pérennisation revient à appréhender les supports et les enjeux de la mémoire, au carrefour de la sémiologie et de la phénoménologie, dans une perspective profondément patrimoniale. La médiologue Louise Merzeau insiste ainsi sur le fait que le stockage et la conservation de données ne sauraient constituer, à eux-seuls, une mémoire¹. Bien au contraire, l'immobilité de cet environnement exposerait les données à une paralysie délétère à leur amplitude testimoniale au regard d'une histoire socio-culturelle à transmettre. La pérennisation est donc moins traversée d'un inachèvement que d'un mouvement perpétuel entre les « signes » que sont les données, et les « traces » comprises comme des empreintes mémorielles² que peuvent offrir ces dernières. Cette circulation perceptive est précisément ce qui anime et vitalise la conservation en lui apposant la marque de la conscience. L'architecte de la pérennisation doit par conséquent sans cesse s'interroger sur la concordance des données en tant que signes organisés avec l'environnement sémiologique chargé de les relier à leur fonction mémorielle et patrimoniale, ces deux dimensions se confondant aisément. Les supports

1. Louise Merzeau, « Du signe à la trace », *Médium*, 18, 2009, p. 21-36.

2. Paul Ricoeur, « Mémoire, Histoire, Oubli », *Esprit*, 3, 2006, p. 27.

technologiques, les formats et l'organisation des données doivent donc évoluer symbiotiquement avec les fluctuations phénoménologiques, lesquelles doivent être perçues comme les variations de la conscience à l'égard de ce qui mérite, au regard de la mémoire et de l'histoire, l'« éternité numérique »³. Si l'on considère l'étendue des données en sciences du patrimoine, dont AGLAE ne forme qu'une parcelle, on devine que cette médiation opérée au moyen de l'intelligence, de la capacité de discrimination et d'interprétation humaines, ne peut couvrir uniformément l'ensemble des « signes » conservés. Elle doit donc *a minima* œuvrer à la « traçabilité »⁴ de leur unité signifiante⁵, tant technologique que cognitive. C'est ainsi que la « trace », avec son ancrage perceptif et culturel, l'emporte finalement sur le « signe » dans les traductions concrètes de l'effort de préservation au long cours. La pérennisation n'est autre que la traduction logistique, technique, organisationnelle et conceptuelle de cette quête de traçabilité, marquée par une convergence toujours perfectible entre le signifiant et le signifié.

Cette entreprise est fondamentale si l'on en croit Yves Jeanneret, spécialiste en sciences de l'information et de la communication, qui affirme dans sa *Critique de la trivialité* : « la trace ne fonctionne pas sans le signe »⁶. La trace est en effet simultanément une résurgence du passé et une inscription dans une interaction actualisée. Dans le monde numérique, elle n'est pas un simple produit cognitif mais elle est également une entité soutenue par une base technologique qui en assure la visibilité humaine. Elle tend alors à devenir un « signe-trace »⁷ où la perte de l'un entraîne irrémédiablement la disparition de l'autre. Notons que le support apportant à la forme d'expression sa manipulabilité, la donnée ne peut être émancipée, dans sa préservation, de son format, du logiciel ou du matériel nécessaire à sa lecture et de sa structuration⁸. C'est pourquoi la prise en charge de la migration pérenne des supports et des formats des données d'AGLAE ne saurait être considérée comme un effort potentiellement stérile tacitement dicté par l'importance de conserver les résultats d'un instrument aussi coûteux qu'un accélérateur de particules. Il est bien question d'une maintenabilité concomitante des données ainsi que des informations de représentation et de contenu qui leur sont liées afin d'assurer la permanence d'un « signe-trace », patrimonial en raison de la valeur attribuée par sa fonction. De cette constance par le mouvement et le changement dépend l'insertion réussie des sciences du

3. Hélène Bourdeloie, Christine Chevret-Castellani, *L'Impossible patrimoine numérique ? Mémoire et Traces*, Lormont, Éd. Le Bord de l'eau, 2019, p. 103. Le principe d'éternité numérique doit être compris comme un horizon idéal et non comme une réalité.

4. Pour approfondir la théorie de la traçabilité numérique : Michel Arnaud et Louise Merzeau, « Introduction », *Hermès*, 53, 2009, Paris, CNRS Éditions, p. 9-12.

5. Il s'agit le plus souvent du « paquet d'informations » au sens archivistique et conformément au modèle OAIS.

6. Yves Jeanneret, *Critique de la trivialité. Les médiations de la communication, enjeu de pouvoir*, Paris, Éd. Non Standard, p. 469.

7. Béatrice Galinon-Méléne, « À la recherche de la trace », *Communication et organisation*, 47, 2015, p. 35.

8. Bruno Bachimont, *Patrimoine et numérique, Technique et politique de la mémoire*, Paris, INA, 2017, p. 30.

patrimoine dans un monde de l'information dont elles dépendent désormais pour leur légitimité sociale.

9.2 La rematérialisation perpétuelle pour maintenir la conscience

D'un point de vue sans doute plus métaphysique que culturel et mémoriel, nous pouvons également affirmer que la lutte perpétuelle contre l'obsolescence matérielle des supports de contenus engage aussi la vitalité de la conscience humaine pour contrer sa condition et garantir la permanence de son flux en transcendant ses caractères éphémères. Bien entendu, l'orientation prise par la pérennisation numérique à travers des organismes tels que le CINES vise avant tout à répondre à des besoins-métiers, rationnels voire pragmatiques. Mais si nous envisageons le processus à travers le prisme de la patrimonialisation, la rationalité des usages par les chercheurs se double d'une envergure que nous serions tentés de qualifier d'existentialiste. En contrant sans cesse la déchéance de l'environnement et des supports physiques des données, les acteurs de la pérennisation œuvrent contre la fatalité de l'oubli et le déterminisme du temps, en préservant un patrimoine qui en porte intrinsèquement les germes par sa « matérialité immatérielle », mais qui participe également de la construction d'une représentation renouvelée de l'Homme, dans laquelle l'esprit peut désormais se confondre étroitement à la matière. Les aspects technologiques et techniques de la pérennisation ne sont ainsi que la partie émergée d'une conquête qui se fond dans une évolution civilisationnelle où l'être humain s'efforce de dompter la machine pour la rendre réductible à ses besoins et ne pas être débordé par elle. En cela, le stockage, la préservation et l'accès à long terme aux données d'AGLAE constituent une entreprise culturellement et socialement utile, mais ontologiquement nécessaire. La nuance mérite d'être approfondie : la pérennisation des données d'AGLAE est bénéfique à l'édifice culturel et historique de nos sociétés, qui s'en trouve enrichi et redimensionné sans être complètement dépendant d'elle, en raison de l'existence physique des objets patrimoniaux. En revanche, la question de l'être, de son sens et de sa place dans un environnement fortement dématérialisé impose la maîtrise des éléments numériques de son patrimoine culturel, et la préséance de la pensée humaine sur la technique. En ce sens, l'inachèvement inhérent à la pérennisation démontre que les enjeux du processus sont prédominants sur sa finalité, inatteignable par ailleurs :

À cet instant subtil où l'homme se retourne sur sa vie, Sisyphe revenant vers son rocher, dans ce léger pivotement, il contemple cette suite d'actions sans lien qui devient son destin, créé par lui, uni sous le regard de sa mémoire et bientôt scellé par sa mort. Ainsi, persuadé de l'origine tout humaine de ce qui est humain, aveugle qui désire voir et qui sait que la nuit n'a pas de fin,

il est toujours en marche. Le rocher roule encore [...] Cet univers désormais sans maître ne lui paraît ni stérile ni futile. Chacun des grains de cette pierre, chaque éclat minéral de cette montagne pleine de nuit, à lui seul, forme un monde. La lutte elle-même vers les sommets suffit à remplir un cœur d'homme. Il faut imaginer Sisyphe heureux.⁹

9. Albert Camus, *Le Mythe de Sisyphe...*, p. 168.

Quatrième partie

Euphrosyne-data : exposer et
disséminer pour l'*Open Science*

Protégés de l'épreuve du temps au moyen de leur pérennisation, les jeux de données d'AGLAE pourront ensuite être exposés et valorisés. Lors de la présentation d'audience d'Euphrosyne du 8 avril dernier pour l'obtention du statut de start-up d'État, la volonté d'ouvrir, d'exposer et de disséminer les données d'AGLAE a ainsi innervé l'ensemble de l'argumentaire. Si la création d'un accès à distance pour les utilisateurs d'AGLAE constitue la première phase de déploiement d'Euphrosyne, la réutilisation des données d'AGLAE par d'autres chercheurs dans le cadre de l'*Open Science* est au cœur de la seconde. Une troisième et dernière phase renvoie à la réutilisation du modèle d'Euphrosyne par d'autres plateformes ouvrant les données d'instruments de sciences du patrimoine dans le cadre d'IPERION-HS/E-RIHS. Il ne s'agit pas seulement d'un effort de mise en conformité avec la politique d'ouverture promue par l'Union Européenne, mais également de mutualiser des ressources relativement proches par leur nature et leur utilisation dans un mouvement de coopération et de fédération déjà bien amorcé dans le cadre de FIXLAB et de MOLAB. Par une forme d'articulation entre l'exposition historique et patrimonialisante des données et les besoins de recherche des utilisateurs, l'intégration interopérable des données d'AGLAE dans les circuits de diffusion de l'*Open Science* ouvre des voies de conciliation entre une forme de muséification numérique et un rapport utilitaire aux données. L'exposition et la diffusion des données exigent le partage d'une syntaxe et d'un vocabulaire communs.

Chapitre 10

Construire un catalogue : sacrement patrimonial des données ?

10.1 Le catalogue d'Euphrosyne-data : viatique de patrimonialisation

Ce vocabulaire commun doit venir alimenter un catalogue, véritable pierre angulaire d'Euphrosyne-data puisqu'il permettra aux utilisateurs de l'interface de chercher et de trouver les contenus recherchés en fonction de champs spécifiques d'interrogation. Dans son ouvrage intitulé *Vers de nouveaux catalogues*, Emmanuelle Bermès rappelle que le catalogue est une entité mouvante qui a endossé plusieurs fonctions au fil du temps :

Outil de gestion des collections pour les professionnels, système informatisé permettant d'automatiser les processus métiers tels que le catalogage et la circulation des documents, le catalogue est aussi, avant tout, l'interface proposée aux lecteurs pour chercher et trouver les documents et ressources de la bibliothèque.¹

À l'instar du désherbage et de l'archivage, la notion de catalogue recouvre une pluralité fonctionnelle et sémantique, d'autant plus complexe qu'elle a d'abord été prévue pour les bibliothèques traditionnelles et non pour les ressources et les usages numériques. Transposé au web, le catalogue est concurrencé par les moteurs de recherche, lesquels fonctionnent sur des principes relativement similaires mais avec une envergure technologique adaptée :

Rapidité de réponse, classement des résultats par pertinence, point d'entrée unique pour toutes les recherches, ergonomie simple et intuitive sont devenus des exigences naturelles à la hauteur desquelles le catalogue doit se hisser s'il souhaite continuer à exister en tant que tel.²

1. Emmanuelle Bermès (sous la dir. de), *Vers de nouveaux catalogues*, Paris, Éditions du Cercle de la Librairie, 2016, p. 9.

2. *Ibid.*

Dès lors, le catalogue trouve sa légitimité et sa force dans la pertinence et l'efficacité du signalement des ressources auxquelles il donne accès. Construit à la fois pour des usages-métiers connus et dans un but d'exposition culturelle et didactique, il offre l'avantage d'une clé d'accès sur mesure aux contenus signalés. Parce qu'il procède d'une démarche double d'inventaire raisonné et de médiation, le catalogue peut être perçu comme un véritable viatique de patrimonialisation en permettant certains gestes constitutifs du processus que sont la classification et la célébration d'un objet par son exposition³.

Cette assertion se justifie particulièrement dans le cadre d'Euphrosyne-data qui a vocation à proposer des jeux de données savamment signalés et indexés, dont la classification est intrinsèquement déterminée par la finalité de valorisation. L'objectif est que les utilisateurs de l'interface puissent accéder aux ressources souhaitées comme un chercheur accéderait à un objet patrimonial manuscrit précis après en avoir fait la requête par l'intermédiaire d'un catalogue des fonds. Aucun précédent n'existe pour les données issues de l'analyse par faisceau d'ions, ni même pour les autres techniques scientifiques du C2RMF. La création du catalogue d'Euphrosyne-data signe une importante transition des pratiques et des mentalités d'une communauté largement scientifique, qui s'approprie les apports bibliothéconomiques récents ainsi que le défi de l'interopérabilité informatique et sémantique dont elle est d'ordinaire peu familière. Les implications de cette évolution sont multiples et, pour beaucoup, encore inconnues. Les enquêtes menées auprès d'utilisateurs et de non-utilisateurs d'AGLAE spécialisés en sciences du patrimoine ont esquisé quelques vagues perspectives mais l'horizon reste flou. Certains imaginent la création d'un catalogue mutualisable à d'autres et donnant accès à de grandes bases de données thématiques et transdisciplinaires, telles que la pigmentothèque pour la Préhistoire⁴ évoquée par Émilie Chalmin, maître de conférence à l'Université de Savoie.

Notons que l'interconnexion de catalogues n'a rien de neuf pour les fonds documentaires et patrimoniaux. Le C2RMF participe d'ailleurs, par l'intermédiaire des projets PARCOURS et SoCoRe évoqués dans notre premier chapitre, à l'alimentation de Joconde, catalogue collectif des collections des musées de France, riche de plus de 600 000 notices d'objets de toutes natures, en majorité illustrées⁵. Néanmoins, il n'existe pas de catalogues de données issues des analyses scientifiques pratiquées sur ces objets. Dans SoCore, encore en phase de développement, seul est indiqué un historique de ces analyses dans une rubrique attachée à la notice.

La création d'un catalogue pour les données d'AGLAE doit donc être considérée

3. Pour approfondir cette théorie, se reporter à l'ouvrage suivant : Jean Davallon, *Le Don du patrimoine : une approche communicationnelle de la patrimonialisation*, Paris, Hermès science publications, 2006.

4. Une présentation de la pigmentothèque en construction est accessible en ligne : <https://www.researchgate.net/project/Pigmentotheque> (visité le 2 août 2021).

5. La base Joconde est accessible sur la Plateforme ouverte du patrimoine (POP) du Ministère de la Culture : <https://www.pop.culture.gouv.fr/search/list?base=%5B%22Collections%20des%20mus%C3%A9es%20de%20France%20%28Joconde%29%22%5D>

comme une initiative expérimentale complexe, qui doit être accordée aux outils déjà existants d'exposition et de valorisation numériques dédiés au patrimoine. Cette interopérabilité nécessaire ne devra toutefois pas aller à l'encontre de la spécificité et de la singularité sémiologique, conceptuelle et informationnelle du signalement des données d'AGLAE.

10.2 Le défi du signalement des contenus numériques d'AGLAE

Le défi de ce signalement réside dans son caractère inédit mais également dans l'interdisciplinarité intrinsèque aux données d'AGLAE. Ni tout à fait scientifiques, ni complètement historiques et culturelles, elles se situent dans une marge grise de l'identification et de l'information documentaire. Multiples, leurs usages sont extrêmement différents selon que les données soient destinées à des conservateurs et des historiens ou à des spécialistes en sciences des matériaux. Les premiers et les seconds n'auront pas tout à fait recours à la même grille de référentiels lors d'une requête dans Euphrosyne-data. Il s'agit donc de couvrir, par le signalement, l'ensemble des possibles en matière d'interrogation de données, ce qui impose un effort de généricité.

Construire un catalogue revient à nommer, identifier et distinguer des entités particulières au sein de la masse d'informations émises à partir d'un jeu de données. Précisons que le choix du terme d'« entité » est fait à dessein car Euphrosyne-data repose sur un système de représentation inspiré du modèle entité-relation, utilisée en informatique depuis les années 1970 pour la conception des bases de données. Il est notamment la clé de voûte du modèle FRBRer (*Functional Requirements for Bibliographic Records*)⁶, les deux dernières lettres renvoyant à « entité-relation ». Si les entités pertinentes pour les catalogues de bibliothèque sont définies par le modèle FRBR-LRM (*Library Reference Model*), il n'existe pas d'équivalent pour un catalogue dédié aux ressources de l'analyse par faisceau d'ions appliquée au patrimoine. Nous approchons ainsi de l'une des multiples facettes du catalogue, considéré ici comme une représentation, un système de signes chargé de décrire un monde. Cette approche vise à assurer une compatibilité au moins partielle des données avec celles des institutions de DIGILAB ayant structuré et signalé leurs ressources de façon interopérable. Elle doit également permettre de préparer l'introduction des données d'AGLAE dans le Web sémantique et le Web de données liées (Linked Open Data). L'interconnexion avec d'autres bases internes – comme EROS – et externes existantes associée aux apports du *Linked Open Data* favoriseront en effet l'enrichissement.

6. Ce modèle conceptuel de données bibliographiques élaboré par un groupe d'experts de l'IFLA de 1991 à 1997, décrit les informations d'une notice bibliographique d'un point de vue logique. Il constitue une véritable révolution dans le domaine de l'information documentaire car il amorce la mue d'un écosystème orienté notice vers un écosystème structuré par une information atomique. [Fonctionnalités requises des notices bibliographiques, IFLA, 1998. Accessible en ligne : http://www.bnf.fr/documents/frbr_rapport_final.pdf

chissement des résultats des requêtes. Précisons d’emblée que cette interopérabilité sera à relativiser, tant les technologies du web sémantique ont jusqu’ici été mal été appropriées par les développeurs et tant les données sont hétérogènes par leur mode d’acquisition et leur utilisation⁷.

Actuellement en construction, Euphrosyne-data devra multiplier les points d'accès à l'information en prenant en compte la diversité des utilisateurs. En cela, le choix des champs de métadonnées, de leur structure et de leur format a occupé une place déterminante dans la réflexion conduite autour de l'interface. Ces champs seront ceux renseignés à la fois manuellement et au moyen d'une extraction automatique par l'analyste responsable de l'expérience dès la préparation de la manipulation voire l'acceptation du PROPOSAL. L'élaboration du catalogue d'Euphrosyne-data induit donc un processus initié dès le début du cycle de vie des données d'AGLAE. Dans une dynamique transverse, il y aura bien une continuité, assurée technologiquement et conceptuellement, entre les deux volets d'Euphrosyne. Il subsiste toutefois un important travail d'alignement à réaliser entre les entrées envisagées du catalogue d'Euphrosyne-data et les métadonnées des jeux de données expérimentaux. La durée du stage et l'amplitude de ses missions n'ont malheureusement pas permis d'atteindre cette étape de *mapping*.

Sans être centrale, la quête de sérendipité affleure par touches dans la conception d'Euphrosyne-data. S'il est avant tout question de fournir un contenu pertinent à l'utilisateur en fonction d'une requête spécifique, il y a aussi une volonté de mettre en relief d'autres ressources susceptibles d'élargir et d'enrichir la recherche initiale. Cela impose une mise en relation automatique des contenus, ainsi que l'emploi d'algorithmes dédiés à l'analyse de la répartition et de la fréquence des mots. Ce travail de développement sera certainement réalisé, du moins partiellement, par l'Atelier Numérique et le(s) prestataire(s) recruté(s) pour cette mission. Nous constatons là aussi une implication directe relative au statut de start-up d'État du Ministère de la Culture d'Euphrosyne, avec le recours à des prestations limitées voire brèves pour un travail d'ingénierie au long cours.

7. Gautier Poupeau, « La donnée, nouvelle perspective pour les bibliothèques » dans Emmanuelle Bermès (sous la dir.), *Vers de nouveaux catalogues...*, p. 166-167.

Chapitre 11

Définir le référentiel de l'analyse par faisceau d'ions : enjeux d'une classification singulière

En tant qu'opération intellectuelle d'identification et de signalement, le catalogage des données d'AGLAE doit s'appuyer sur un ensemble de références conçues et validées transversalement, à la fois par la communauté-métier spécialiste des contenus et par les spécialistes des systèmes d'information et de documentation. On parle alors de référentiel, terme générique recouvrant une réalité protéiforme, de la simple liste de mots au thésaurus. Il s'agit dans un premier temps de contrôler les formes des termes utilisés lors du catalogage afin d'éviter les redondances d'information et d'offrir des clés permettant la classification ainsi que la recherche aisée des contenus¹. Le référentiel peut notamment prendre la forme d'un thesaurus, vocabulaire contrôlé hiérarchique le plus fréquent. Avant d'opter pour un référentiel spécifique, il est d'abord question d'une entente sémiologique et sémantique à créer à la fois au sein de la communauté élargie d'AGLAE – composée à la fois des spécialistes en sciences des matériaux et des acteurs des institutions patrimoniales - , et parmi la communauté plus étendue de l'analyse par faisceau d'ions.

11.1 Créer une entente sémiologique et sémantique

Dans le cadre d'Euphrosyne, la définition d'une classification à base de vocabulaires partagés est le préalable indispensable à toute entreprise de catalogage et de mise en accès des jeux de données expérimentaux. Ce travail est rendu particulièrement complexe par la multiplicité des objectifs poursuivis à travers Euphrosyne-data (interopérabilité, réuti-

1. Maxime Challon, « Les référentiels en institutions patrimoniales : évolution des pratiques et repositionnement. L'exemple des référentiels de l'Institut National de l'Audiovisuel », mémoire de master « Technologies numériques appliquées à l'histoire », dir. Gautier Poupeau, École nationale des Chartes, 2020, p. 17.

lisation scientifique, ouverture universelle, etc.), mais également par l'hétérogénéité des communautés concernées. Parce qu'elles sont produites par un accélérateur de particules, l'amplitude des données d'AGLAE déborde du cadre des sciences appliquées au patrimoine pour toucher celui de la physique nucléaire, de la physique atomique, de l'électromagnétisme, etc. Autant de disciplines qui possèdent leur propre lexique, voire « jargon », terme familier mais approprié pour les notions écloses au cœur des pratiques mais non issues directement de la terminologie officielle et académique. Selon une dynamique centrifuge, il a semblé naturel et opportun de s'appuyer d'abord sur le premier cercle formé par la communauté d'AGLAE sans toutefois l'émanciper complètement de la sphère de l'analyse par faisceau d'ions à laquelle elle appartient. Précisons d'emblée qu'en dépit des attentes explicitement et récemment formulées d'une partie des membres de cette sphère en la matière², il semble intellectuellement impossible de couvrir en un seul référentiel l'ensemble des activités et des résultats extrêmement divers produits par l'analyse par faisceau d'ions. La classification réalisée pour les données d'AGLAE est donc vouée à être un sous-ensemble d'une classification mosaïque plus vaste qui reste aujourd'hui à construire.

S'entendre sur le sens et le signe à conférer aux données d'AGLAE : ce credo engage deux voies de prospection qui ne cessent de s'entrecroiser sur le fond et sur la forme. Ce mouvement d'abstraction et de formalisation ne saurait être réduit à un simple alignement de termes utilisés au sein d'une communauté. Au-delà des mots choisis, c'est bien l'épaisseur cognitive même des « choses » qui est explorée, défrichée jusqu'à atteindre un noyau saisissable par l'ensemble des subjectivités. Dans le cas de données aussi complexes que celles d'AGLAE, ce centre de gravité conceptuel ne se découvre qu'à la lueur de la transdisciplinarité et dans l'effeuillage de la polysémie. Pour illustrer notre propos, nous pouvons prendre pour exemple la définition de l'entité « DOSE », particulièrement significante. Nous avons retenu que celle-ci correspondait à « la valeur mesurée proportionnelle au nombre de particules envoyées sur la cible »³. Avant de parvenir à cette définition, incomplète par ailleurs, nous avons dû revenir au processus expérimental concret pour apprêhender avec précision les différents éléments qui aboutissent au renseignement de la DOSE. En effet, les propriétés de cette dernière sont intrinsèquement dynamiques et ne peuvent être réduites à la simple valeur d'un résultat, contrairement à la proposition de définition faite initialement. La DOSE correspond à la fois à une réaction physique des particules, à un paramètre défini par l'utilisateur, à un phénomène mécanique et à une valeur mesurée. Le choix d'une définition trop restrictive imposerait ainsi déjà une prépondérance accordée à un usage-métier sur les autres. À l'inverse, une signification

2. Échanges menés dans le cadre de la 8^e Rencontre « Ion Beam Applications Francophone », organisée du 5 au 7 juillet 2021 par la Société Française du Vide.

3. La valeur mesurée est celle de l'aire du pic de silicium sur le spectre PIXE de la fenêtre d'extraction faite en S3N4. Il n'y a qu'à AGLAE que l'on mesure la dose comme cela. Cette mesure est proportionnelle au nombre de particules qui atteignent la cible. Ce nombre de particules est lui-même proportionnel au courant (exprimé en ampères) et ce dernier est une valeur systématiquement indiquée dans les conditions expérimentales des manipulations IBA mondiales.

excessivement large et consensuelle risquerait de dissoudre la pertinence du sens. Si la définition retenue est fiable du point de vue de l'expérience, elle ne rend pas compte de l'usage de la DOSE qui lui donne pourtant son sens fondamental. Par une réaction connue et maîtrisée des analystes, les particules projetées sur la cible entraînent l'émission de rayons qui se répercutent au niveau d'un détecteur. Le choix de leur nombre détermine à la fois la durée de l'analyse et les résultats en les rendant plus ou moins signifiants selon la DOSE choisie. C'est donc également un paramètre essentiel à la circonscription d'un contexte expérimental, lourd en enjeux scientifiques. La plupart des termes inhérents à AGLAE partagent cette structure sémantique feuillettée, étoffée dans la complexité cognitive, scientifique et technologique de l'expérience, et font référence tout autant à un processus qu'à une valeur mathématique. Nous pouvons ainsi également citer la « Taille du pinceau du faisceau ». D'abord définie par ignorance comme la « dimension du faisceau de sortie », il s'est ensuite avéré que cette entité fait référence à une configuration complexe ayant pour but d'optimiser le balayage par le faisceau lors de la réalisation d'une cartographie. Là encore, la définition retenue écarte sa justification et mériterait d'être révisée : « Distance sur laquelle est défléchi le faisceau verticalement pour faire une cartographie, exprimée en micromètres ».

Notre prudence conceptuelle s'explique sans doute par un excès d'essentialisation. Notre travail de classification, réalisé avec l'équipe et certains utilisateurs d'AGLAE doit désormais être présenté aux partenaires d'IPERION-HS dans le cadre des groupes de travail dédiés à l'élaboration de référentiels et de syntaxes partagés. Notre classification ne peut, à ce jour, être considérée comme un ensemble de référence, faute d'avoir encore été débattue et validée par l'ensemble de la communauté d'AGLAE et, plus largement, par les spécialistes des différentes disciplines concernées.

Notons que la spécificité de la terminologie d'AGLAE est telle qu'elle offre au moins l'avantage de ne guère offrir la possibilité de synonymes ou de fluctuation lexicale. Les variations concernent ainsi plutôt les objets de l'expérience. Chaque élément de cette terminologie nécessite d'être décomposé conceptuellement en vue d'une représentation logique adaptée à la granularité du signalement des (méta) données. L'émulation collective et les circulations conceptuelles diffusés à travers les canaux de DIGILAB, d'ESPADON et de SSHOC (*Social Sciences and Humanities Open Cloud*)⁴ ont permis d'ajuster et d'alimenter notre essai de classification en nous inspirant des plus récents travaux en matière de référentiel et d'organisation sémantique des données des sciences du patrimoine.

4. *Social Sciences & Humanities Open Cloud* (SSHOC) est un projet financé par le programme-cadre de l'Union Européenne Horizon 2020. Il réunit une vingtaine d'organisations partenaires – notamment des organismes patrimoniaux - afin de développer un *cloud* ouvert et accessible dédié aux ressources en sciences humaines et sociales. Les partenaires possèdent tous une expertise sur l'ensemble du cycle de vie des données, de leur production à leur conservation en passant par leur réutilisation et leur pérennisation. Site officiel : <https://sshopencloud.eu/> (visité le 07 juillet 2021).

11.2 Organiser : la réalisation du modèle sémantique

L'ontologie CIDOC-Crm Cr élaborée dans le cadre de PARCOURS, présentée brièvement dans notre premier chapitre, esquisse plusieurs pistes d'introduction des données produites par le laboratoire du C2RMF dans l'environnement sémantique des sciences du patrimoine. Parce qu'elle conjugue les apports de deux ontologies existantes que sont CIDOC-Crm et CIDOC-Crm Sci tout en les enrichissant de nouveaux prolongements, CIDOC-Crm Cr a offert un socle de base à notre réflexion. Rappelons que le CIDOC-Crm (Comité International pour la Documentation – *Conceptual Reference Model*)⁵ est un modèle sémantique de référence dédié au patrimoine culturel, orienté objet⁶ et chargé de pouvoir décrire les interactions de chaque objet avec d'autres entités. Il doit permettre aux musées de structurer leurs données et d'échanger des informations. Le CIDOC-Crm Sci (*The Scientific Observation model*), qui en dérive, est une ontologie de domaine conçue pour l'intégration des métadonnées relatives aux observations scientifiques, aux mesures et aux données traitées dans les sciences descriptives et empiriques telles que la biodiversité, la géologie, la géographie, l'archéologie, la conservation du patrimoine culturel, etc. Son objectif principal est de faciliter la gestion, l'intégration, la médiation, l'échange et l'accès aux données de la recherche par la description des relations sémantiques⁷. Synthèse de ces deux ontologies, spécialement adaptée aux données de la conservation-restauration, CIDOC-Crm Cr est une ontologie d'application prévue pour diffuser les ressources issues des analyses scientifiques pratiquées sur le patrimoine à des fins de conservation préventive, de restauration et de découverte.

Elle approfondit les ressources proposées par le CIDOC-Crm Sci en intégrant les thesaurus de la conservation-restauration. Elle ébauche un début de représentation générale des procédures expérimentales inhérentes aux techniques et aux différents instruments en affinant la profondeur sémantique des entités relatives aux matériaux étudiés. Réalisée pour des besoins d'échanges et de diffusion de données spécifiques, CIDOC-Crm Cr s'appuie sur l'organisation et les bases de données suivantes : celles du C2RMF, qui met à disposition ses corpus de données constituées de fiches descriptives, d'imagerie scientifique et documentaire et de mesures ; du LRMH (Laboratoire de Recherche des Monuments Historiques) et du CRC/CRCC (Centre de recherche sur la conservation),

5. CIDOC-CRM, site officiel : <http://www.cidoc-crm.org/> (visité le 16 mai 2021). La première version du CIDOC-Crm est achevée en 1999. En 2006, ce modèle a fait l'objet d'une publication ISO, en devenant une norme internationale : ISO 21127 : 2006.

6. Un objet représente : « un concept, une idée ou toute entité du monde physique. Il possède une structure interne, un comportement et il sait interagir avec ses pairs. Dans une approche "orientée objet" il s'agit de représenter ces objets et leurs relations. L'interaction entre les objets via leurs relations permet de concevoir des fonctionnalités et de mieux résoudre les problèmes. » [Eleonora Moiraghi, Construire l'interopérabilité, entre conservation et changement : projet de valorisation des données de la collection photographique de l'Institut Warburg, mémoire de master « Technologies numériques appliquées à l'histoire », dir. Gautier Poupeau, École nationale des chartes, 2017, p. 41].

7. Site officiel : <http://www.cidoc-crm.org/crmsci/home-1> (visité le 16 mai 2021).

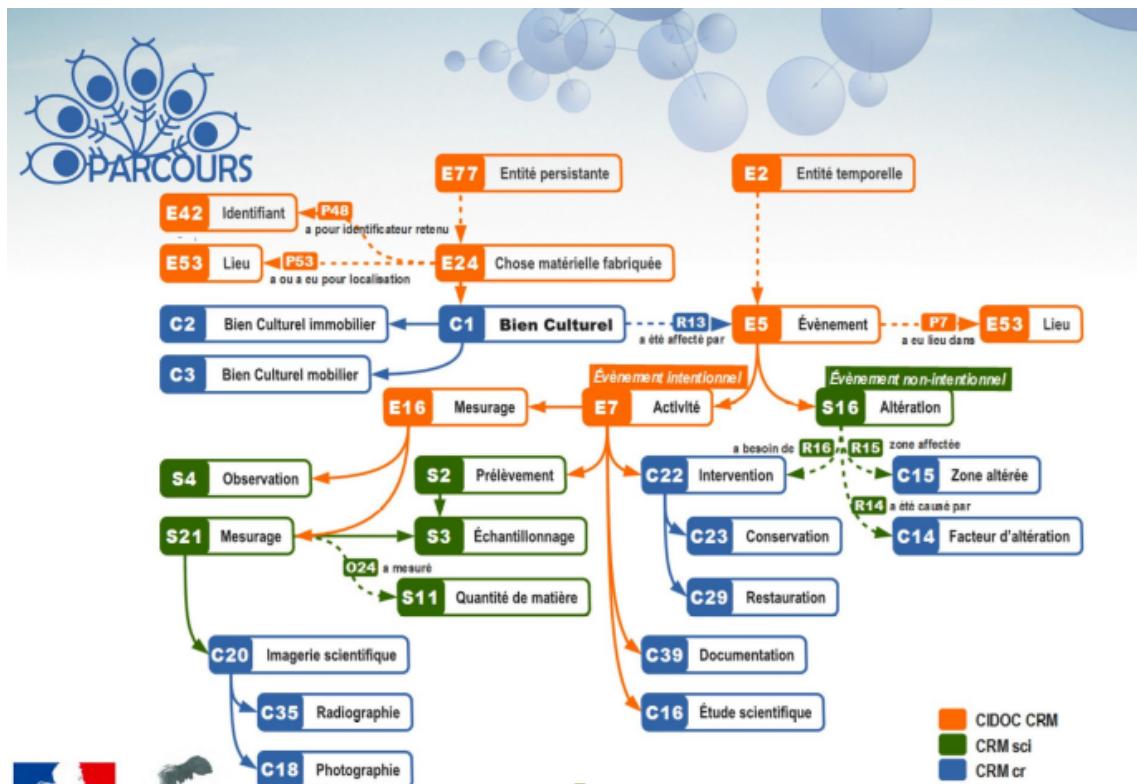


FIGURE 11.1 – Capture d'écran d'un extrait de la présentation du projet PAR-COURS, avril 2021. @Luc Bouiller-C2RMF

expert dans la conservation-restauration des matériaux constitutifs des biens culturels. Malheureusement, CIDOC-Crm Cr n'a pas encore été publiée et il n'a pas été possible d'accéder à sa version complète et réactualisée. Aussi, nous devons d'emblée émettre une réserve nécessaire aux considérations portées sur cette ontologie et ses éventuelles ellipses au regard des (méta) données d'AGLAE. À partir des éléments du CIDOC-Crm Cr dont nous avons pu prendre connaissance, il est apparu que cette ontologie s'arrête au seuil de la procédure expérimentale, sans en représenter les flux propres à chaque technique, les objets numériques qui en résultent et leurs interactions. Cela peut s'expliquer par le fait qu'elle a été construite en fonction de ressources et non de processus.

Plusieurs institutions partenaires de DIGILAB sont allés plus loin en modélisant à la fois les dynamiques expérimentales, leurs objets et les ressources produites par l'analyse scientifique du patrimoine, jusqu'à intégrer les apports de technologies récentes telles que IIIF (*International Image Interoperability Framework*)⁸. Elles ont élaboré des modèles sémantiques extrêmement détaillés de façon à lier leurs données à celles moissonnées à l'aide des moteurs de recherche, et afin d'optimiser leur réutilisation. L'exemple le plus inspirant est sans nul doute celui de la National Gallery (NG) avec le travail réalisé par

8. IIIF est un ensemble de standards qui définissent un cadre d'interopérabilité pour la diffusion des images numériques sur le Web. Il doit créer un cadre technique commun grâce auquel les bibliothèques numériques peuvent diffuser leurs images de manière standardisée sur le Web afin de les rendre consultables, manipulables et annotables par n'importe quelle application ou logiciel compatible. [« Qu'est-ce que IIIF ? », *Biblissima*, <https://iiif.biblissima.fr/> (visité le 23 juillet 2021)]

l'équipe de Joseph Padfield, spécialiste de la conservation patrimoniale. Après avoir œuvré à la FAIRisation et au signalement de plusieurs jeux de données de la NG, elle a procédé à un *mapping* des référentiels avec ceux du CIDOC-Crm ainsi qu'avec les standards pour les galeries et musées de l'ICOM (*International Council of Museums*)⁹. Un identifiant à la syntaxe interopérable a été attribué à chaque ressource afin de les connecter à des contenus externes tels que ceux de Wikidata et Getty Art & Architecture Thesaurus. Une base de données baptisée *The Grounds Database*, produite dans le cadre d'un projet IPERION-CH a par ailleurs été réalisée en s'inspirant du CIDOC-Crm ainsi que de ses extensions CRMdig et CRMscl. Elle inclut des informations spécifiques à l'imagerie, à l'échantillonnage et à l'analyse d'œuvres picturales conservées dans plusieurs institutions muséales européennes. Les triplets RDF associés à ces informations résultent d'un *mapping* semi-automatique réalisé à l'aide d'un script Python utilisant RDFlib¹⁰. Ce dernier parse le contenu des anciennes bases de données et le compile en un graphe compatible avec les ontologies retenues. Les triplets sont stockés et accessibles dans un triple store Blazegraph, et peuvent faire l'objet d'une requête SPARQL. Un exemple produit à partir des ressources de recherche sur l'œuvre de Raphaël nous a permis de voir toute l'étendue conceptuelle couverte par la modélisation de la NG. Dans cette dernière, un créateur/acteur est une personne, un groupe ou une organisation impliquée ou responsable des activités liées à un événement donné. Cela peut inclure des artistes, des auteurs et même des scientifiques de la conservation qui peuvent prélever un échantillon d'un objet ou effectuer un travail analytique.

Une gamme extrêmement riche d'événements possibles a été représentée, notamment les modifications ou traitements induits par des opérations de conservation¹¹, des changements de propriété, des altérations non-intentionnelles, etc. Les objets numériques tels que les fichiers et les outils logiciels sont également modélisés¹², selon une granularité d'information extrêmement poussée¹³.

Parce qu'il entremèle les enjeux de pérennisation, d'interopérabilité, de recherche scientifique et de conservation patrimoniale, l'exemple de la NG nous a inspiré le choix

9. « Par le biais de ses normes et lignes directrices, l'ICOM diffuse les meilleures pratiques aux professionnels des musées en termes d'acquisition d'objets, de documentation des collections, de description, de terminologie, de conservation, etc. Ces normes sont développées par les comités de l'ICOM et sont approuvées par le Comité exécutif de l'ICOM. L'ICOM soutient et contribue également au développement d'autres normes internationales, à l'image de l'*Object ID Recommandation* de 2015 de l'Unesco concernant la protection et la promotion des musées et des collections, de leur diversité et de leur rôle dans la société. » [Site officiel de l'ICOM : <https://icom.museum/fr/ressources/normes-et-lignes-directrices/normes/> (visité le 03 juillet 2021)]

10. RDFlib est un ensemble de paquets qui permettent aux utilisateurs d'interroger, de construire, de stocker et de manipuler des jeux de données structurés en RDF. Il permet notamment l'association simultanée de nombreux jeux de données.

11. Voir Figure 11.2

12. Voir Figure 11.4

13. Présentation des exemples inspirés de CIDOC-Crm, CIDOC-Crm Sci et de CIDOC-Dig pour la modélisation conceptuelle des ressources inhérentes aux recherches menées sur l'œuvre de Raphaël [Site officiel : <https://jpadfield.github.io/sshoc-ng/Raphael%20Examples.html> (visité le 15 juin 2021)]

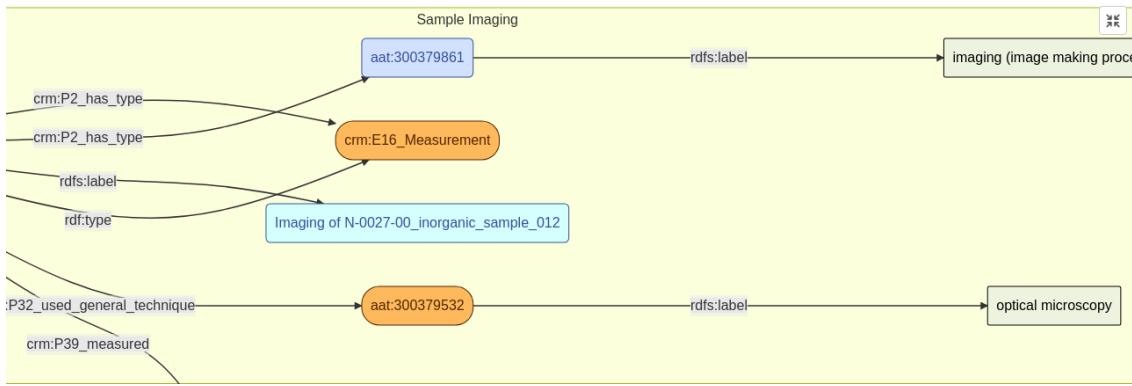


FIGURE 11.2 – Capture d’écran de la modélisation sémantique pour *Raphael Research Resource*, juillet 2021. @National Gallery

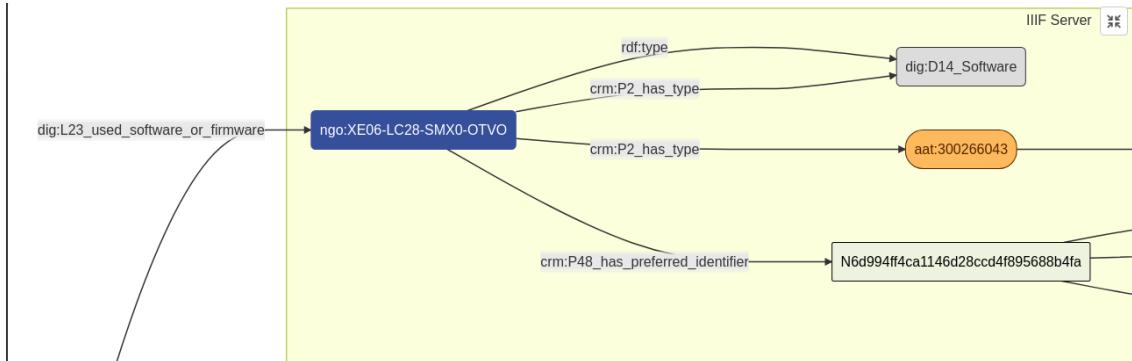


FIGURE 11.3 – Capture d’écran de la modélisation sémantique pour *Raphael Research Resource*, juillet 2021. @National Gallery

d'un certain nombre d'objets, ressources et prédicats pour notre propre modèle, lequel n'a rien d'un simple exercice de conceptualisation. Nous n'avons en effet jamais omis la visée d'implémentation de notre représentation ainsi que les besoins auxquels il se doit de répondre. Il s'agit pour AGLAE de pouvoir partager ses jeux de données expérimentaux tout en les reliant prioritairement :

1. À des ressources internes relatives aux objets patrimoniaux analysés et aux résultats des autres techniques d'analyse du C2RMF, actuellement répartis entre la base EROS et les différents silos du laboratoire.
2. À des ressources mutualisées dans le cadre d'ESPADON, de DIGILAB et plus largement d'E-RIHS.
3. À des bases de données dédiées au patrimoine culturel telles que Joconde.
4. À des bases de données-métiers consacrées à l'analyse par faisceau d'ions et aux grands instruments de physique des particules.
5. Aux ressources et outils des technologies IIIF.

Cette feuille de route nous a incité à raffiner notre modélisation selon une triple approche, relativement similaire à celle de la NG : la représentation contextualisée de l'objet patrimonial à l'aide des ressources du CIDOC-Crm, largement utilisées et déclinées par les

différentes structures de conservation et de médiation patrimoniales françaises et étrangères ; celle des procédures expérimentales appliquées au patrimoine à partir de certains éléments du CIDOC-Crm Sci mais également de l'ontologie appliquée à la biochimie¹⁴ ; celle des objets numériques inhérents aux résultats de ces procédures.

Entre la proposition du CIDOC-Crm Cr et la piste hybride ouverte par la NG, nous avons ainsi esquissé une troisième voie, adaptée à la singularité des données d'AGLAE ainsi qu'aux usages actuels et futurs prévus pour elles. Notre principal apport réside dans le niveau de granularité de l'expérience, décomposée jusqu'à atteindre son objet à l'échelle de l'infiniment petit : l'élément chimique. Nous nous sommes efforcés de représenter ce processus dans le respect de sa double dynamique, à savoir le permanent et l'occurent¹⁵, en lui conférant la souplesse conceptuelle nécessaire pour s'adapter à l'imprévisibilité et à la discontinuité temporelle. L'ontologie appliquée à la biochimie a pour cela été particulièrement utile à notre réflexion, puisqu'elle est organisée selon deux axes principaux : « continuant », qui regroupe l'entité matérielle analysée et ses dépendances ; et « process » qui comprend les événements liés à l'expérience tels que les accidents de laboratoire, la planification du processus expérimental et ses éventuels réajustements. Nous nous en sommes inspirés pour la modélisation générique d'un contexte expérimental à AGLAE avec la création d'un triplet, directement relié par sa ressource à celui consacré à la définition de l'événement général (crm : E5 Event / is defined / by BCO : experimental_planning_process) :

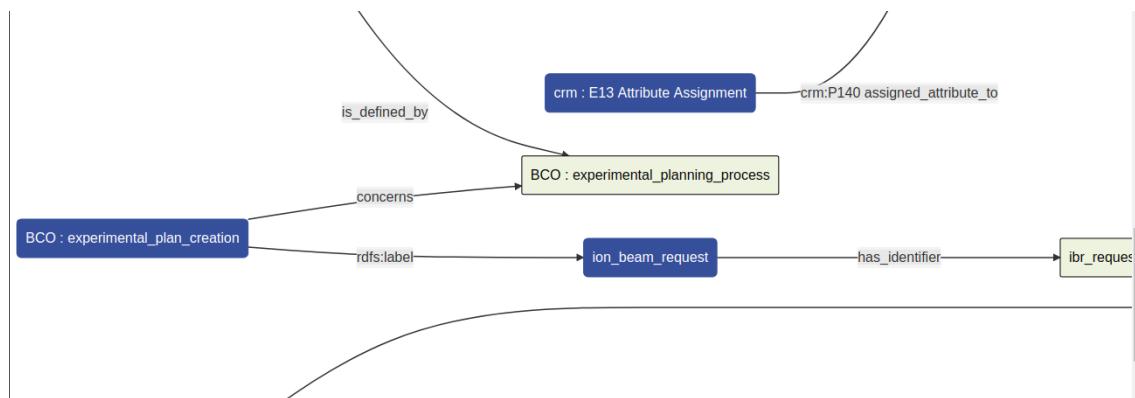


FIGURE 11.4 – Capture d'écran de la modélisation sémantique réalisée pour une expérience à New AGLAE. @Pauline Breton-Chauvet-New AGLAE

La difficulté majeure de notre conceptualisation a résidé dans notre volonté de restitution étroite de l'interdépendance des ressources numériques (crm : dig) avec les flux de

14. BioPortal, portail consacré aux ontologies appliquées à certains sous-domaine de la biologie : <https://bioportal.bioontology.org/ontologies/CHMO/?p=classes&conceptid=root> (visité le 14 avril 2021)

15. Les catégories de processus et d'événement incarnent la dynamicité du monde en correspondant à des entités qui « occurrent » ou « surviennent ». Elles s'opposent aux objets et à la matière dont ceux-ci sont constitués, ces entités incarnant la stabilité du monde en « endurant » ou « continuant ». [Gilles Kassel, « Une alternative à la distinction "continuant" vs "occurrent" », 29^e Journées Francophones d'Ingénierie des Connaissances, AFIA, 2018, p. 147-162. fhal-01839618f]

l’expérience elle-même et la matérialité complexe et protéiforme des objets patrimoniaux étudiés. En effet, les données d’AGLAE trouvent leur substance cognitive dans cette jonction tant technologique que scientifique et culturelle. La structure en graphe en a facilité la visualisation et la répartition interconnectée des différents objets qui en découlent. Toutefois, notre modèle aurait dû traiter avec davantage de finesse les ressources relatives à certains paramètres de l’analyse telles que la DOSE ou le facteur de calibrage du spectre. Le chevauchement sémantique de ces éléments, qui relèvent à la fois de l’expérience et de l’interprétation de ses résultats, ainsi qu’une maîtrise relativement superficielle de leur contenu scientifique – particulièrement complexe pour un individu non-initié – ont justifié cette prudence temporaire, dans l’attente de développements futurs en concertation avec des spécialistes aguerris.

Nous devons enfin reconnaître le caractère incertain lié à la dimension prospective de cette base d’essai, particulièrement dans son ambition de construction d’une interopérabilité externe. Si des institutions culturelles et patrimoniales majeures – tant françaises qu’européennes - usent des technologies du web sémantique, nous incitant ainsi à en faire de même, il est indispensable de relativiser ce qui peut ressembler à un « effet de mode » en faisant preuve de distance critique. Les opportunités du web sémantique sont actuellement utiles à la diffusion et à la valorisation des données d’AGLAE, en raison des circuits déjà en place pour l’échange et l’accès aux ressources à vocation patrimoniale. Toutefois, rappelons les limites de ces technologies, qui ont pour contrepartie : un risque de perte d’intégrité dû à la gestion des transactions des données stockées en RDF ; une réduction de la scalabilité et de la performance dans les réponses à certaines requêtes¹⁶. De plus, la structure du modèle RDF induit des problèmes d’opacité quant à la provenance des informations et à la contextualisation du triplet, ce qui peut être préjudiciable pour les ressources d’AGLAE, dont l’intelligibilité et la valeur sont dépendantes d’un contexte d’identification et de signalement extrêmement spécifique. Le recours aux technologies du web sémantique doit donc être considéré comme l’opportunité d’une publicité large, bien qu’imparfaite, des données d’AGLAE à travers le web, qui ne saurait faire l’économie d’un renvoi à la base de données que sera Euphrosyne.

16. Gautier Poupeau, « Au-delà des limites, que reste-t-il concrètement du Web sémantique ? », Les Petites Cases, 6 oct. 2018, <http://www.lespetitescases.net/au-delà-des-limites-que-reste-t-il-concrètement-du-web-sémantique> (visité le 27 avril 2021).

Chapitre 12

L'*Open Science* : du carcan prescriptif au vecteur de moyens pour AGLAE

Open Science ou Science Ouverte : deux mots qui cristallisent à eux-seuls le contexte actuel de diffusion des données de la recherche scientifique, incubateur de tensions, mais également terreau fertile d'opportunités et d'innovations pour le partage et la dissémination des savoirs, ainsi que ceux de leurs outils d'acquisition. Dynamiques à la genèse éminemment politique et économiquement stratégique, l'ouverture, la diffusion et le libre accès aux données issues des différentes disciplines scientifiques ont d'abord pu apparaître comme une prescription imposée verticalement à des communautés scientifiques dont les pratiques s'inscrivaient – et s'inscrivent toujours pour certaines - à contre-sens du credo de la dissémination. C'est particulièrement le cas de celles, comme AGLAE, qui se trouvent à l'intersection des données de la recherche et des données des sciences du patrimoine, traversées par des courants antagonistes d'exigence de libre-accès et d'incitation au repli conservatoire.

12.1 Dans l'entrecroisement des données de la recherche et des données du patrimoine : une lente et difficile ouverture

Officiellement, l'*Open Science* est présentée comme :

[...] la diffusion sans entrave des résultats, des méthodes et des produits de la recherche scientifique. Elle s'appuie sur l'opportunité que représente la mutation numérique pour développer l'accès ouvert aux publications et – autant que

possible – aux données, aux codes sources et aux méthodes de la recherche.¹

Il s'agit donc à la fois d'un cadre politique, de ressources et d'outils ancrés opportunément dans le numérique et ses technologies, seuls à pouvoir offrir les supports et les circuits nécessaires aux impératifs d'ouverture et de diffusion. Il est indispensable de rappeler quelle fut la genèse de l'appropriation française de l'*Open Science* afin de comprendre le caractère lent et laborieux de son introduction au sein de la communauté d'AGLAE et, plus largement, de la sphère scientifique de l'analyse par faisceau d'ions.

En juillet 2012, la Commission européenne émet une recommandation relative à la publication en accès ouvert – gratuit pour le lecteur – des résultats de la recherche scientifique financée sur fonds publics, dans le but de renforcer la visibilité de la recherche européenne à l'horizon 2020². Le 25 avril 2018, cette recommandation est enrichie et substituée par celle relative à l'accès aux informations scientifiques et à leur conservation³. Avant tout socle déontologique dédié à la mise en œuvre d'une « économie compétitive des données et de la connaissance en Europe »⁴, elle promeut la pérennisation, l'ouverture, le libre-accès, la diffusion et la réutilisation des données scientifiques, en particulier celles obtenues au moyen de fonds publics. Parmi les différents axes développés, relevons celui qui sous-tend l'agrégation et la synergie au bénéfice d'infrastructures fédératrices :

Les États membres devraient définir et mettre en œuvre des politiques claires (telles qu'exposées dans les plans d'action nationaux) visant, d'une part, à développer davantage les infrastructures sous-tendant le système permettant d'accéder aux informations scientifiques, de les conserver, de les partager et de les réutiliser et, d'autre part, à promouvoir leur fédération au sein du nuage européen pour la science ouverte [...] Ils devraient veiller à créer des synergies entre les infrastructures nationales, avec le nuage européen pour la science ouverte et avec d'autres initiatives mondiales.⁵

Cette incitation repose particulièrement sur : la définition de normes communes pour les données et les services accessibles ; l'interopérabilité des infrastructures récemment développées ou adaptées aux exigences de la science ouverte pour éviter l'apparition de cloisonnements, tant technologiques que disciplinaires et géographiques.

Quelques mois après la publication de la Recommandation de la Commission européenne, Frédérique Vidal, alors Ministre de l'Enseignement supérieur, de la Recherche et

1. Site Ouvrir la science : https://www.ouvrirlascience.fr/category/science_ouverte/ (visité le 19 juin 2021).

2. Recommandation de la Commission européenne du 17 juillet 2012 relative à l'accès aux informations scientifiques et à leur conservation. Accessible en ligne : <https://eur-lex.europa.eu/legal-content/FR/TXT/PDF/?uri=CELEX:32012H0417&from=FR>. Voir Annexe : D.2

3. Recommandation de la Commission européenne 2018/790 du 25 avril 2018 relative à l'accès aux informations scientifiques et à leur conservation. Accessible en ligne : <https://eur-lex.europa.eu/legal-content/FR/TXT/PDF/?uri=CELEX:32018H0790&from=GA>. Voir Annexe : D.3

4. *Ibid.*, p. 1.

5. *Ibid.*, p. 5.

de l’Innovation, annonce le lancement du premier plan national pour la science ouverte, fondé par trois arguments généraux⁶ :

1. L’engagement républicain d’ouverture et de rapprochement de la science vers la société.
2. L’inclusion stratégiquement indispensable de la France dans la science ouverte promue à l’échelle mondiale.
3. Le caractère jugé délétère des pratiques de fermeture avec l’idée qu’« un savoir enfermé est un savoir stérile »⁷.

Elle dénonce par ailleurs le caractère onéreux et l’inaccessibilité de nombreuses publications et données de la recherche, qu’elle qualifie de « profonds obstacles aux progrès d’une science cumulative »⁸. Toutefois, avant de défendre le principe d’une ouverture universelle émancipée de toute période d’embargo, Frédérique Vidal insiste sur la précarité et la vulnérabilité des jeux de données à moyen et long terme, en raison d’un déficit de saisie des enjeux et des modalités de leur pérennisation :

Lorsqu’on revient sur une recherche menée il y a dix ans, et qu’il n’est plus possible de recomposer les données qui ont permis des conclusions importantes, il y a lieu de s’inquiéter. À un moment où l’intégrité scientifique est si importante pour défendre le rôle de la science dans la société, il est indispensable de maîtriser les données dans la durée. Ainsi, la première démarche est de structurer et conserver les données, en préalable à leur ouverture. Cela se fait très bien dans environ un cinquième de la communauté scientifique, mais le reste est largement laissé à l’appréciation individuelle et aux hasards de la vie des clés USB, des vols d’ordinateurs portables dans les coffres de voitures ou des défaillances des disques durs individuels...⁹

À travers ces constats – non sans écho avec les usages en cours à AGLAE - et ces prescriptions multiscalaires véhiculées de façon descendante jusqu’aux institutions et organismes de recherche, nous relevons une tension intrinsèque à la genèse-même de l’*Open Science* entre les pratiques des chercheurs et les impératifs propres à la compétitivité économique et aux lignes de force politiques. Ainsi, l’*Open Science* est d’abord apparue en « contre », dans la lueur des contrastes avec l’existant qu’elle entendait déconstruire et remodeler en profondeur. Cela explique qu’une large proportion des différentes communautés scientifiques, y compris celles d’AGLAE, ait pu avoir, dans un premier temps, le sentiment d’une révolution à marche forcée au nom d’un vague objectif d’universalité

6. Ministère de l’Enseignement supérieur, de la Recherche et de l’Innovation, *Discours Science Ouverte à LIBER de Mme la Ministre Frédérique Vidal*, prononcé le 4 juillet 2018.

7. « Qui a peur de l’open access ? », tribune collective de responsables universitaires, d’éditeurs et d’enseignants-chercheurs, *Le Monde*, 15 mars 2013.

8. Ministère de l’Enseignement supérieur, de la Recherche et de l’Innovation, *Discours Science Ouverte à LIBER... , p. 2.*

9. *Ibid.*, p. 3.

dissimulant mal la prépondérance des enjeux économiques. La défiance s'est donc exprimée autant à l'égard de l'intention que de l'armature technologique de la science ouverte, perçue comme potentiellement destructrice de l'environnement éditorial spécifique dont dépend la recherche scientifique, tant en sciences appliquées qu'en sciences humaines et sociales¹⁰. Outre les risques de pillage et de détournement évoqués en filigrane des chapitres précédents, les chercheurs et spécialistes en sciences du patrimoine se sont interrogés sur l'utilité – et ce faisant, sur la justification sociale – d'une accessibilité universelle aux données d'instruments dont la compréhension est l'apanage des initiés. Au-delà des arguments avancés par les promoteurs de l'*Open Science*, ces chercheurs ont donc dû élaborer leurs propres réponses aux questions fondamentales de l'ouverture et de la diffusion sans restriction des données qu'ils ont contribué à produire : « Pour qui ? » et « Pourquoi ? ».

Pour la communauté d'AGLAE, l'interdépendance de ces deux interrogations aboutit à une réponse englobante : le libre-accès et la dissémination des contenus peuvent favoriser un élargissement vertueux aux communautés scientifiques adjacentes en raison de l'utilisation d'instruments, d'objets d'analyse et/ou de méthodes relativement proches voire similaires. L'agrégation et éventuellement la mutualisation synergique des ressources, des outils et des résultats à l'échelle européenne optimiseraient les pratiques expérimentales en augmentant leur efficacité, leur ingéniosité et leur fiabilité en réduisant le risque d'erreur. Appliquée au patrimoine, rappelons que l'analyse par faisceau d'ions est relativement récente à l'échelle de l'histoire des sciences et qu'elle doit encore conquérir de nouvelles parcelles expérimentales. Extrêmement rodée pour les matériaux inorganiques tels que les céramiques, les pierres, les métaux et les verres, elle s'applique encore mal à ceux mêlant l'organique et l'inorganique. C'est pourquoi les œuvres picturales, en particulier celles dont les pigments sont mélangés à du blanc de plomb – appelé également blanc d'argent ou céruse –, ne sont pas encore analysées par le faisceau d'AGLAE. Les expériences menées sur des échantillons ont en effet révélé des dégradations dont la réversibilité est encore méconnue, dues au bouleversement de la matière provoqué par les particules projetées¹¹. La future ligne de faisceau installée dans le cadre de l'EquipEx ESPADON, dédiée à l'étude d'objets de grande dimension, aura notamment pour fonction de transcender cet obstacle majeur et persistant à l'étude exhaustive des différents matériaux du patrimoine. Cette entreprise s'appuiera sur une recherche « en réseau » déjà existante à travers FIXLAB et DIGILAB. Ainsi, la mutualisation et la comparaison des données obtenues avec celles d'autres instruments et d'autres techniques grâce aux canaux et aux outils de la science ouverte participent de l'adhésion progressive à cette dernière. Nous devons reconnaître que les bénéfices attendus sont ici bien plus scientifiques et techniques que pédagogiques et culturels, l'idée d'une médiation des ressources d'AGLAE auprès de

10. « Qui a peur de l'open access ? »...

11. « Le blanc de plomb dans les œuvres peintes », C2RMF, <https://c2rmf.fr/nos-activites/peinture/le-blanc-de-plomb-dans-les-oeuvres-peintes> (visité le 25 août 2021).

la société au sens large à travers l'*Open Science* n'étant pas encore envisagée. C'est sans doute là un écueil fondamental des sciences appliquées au patrimoine : vouées à la compréhension et à la préservation de biens à la valeur universelle et intemporelle, elles ne parviennent pas véritablement à s'extraire de la gangue de l'expertise pour diffuser leurs résultats, pourtant potentiellement patrimoniaux eux-mêmes, à tous les publics.

12.2 De l'agrégation à la dissémination : une inégale prise en compte du cycle de vie des données

Si l'articulation interopérable des ressources et des outils des sciences du patrimoine, dont ceux d'AGLAE, semble en cours de réalisation grâce aux efforts collectifs menés dans de grandes infrastructures telles qu'E-RIHS mais également grâce aux déclinaisons « locales » telles qu'Euphrosyne, la dissémination au sens anglo-saxon du terme en est encore à ses fondations, par ailleurs inégales selon les différentes étapes du cycle de vie des données. Nous entendons par dissémination une activité de diffusion large à travers des réseaux aux multiples ramifications, ce qui suppose une forme d'éparpillement fécond à travers une toile numérique. Elle est en quelque sorte une diffusion poussée à son paroxysme, propice à l'universalité et à la sérendipité. Les données publiées sont à ce jour largement favorisées, depuis la phase d'agrégation jusqu'à celle de la circulation ouverte. Recherche Data Gouv (RDG), plateforme nationale fédérée des données de la recherche dont la disponibilité est prévue pour la fin du premier trimestre 2022, en est un exemple particulièrement significatif. Elle a ainsi vocation à traiter et à valoriser en priorité les données de la recherche associées aux publications en proposant :

1. Un service de dépôt et de diffusion pour les données ne bénéficiant d'aucun entrepôt disciplinaire adapté.
2. Un catalogue des données de la recherche française qui signale les données déposées dans des entrepôts nationaux ou internationaux thématiques et disciplinaires.
3. Des services d'accompagnement à la donnée tels que des ateliers pour les chercheurs dédiés à la préparation et à la diffusion de leurs données, ainsi que des centres de référence thématiques chargés de promouvoir les référentiels des différents domaines scientifiques.

Le principe de réutilisation des données par des acteurs de la même communauté disciplinaire ou d'autres communautés est prédominant. Cette plateforme ne renvoie ni à un système d'archivage ni à une solution de gestion du cycle de vie des données, malgré un accompagnement aux méthodes relatives à leur curation¹² dans le cadre des ateliers. Le

12. D'après une définition de l'Institut de l'information scientifique et technique (INIST) français, la curation désigne : « L'ensemble des activités et opérations nécessaires à une gestion active des données de recherche numériques, tout au long de leur cycle de vie. L'objectif est de les rendre accessibles, partageables

C2RMF a pu assister à la genèse de la construction de la plateforme par l'intermédiaire de réunions de présentation¹³, et étudie actuellement l'opportunité d'utiliser Recherche Data Gouv, moins pour sa solution d'entrepôt de données que pour ses services aux chercheurs, lesquels ont actuellement besoin d'être formés et accompagnés en vue d'une adaptation à la mue – prévue ou déjà en cours - de leurs données. Par ailleurs, le lien qui sera établi entre RDG et les publications déposées dans HAL est un bénéfice non négligeable en raison de la forte présence éditoriale au sein de cette plateforme des chercheurs du C2RMF, dont ceux d'AGLAE. En ce qui concerne les données brutes et les données intermédiaires, la formation et le suivi en vue d'une conformité avec l'exigence d'accès, de diffusion et de réutilisation, devront dans un premier temps être assurés par l'équipe d'AGLAE et du SAED, eux-mêmes sensibilisés par les spécialistes de la médiation numérique, des archives et de l'ingénierie des données.

et réutilisables de façon pérenne. Trois intervenants peuvent être identifiés dans le cycle de vie de données : les créateurs, le plus souvent les chercheurs, les curateurs et les utilisateurs. » [Voir le glossaire de l'INIST : <https://openaccess.inist.fr/glossaire/>]

13. Présentation d'Isabelle Blanc, administratrice ministérielle des données, des algorithmes et des codes sources de la recherche, de l'enseignement supérieur et de l'innovation, 11 mai 2021.

Chapitre 13

Conclusion

L’ambition originelle de cet argumentaire était élevée et, à bien des égards, périlleuse : appréhender les enjeux et les modalités d’un processus de patrimonialisation supposé, donc non avéré, des données expérimentales d’un accélérateur de particules à travers ses usages scientifiques, heuristiques et culturels. Le défi était intrinsèquement transversal, technologiquement complexe et conceptuellement exploratoire.

Il s’agissait notamment de raccorder le pouls naissant d’Euphrosyne aux opportunités ouvertes par un contexte de refonte stratégique, technologique, sémiologique et fonctionnelle des (méta) données produites au cours des analyses pratiquées par AGLAE sur des objets patrimoniaux. Traduite par une mise en conformité avec les principes FAIR, elle participe d’une forme de révolution feutrée vers une autre culture de la donnée, qui émancipe cette dernière de son écrin purement didactique pour la porter vers des dimensions sociales, culturelles et métaphysiques inédites et jusqu’alors inenvisagées. L’élaboration de ce mémoire a été l’occasion de mettre en lumière ce limon fertile mais difficilement saisissable qui a nourri les racines conceptuelles et fonctionnelles d’Euphrosyne, sous des aspects extrêmement pratiques et technologiques. Aussi, ce sont finalement deux fils d’Ariane interdépendants qui ont guidé notre progression en entremêlant l’ouvrage d’ingénierie informatique et documentaire qu’est Euphrosyne à ses implications socio-culturelles. Cette articulation a constitué une ligne de funambule sur laquelle nous avons tenté de trouver un équilibre perfectible et parfois insatisfaisant.

Dans un premier temps, la coupe « anatomique » et la circonscription typologique des données d’AGLAE ont livré de précieux indices sur leur valeur et leur fonction au regard de l’objet patrimonial, nous obligeant à raffiner et à clarifier le contenu polysémique dissimulé par la généricté terminologique de la donnée mais également du patrimoine. Ce n’est qu’à travers une cartographie contextualisée et centripète d’un jeu de données existant, depuis le répertoire utilisateur jusqu’à l’atomicité du signe contenu dans le fichier, que nous avons pu déduire le rôle majeur d’une considération et d’un traitement holistiques jusqu’au plus fin niveau de granularité de l’information dans l’acquisition de sa souche patrimoniale. L’anamorphose de la donnée expérimentale jusqu’à cette dernière

est visible à travers un jeu d’imbrication qui lui confère toute sa portée dès lors qu’elle est entrecroisée et mise en perspective avec les éléments produits par d’autres techniques scientifiques appliquées au même objet. Ainsi, la donnée en tant que résultat chiffré et/ou visuel relatif à une étude chimique n’a pas de valeur patrimoniale en elle-même. Elle ne l’acquiert que par la puissance de transfiguration cognitive de l’association, de la comparaison et d’une nécessaire discrimination sélective. Ce faisant, ce mécanisme ambivalent fait de subjectivité et d’application cartésienne révèle la force de saisie de la matérialité patrimoniale par l’esprit humain, lequel est seul à pouvoir lui offrir son sens et son amplitude à travers le temps. Par conséquent, les données d’AGLAE ne sont ni une adjonction périphérique à l’objet patrimonial, ni une augmentation de celui-ci, en dépit des arguments de communication développés en ce sens par les programmes et les équipements dédiés à la gestion mutualisée des instruments et des données obtenues. Sous réserve de l’efficience et de la performance des outils technologiques nécessaires à leur traitement, elles forment à la fois le vecteur et l’expression d’une conquête heuristique qui accroît l’intelligibilité, et simultanément l’épaisseur testimoniale, du patrimoine analysé. Elles constituent donc une « mise en profondeur » de ce dernier, à condition d’être présentées et structurées dans une unité signifiante, directement raccordée à l’objet ou au lot patrimonial auquel elle se rapporte. Cette unité reste un idéal à atteindre, car elle est complètement dépendante de la création d’un environnement commun de stockage, d’accès, de signalement documentaire et d’interrogation qui en est encore à sa genèse pour le laboratoire du C2RMF.

Par ailleurs, l’ébauche des contours d’un jeu de données pour AGLAE, dans le cadre de la première phase de construction d’Euphrosyne et à partir de la restitution des flux de données, a mis en évidence la difficulté de regrouper en une unité cohérente des éléments certes, réunis autour de l’épicentre qu’est l’expérience, mais néanmoins marqués par une hétérogénéité de contenu, de structure, d’organisation et de représentation. Le format conteneur HDF5 a semblé être une voie de consensus entre le respect de l’intégrité scientifique des données, caractérisées par une importante plasticité, et les besoins documentaires et sémantiques. Toutefois, si cette option semble particulièrement appropriée et aisée pour les résultats de l’imagerie, elle soulève davantage d’obstacles pour ceux de l’analyse ponctuelle, en raison de l’incompatibilité de ses outils logiciels avec HDF5. Cela nécessite une migration des données *a posteriori* de la génération du flux, à l’aide d’un programme qui doit encore être développé. Par ailleurs, ce format initialement prévu pour des calculs à partir de données massives n’est pas encore pris en charge par certaines infrastructures telles qu’Huma-Num, qui offre cependant des services et des outils intéressants pour le stockage, l’accès et l’interrogation des données intermédiaires et définitives. Il a donc été nécessaire de lever ces pierres d’achoppement en hiérarchisant les priorités au prisme des usages et du calendrier contraint d’Euphrosyne, qui devait faire l’objet d’une livraison rapide de prototype durant l’été. Ce travail de planification et de gestion de risques a ainsi laissé plusieurs arbitrages momentanément en suspens – ainsi du format HDF5 –

pour en privilégier d'autres aptes à garantir le maintien viable de l'activité d'AGLAE.

La construction d'Euphrosyne ainsi que l'élaboration d'une méthodologie FAIR pour les données de l'accélérateur ont ouvert de multiples zones de friche, inévitables dans le cadre d'une refonte aussi ambitieuse, où les besoins immédiats se sont parfois heurtés à l'effort prospectif. Cet écueil apparaît inhérent à la prise en charge concomitante de toutes les étapes du cycle de vie des données, ainsi que des informations produites à partir d'elles. Si elle a ralenti le chantier du premier prototype d'Euphrosyne, cette visée holistique a imposé une anticipation bénéfique en nous incitant à une forte souplesse dans l'architecture des fondations théoriques et fonctionnelles de l'interface, notamment lors de la création du modèle conceptuel puis relationnel de données. Ce dernier, qui a fait l'objet de nombreuses itérations avec l'équipe d'AGLAE et l'Atelier Numérique, a été conçu pour s'adapter à presque tous les cas d'usage ainsi que pour recueillir les éléments de la quasi-intégralité des différents flux de données possibles. Son caractère presque minimaliste a néanmoins pour revers son exclusion de quelques cas expérimentaux rares mais existants, comme les associations peu habituelles entre certaines données des différentes méthodes d'analyse que sont PIXE, PIGE, RBS et IBIL. Par ailleurs, notre travail s'est arrêté au seuil de l'implémentation, qui relève des missions de l'Atelier Numérique et de ses prestataires.

Avant d'être relayé et concrétisé par un système sémiologique et une gestion spécifique, le rouage élémentaire du processus de patrimonialisation n'est finalement autre que la pensée humaine, enserrée dans un tissu de représentations, de contraintes normatives et influencée par des pratiques spécifiques aux différentes collectivités voire communautés scientifiques. À partir des enquêtes réalisées lors de la phase d'investigation d'Euphrosyne et auprès des utilisateurs au cours de la période du stage, nous avons démontré l'incidence déterminante de la formation de la communauté d'AGLAE, à la faveur des derniers développements de l'instrument et du fonctionnement en réseau de la recherche en sciences du patrimoine, dans la reconnaissance et l'identification indispensables à la patrimonialisation de certains agrégats de données expérimentales. Ces deux étapes-clés forment la clé de voûte de l'acquisition de leur valeur et de leur utilité socio-culturelles, faisant de la communauté d'AGLAE, ou plutôt de ses pratiques partagées, une véritable matrice de patrimonialisation. Nous avons constaté l'importance cruciale de l'autonomie et de la proximité des utilisateurs avec l'instrument lors de l'événement expérimental dans l'émergence et la consolidation d'une acculturation harmonieuse, indispensable à l'éclosion de normes et de référentiels communs. Toutefois, cette assimilation est à géométrie variable, inégale selon les profils des utilisateurs et laissant l'avantage aux spécialistes en sciences des matériaux. Les nombreuses journées d'étude, ateliers collaboratifs et webinaires organisés par DIGILAB, SSHOC, la Société Française du Vide et ESPADON ont été l'occasion d'entrevoir par touches, au cours du stage, l'état actuel de cette construction collective normative et terminologique.

Celle-ci laisse apparaître un net déséquilibre à la faveur des sciences du patrimoine, lesquelles bénéficient d'un cadre conceptuel relativement abouti et robuste grâce à l'antériorité déjà ancienne de modèles de référence réappropriés et déclinés en ontologies d'application pour des domaines plus spécifiques, à l'instar du CIDOC-Crm Sci ou du CIDOC-Crm Cr. L'analyse par faisceau d'ions appliquée au patrimoine se trouve ainsi dans une marge de chevauchement avec une représentation conceptuelle particulièrement élaborée et disponible jusqu'au seuil de l'événement expérimental, mais qui reste à faire pour le champ propre aux différentes méthodes IBA (Ion Beam Analysis). De façon encore élémentaire et modeste, nous avons donc élaboré un modèle sémantique pour les ressources d'AGLAE, moins dans un objectif d'interopérabilité externe d'Euphrosyne qu'en vue d'élaborer une classification et une organisation conceptuelle à soumettre au débat et à l'itération des groupes de travail récemment mis en place dans le cadre de DIGILAB. Réalisé à la fin du stage, en concordance avec les dernières propositions des institutions à vocation scientifique et patrimoniale telles que la National Gallery, il n'a pas encore été présenté aux partenaires d'AGLAE et nous ignorons donc encore son degré de viabilité et de pertinence.

La construction d'un langage et d'un modèle de représentation communs, toujours en cours pour la communauté élargie d'AGLAE¹, se déploie à travers un espace de subjectivité fortement mosaïque caractérisé cependant par une préoccupation unanime à l'égard du risque de perte. La reconnaissance et la signalisation des données par la communauté d'AGLAE lui confèrent une légitimité scientifique autant que culturelle, et initient le processus de patrimonialisation en préfigurant la quête de leur conservation au long cours. La réflexion autour de la pérennisation des données d'AGLAE s'est ainsi développée en filigrane des différents volets méthodologiques de l'optimisation de l'accès, de l'ouverture et de la réutilisation des jeux de données expérimentaux. En concertation avec le SAED du C2RMF, nous avons esquissé une proposition de stratégie d'archivage conforme au modèle OAIS, alimentée par nos prospections auprès du CINES et d'Huma-Num, particulièrement éclairantes sur les mutations en cours des plateformes externalisées. Nous avons ainsi perçu la convergence de nos besoins avec l'émergence d'une politique d'archivage apte à intégrer et à traiter les différents états des données en une seule infrastructure. Des partenariats étroits contractualisés entre le CINES et des organismes/structures dédiés à l'archivage des données courantes et intermédiaires ouvrent des perspectives sur une solution holistique capable d'offrir un cadre, des moyens et une sécurité aux besoins de stockage, d'accès, de réutilisation et de pérennisation des données produites dans le cadre d'Euphrosyne. Avant d'envisager sa mise en œuvre, il est cependant indispensable d'approfondir concomitamment la politique de gestion de données et celle de l'archivage pérenne de l'ensemble des contenus générés par les analyses pratiquées au C2RMF. Les aspects

1. Cet élargissement inclut les domaines de la physique des particules et de la chimie, hors de leur application au patrimoine.

fonctionnels et organisationnels inhérents au processus informationnel et documentaire de l’archivage, mis en relief au cours du stage, soulèvent notamment des problématiques liées aux ressources humaines et logistiques disponibles, actuellement à l’étude.

Les enjeux et les modalités de la pérennisation des données d’AGLAE ont été davantage appréhendés que ceux de leur exposition et de leur diffusion, en particulier à travers les supports et les circuits de l’*Open Science*. Les chercheurs d’AGLAE participent depuis longtemps à cette dernière par leur contribution à l’édition scientifique ouverte et gratuite, mais un large chantier demeure pour l’ouverture et la diffusion sans délai des données courantes et intermédiaires, en particulier celles traitées au moyen de paramètres savamment sélectionnés à l’aide des logiciels spécifiques. Conscients de la rareté et de la préciosité de ces données en raison de leur objet et de leur contexte de production, ils tendent encore à les cloisonner comme des trésors à dérober à la vue d’esprits mal intentionnés. Ce faisant, ils créent un obstacle à l’exposition nécessaire à leur consécration patrimoniale, mais également à l’application des principes FAIR d’accès, de signalement et de réutilisation. Néanmoins, les infrastructures, les réseaux et les outils bâtis sur le socle déontologique de l’*Open Science* livrent progressivement la preuve de leurs bénéfices pour la valorisation et l’évolution de la recherche en sciences du patrimoine, dont le périmètre est universel. Par conséquent, si l’*Open Science* ne saurait constituer à elle-seule une justification des coûts multiples de pérennisation et d’exposition de cette masse numérique toujours croissante, elle apparaît de plus en plus aux chercheurs comme le véhicule de moyens indispensables à la conservation et à la transmission inaliénables de « cet héritage du passé aux générations futures », à la nature et aux caractères hétérogènes.

Ce stage s’est finalement surtout apparenté à une intense prospection protéiforme et transdisciplinaire, dans l’entrecroisement des pratiques et des concepts scientifiques, patrimoniaux, informatiques et archivistiques, pour contribuer à l’accompagnement d’une communauté dans sa transition vers une culture du signe et de l’information sinon universelles, du moins résolument testimoniales.

Annexes

Annexe A

Inventaire de l'existant

Questionnaire AGLAE

- 1) Utilisez-vous l'ontologie CRMcr ? Si oui, pouvez-vous en préciser exactement les usages ?
- 2) AGLAE bénéficie t-il d'une/de base(s) métier(s) ainsi que d'un silo local RDF ?
- 3) Pouvez-vous indiquer le type, le format, le nom et le support de stockage des données à chacune des étapes de leur cycle ?
- 4) Quels sont les logiciels utilisés pour les différents traitements des données ? (précisez les noms de logiciels, si possible leurs licences, leurs usages spécifiques et les types de données auxquels ils s'appliquent)
- 5) Quelle est l'infrastructure de stockage des données d'AGLAE (type de serveur, scalabilité horizontale ou verticale, coût économique de maintenance, nombre de silos et types de données associées)
- 6) Quel est l'état actuel des métadonnées associées aux données d'AGLAE ? (format, usages immédiats et différés, type et lieu de stockage, nom éventuel de fichier lors de leur génération)
- 7) Les données et métadonnées d'AGLAE sont-elles associées à un identifiant numérique spécifique ?
- 8) Existe t-il des données de provenance détaillées (personne responsable de la manipulation, date, etc.) ? Si oui, selon quels format, support et durée de stockage ?
- 9) Les données d'AGLAE sont-elles mises en relation avec des données issues d'autres laboratoires du C2RMF (ou autre) ?
- 10) Pouvez-vous décrire le plus précisément possible l'infrastructure informatique d'enregistrement et d'indexation des données ?
- 11) Sous quelle(s) licence(s) les données d'AGLAE sont-elles actuellement publiées ?

- 12) Pourriez-vous décrire vos propres usages des données et éventuelles métadonnées d'AGLAE depuis l'instant 0 de la manipulation jusqu'à leurs interprétations et analyses scientifiques ? **Question adressée à chaque personne de l'équipe pour bien déterminer les usages métiers spécifiques.**
- 13) Décrire précisément les interactions entre les (méta)données d'AGLAE et EROS. Quelles seraient les évolutions souhaitées sur ce point ?
- 14) Pouvez-vous décrire avec précision l'accès distant via le réseau Renater mis en place depuis le premier confinement (moteur de recherche, données accessibles, volumétrie, degré de scalabilité (avec le double enjeu du respect de l'intégrité des données et de la montée en charge due aux connections utilisateurs), etc.
- 15) Pourriez-vous expliquer les modalités et les implications de l'architecture en fédération de nœuds de recherche et si cette dernière vous semble viable pour les données d'AGLAE ?
- 16) Quelles sont les utilisations envisagées du référentiel IBA ?
- 17) Quel est le système de gestion de base de données (SGBD) d'EROS ?

FIGURE A.1 – Questionnaire à destination des membres de l'équipe d'AGLAE et du SAED, avril 2021.

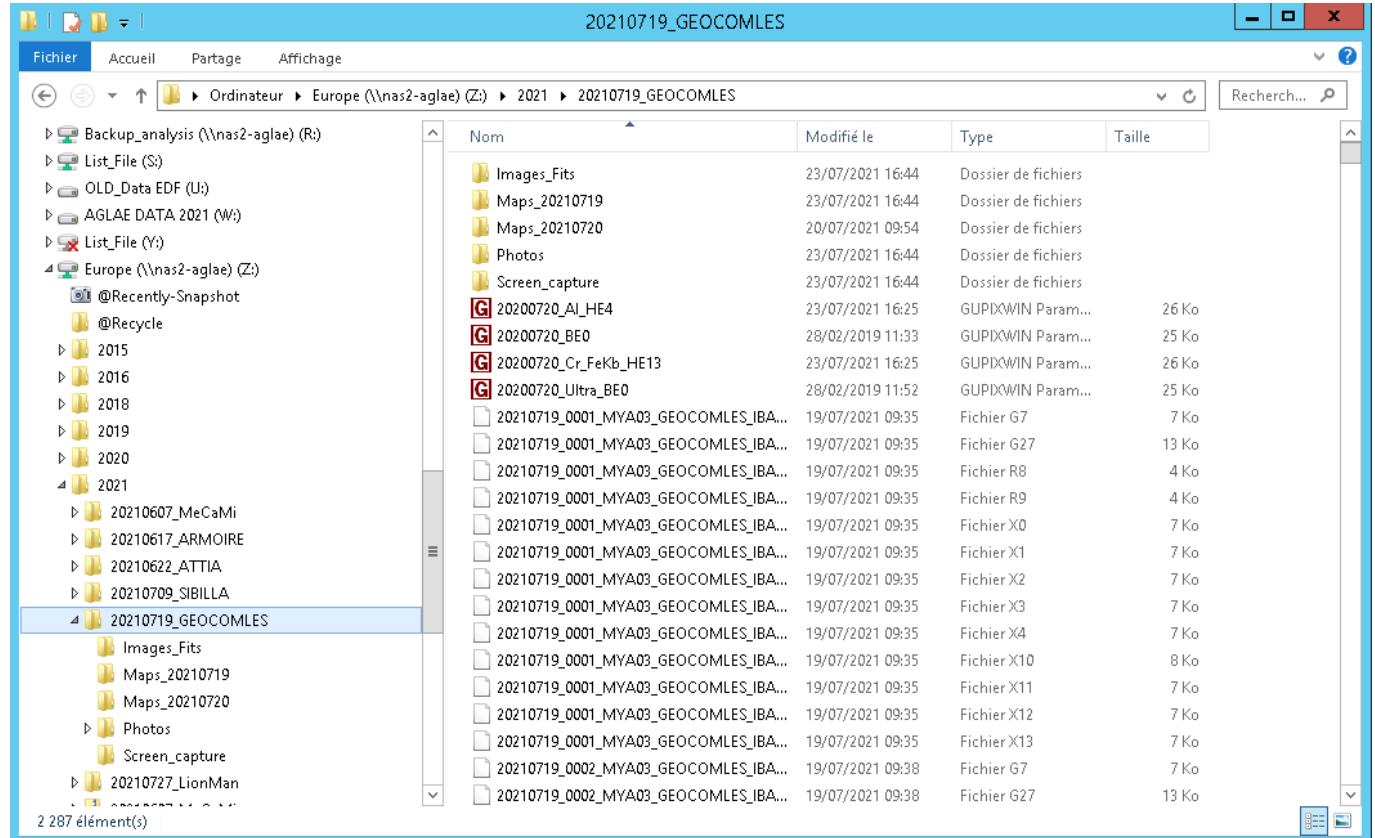
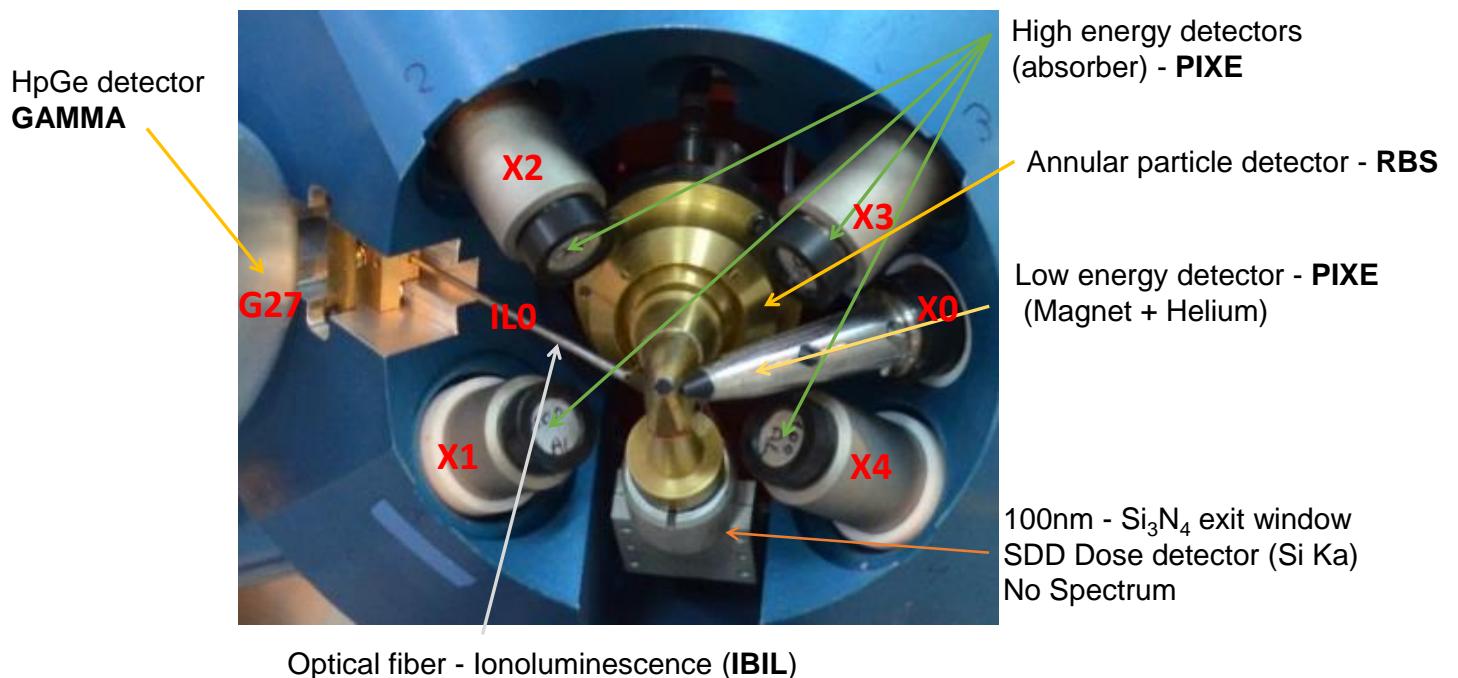
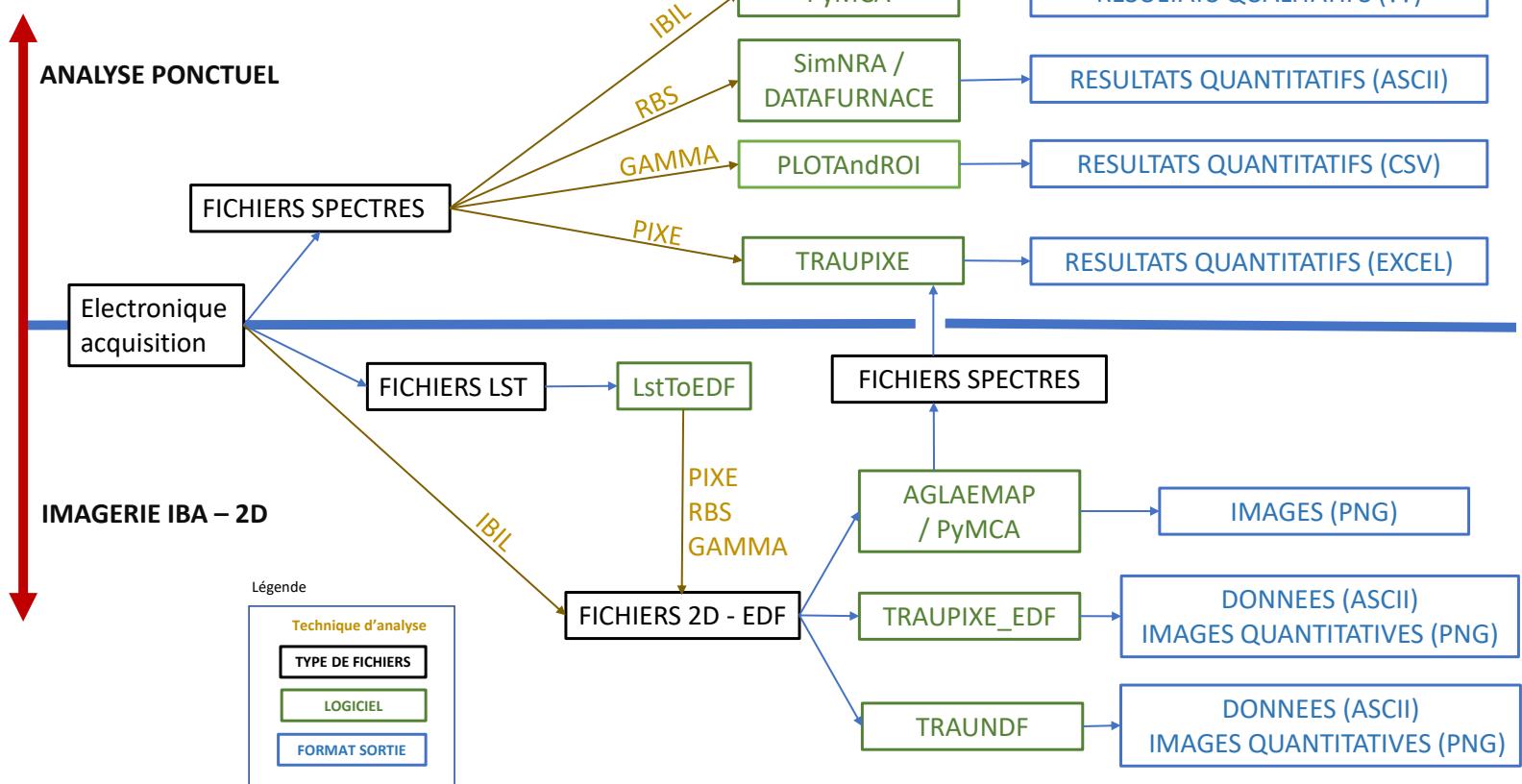


FIGURE A.2 – Capture d'écran d'un exemple d'arborescence de dossiers et de fichiers d'AGLAE depuis le répertoire de l'année 2021.

External NMP end-station





MODE CARTOGRAPHIE

Données des détecteurs

1) Format brut : Fichier LST (*.lst)

Format des données documenté FastComtec GmbH
Output : 1 Fichier LST, 10 fichiers des spectres bruts
(x0, x1, x2, x3, x4, x10, x11, x12, x13, rbs, g7)



2) Logiciel faisant l'enregistrement des données des détecteurs caché à l'utilisateur

2) Convertisseur LST To EDF

Logiciel maison (GUI Labview,
Moteur extraction via un exécutable)



3) LSTToEDF - Logiciel utilisateur

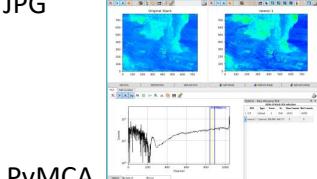
4) Visualisation des cartographies

4.1) AGLAEMAP , logiciel maison GUI Labview

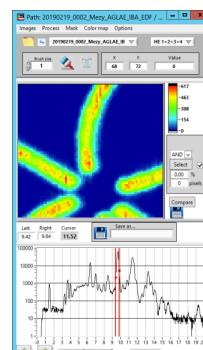
Output : Fichiers des spectres de zones d'intérêts (.x0, .rbs, .g7, ...) et images JPG

4.2) PyMCA , Logiciel ESRF (European Synchrotron Radiation Facility), Free Python code

Output : Images JPG



PyMCA

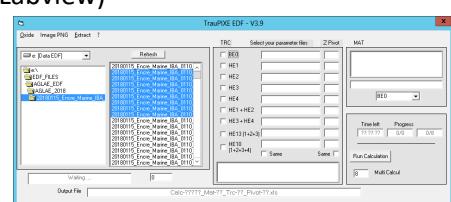


AGLAEMAP

5) Traitement quantitatif des cartographies

TRAUIPIX_EDF , Logiciel maison, code VB, moteur de calcul GUPIX (pixwin.exe)

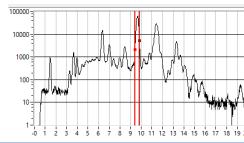
Output : Images quantitatives , tableau 2D ASCII
Visualisation via DataImaging (Logiciel Maison, GUI Labview)



TRAUIPIX_EDF

MODE PONCTUEL / SPECTRE TOTAL - PIXE

1) INPUT : Fichiers ASCII des spectres

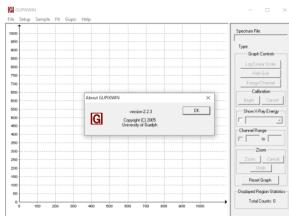


3) Traitement des données PIXE

3.1) GUPIX : Configuration des fichiers de paramètres (Energie faisceau; type de particules; filtres; éléments chimique recherchés, ...)

Logiciel University of Guelph, Canada , code FORTRAN
Licence payante

Output : Fichiers .par , ASCII



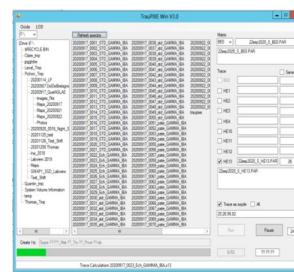
3.1) GUPIX

3.2) TRAUIXE : Batch l'ensemble des spectres.

Output : 1 Fichiers Excel contenant la concentration des éléments chimique.

Logiciel Maison, code VB , Moteur de calcul GUPIX (pixwin.exe)

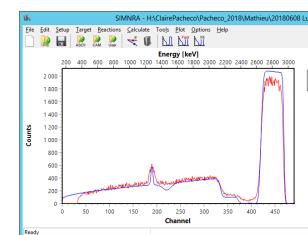
Output : 1 fichiers Excel, fichiers JPG



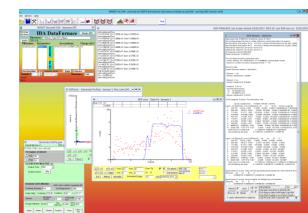
MODE PONCTUEL / CARTOGRAPHIES – RBS

1) INPUT : Fichiers ASCII, Format RBS

2) SimNRA, Max-Planck-Institut für Plamaphysik,
Garching, Germany,
Licence payante
Output : Fichier ASCII

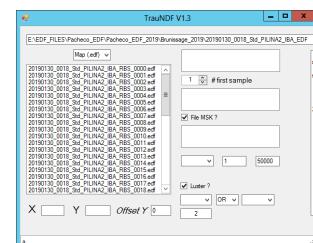


3) NDF-DataFurnace, The UK National Ion Beam
Centre, Surrey, Licence payante
Output : Fichier ASCII résultat ou fichiers paramètres
pour TRAUNDF



4) TRAUNDF , Traitement des cartographies RBS

Logiciel maison, code VB, Moteur de calcul Datafurnace (nfd.exe)
INPUT : Fichiers EDF
Output : Fichier ASCII 2D
Visualisation via DataImaging (Logiciel Maison, GUI Labview)



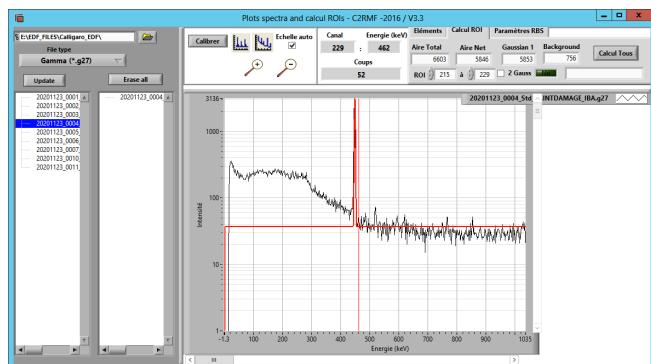
MODE PONCTUEL / GAMMA

1) INPUT : Fichiers ASCII, Format Gamma

2) PlotAndROI : Calcul l'aire d'un pic Gamma

Logiciel Maison , GUI Labview

Output : Fichier ASCII, CSV



FORMAT LST

HEADER ASCII suivi des données en format binaire

```
..  
cmline2= Map size:10000,1000,20,20,500,-1998,50000  
cmline3= Exp.Info:Proton,3002keV,4mmHe,50umAl,OFF,50umAl,50umAl  
..
```

Cmline2 = Map size: « TailleX », « TailleY », « Pixel size X », « Pixel size Y », « Taille Pinceau du faisceau », « Vitesse de déplacement de l'échantillon », « DOSE »
Cmline3 = Exp. Info : « Type de particules », « Energie Faisceau », « Filtre détecteur 0; « Filtre Détecteur 1 », « Filtre Détecteur 2 », « Filtre Détecteur 3 », « Filtre Détecteur 4 »

FORMAT EDF

HEADER ASCCI suivi des données en format binaire

```
HeaderID = EH:000001:000000:000000 ;  
Image = 1 ;  
ByteOrder = LowByteFirst ;  
DataType = UnsignedLong ;  
Size = 16068944 ;  
Dim_1 = 512 ; # Dimension d'un spectre  
Dim_2 = 75 ; # Nb pixels dans le fichier  
MCA a = 1 ; # Facteur calibrage du spectre  
MCA b = 1 ;  
MCA c = 0.0 ;  
COMMENTS =Map size X*Y (um) = 1500*1500 / Pixel size X*Y (um) = 20*20 / PenSize (um) = 250 / DoseRow = 4444 / TotalDose = 2000000;
```

FORMAT PIXE & GAMMA

.x0, .x1,, Format ASCII entête de 2 Lignes suivi des données ASCII

1^{ere} ligne : Nombre de canaux du spectre (2048)

2^{eme} ligne : « Année » « mois » « nombre de secondes depuis minuit » « temps d'acquisition » « somme du spectre », « Commentaire écrit par l'utilisateur (Mica-Mg – Gamma bougé), suivi des infos expérimentales.

```
2048 1
2020 09 40377 243 9450139 'Mica-Mg - Gamma bougé,1000,1000,20,20,500,40,500000,Proton , 3002 keV , 4mm He , 50 um Al , OFF , 50 um Al , 100 um Al'
0
0
48
59
..
```

FORMAT RBS

.r9, format ASCII, Header suivi des données ASCII

ligne NAME : « Commentaire écrit par l'utilisateur (Ech2 –Blanc) , suivi des infos expérimentales.

```
[DISPLAY]
VER=244
TYPE=0
DATALEN=2
XPARAM=2
YPARAM=0
XRANGE=511
YRANGE=
NAME=Ech2 - Blanc,1000,1000,20,20,500,40,500000,Proton , 2999 keV , 4mm He , 50 um Al , OFF , 50 um Al , 100 um Al
XTITLE=Coups
YTITLE=

[DATA]
0      0
1      0
```

AUTRES METADATA

1- Fichier Excel / expérience– renseigné automatiquement

File N°	Sample info	Ref Objet	Project name	Dose /Sec.	Analyse time	LE 0	HE 1	HE 2	HE 3	HE 4	Gamma	RBS	Flux He (l/min)	Map Size	Time	Beam energy	Particle	Filters (LE0; HE1; HE2; HE3; HE4)	
20200917_0001 _STD_GAMMA_ IBA	Dr-N	STD	GAMMA	50064	2080	00:03:54	8357	16590	0	18222	4009	3961	0	1,0	1000,1000, 20,20,500, 40,50000	10:47:12	3002	Proton	4mm He , 50 um Al , OFF , 50 um Al , 100 um Al

2- Photo de la zone analysée (JPG)



FIGURE A.3 – Cartographie générale des flux de données produits au sein d'AGLAE.
©New AGLAE

Annexe B

Contexte expérimental-AGLAE



Description of the project

To upload the file online, please save the file in pdf and rename it with the acronym of the proposal (acronym.pdf)

1. *Project summary¹ (max 300 words)**

2. *Scientific background (max 500 words)**

3. *Description of the planned work and experimental methods (if it is the case)² (max 600 words) **

4. *Previous analysis on the artefact (max 300 words)**

5. *Expected achievements (max 400 words)**

6. *Impact and dissemination plan (max 400 words)**

7. *References (min 5 - max 10)**

¹ Be sure to include the novelty and/or added value of the proposed work.

² The experimental technique(s) requested with justification, required set-up(s), measurement strategy, sampling area/point details (number, location, etc.).



FIGURE B.1 – *Template* d'une description de projet pour une demande de temps de faisceau dans le cadre d'IPERION-HS.

Please add as many users as necessary, via copied extra pages

5) AGLAE Local Contact (if any)

Quentin Lemasson

6) Beam characteristics

Particle	<input type="checkbox"/> Proton	<input type="checkbox"/> Alpha	<input type="checkbox"/> Other
Energy	3 Mev		
Intensity	1,5 nA		
Size	Ø ~0,5 mm		

7) Method(s) implemented

PIXE PIGE RBS NRA ERDA IBIL

8) Experimental set-up

Detectors	<input type="checkbox"/> Low energy X	<input type="checkbox"/> High energy X (4 detectors)		
Filter	no	Al - 50 µm		
Diaphragm	no	no		

Charged particles external Charged particles annular Gamma

9) Target description

Number approximately 50 geological samples
 approx. 20 polished axes

Type _____

Size 15x5x5 cm max _____

10) Requested preparation

____ No _____

11) Possible risks : no

12) Project description (Scientific background - Experimental methods - Results expected – References)

Description du projet (Contexte scientifique - Méthodes expérimentales - Résultats attendus – Bibliographie)

max. 3 pages

Color relationship and chemical composition of fibrolite, application to the determination of the origins of Neolithic polished axes from France.

Scientific background;

The FIBROLITE programme is a multidisciplinary programme in both the Human Sciences and Earth Sciences. He is interested in Neolithic polished axes in fibrolite discovered mainly in the tombs of the great tumuli of the region of Carnac in Brittany dating from the 5th to the 3rd millennium BC. These are prestigious objects that are often associated with other precious objects such as jadeite axes from the Alps and variscite pearls from the Iberian Peninsula.

The objective of the project is to determine the nature, origin and circulation of fibrolite axes during prehistoric times in Western Europe.

The term fibrolite refers to the fibrous habitus of sillimanite, alumina silicate but also fibre concentrations in the form of nodule, kidney or platelet. Fibrolite is a common mineral phase in silico-aluminous metamorphic rocks such as gneisses or micaschists. It is often present in diffuse form in rocks and much more rarely in the form of nodules exploitable by neolithics. Potential geological sources of this rock for axe processing are therefore rare. There are three known deposits in the armorican Massif. Other deposits are located in the Massif Central whose precise location is in progress. We also plan to incorporate the deposits of the Iberian Peninsula into our programme. The hypothesis of an Iberian origin of the polished fibrolite axes is often put forward by prehistorians, knowing that the variscite pearls associated with these objects come from this region.

Two types of fibrolite have been distinguished during our previous studies of armorican deposits (Guiavarc'h and Querré, 2012). P-type fibrolite occurs as massive blocks with a microgranular appearance. The material is usually greenish in colour, some facies are more red, grey, black or white. Type K fibrolite comes in the

form of small, stringy platelets. It is characterized by brown to burgundy marbling with a whitish to grey matrix. (cf photo)

In the FIBROLITE programme, we plan to carry out an exhaustive study of the raw material of the French and Spanish deposits and at the same time of all the polished axes from the tombs of the Carnac region in Brittany.

The three fundamental objectives of the FIBROLITE project are:

- Characterize the origin of the fibrolite color by determining the chromophore elements. Evaluate the influence of inclusions in colour variations.
- Characterize the Armorican Massif and Massif Central fibrolite geological deposits using the PIXE method, in addition to other analytical techniques such as X-ray fluorescence spectrometry.
- Using the PIXE, characterize the material of the fibrolite axes of western France, centred on the museums of Carnac, Vannes and the Musée d'Archéologie Nationale de Saint Germain-en-Laye.

Experimental methods

Three investigation methods are used in the FIBROLITE project: macroscopic observation, binocular magnifying glass observation and PIXE extracted beam geochemical analysis. These methods integrate the constraints linked to the study of precious objects. Completely harmless, the PIXE method allows rapid chemical analysis of the mineral characteristics of polished axes and geological deposits.

Moreover, it is particularly indicated here because it covers a wide spectrum of major, minor and trace elements that make up fibrolite: Na, Mg, Al, Si, P, K, Ca, Ti, Mn, Fe, V, Cr, Co, Ni, Cu, Zn, Ga, As, Rb, Sr, Y, Zr, Nb and U, all parameters used to determine the origin of materials. The presence of light chemical elements in fibrolite, in particular sodium, does not allow the use of X-ray fluorescence but PIXE at atmospheric pressure.

The Fibrolite project is particularly interested in the chemical mapping developed on the AGLAE accelerator that we have already had the opportunity to test as part of the Equipex "New ALAE" project. We would like to be able to combine chemical composition, colorimetry and ionoluminescence for the same analysis surface.

Expected Results

The objectives of the polished axe and geological reference analyses are threefold. On the one hand, it is a question of characterizing the mineralogical nature of the materials constituting these archaeological pieces, i.e. the fibrolitic matrix and the other mineral phases present in the form of inclusion.

Based on previous determinations and depending on the geological context of the archaeological sites from which the axes originate, the second objective is to confirm or refute the allochthonous origin of the axes.

Finally, the last point aims to determine their geological and geographical origin. This type of research concerning the origin of objects and materials can only be carried out with reference to a homogeneous and

coherent database. The establishment of this database began with the analysis of geological samples obtained on AGLAE (Guiavarc'h et al. 2016).

We would also like to pursue research in specific areas in order to determine:

- if there are several origins within the same archaeological site,
- if the typology of the axes is linked to their origin,
- the evolution of the exploitation of each of the geological resources during the 2.5 millennia of use,
- the functioning of the network(s) and their evolution from the 5th to the 3rd millennium to address human issues: breakdown of relations with a given region in favour of another resulting either from a conflict in a region blocking communication routes, or from the drying up of a field or another reason to be determined.

For this, it is essential to significantly expand our reference base by analyzing several geological samples representative of each French and Iberian deposit and armorican archaeological objects.

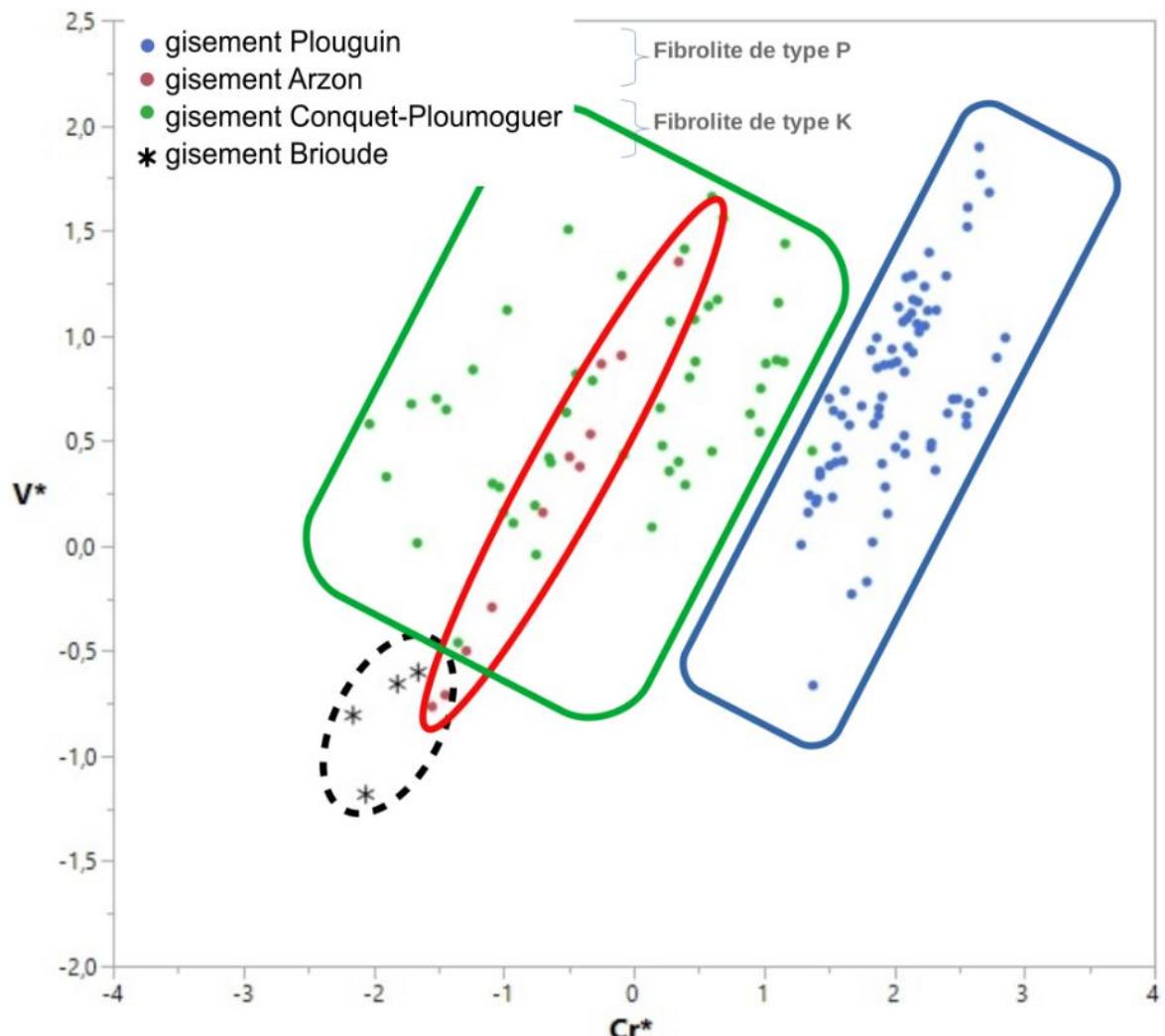
This work will be the subject of several presentations in international colloquia and international journals. A joint paper at the International Archaeometry Symposium to be held in Montreal in May 2019 will provide an initial synthesis of all the data acquired on fibrolite in France.

This project will lead to further requests for analysis on AGLAE as part of a forthcoming european programme in collaboration with Salvador Dominguez-Bella and his team at the University of Cadiz in Spain.

First Results

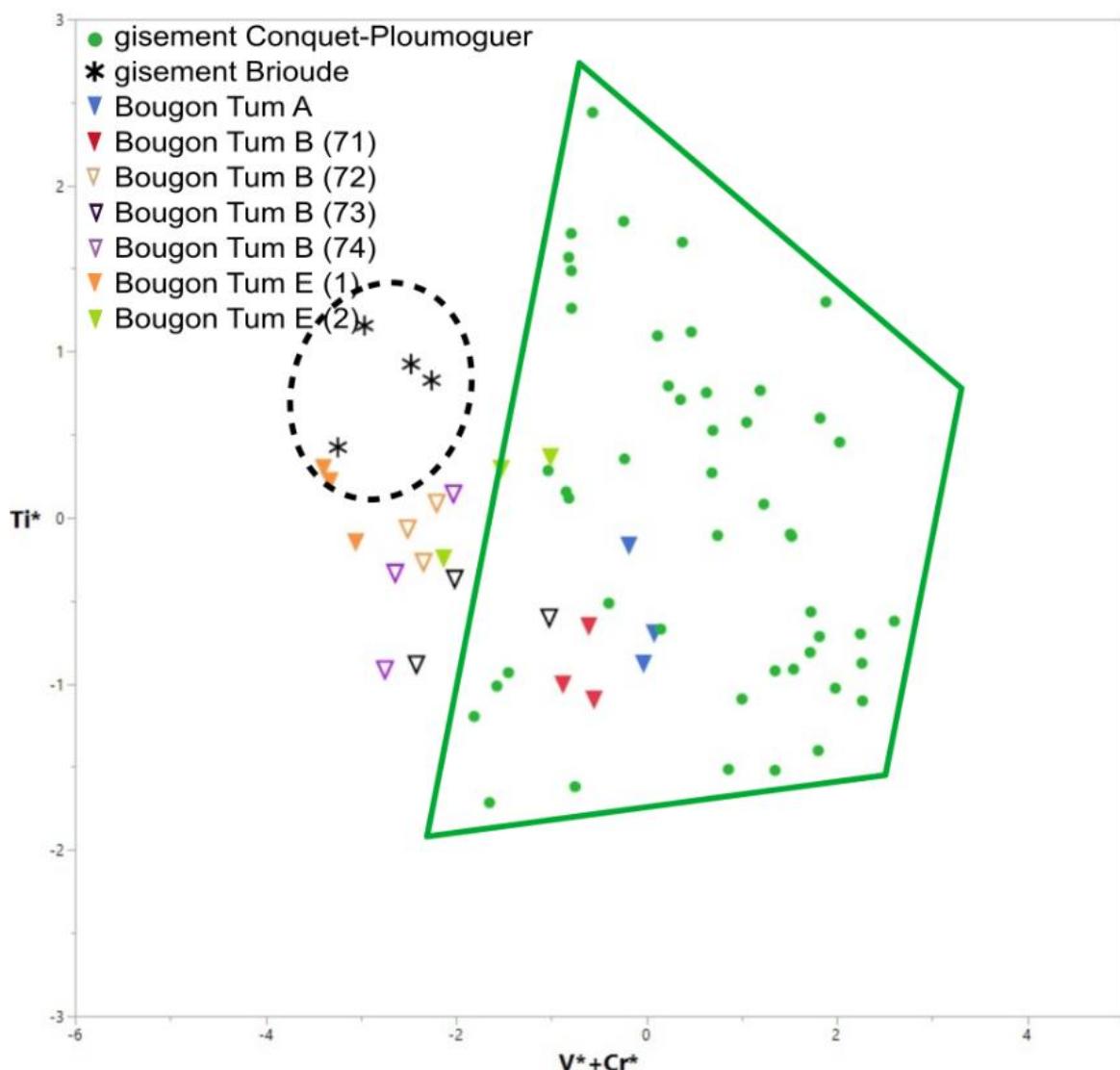
Fibrolite deposits in the Armorican Massif

The last two analysis sessions using the Pixe method have made it possible to determine the chemical elements that discriminate the different fibrolite deposits in the Armorican Massif (see figure). The elements Cr and V are particularly interesting to use. Particularly within the Plouguin deposit itself, these elements make it possible to differentiate between exploitation zones 1 and 2. For the moment, we are unable to distinguish between zones 2 and 3.



Polished fibrolite axes between the Loire and the Pyrenees

The first chemical characterisations made it possible to propose the origins of polished blades found in the geographical area between the Loire and the Pyrenees (Marticorena et al., in press). Some axes have chemical characteristics similar to the Leconquet-Ploumoguer deposit in Finistère. (cf. figure).



Prospects for research on the origin of polished fibrolite axes.

The next analysis sessions will be partly devoted to methodological development. The main objective of this development is to determine as precisely as possible the different areas of exploitation of the Plouguin deposit. To do this, we plan to carry out chemical and colorimetric mapping around mica inclusions. In this configuration, we will be able to have not only the chemical composition of the fibrolite but also that of the micas. In this way, we will increase the possibilities of discriminating areas of exploitation of the Plouguin fibrolite.

Within the framework of the study of the origins of the polished axes from the south-west of France, other fibrolite blades will be analysed. These objects have the particularity of being associated with variscite beads in their archaeological context. In the continuity of the Callaïs programme, we therefore plan to analyse these pearls as well in order to compare their provenance with that of the fibrolite axes.

Bibliography

GUIAVARC'H M., QUERRÉ G., LEMASSON Q. (2016) Characterization by PIXE elemental analysis of fibrolite: application to the study of neolithic polished axes provenance. Poster, Colloque "Raw materials exploitation in Prehistory: Sourcing, processing and distribution 10-12 March 2016, Faro,Portugal

GUIAVARC'H, M., et QUERRÉ, G. (2012) La fibrolite du Massif armoricain: état des recherches sur les gisements et la caractérisation des haches polies néolithiques, in Roches et sociétés en préhistoire entre massifs cristallins et bassins sédimentaires (eds. G. Marchand and G. Querré), 291.306, Archéologie & Culture, Presses Universitaires de Rennes, Rennes.

CASSEN, S., BOUJOT, C., DOMINGUEZ-BELLA, S., GUIAVARC'H, M., LE PENNEC, C., PRIETO MARTINEZ, M. P., & VIGIER, E, (2012) Dépôts bretons, tumulus carnacéens et circulations à longue distance. In Jade. Grandes haches alpines du Néolithique européen. V e et IV e millénaires av. J.-C. Cahiers de la MSHE CN Ledoux, Besançon, Presses Universitaires de Franche-Comté et Centre de Recherche Archéologique de la Vallée de l'Ain, 2, 918-995.

CASSEN S., PÉTREQUIN P., BOUJOT C., DOMINGUEZ BELLA S., GUIAVARC'H M., M.P., QUERRÉ G. (2011) Measuring distinction inside the architectures of the Carnac region. From sign to material. In :J. Müller, M. Furholt (Ed.), Megaliths and Identities ; European Megalithic Studies Group.

MARTICORENA P., GUIAVARC'H M., BOSCUS S., QUERRE G. (in press) Entre Loire et Pyrénées, les lames polies en fibrolite.

FIGURE B.2 – Extrait de la demande de temps de faisceau (PROPOSAL) en vue de l'analyse d'échantillons et d'objets en fibrolite, janvier 2021.

Please add as many users as necessary, via copied extra pages

5) AGLAE Local Contact (if any)

Claire Pacheco, Quentin Lemasson, Brice Moignard et Laurent Pichon

6) Beam characteristics

Particle	<input checked="" type="checkbox"/> Proton	<input type="checkbox"/> Alpha	<input type="checkbox"/> Other
Energy	3 MeV		
Intensity	Maximum possible		
Size	30 µm		

7) Method(s) implemented

PIXE PIGE RBS NRA ERDA IBIL

8) Experimental set-up

Detectors	<input checked="" type="checkbox"/> Low energy X	<input checked="" type="checkbox"/> High energy X (4 detectors)			
Filter		50 µm Al	50 µm Al	50 µm Al	50 µm Al
Diaphragm					

Charged particles external Charged particles annular Gamma

9) Target description

Number about 10
Type Ceramic sherds and cross-sections

Size centimetric

10) Requested preparation

None

11) Possible risks

None identified

12) Project description (Scientific background - Experimental methods - Results expected – References)

Description du projet (Contexte scientifique - Méthodes expérimentales - Résultats attendus – Bibliographie)
max. 3 pages

The continuation of this study focuses mainly on iron-rich glazes of Chinese ceramics. Chinese glazes coloured with iron can display an outstandingly rich palette, ranging from light blue and green to glossy black. Recently, brown-to-black glazes have attracted the attention of researchers, as recent studies of iron-rich glaze samples analyses revealed that a rare metastable iron oxide (the ϵ -Fe₂O₃ phase) was present on Jian ware samples (Dejoie et al, 2014). Analyses of other types of iron-rich Chinese glazed material, such as sauce glaze, also revealed the presence of this rare phase on the glaze (Liu et al, 2018; Wang et al, submitted). Observation and analyses of these porcelains also revealed that the glaze is often layer-structured. For instance macroscopic observation of brown glaze ceramic reveals two layers: a brown and a black.

Recent analyses carried out on samples of brown glaze porcelain show an enrichment in iron oxide on the top brown layer (Wen et al, 2019; Wang et al, submitted). Wen et al. also underlined the possible importance of the titanium content to explain the coloring of the glaze. TEM observation of purple-gold glaze, another type of Chinese porcelain, also highlighted a layered-structure of the glaze when looking at the repartition of ϵ -Fe₂O₃ crystals (Liu et al, 2018).

The first experiments implemented at the AGLAE facility last year perfectly showed the enrichment in iron at the surface of the cross-sections of archaeological samples. In addition, it also revealed a significant increase of the calcium concentration at the surface (cf. Figure bellow and document Rapport-CP2CGC.pdf sent after the experiment). This diffusion can be due to outward diffusion resulting from the crystallisation mechanism of ϵ -Fe₂O₃ crystals at the surface of the glaze or to inward diffusion due to the weathering.

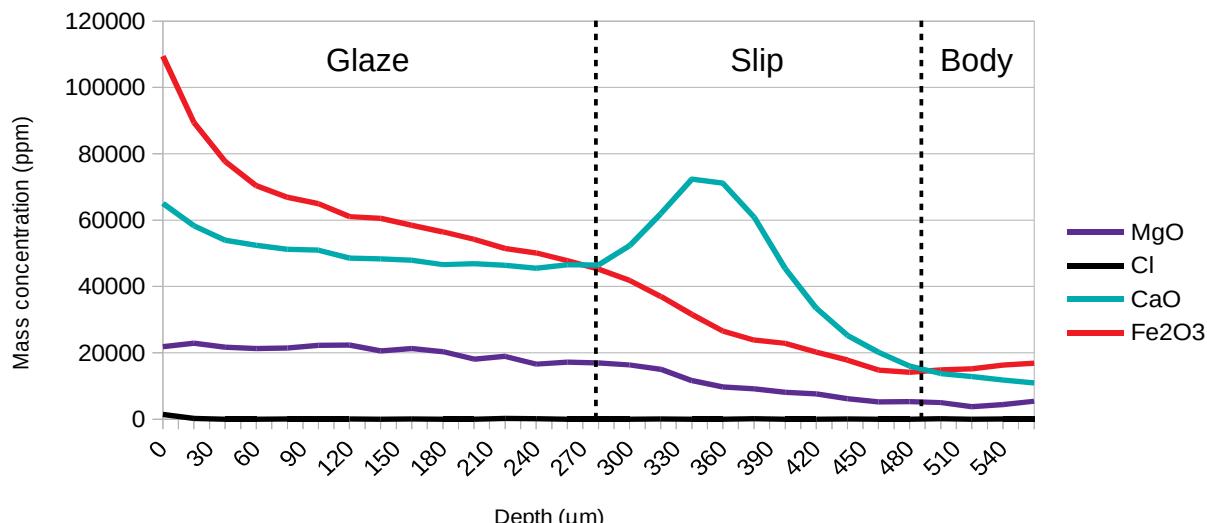


Figure: Concentration profiles of divalent ions in the cross-section of BG02 (Cl monitored for the distinction of the interface between the glaze and the resin)

To assess the origin of the Ca diffusion, analyses of cross-sections of control samples taken at different times during a traditional firing of brown glazes in China. PIXE cartography would be a great tool to investigate this diffusion, as it allows the investigation of millimetric areas impossible to assess correctly by EDS. Furthermore the AGLAEMAP and DataImaging softwares would make it possible to easily transform the maps into profiles by averaging the pixels on each line and give a pertinent overview of the diffusion without the influence of the crystallised phases which prevent averaging with the EDS studies.

In addition to the study of the control samples mentioned above, analyses of cross-sections of other ancient samples that were not investigated are planned to complement the previous study. Quantitative PIXE cartographies of large surfaces of shards overnight to investigate the iron distribution according to the patterns would also complement greatly the study of the cross sections.

Bibliography :

- C. Dejoie, P. Sciau, W. Li, L. Noé, A. Mehta, K. Chen, H. Luo, M. Kunz, N. Tamura, Z. Liu, Learning from the past, Rare ϵ -Fe₂O₃in the ancient black-glazed Jian (Tenmoku) wares, **2014**, Scientific Reports 4, 4941
- Z. Liu, C. Jia, L. Li, X. Li, L. Ji, L. Wang, Y. lei, X. Wei, The morphology and structure of crystals in Qing Dynasty purple-gold glaze from the Forbidden City, **2018**, J. Am. Ceram. Soc. 11, 5229-5240
- T. Wang, Z. Ren, C. Holé, P. Zhang, P. Shi, J. Zhu, H. Luo, F. Wang, P. Sciau, Morphological and structural study of crystals in black-to-brown glazes of Yaozhou ware (Song dynasty) using imaging and spectroscopic techniques, submitted in January 2021, J. Eu. Ceram. Soc.
- R. Wen, D. Wand, L. Wang, Y. Dang, The colouring mechanism of the Brown glaze porcelain of the Yaozhou Kiln in the Northern Song Dynasty, **2019**, Ceramics International 45, 10589-10595

I accept that all the necessary actions regarding the safe transportation of samples/objects back and forth to the facility, insurance of the samples/objects, etc. are the responsibility of the users, and no associated costs can be refund within the project.

I further confirm that, I will send a Summary Report within 2 months after the experiment

Date 06/01/2021

Group Leader Signature



FIGURE B.3 – Extrait d'une demande de temps de faisceau effectuée dans le cadre du programme MOLAB, janvier 2021.

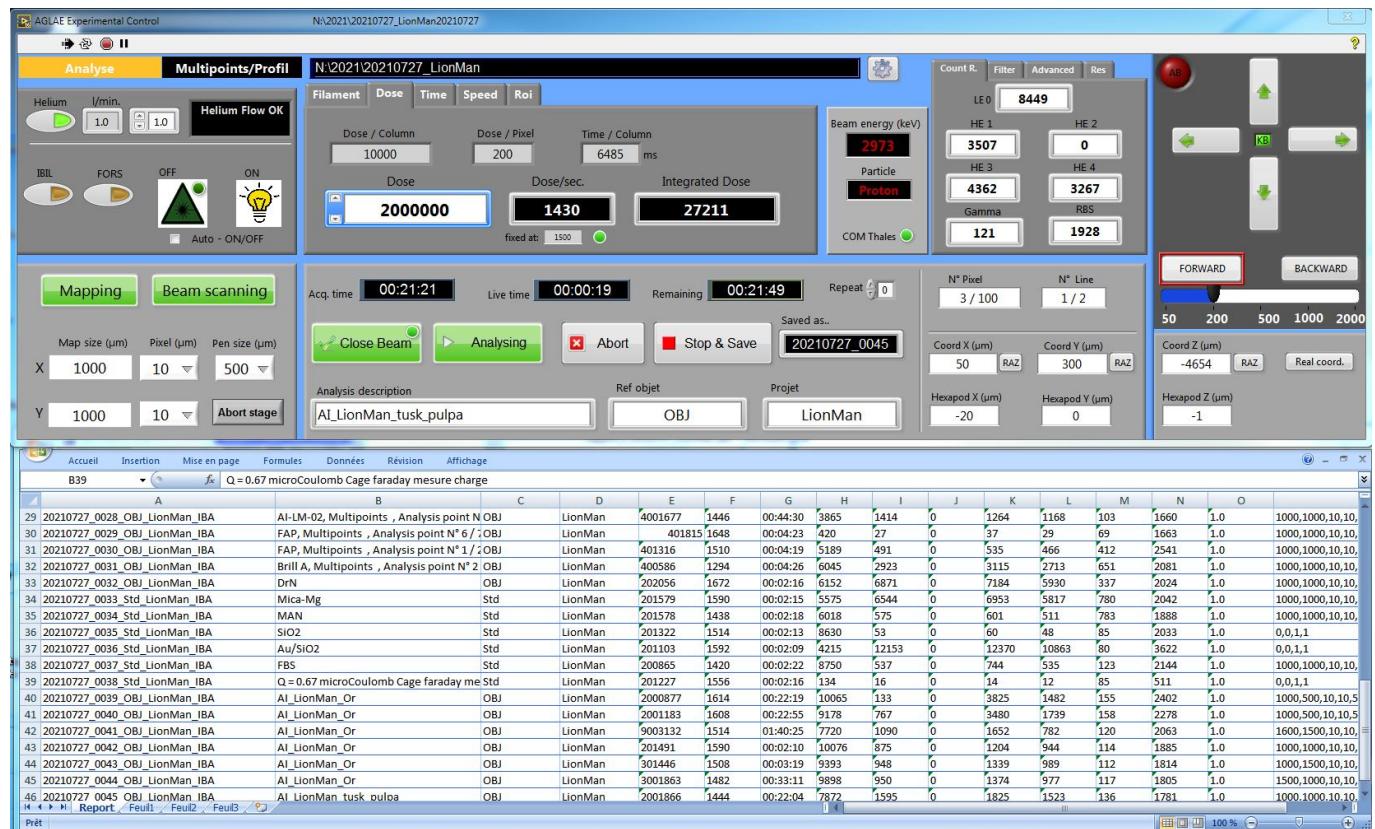


FIGURE B.4 – Capture d'écran de l'interface utilisateur d'AGLAE, avec les paramètres et le fichier Excel de sortie visibles.



FIGURE B.5 – Photographie d'une plaque contenant plusieurs standards, utilisés lors des expériences.

Annexe C

L'utilisation des logiciels d'AGLAE

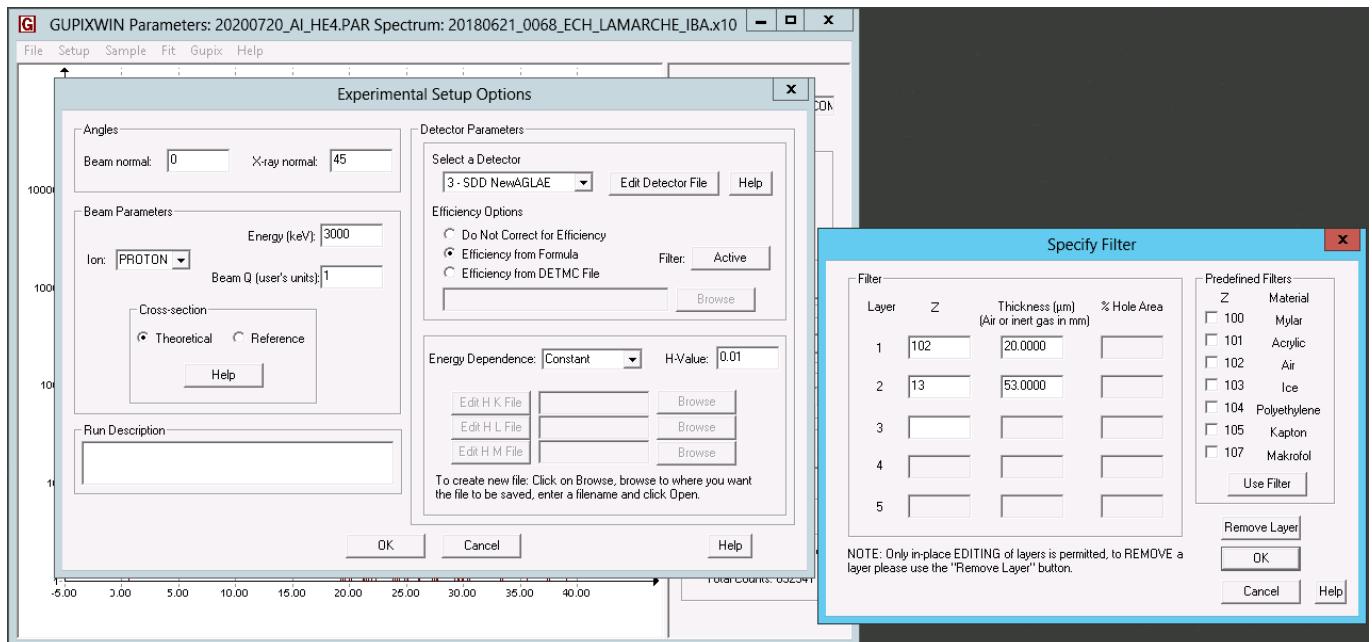


FIGURE C.1 – Capture d'écran de la définition des paramètres dans GUPIXWin (analyse PIXE).

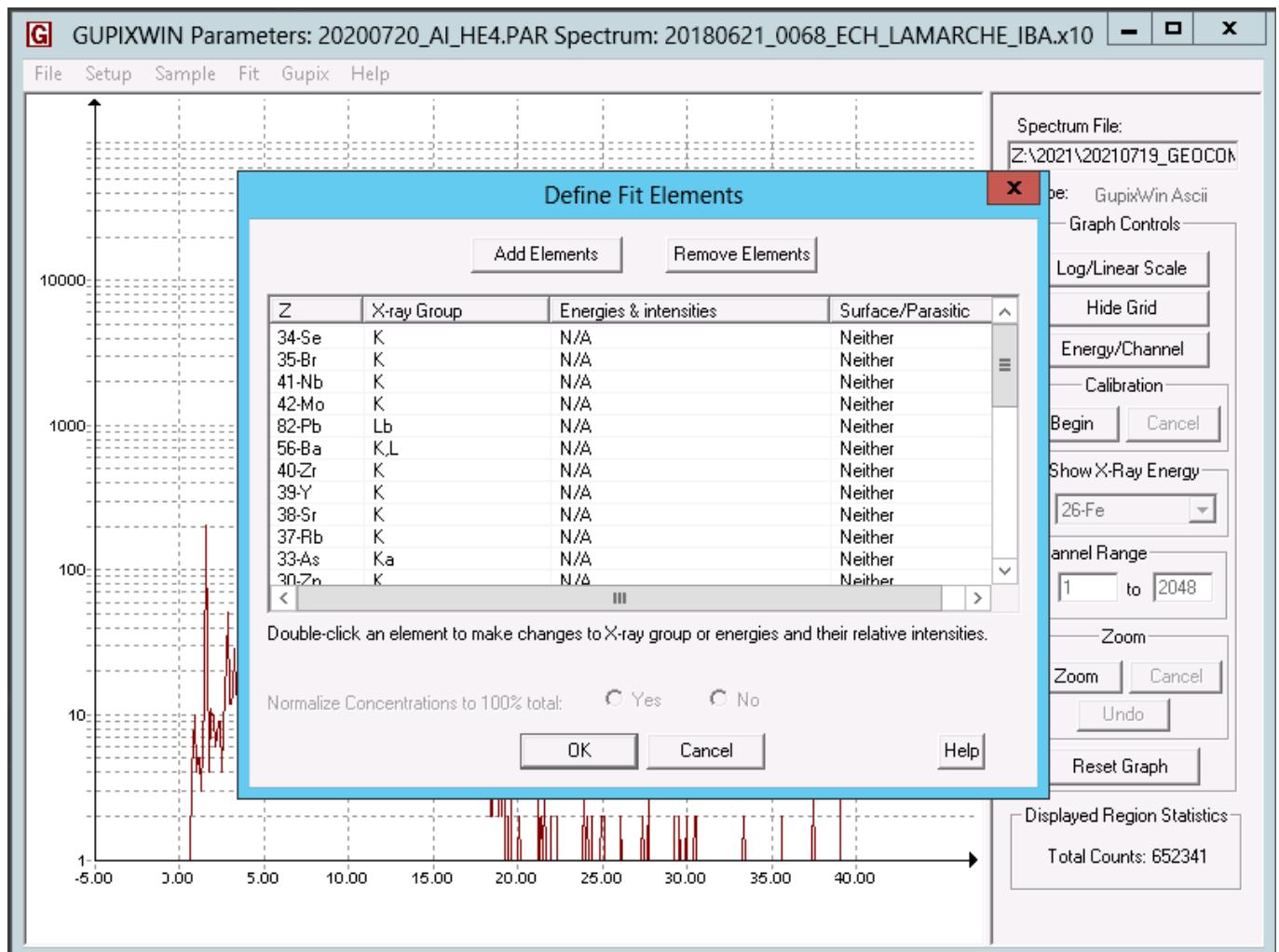


FIGURE C.2 – Capture d'écran de la définition des paramètres dans GUPIXWin (analyse PIXE).

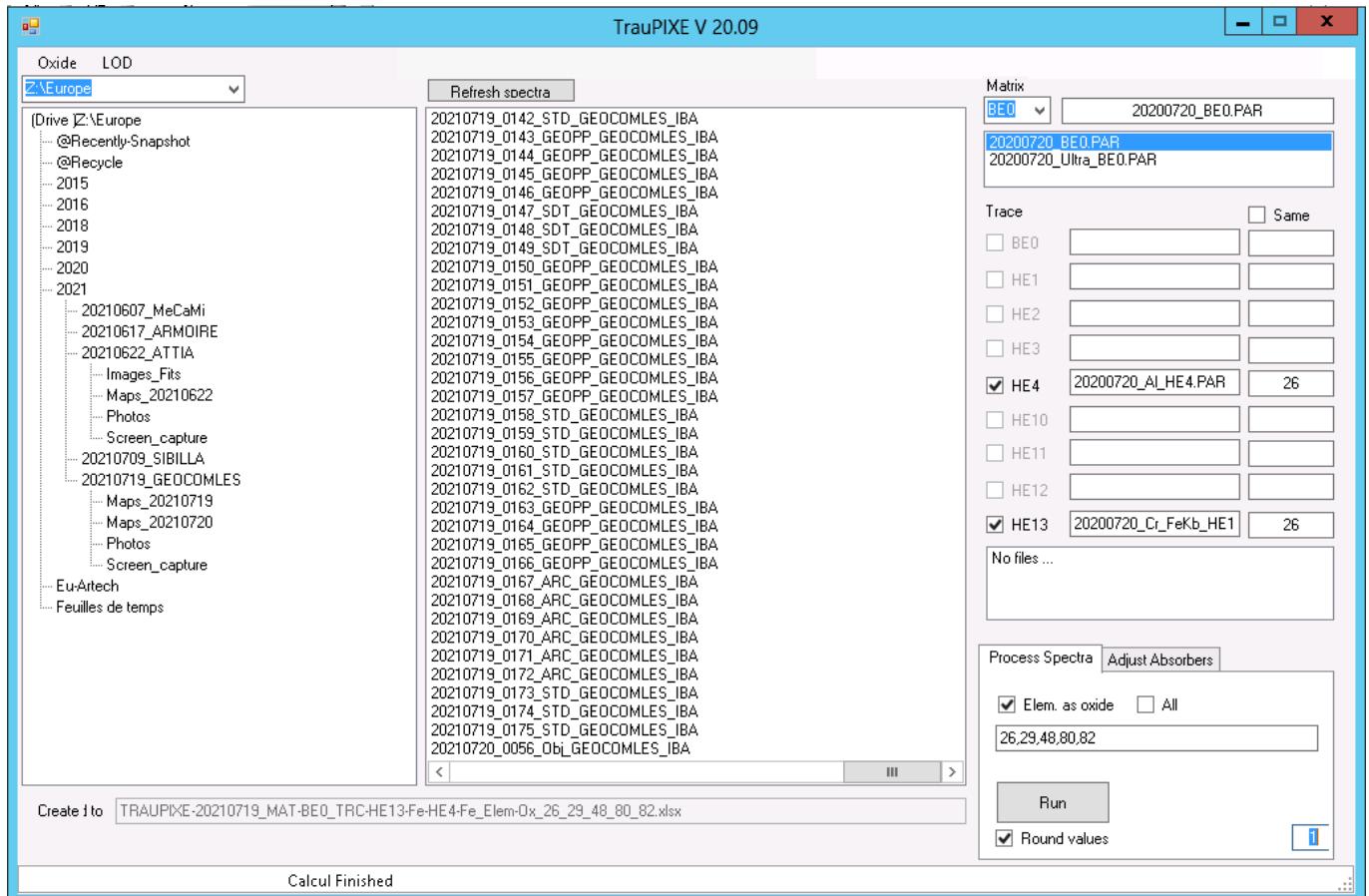


FIGURE C.3 – Capture d'écran de TRAUIXE lors d'une sélection de fichiers .par.

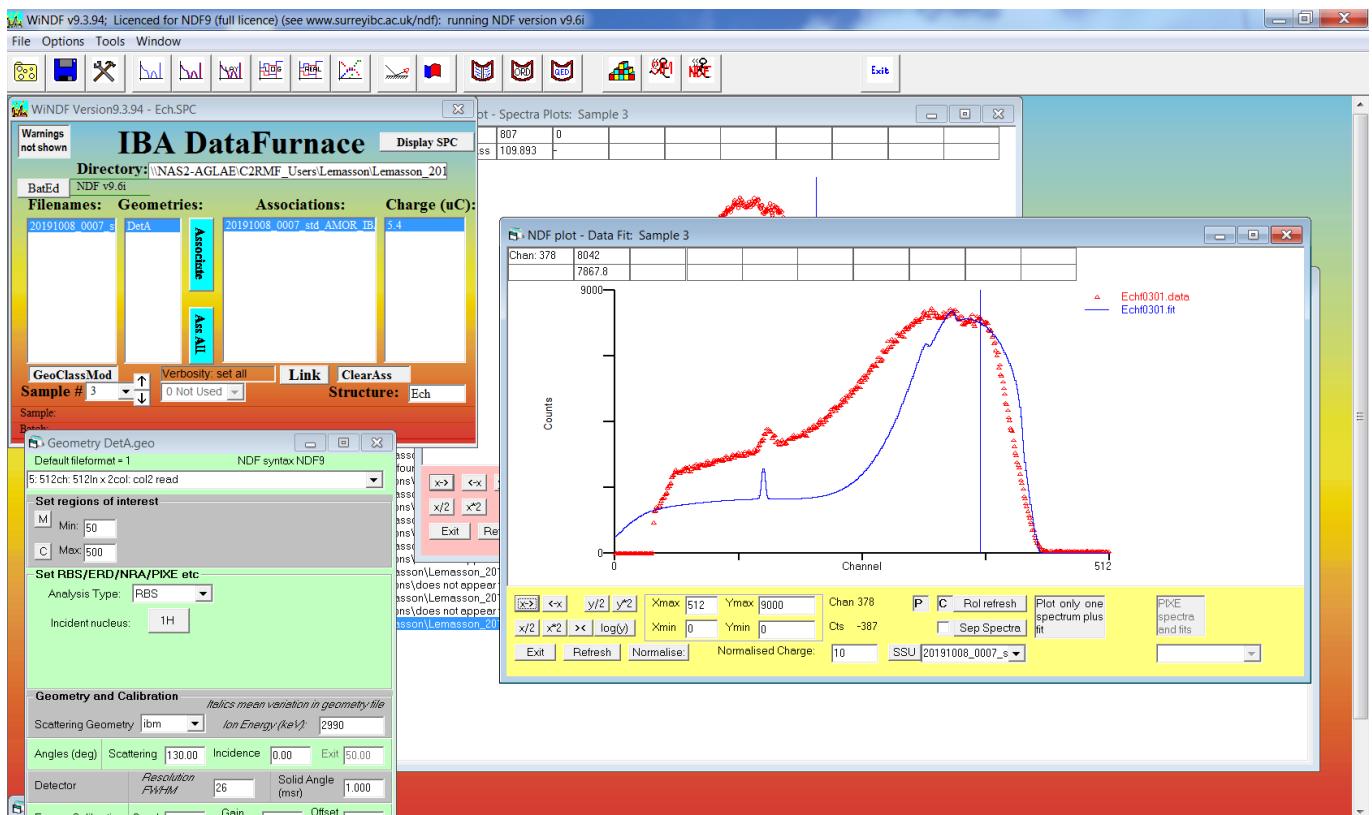


FIGURE C.4 – Capture d'écran de DataFurnace pour l'analyse RBS d'un matériau.

Annexe D

Pérennisation et diffusion

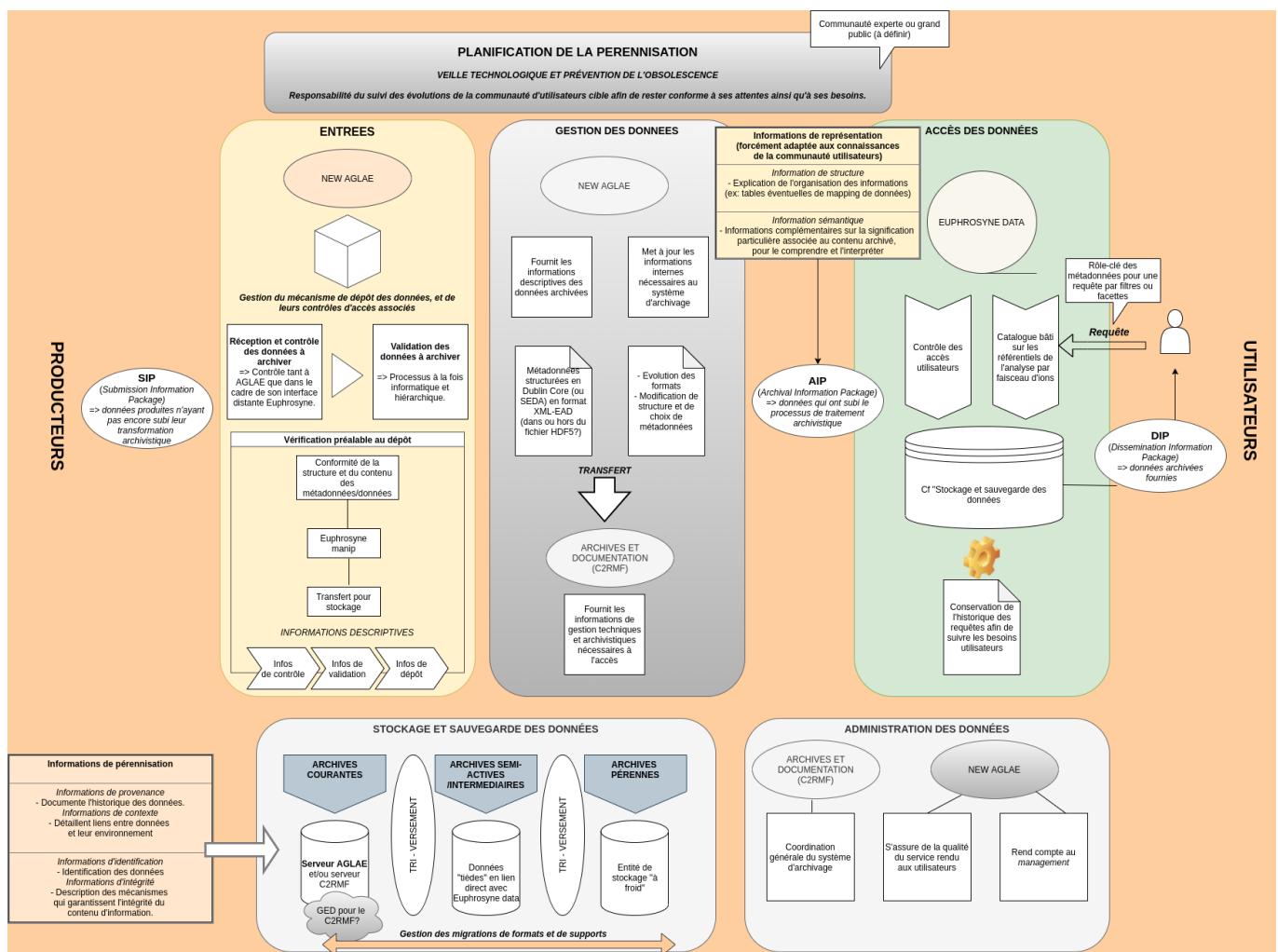


FIGURE D.1 – Présentation des acteurs et des fonctions envisagés pour l'archivage OAIS des données d'AGLAE.

RECOMMANDATIONS

RECOMMANDATION DE LA COMMISSION

du 17 juillet 2012

relative à l'accès aux informations scientifiques et à leur conservation

(2012/417/UE)

LA COMMISSION EUROPÉENNE,

vu le traité sur le fonctionnement de l'Union européenne, et notamment son article 292,

considérant ce qui suit:

- (1) Dans sa communication «Europe 2020 – une stratégie pour une croissance intelligente, durable et inclusive»⁽¹⁾, la Commission fait du développement d'une économie fondée sur la connaissance et l'innovation une priorité.
- (2) Les objectifs fixés par la stratégie Europe 2020 sont présentés plus en détail, notamment, dans les communications relatives aux initiatives phares «Une stratégie numérique pour l'Europe»⁽²⁾ et «Une Union de l'innovation»⁽³⁾. Parmi les actions à engager dans le cadre de la stratégie numérique, il y a lieu d'assurer à la recherche financée par des fonds publics une large diffusion par la publication de données et d'articles scientifiques en libre accès. L'initiative phare «Une Union de l'innovation» prône la mise en place d'un cadre pour l'Espace européen de la recherche (EER) afin d'éliminer les obstacles à la mobilité et à la coopération transnationale. Elle indique qu'il convient de promouvoir le libre accès aux publications et aux données de la recherche publique et de faire du libre accès aux publications un principe général applicable aux projets financés par les programmes-cadres de recherche de l'UE.
- (3) Le 14 février 2007, la Commission a adopté une communication «sur l'information scientifique à l'ère numérique: accès, diffusion et préservation»⁽⁴⁾, qui était accompagnée d'un document de travail des services de la Commission. Cette communication donnait un aperçu de

la situation en Europe en matière de publication scientifique et de conservation des résultats de recherche, et examinait les aspects organisationnels, juridiques, techniques et financiers pertinents.

(4) Elle a été suivie, en novembre 2007, par des conclusions du Conseil sur l'information scientifique à l'ère numérique: accès, diffusion et conservation. Dans ses conclusions, le Conseil invitait la Commission à expérimenter le libre accès aux publications scientifiques issues de projets financés par les programmes-cadres de recherche de l'UE et proposait une série d'actions à mener par les États membres. Des avancées ont été réalisées dans certains des domaines abordés dans les conclusions du Conseil, mais tous les objectifs n'ont pas été atteints et les progrès se sont révélés inégaux d'un État membre à l'autre. Une action de l'UE est nécessaire pour exploiter pleinement le potentiel de l'Europe en matière de recherche.

(5) Les politiques de libre accès visent à fournir aux lecteurs un accès gratuit, au stade le plus précoce du processus de diffusion, aux publications scientifiques évaluées par des pairs et aux données de la recherche, et à permettre l'utilisation et la réutilisation des résultats de la recherche scientifique. La mise en œuvre de ces politiques devrait tenir compte de la question des droits de propriété intellectuelle.

(6) Les politiques de libre accès aux résultats de la recherche scientifique devraient s'appliquer à toutes les activités de recherche financées par des fonds publics. Ces politiques sont censées améliorer les conditions dans lesquelles s'effectuent les activités de recherche en réduisant la duplication des efforts et en limitant autant que possible le temps passé à rechercher des informations et à y accéder, ce qui contribuera à accélérer le progrès scientifique et facilitera la coopération dans toute l'UE et au-delà. Ces politiques répondront également aux appels de la communauté scientifique à un meilleur accès aux informations scientifiques.

(7) Donner aux acteurs de la société les moyens d'interagir dans le cycle de la recherche permet d'accroître la qualité, la pertinence, l'acceptabilité et la durabilité des résultats de l'innovation en y intégrant les attentes, les besoins, les intérêts et les valeurs de la société. Le libre accès est un élément clé des politiques des États membres en faveur d'une recherche et d'une innovation responsables: il permet de mettre les résultats de recherche à la disposition de tous et facilite la participation de la société.

⁽¹⁾ COM(2010) 2020 final du 3.3.2010. Communication disponible à l'adresse suivante (<http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2010:2020:FIN:FR:PDF>).

⁽²⁾ COM(2010) 245 final/2 du 26.8.2010. Communication disponible à l'adresse suivante (<http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2010:0245:FIN:FR:PDF>).

⁽³⁾ COM(2010) 546 final du 6.10.2010. Communication disponible à l'adresse suivante (http://ec.europa.eu/research/innovation-union/pdf/innovation-union-communication_fr.pdf#view=fit&pagemode=none).

⁽⁴⁾ COM(2007) 56 final du 14.2.2007. Communication disponible à l'adresse suivante (<http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:52007DC0056:FR:NOT>).

(8) Un accès plus large aux résultats de la recherche scientifique profitera également aux entreprises, et notamment aux petites et moyennes entreprises, qui pourront améliorer leur capacité d'innovation. Les politiques d'accès aux informations scientifiques devraient donc également permettre aux entreprises privées d'accéder plus facilement à ce type d'informations.

(9) L'internet a bouleversé le monde scientifique et la recherche. Par exemple, les communautés de recherche ont expérimenté des nouveaux moyens de répertorier, de certifier, de diffuser et de conserver les publications scientifiques. Il faut adapter les politiques de recherche et de financement à ce nouvel environnement. Il convient de recommander aux États membres d'adapter et d'étoffer leurs politiques de libre accès aux publications scientifiques.

(10) Le libre accès aux données de la recherche scientifique améliore la qualité des données, réduit le besoin de duplication des efforts de recherche, accélère le progrès scientifique et contribue à la lutte contre la fraude scientifique. Dans son rapport final intitulé «Riding the wave: How Europe can gain from the rising tide of scientific data»⁽¹⁾ et publié en octobre 2010, le groupe d'experts à haut niveau sur les données scientifiques a souligné l'importance essentielle que revêtent l'échange et la conservation de données fiables issues du processus scientifique. Il est donc urgent de mener une action politique en matière d'accès aux données, et il y a lieu d'adresser une recommandation aux États membres à cet égard.

(11) Il est dans l'intérêt public de conserver les résultats de la recherche scientifique. Les bibliothèques, et plus particulièrement les bibliothèques nationales de dépôt légal, ont toujours été responsables de leur conservation. Le volume des résultats générés par la recherche augmente de manière spectaculaire. Il faudrait mettre en place des mécanismes, des infrastructures et des solutions logicielles pour permettre la conservation à long terme des résultats de recherche sous forme numérique. Les coûts liés à la curation des contenus numérisés étant toujours relativement élevés, il est fondamental de garantir le financement pérenne de la conservation. Étant donné l'importance de la conservation pour l'utilisation future des résultats de recherche, il y a lieu de recommander aux États membres d'établir des politiques en la matière ou de les renforcer.

(12) Les politiques à élaborer par les États membres devraient être définies aux niveaux national ou infranational, selon la situation constitutionnelle de chaque État et la répartition des responsabilités en matière de définition de sa politique de recherche.

(13) Des infrastructures électroniques robustes sous-tendant le système d'information scientifique permettront d'améliorer l'accès aux informations scientifiques et contribueront à leur conservation à long terme, ce qui peut dynamiser la recherche conjointe. Conformément à la communication de la Commission intitulée «Infrastruc-

tures TIC pour la science en ligne»⁽²⁾, une infrastructure électronique est «un environnement dans lequel les moyens de recherche (matériel, logiciels et contenu) peuvent être aisément partagés et utilisés chaque fois que c'est nécessaire pour obtenir des résultats plus probants». Il y a donc lieu de recommander le développement plus poussé de ce type d'infrastructures et leur interconnexion à l'échelle européenne.

(14) L'évolution vers le libre accès est une entreprise mondiale, comme en attestent le projet révisé de stratégie relatif à la contribution de l'Unesco à la promotion de l'accès libre à l'information et à la recherche scientifiques⁽³⁾, ainsi que la déclaration de l'OCDE sur l'accès aux données de la recherche financée par des fonds publics⁽⁴⁾. Les États membres devraient participer à cette entreprise et faire figure d'exemple en favorisant la mise en place d'un environnement de recherche ouvert et collaboratif fondé sur la réciprocité.

(15) Le secteur de l'édition se trouvant actuellement dans une phase de transition, les parties prenantes doivent s'associer pour prendre des mesures d'accompagnement appropriées et chercher des solutions durables pour le processus de publication scientifique.

(16) Le 12 décembre 2011, la Commission a adopté un ensemble de mesures constitué d'une communication sur l'ouverture des données, d'une proposition de directive modifiant la directive 2003/98/CE du Parlement européen et du Conseil du 17 novembre 2003 concernant la réutilisation des informations du secteur public⁽⁵⁾ et de nouvelles règles de la Commission sur les documents dont elle dispose. Cet ensemble de mesures présentait la stratégie de la Commission en matière d'ouverture des données dans un cadre unique et cohérent qui englobait des actions incluant, notamment, la présente recommandation.

(17) La présente recommandation est accompagnée d'une communication dans laquelle la Commission définit sa politique et présente sa vision en matière de libre accès aux résultats de recherche. Elle y expose les actions qu'elle entend mener en tant qu'organisme de financement de la recherche scientifique au titre du budget de l'Union.

(18) Parallèlement à la présente recommandation et à la communication qui l'accompagne, la Commission adopte également une communication intitulée «Un partenariat de l'Espace européen de la recherche renforcé pour promouvoir l'excellence et la croissance», dans laquelle elle définit les grandes priorités pourachever la réalisation de l'Espace européen de la recherche, l'une d'entre elles consistant à optimiser la diffusion et le transfert des connaissances scientifiques, ainsi que leur accès,

⁽²⁾ COM(2009) 108 final.

⁽³⁾ «Revised draft strategy on UNESCO's contribution to the promotion of open access to scientific information and research», disponible en anglais à l'adresse suivante (<http://www.unesco.org/new/fileadmin/MULTIMEDIA/HQ/CI/CI/images/GOAP/OAF2011/213342e.pdf>).

⁽⁴⁾ (<http://acts.oecd.org/Instruments>ShowInstrumentView.aspx?InstrumentID=157&Lang=fr&Book=False>).

⁽⁵⁾ JO L 345 du 31.12.2003, p. 90.

⁽¹⁾ Rapport disponible en anglais à l'adresse suivante (<http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/hlg-sdi-report.pdf>).

RECOMMANDÉ AUX ÉTATS MEMBRES:

Libre accès aux publications scientifiques

1. de définir des politiques claires en matière de diffusion des publications scientifiques issues de la recherche financée par des fonds publics et du libre accès à ces dernières. Ces politiques devraient prévoir:

- des objectifs et des indicateurs concrets permettant de mesurer les progrès accomplis,
- des plans de mise en œuvre, incluant la répartition des responsabilités,
- la programmation financière correspondante;

dans le prolongement de ces politiques, de veiller:

- à ce que les publications issues de la recherche financée par des fonds publics soient librement accessibles dans les meilleurs délais, de préférence immédiatement et, dans tous les cas, au plus tard six mois après leur date de publication, et au plus tard douze mois pour les publications dans les domaines des sciences sociales et humaines,
- à ce que les systèmes d'octroi de licences contribuent, de façon équilibrée, au libre accès aux publications scientifiques issues de la recherche financée par des fonds publics, dans le respect et sans préjudice de la législation applicable en matière de droit d'auteur, et encouragent les chercheurs à conserver leurs droits d'auteur tout en concédant des licences aux éditeurs,

- à ce que le système des carrières universitaires soutienne et récompense les chercheurs qui adhèrent à une culture de partage de leurs résultats de recherche, notamment en garantissant le libre accès à leurs publications et en élaborant, en encourageant et en utilisant de nouveaux modèles, critères et indicateurs alternatifs pour l'évaluation des carrières,

- à ce que la transparence soit améliorée, notamment en informant le grand public sur les accords conclus entre organismes publics, ou groupes d'organismes publics, et éditeurs pour la fourniture d'informations scientifiques. Cela comprend les accords portant sur les offres groupées, à savoir un prix réduit pour une formule d'abonnement aux versions papier et électronique des revues,

- à ce que les petites et moyennes entreprises et les chercheurs indépendants disposent de l'accès le plus large et le moins cher possible aux publications des résultats de la recherche financée par des fonds publics;

2. de veiller à ce que les organismes de financement de la recherche chargés de gérer le financement public de la recherche et les établissements universitaires bénéficiaires de financements publics mettent en œuvre les politiques:

- en définissant des politiques institutionnelles assurant le libre accès aux publications scientifiques et leur diffusion, et en élaborant des plans de mise en œuvre au niveau des organismes de financement,

— en mettant à disposition le financement nécessaire pour la diffusion (y compris le libre accès), en permettant des canaux différents, y compris, le cas échéant, des infrastructures électroniques et de nouvelles méthodes pilotes de communication scientifique,

— en adaptant les systèmes de recrutement et d'évaluation de carrière des chercheurs, ainsi que le système d'évaluation pour l'octroi de subventions de recherche, de manière à récompenser les chercheurs qui adhèrent à une culture de partage de leurs résultats de recherche. Les systèmes ainsi améliorés devraient tenir compte des résultats de recherche mis à disposition en libre accès et développer, encourager et utiliser de nouveaux modèles, critères et indicateurs alternatifs pour l'évaluation des carrières,

— en fournissant aux chercheurs des orientations sur la manière de se conformer aux politiques de libre accès et, en particulier, de gérer leurs droits de propriété intellectuelle pour assurer le libre accès à leurs publications,

— en menant des négociations conjointes avec les éditeurs pour obtenir les meilleures conditions possibles d'accès aux publications, y compris l'utilisation et la réutilisation,

— en faisant en sorte que les résultats de la recherche financée par des fonds publics soient facilement identifiables par des moyens techniques adaptés, y compris par l'utilisation de métadonnées associées aux versions électroniques des résultats de recherche;

Libre accès aux données de la recherche

3. de définir des politiques claires en matière de diffusion des données de la recherche financée par des fonds publics et de libre accès à ces dernières. Ces politiques devraient prévoir:

- des objectifs et des indicateurs concrets permettant de mesurer les progrès accomplis,

- des plans de mise en œuvre, incluant la répartition des responsabilités (y compris la concession des licences adéquates),

- la programmation financière correspondante;

dans le prolongement de ces politiques, de veiller:

— à ce que les données de la recherche financée par des fonds publics deviennent accessibles, utilisables et réutilisables par le public au moyen d'infrastructures électroniques. Il y a lieu de tenir dûment compte des questions relatives, notamment, au respect de la vie privée, aux secrets industriels, à la sûreté nationale, aux intérêts commerciaux légitimes et aux droits de propriété intellectuelle. Les données, le savoir-faire et/ou les informations, quelle que soit leur forme ou leur nature, qui sont détenus par des acteurs privés participant à un partenariat public-privé avant les activités de recherche et qui ont été identifiés comme tels ne sont pas soumis à ce type d'obligation,

- à ce que les ensembles de données soient facilement identifiables et puissent être liés à d'autres ensembles de données et publications au moyen de mécanismes adaptés, et à ce que des informations supplémentaires soient fournies afin de permettre une évaluation et une utilisation correctes,
- à ce que les organismes chargés de gérer le financement public de la recherche et les établissements universitaires bénéficiaires de financements publics contribuent à la mise en œuvre des politiques nationales en mettant en place des mécanismes permettant le partage de données de la recherche et le récompensant,
- à ce que les programmes d'études approfondies visant à former des nouveaux profils professionnels dans le domaine des technologies de traitement des données soient promus et/ou mis en œuvre;

Conservation et réutilisation des informations scientifiques

4. de renforcer la conservation des informations scientifiques:
 - en définissant et en mettant en œuvre des politiques, y compris la répartition des responsabilités en matière de conservation des informations scientifiques, ainsi que la programmation financière correspondante, afin de garantir la curation et la conservation à long terme des résultats de recherche (données de la recherche fondamentale et tous les autres résultats, y compris les publications),
 - en garantissant la mise en place d'un système efficace de dépôt des informations scientifiques au format électronique, couvrant les publications d'origine numérique et, le cas échéant, les ensembles de données correspondants,
 - en conservant le matériel informatique et les logiciels nécessaires pour lire les informations dans l'avenir, ou en faisant migrer régulièrement les informations vers de nouveaux environnements matériels et logiciels,
 - en créant les conditions permettant aux parties prenantes d'offrir des services à valeur ajoutée fondés sur la réutilisation d'informations scientifiques;

Infrastructures électroniques

5. de développer davantage les infrastructures électroniques sous-tendant le système de diffusion des informations scientifiques:
 - en soutenant les infrastructures de données scientifiques pour la diffusion des connaissances, les instituts de recherche et les organismes de financement pour couvrir toutes les étapes du cycle de vie des données. Ces étapes devraient comprendre l'acquisition, la curation, les métadonnées, la provenance, les identificateurs persistants, l'autorisation, l'authentification et l'intégrité des données. Il convient d'élaborer des approches pour permettre une compréhension commune de la découverte de données dans toutes les disciplines, ce qui permettra de réduire la courbe d'apprentissage requise pour atteindre les objectifs de productivité,
 - en soutenant la formation et le perfectionnement de nouvelles cohortes d'experts en sciences informatiques

- faisant un usage intensif de données, parmi lesquels des spécialistes des données, des techniciens et des gestionnaires de données,
- en mobilisant davantage les ressources existantes et en s'appuyant sur celles-ci pour garantir l'efficacité sur le plan économique et innover dans les domaines suivants: outils d'analyse, visualisations, aide à la prise de décision, modèles et outils de modélisation, simulations, nouveaux algorithmes et logiciels scientifiques,
- en renforçant les infrastructures pour l'accès aux informations scientifiques et leur conservation au niveau national, et en y affectant les ressources financières nécessaires,
- en garantissant la qualité et la fiabilité des infrastructures, y compris par l'utilisation de mécanismes de certification pour les archives,
- en garantissant l'interopérabilité des infrastructures électroniques aux niveaux national et mondial;

6. de créer des synergies, aux niveaux européen et mondial, entre les infrastructures électroniques nationales:
 - en contribuant à l'interopérabilité des infrastructures électroniques, notamment en ce qui concerne l'échange de données scientifiques, tout en tenant compte de l'expérience acquise avec les projets existants et les infrastructures et logiciels développés aux niveaux européen et mondial,
 - en encourageant les efforts de coopération transnationaux qui promeuvent l'utilisation et le développement de l'infrastructure des technologies de l'information et de la communication pour l'enseignement supérieur et la recherche;

Dialogue multilatéral aux niveaux national, européen et international

7. de prendre part à des dialogues établis entre les différentes parties prenantes aux niveaux national, européen et/ou international pour étudier la manière de promouvoir le libre accès aux informations scientifiques et leur conservation. Les participants devraient notamment examiner:
 - la possibilité de lier les publications aux données sous-jacentes,
 - la possibilité d'améliorer l'accès et d'assurer le contrôle des coûts, par exemple en menant des négociations conjointes avec les éditeurs,
 - de nouveaux indicateurs de la recherche et de nouveaux indicateurs bibliométriques englobant non seulement les publications scientifiques, mais aussi les ensembles de données et d'autres types de résultats issus des activités de recherche et les réalisations de chaque chercheur,
 - de nouveaux systèmes et structures de récompense,
 - la promotion des principes du libre accès et leur mise en œuvre à l'échelle internationale, particulièrement dans le cadre d'initiatives de coopération bilatérale, multilatérale et internationale;

Coordination structurée des États membres à l'échelle de l'UE et suivi de la recommandation

8. de désigner, avant la fin de l'année, un point de référence national qui aura pour mission:

- de coordonner les mesures énumérées dans la présente recommandation,
- de faire office d'interlocuteur avec la Commission européenne sur des questions relatives à l'accès aux informations scientifiques et à leur conservation, et notamment à l'amélioration des définitions des normes et principes communs, à la mise en œuvre des mesures et aux nouveaux modes de diffusion et de partage des résultats de recherche dans l'Espace européen de la recherche,
- d'établir des rapports sur le suivi de la présente recommandation;

Examen et communication d'informations

9. d'informer la Commission, dix-huit mois après la publication de la présente recommandation au *Journal officiel de l'Union européenne*, puis tous les deux ans, des mesures prises pour donner suite aux différents éléments de la présente recommandation, conformément aux modalités qui seront définies et adoptées. Sur la base des informations communiquées, la Commission examinera les progrès accomplis dans l'UE afin de déterminer si de nouvelles mesures s'imposent pour atteindre les objectifs énoncés dans la présente recommandation.

Fait à Bruxelles, le 17 juillet 2012.

Par la Commission

Neelie KROES

Vice-président

FIGURE D.2 – Recommandation de la Commission européenne relative à l'accès aux informations scientifiques et à leur conservation, 17 juillet 2012.

RECOMMANDATIONS

RECOMMANDATION (UE) 2018/790 DE LA COMMISSION

du 25 avril 2018

relative à l'accès aux informations scientifiques et à leur conservation

LA COMMISSION EUROPÉENNE,

vu le traité sur le fonctionnement de l'Union européenne, et notamment son article 292,

considérant ce qui suit:

- (1) En juillet 2012, la Commission européenne a adopté un paquet sur l'accès aux informations scientifiques composé de la communication intitulée «Pour un meilleur accès aux informations scientifiques: dynamiser les avantages des investissements publics dans le domaine de la recherche»⁽¹⁾ et de la recommandation 2012/417/UE de la Commission⁽²⁾. Dans sa recommandation 2012/417/UE, la Commission indique qu'elle examinera les progrès accomplis dans l'Union afin de déterminer si de nouvelles mesures s'imposent pour atteindre les objectifs fixés dans la recommandation.
- (2) La communication intitulée «Stratégie pour un marché unique numérique en Europe»⁽³⁾ souligne l'importance de la diffusion des données en tant que catalyseur de croissance économique, d'innovation et de conversion au numérique dans tous les secteurs économiques, en particulier pour les petites et moyennes entreprises (et les start ups) et pour la société dans son ensemble. Elle reconnaît que les mégadonnées et le calcul à haute performance font évoluer les pratiques dans le domaine de la recherche et du partage des connaissances, participant d'une transition vers une «science ouverte» plus performante et réactive⁽⁴⁾. Dans sa communication, la Commission annonce qu'elle encouragera l'accès aux données publiques pour stimuler l'innovation et qu'elle travaillera à la création d'un nuage pour la recherche consacré à la science ouverte dans le cadre de l'initiative européenne sur l'informatique en nuage. Dans son examen à mi-parcours de la mise en œuvre de la stratégie pour le marché unique numérique⁽⁵⁾, la Commission annonce son intention d'améliorer encore «l'accessibilité et la réutilisation des données du secteur public et des données obtenues au moyen de fonds publics».
- (3) Dans sa communication sur l'initiative européenne sur l'informatique en nuage «Bâtir une économie compétitive des données et de la connaissance en Europe»⁽⁶⁾, la Commission présente le plan général et rationnel visant à développer le nuage européen pour la science ouverte en tant qu'environnement fiable et ouvert permettant à la communauté scientifique de stocker, de partager et de réutiliser les données et résultats scientifiques. Elle y annonce également qu'elle procédera à un réexamen de la recommandation 2012/417/UE relative à l'accès aux informations scientifiques et à leur conservation pour encourager le partage des données scientifiques et la création de mécanismes d'incitation, de systèmes de récompenses et de programmes d'enseignement et de formation afin que chercheurs et entreprises partagent leurs données. Le document de travail des services de la Commission intitulé «Implementation Roadmap for the European Open Science Cloud»⁽⁷⁾ (Feuille de route pour la mise en œuvre du nuage européen pour la science ouverte) présente les résultats de l'examen, mené en coopération avec les États membres et les parties prenantes, des mécanismes de gouvernance et de financement envisageables pour le nuage européen pour la science ouverte et précise davantage les lignes d'actions pour développer ce nuage sur le modèle d'une fédération des infrastructures de données de la recherche.
- (4) La directive 2003/98/CE du Parlement européen et du Conseil⁽⁸⁾ établit le principe selon lequel toutes les données accessibles détenues par un organisme du secteur public doivent aussi pouvoir être réutilisées à des fins commerciales et non commerciales par toutes les parties intéressées, dans des conditions non discriminatoires pour des catégories comparables de réutilisation et à des prix qui n'excèdent pas les coûts marginaux de la diffusion des données.

⁽¹⁾ COM(2012) 401 final du 17 juillet 2012.

⁽²⁾ Recommandation 2012/417/UE de la Commission du 17 juillet 2012 relative à l'accès aux informations scientifiques et à leur conservation (JO L 194 du 21.7.2012, p. 39).

⁽³⁾ COM(2015) 192 final du 6 mai 2015.

⁽⁴⁾ La science ouverte représente une nouvelle approche du processus scientifique fondée sur le travail coopératif et les nouveaux modes de diffusion des connaissances, qui améliore l'accessibilité et la réutilisabilité des résultats de recherche en utilisant des technologies numériques et de nouveaux instruments de collaboration.

⁽⁵⁾ COM(2017) 228 final du 10 mai 2017.

⁽⁶⁾ COM(2016) 178 final du 19 avril 2016.

⁽⁷⁾ SWD(2018) 83 final du 14 mars 2018.

⁽⁸⁾ Directive 2003/98/CE du Parlement européen et du Conseil du 17 novembre 2003 concernant la réutilisation des informations du secteur public (JO L 345 du 31.12.2003, p. 90).

- (5) Les politiques de libre accès⁽¹⁾ visent à fournir aux chercheurs et au grand public un accès gratuit, de manière ouverte et non discriminatoire et au stade le plus précoce du processus de diffusion, aux publications scientifiques évaluées par des pairs, aux données de la recherche et à d'autres résultats de recherche, et à permettre l'utilisation et la réutilisation des résultats de recherches scientifiques. Le libre accès permet de renforcer la qualité, de réduire la nécessité de duplication inutile des efforts de recherche et d'accélérer le progrès scientifique, il contribue à la lutte contre la fraude scientifique et peut, de manière globale, favoriser la croissance économique et l'innovation. Outre le libre accès, la planification de la gestion des données devient une pratique scientifique courante.
- (6) Le libre accès est un moyen de diffusion pour les chercheurs qui décident de publier leurs travaux, notamment ceux menés dans le cadre de la recherche financée par des fonds publics. Les solutions en matière d'octroi de licences devraient viser à faciliter la diffusion et la réutilisation des publications scientifiques.
- (7) Il est dans l'intérêt public de conserver les résultats de recherches scientifiques. Les archives ou les bibliothèques, et plus particulièrement les bibliothèques nationales de dépôt légal, ont toujours été responsables de leur conservation. Le volume des résultats de recherche générés ne cesse d'augmenter. Il faudrait mettre en place des mécanismes, des infrastructures et des solutions logicielles pour permettre la conservation à long terme des résultats de recherche sous forme numérique. Les coûts liés à l'organisation des contenus numérisés étant toujours relativement élevés, il est fondamental de garantir le financement pérenne de la conservation. Étant donné l'importance que revêt la conservation pour l'utilisation future des résultats de recherche, il y a lieu de recommander aux États membres d'établir des politiques en la matière ou de les renforcer.
- (8) Le progrès technologique a permis aux gouvernements nationaux, aux universités ou aux organismes de recherche de mettre en place des infrastructures de recherche en ligne. En aidant les chercheurs à gérer les résultats de leurs travaux de recherche et en facilitant leur diffusion, ces infrastructures contribuent à la réalisation des objectifs de la présente recommandation. Dans sa communication relative à l'initiative européenne sur l'informatique en nuage, la Commission annonçait que «[l]e nuage européen pour la science ouverte commencera par fédérer les infrastructures de données scientifiques existantes, aujourd'hui dispersées entre les différents domaines et les différents États membres». Il convient de recenser et de recommander les mesures à l'échelle nationale qui devraient permettre le bon fonctionnement et la bonne utilisation du nuage européen pour la science ouverte.
- (9) Le progrès technologique a, au fil du temps, entraîné un changement radical dans le monde scientifique, en le poussant à adopter des méthodes de plus en plus collaboratives, et a constamment contribué à augmenter le volume d'informations scientifiques. Compte tenu du caractère de plus en plus collaboratif et transparent de l'approche scientifique, il convient de veiller à ce que les chercheurs aient accès, à tous les stades de leurs études et de leur carrière, au perfectionnement professionnel, notamment grâce à des programmes d'enseignement supérieur. Il faudrait également leur donner la possibilité d'acquérir les aptitudes nécessaires pour s'engager pleinement dans la science ouverte, comme indiqué dans la communication concernant le plan d'action en matière d'éducation numérique⁽²⁾.
- (10) Les incitations et les récompenses constituent des aspects importants d'une carrière professionnelle. Bien que les chercheurs soient encouragés à travailler dans d'autres pays ou dans différents secteurs et disciplines, et à adhérer à une culture de partage de leurs résultats, leurs efforts en la matière ne sont que rarement récompensés ou reflétés en termes d'évolution de carrière. Des indicateurs transparents et utilisés de manière responsable sont en cours d'élaboration pour soutenir la mise en œuvre des pratiques en matière de science ouverte dans les universités modernes. Des mécanismes de récompenses actualisés qui tiennent compte de métriques d'évaluation de nouvelle génération pourraient être utilisés pour mieux apprécier la qualité de la recherche européenne et inciter les chercheurs à partager les résultats de leurs recherches et les universités à développer une démarche entrepreneuriale tout en encourageant la concurrence dans le marché intérieur.
- (11) Les États membres devraient continuer à soutenir la science ouverte et le libre accès, comme indiqué dans les conclusions du Conseil relatives à une recherche ouverte, en réseau et à forte intensité de données, qui constitue le moteur d'une innovation plus rapide et plus large⁽³⁾ et dans les conclusions du Conseil sur la transition vers un système de science ouverte⁽⁴⁾.
- (12) L'évolution vers le libre accès est une démarche d'envergure mondiale. Les États membres ont participé à cette entreprise et devraient être soutenus dans leurs efforts pour favoriser la mise en place d'un environnement de recherche ouvert et collaboratif fondé sur la réciprocité au niveau mondial. La science ouverte est un élément clé des politiques des États membres pour une recherche responsable et une innovation ouverte. Il faudrait adapter les politiques de recherche et de financement à mesure que les nouvelles technologies numériques deviennent disponibles.

⁽¹⁾ Le libre accès désigne la possibilité d'accéder à des résultats de recherche numériques et de les réutiliser avec le moins de limitations possible.

⁽²⁾ COM(2018) 22 final.

⁽³⁾ Conclusions du Conseil 9360/15 du 29 mai 2015.

⁽⁴⁾ Conclusions du Conseil 9526/16 du 27 mai 2016.

- (13) La Commission a montré l'exemple en élargissant autant que possible l'accès aux résultats de recherche obtenus dans un environnement de science ouverte et leur réutilisation, y compris dans les programmes cadres, et en appliquant une politique d'ouverture des données aux données de la recherche du Centre commun de recherche de la Commission.
- (14) De nombreuses avancées ont été réalisées dans les domaines abordés dans la recommandation 2012/417/UE et dans les autres documents mentionnés dans les considérants précédents, mais tous les objectifs n'ont pas été atteints et les progrès se sont révélés inégaux d'un État membre à l'autre. Les États membres devraient tous déployer davantage d'efforts pour exploiter au mieux le potentiel de recherche et d'innovation de l'Europe.
- (15) La présente recommandation se fonde sur la recommandation 2012/417/UE et la remplace,

A ADOPTÉ LA PRÉSENTE RECOMMANDATION:

Libre accès aux publications scientifiques

1. Les États membres devraient définir et mettre en œuvre des politiques claires (telles qu'exposées dans les plans d'action nationaux) pour la diffusion des publications scientifiques issues de la recherche financée par des fonds publics et le libre accès à ces dernières. Ces politiques et plans d'action devraient établir:

- des objectifs et des indicateurs concrets permettant de mesurer les progrès accomplis,
- des plans de mise en œuvre, incluant la répartition des responsabilités et l'octroi de licences adéquates,
- la programmation financière correspondante.

Dans le prolongement de ces politiques ou plans d'action, les États membres devraient, conformément à l'acquis de l'Union européenne en matière de droit d'auteur et de droits voisins, veiller:

- à ce que toutes les publications scientifiques issues de la recherche financée par des fonds publics soient mises à disposition en libre accès à compter de 2020 au plus tard,
- quel que soit le canal de publication (revue scientifique, infrastructure numérique, canal multimédia, ou toute nouvelle méthode pilote de communication scientifique), à ce que le libre accès aux publications issues de la recherche financée par des fonds publics soit accordé aussi rapidement que possible, de préférence au moment de la publication et, en tout état de cause, au plus tard six mois après la date de publication (au plus tard douze mois pour les sciences sociales et humaines),
- compte tenu des évolutions technologiques, à ce que les conditions d'octroi de licences appliquées sur le marché ne restreignent pas indûment la fouille de textes et de données des publications issues de la recherche financée par des fonds publics, dans le respect et sans préjudice de la législation applicable en matière de droit d'auteur,
- à ce que les chercheurs, lorsqu'ils passent des accords contractuels avec des éditeurs de publications scientifiques, conservent les droits de propriété intellectuelle nécessaires, notamment, au respect des exigences des politiques de libre accès. Cela concerne notamment l'auto-archivage et la réutilisation (au moyen, par exemple, de la fouille de textes et de données),
- à ce que des informations soient publiées sur les accords passés entre organismes publics, ou groupes d'organismes publics, et éditeurs concernant la fourniture d'informations scientifiques, en vue de renforcer la transparence des marchés et une concurrence loyale, sans préjudice de la protection du savoir-faire et des informations commerciales (secrets industriels). Cela devrait inclure toutes sortes d'accords portant en particulier sur les offres dites «groupées» (à savoir un prix réduit pour une formule d'abonnement aux versions papier et électronique des revues) et les accords de compensation connexes visant à obtenir une réduction des redevances de publication liées au libre accès pour les consortiums,
- à ce que les entreprises innovantes, en particulier les petites et moyennes entreprises, les chercheurs indépendants (par exemple, les scientifiques amateurs), le secteur public, la presse et les citoyens dans leur ensemble disposent, de manière transparente et non discriminatoire, de l'accès le plus large possible aux publications scientifiques des résultats de la recherche financée par des fonds publics en vue de débloquer le potentiel d'innovation, de renforcer les moyens d'action du secteur public et d'informer les citoyens.

2. Les États membres devraient veiller à ce que les organismes de financement de la recherche chargés de gérer le financement public de la recherche et les établissements universitaires bénéficiaires de financements publics mettent en œuvre, au niveau national et de manière coordonnée, les politiques et plans d'action nationaux visés au point 1:

- en mettant en place des politiques assurant le libre accès aux publications scientifiques et leur diffusion, et en élaborant des plans de mise en œuvre,

- en prévoyant des exigences de libre accès à respecter pour l'octroi d'une convention de subvention ou de tout autre soutien financier à des projets de recherche, ainsi que des mécanismes de suivi du respect de ces exigences et des actions de suivi pour remédier aux cas de non-respect,
- en mettant à disposition le financement nécessaire pour la diffusion (y compris le libre accès et la réutilisation) de manière transparente et non discriminatoire, en permettant le recours à des canaux différents, y compris, le cas échéant, à des infrastructures numériques, et à de nouvelles méthodes pilotes de communication scientifique,
- en fournissant aux chercheurs des orientations sur la manière de se conformer aux politiques de libre accès, et en les encourageant à le faire, en particulier en ce qui concerne la gestion de leurs droits de propriété intellectuelle pour garantir le libre accès à leurs publications,
- en menant des négociations conjointes avec les éditeurs pour obtenir des conditions d'accès aux publications les plus transparentes et les meilleures possibles, y compris pour leur utilisation et leur réutilisation,
- en faisant en sorte que les publications issues de la recherche financée par des fonds publics soient facilement identifiables par des moyens techniques adaptés, y compris par l'utilisation de métadonnées associées aux versions électroniques des résultats de recherche et d'identificateurs persistants.

Gestion des données de la recherche, y compris libre accès à celles-ci

3. Les États membres devraient définir et mettre en œuvre des politiques claires (telles qu'exposées en détail dans les plans d'action nationaux) en matière de gestion des données de la recherche financée par des fonds publics, y compris de libre accès à celles-ci. Ces politiques et plans d'action devraient établir:

- des objectifs et des indicateurs concrets permettant de mesurer les progrès accomplis,
- des plans de mise en œuvre, incluant la répartition des responsabilités et l'octroi de licences adéquates,
- la programmation financière correspondante.

Dans le prolongement de ces politiques ou plans d'action, les États membres devraient veiller:

- à ce que la planification de la gestion des données devienne une pratique scientifique courante dès le début du processus de recherche lorsque des données sont générées ou collectées, notamment par l'obligation de prévoir des plans pour la gestion des données,
- à ce que les données de la recherche financée par des fonds publics deviennent et demeurent faciles à trouver, accessibles, interopérables et réutilisables (principes FAIR) dans un environnement sécurisé et fiable, par l'intermédiaire d'infrastructures numériques (y compris celles fédérées au sein du nuage européen pour la science ouverte, le cas échéant), sauf si cela se révèle impossible ou incompatible avec la poursuite de l'exploitation des résultats de recherche («aussi ouvert que possible, mais aussi fermé que nécessaire»). Il peut y avoir plusieurs raisons à cela, dont notamment le respect de la vie privée, les secrets industriels, la sûreté nationale, les intérêts commerciaux légitimes et les droits de propriété intellectuelle détenus par des tiers. Ces politiques ou plans d'action nationaux ne devraient pas avoir d'incidence sur les données, le savoir-faire et/ou les informations, quelle que soit leur forme ou leur nature, qui sont détenus par des acteurs privés participant à un partenariat public-privé avant les activités de recherche,
- compte tenu des évolutions technologiques [y compris en matière de données dynamiques (en temps réel)], à ce que les conditions d'octroi de licences appliquées sur le marché ne restreignent pas indûment la fouille de textes et de données pour les données issues de la recherche financée par des fonds publics, dans le respect et sans préjudice de la législation applicable en matière de droit d'auteur,
- à ce que les entreprises innovantes, en particulier les petites et moyennes entreprises, les chercheurs indépendants (par exemple, les scientifiques amateurs), le secteur public, la presse et les citoyens dans leur ensemble disposent, de manière transparente et non discriminatoire, de l'accès le plus large possible aux données de la recherche financée par des fonds publics en vue de débloquer le potentiel d'innovation, de renforcer les moyens d'action du secteur public et d'informer les citoyens.

4. Les États membres devraient veiller à ce que les organismes de financement de la recherche chargés de gérer le financement public de la recherche et les établissements universitaires bénéficiaires de financements publics mettent en œuvre, au niveau national et de manière coordonnée, les politiques et plans d'action nationaux visés au point 3:

- en mettant en place des politiques assurant la gestion des données de la recherche, et en élaborant des plans de mise en œuvre,

- en prévoyant, dans les conventions de subvention et autres mécanismes de soutien financier accordé à des projets de recherche, des obligations en matière de plans de gestion des données et l'introduction du principe de libre accès aux données de la recherche («aussi ouvert que possible, mais aussi fermé que nécessaire») pour les projets générant des données de la recherche, ainsi que des systèmes de suivi du respect de ces exigences et des actions de suivi pour remédier aux cas de non-respect,
- en mettant à disposition le financement nécessaire pour la gestion des données,
- en fournissant aux chercheurs des orientations sur la manière de se conformer aux politiques en matière de gestion des données de la recherche, et en les encourageant à le faire, en particulier en ce qui concerne le développement d'aptitudes adéquates en matière de planification de la gestion des données et d'infrastructures numériques qui favorisent l'accès aux données de la recherche et leur conservation,
- en veillant à ce que les ensembles de données soient facilement identifiables grâce à des identificateurs persistants et puissent être liés à d'autres ensembles de données et publications au moyen de mécanismes adaptés, et à ce que des informations supplémentaires soient fournies pour en permettre l'évaluation et l'utilisation correctes.

Conservation et réutilisation des informations scientifiques

5. Les États membres devraient définir et mettre en œuvre des politiques claires (telles qu'exposées dans les plans d'action nationaux) en matière de renforcement de la conservation et de la réutilisation des informations scientifiques (publications, ensembles de données et autres résultats de recherche). Ces politiques et plans d'action devraient établir:

- des objectifs et des indicateurs concrets permettant de mesurer les progrès accomplis,
- des plans de mise en œuvre, incluant la répartition des responsabilités et l'octroi de licences adéquates,
- la programmation financière correspondante.

Dans le prolongement de ces politiques ou plans d'action, les États membres devraient veiller:

- à ce que les établissements universitaires bénéficiaires de financements publics élaborent des politiques sur la conservation de leurs résultats scientifiques,
- à ce qu'un système efficace de dépôt des informations scientifiques au format électronique soit mis en place, qui couvre les publications d'origine numérique et les résultats de recherche correspondants,
- à ce que les informations scientifiques sélectionnées pour la conservation à long terme soient organisées de manière appropriée, et à ce que le matériel informatique et les logiciels nécessaires à la réutilisation de ces informations soient disponibles,
- à ce que l'identification unique (interconnexion entre les résultats de recherche, les chercheurs, leur rattachement, leurs bailleurs de fonds et les contributeurs) soit favorisée par toute une série d'identificateurs persistants, afin de permettre la repérabilité, la reproductibilité et la conservation à long terme des résultats de recherche,
- à ce que des systèmes et conditions d'octroi de licences lisibles par machine soient en place, compatibles avec les licences ouvertes déjà existantes, et autorisent la réutilisation d'informations scientifiques issues de la recherche financée par des fonds publics, dans le respect et sans préjudice de la législation applicable en matière de droit d'auteur, afin de permettre la réutilisation et la conservation en toute légalité,
- à ce que les conditions permettant aux parties prenantes d'offrir des services à valeur ajoutée fondés sur la réutilisation d'informations scientifiques soient créées.

Infrastructures pour la science ouverte

6. Les États membres devraient définir et mettre en œuvre des politiques claires (telles qu'exposées dans les plans d'action nationaux) visant, d'une part, à développer davantage les infrastructures sous-tendant le système permettant d'accéder aux informations scientifiques, de les conserver, de les partager et de les réutiliser et, d'autre part, à promouvoir leur fédération au sein du nuage européen pour la science ouverte. Ces politiques et plans d'action devraient établir:

- des objectifs et des indicateurs concrets permettant de mesurer les progrès accomplis,
- des plans de mise en œuvre, incluant la répartition des responsabilités et l'octroi de licences adéquates,
- la programmation financière correspondante.

Dans le prolongement de ces politiques ou plans d'action nationaux, les États membres devraient veiller:

- à ce que les ressources soient affectées, mobilisées et créées de manière à garantir l'efficacité sur le plan économique et à innover tout en encourageant la concurrence dans le marché intérieur,

- à ce que la qualité et la fiabilité des infrastructures soient garanties, notamment par l'utilisation de mécanismes, de spécifications et de normes de certification largement reconnus,
 - à ce que les chercheurs disposent d'un meilleur accès, de manière transparente et non discriminatoire, aux moyens et aux services de recherche permettant de stocker, gérer, analyser, partager et réutiliser des informations scientifiques, y compris par l'intermédiaire du nuage européen pour la science ouverte, lorsqu'il sera disponible,
 - grâce à l'utilisation d'indicateurs et de métriques d'évaluation supplémentaires, à ce que les infrastructures soient adaptées à la collecte d'informations sous-tendant les systèmes de suivi et d'évaluation du caractère ouvert et de la science ouverte, ainsi que les systèmes d'évaluation de la recherche et de carrière.
7. Les États membres devraient veiller à créer des synergies entre les infrastructures nationales, avec le nuage européen pour la science ouverte et avec d'autres initiatives mondiales:
- en participant à la définition de normes pour les données et les services accessibles via le nuage européen pour la science ouverte, ainsi que d'indicateurs et de métriques d'évaluation permettant de mesurer l'incidence de la recherche dans le cadre du nuage européen pour la science,
 - en garantissant l'interopérabilité des infrastructures récemment mises au point ou modernisées afin qu'elles tiennent compte de la création du nuage européen pour la science et, partant, qu'elles empêchent l'apparition de cloisonnements, contribuant ainsi à la réduction de la fragmentation et à la promotion de la découverte et de la collaboration scientifiques au-delà des limites disciplinaires et géographiques,
 - en préparant le terrain pour l'utilisation de services et le partage d'informations scientifiques via le nuage européen pour la science ouverte.

Aptitudes et compétences

8. Les États membres devraient définir et mettre en œuvre des politiques claires (telles qu'exposées dans les plans d'action nationaux) en matière d'aptitudes et de compétences requises des chercheurs et du personnel des établissements universitaires en ce qui concerne les informations scientifiques. Ces politiques et plans d'action devraient établir:
- des objectifs et des indicateurs concrets permettant de mesurer les progrès accomplis,
 - des plans de mise en œuvre, incluant la répartition des responsabilités,
 - la programmation financière correspondante.

Dans le prolongement de ces politiques ou plans d'action nationaux, les États membres devraient veiller:

- à ce que la formation et l'enseignement requis soient dispensés dans les domaines du libre accès, de la gestion des données de la recherche, du bon usage des données, de la conservation des données, de l'organisation des données et de la science ouverte, dans le cadre du système d'enseignement supérieur et de formation, à tous les niveaux de carrière, et à ce qu'ils aboutissent à des pratiques exemplaires sur le lieu de travail dans le secteur,
- à ce que la promotion ou la mise en œuvre, ou les deux, des programmes d'études approfondies visant à former des nouveaux profils professionnels dans le domaine des technologies de traitement des données soient prévues,
- à ce que la formation et le perfectionnement d'experts en sciences informatiques faisant un usage intensif de données soient soutenus, y compris pour les spécialistes des données, les techniciens et les gestionnaires de données.

Incitations et récompenses

9. Les États membres devraient définir et mettre en œuvre des politiques claires (telles qu'exposées dans les plans d'action nationaux) pour adapter, en ce qui concerne les informations scientifiques, les systèmes de recrutement et d'évaluation de carrière des chercheurs, le système d'évaluation pour l'octroi de subventions de recherche aux chercheurs, et les systèmes d'évaluation pour les établissements actifs dans la recherche. Ces politiques et plans d'action devraient établir:
- des objectifs et des indicateurs concrets permettant de mesurer les progrès accomplis,
 - des plans de mise en œuvre, incluant la répartition des responsabilités,
 - la programmation financière correspondante.

Dans le prolongement de ces politiques ou plans d'action, les États membres devraient veiller:

- à ce que le système des carrières universitaires soutienne et récompense les chercheurs qui adhèrent à une culture de partage de leurs résultats de recherche, notamment en garantissant la diffusion précoce et le libre accès à leurs publications et autres résultats de recherche,

- à ce que les organismes chargés de gérer le financement public de la recherche et les établissements universitaires bénéficiaires de financements publics contribuent à la mise en œuvre des politiques nationales en mettant en place des mécanismes permettant le partage d'informations scientifiques, le mesurant et le récompensant,
- à ce que les systèmes d'évaluation de la recherche et de carrière soient enrichis par l'introduction d'indicateurs et de métriques d'évaluation supplémentaires permettant d'éclairer l'évaluation du caractère ouvert, y compris, mais pas seulement, de l'incidence sociale plus large de la recherche et, au niveau individuel, du chercheur («métriques d'évaluation de nouvelle génération»).

Dialogue multilatéral sur la science ouverte aux niveaux national, européen et international

10. Les États membres devraient prendre part à des dialogues sur la transition vers la science ouverte établis entre les différentes parties prenantes aux niveaux national, européen et international sur les questions abordées aux points 1 à 9.

Les États membres devraient veiller:

- à ce que ces dialogues renforcent un environnement technologique consacré à la science ouverte qui couvre tous les résultats de recherche à tous les stades du cycle de la recherche (données, publications, logiciels, méthodes, protocoles, etc.),
- à ce qu'un changement systémique en faveur de la science ouverte soit progressivement obtenu et inclue, au-delà du changement et de l'efficacité technologiques, le principe de réciprocité, un changement de culture parmi les chercheurs, ainsi qu'un changement en matière de recherche au sein des établissements universitaires et parmi les bailleurs de fonds en faveur de la science ouverte, y compris des questions telles que l'intégrité et l'éthique dans la recherche, le cas échéant.

Coordination structurée des États membres à l'échelle de l'Union et suivi de la présente recommandation

11. Les États membres devraient avoir un point de référence national qui aura pour mission:

- de coordonner les mesures énumérées dans la présente recommandation,
- de faire office d'interlocuteur avec la Commission sur des questions relatives à l'accès aux informations scientifiques et à leur conservation, et notamment à l'amélioration des définitions des normes et principes communs, à la mise en œuvre des mesures et aux nouveaux modes de diffusion et de partage des résultats de recherche dans l'Espace européen de la recherche,
- d'établir des rapports sur le suivi de la présente recommandation.

12. Les États membres devraient informer la Commission dix-huit mois après la publication de la présente recommandation au *Journal officiel de l'Union européenne*, puis tous les deux ans, des mesures prises pour donner suite aux éléments de la présente recommandation. Sur la base des informations communiquées, la Commission devrait examiner les progrès accomplis dans l'Union afin de déterminer si de nouvelles mesures s'imposent pour atteindre les objectifs proposés dans la présente recommandation.

Fait à Bruxelles, le 25 avril 2018.

Par la Commission

Mariya GABRIEL
Membre de la Commission

Carlos MOEDAS
Membre de la Commission

FIGURE D.3 – Recommandation de la Commission européenne relative à l'accès aux informations scientifiques et à leur conservation, 25 avril 2018.

Table des figures

2.1	Cartographie de 6 x 3 cm² réalisée sur un vitrail de la cathédrale de Chartres du XII-XIII^e siècles. @Claire Pacheco, Quentin Lemasson, <i>et al.</i> , « D'AGLAE à New AGLAE...	13
2.2	Exemple de traitement de données réalisable par le logiciel AGLAE-Map. @Claire Pacheco, Quentin Lemasson, <i>et al.</i> , « D'AGLAE à New AGLAE...	15
2.3	Cartographie de la raie Ka du cuivre obtenue sur le spectre somme PIXE de haute énergie. Le contour d'une lettre se distingue et révèle la lettre Q. @Claire Pacheco, Quentin Lemasson, <i>et al.</i> , « D'AGLAE à New AGLAE...	16
3.1	Capture d'écran d'un extrait de la présentation du projet PAR-COURS, avril 2021. @Luc Bouiller-C2RMF	22
3.2	Capture d'écran de la présentation introductory « DIGILAB and JRA joint meeting », 21 avril 2021. @Sorin Hermon/Vincent Detalle/Victor Etgens	25
4.1	Illustration extraite de la présentation d'audience d'Euphrosyne pour l'obtention du statut de start-up d'État du Ministère de la Culture, 8 avril 2021. @Ned Baldessin/Jonathan Bourguignon/Claire Pacheco	35
5.1	Illustration extraite de la présentation d'audience d'Euphrosyne pour l'obtention du statut de start-up d'État du Ministère de la Culture, 8 avril 2021. @Ned Baldessin/Jonathan Bourguignon/Claire Pacheco	43
7.1	Présentation fonctionnelle du programme VITAM. @CINES	65
7.2	Présentation du processus de versement au sein de la PAC (Plateforme d'archivage au CINES) @CINES	66
7.3	Présentation des six fonctions principales du modèle OAIS, orienté vers les besoins utilisateurs. @CINES	67

7.4 Présentation des acteurs principaux du modèle OAIS, de la production à la diffusion des « paquets d'information ». @CINES	69
11.1 Capture d'écran d'un extrait de la présentation du projet PAR-COURS, avril 2021. @Luc Bouiller-C2RMF	99
11.2 Capture d'écran de la modélisation sémantique pour <i>Raphael Research Resource</i>, juillet 2021. @National Gallery	101
11.3 Capture d'écran de la modélisation sémantique pour <i>Raphael Research Resource</i>, juillet 2021. @National Gallery	101
11.4 Capture d'écran de la modélisation sémantique réalisée pour une expérience à New AGLAE. @Pauline Breton-Chauvet-New AGLAE . .	102
A.1 Questionnaire à destination des membres de l'équipe d'AGLAE et du SAED, avril 2021.	122
A.2 Capture d'écran d'un exemple d'arborescence - répertoire AGLAE	122
A.3 Cartographie générale des flux de données produits au sein d'AGLAE. @New AGLAE	132
B.1 Template d'une description de projet pour une demande de temps de faisceau dans le cadre d'IPERION-HS.	135
B.2 Extrait d'un PROPOSAL	143
B.3 Extrait d'un PROPOSAL-MOLAB	147
B.4 Capture d'écran de l'interface utilisateur d'AGLAE, avec les paramètres et le fichier Excel de sortie visibles.	147
B.5 Photographie d'une plaque contenant plusieurs standards, utilisés lors des expériences.	148
C.1 Capture d'écran de la définition des paramètres dans GUPIXWin (analyse PIXE).	149
C.2 Capture d'écran de la définition des paramètres dans GUPIXWin (analyse PIXE).	150
C.3 Capture d'écran de TRAUIPIXE lors d'une sélection de fichiers .par.	151
C.4 Capture d'écran de DataFurnace pour l'analyse RBS d'un matériau.	151
D.1 Présentation des acteurs et des fonctions envisagés pour l'archivage OAIS des données d'AGLAE.	153
D.2 Recommandation de la CE, 17 juillet 2012	159
D.3 Recommandation de la CE, 25 avril 2018.	167

Table des matières

Résumé	iii
Remerciements	v
Bibliographie	vii
Liste des acronymes	xiii
Introduction	xv
I De la donnée à la recomposition de l'objet patrimonial	1
1 Augmentation ou prolongement de l'objet patrimonial ?	5
1.1 Nature et périmètre d'un « jeu de données »	5
1.2 Enrichissement périphérique ou intrinsèque à l'objet ?	7
2 Lentilles de contemplation de l'inaccessible	9
2.1 Lever le voile de l'invisible	10
2.2 L'ingénierie informatique au service du nécessaire entrecroisement des techniques d'analyse	14
3 Un cadre d'institutionnalisation multiscalaire	19
3.1 ESPADON : un EquipEX en quête de l'« objet patrimonial augmenté » . .	19
3.2 DIGILAB : une infrastructure numérique européenne à construire	23
II Euphrosyne : une communauté d'utilisateurs devenue matrice de patrimonialisation	27
4 De la <i>collectivité</i> à la <i>communauté</i> AGLAE	31
4.1 Un incubateur d'une collectivité singulière	31
4.2 De New AGLAE à Euphrosyne	33

5 Reconnaissance et appropriation	39
5.1 Un processus d'appropriation...	39
5.2 Euphrosyne : creuset d'une symbiose	42
6 Construire une sémiologie préventive	45
6.1 De l'arborescence numérique à la subsidiarité sémantique	45
6.2 Une circulation sémantique fluctuante	50
III La conservation et la pérennisation des données ou le mythe de Sisyphe	57
7 L'archivage OAIS des données d'AGLAE	61
7.1 De l'émergence d'un besoin au choix d'un modèle	61
7.2 Pertinence et tropisme du modèle de référence OAIS	66
8 Ciseler pour conserver	77
8.1 L'esquisse d'un plan de gestion des données pour le C2RMF	78
8.2 Le tri des données d'AGLAE : un dilemme à la Corneille	79
9 Nécessité et utilité d'un inachèvement perpétuel	83
9.1 Du signe à la trace : une sémiologie nourrie du mouvement	83
9.2 Conscience et rematérialisation perpétuelle	85
IV Euphrosyne-data : exposer et disséminer pour l'<i>Open Science</i>	87
10 Le catalogue : sacrement patrimonial des données	91
10.1 Bâtir le catalogue d'Euphrosyne-data	91
10.2 Le défi du signalement des contenus numériques d'AGLAE	93
11 Définir le référentiel de l'analyse par faisceau d'ions	95
11.1 Créer une entente sémiologique et sémantique	95
11.2 Organiser : la réalisation du modèle sémantique	98
12 L'<i>Open Science</i> : du carcan prescriptif au vecteur de moyens pour AGLAE	105
12.1 Une lente et difficile ouverture	105
12.2 De l'agrégation à la dissémination	109
13 Conclusion	111

Annexes	119
A Inventaire de l'existant	119
B Contexte expérimental-AGLAE	133
C L'utilisation des logiciels d'AGLAE	149
D Pérennisation et diffusion	153
Table des figures	169
Table des matières	171