OUVRIR LA SCIENCE !

Rapport du Groupe de Travail sur les Cahiers de Laboratoire électroniques

Réalisé par les membres du GT

Septembre 2021

Table des matières

1 Contexte et objectifs du document **4**
1.1 Contexte . . . 4
1.2 Mission du groupe de travail . . . 4
1.3 Constitution du groupe . . . 5
1.3.1 Pilotes . . . 6
1.3.2 Membres . . . 6
1.4 Fonctionnement du GT . . . 6
1.5 Finalité du rapport . . . 7

2 Définition et périmètre : une vision partagée du cahier de laboratoire électronique **8**
2.1 Définition . . . 8
2.2 Périmètre . . . 8
2.2.1 Personnes concernées . . . 8
2.2.2 Usages . . . 9
2.2.3 Outils connexes . . . 10

3 Recommandations relatives à l'interopérabilité entre différents outils **12**
3.1 Interopérabilité . . . 12
3.2 Recommandations . . . 12
3.2.1 Co-tutelles . . . 12
3.2.2 Intégration . . . 13
3.2.3 Accès aux données . . . 13
3.2.4 Sauvegarde et archivage . . . 13
3.3 Résumé . . . 14

4 Recommandations sur la procédure de choix d'un outil **15**
4.1 Besoins . . . 15
4.1.1 Recueillir les souhaits des différents acteurs . . . 15
4.1.2 Prendre en compte la discipline et les modalités de pratique de la recherche . . . 15
4.1.3 Réfléchir au niveau de sécurité attendu . . . 17
4.1.4 Réfléchir au niveau d'implémentation de l'outil . . . 18
4.1.5 Identifier les systèmes qui seront en interaction avec l'ELN . . . 18
4.2 Établir les critères de choix . . . 19
4.2.1 Caractéristiques techniques et fonctionnelles . . . 19
4.2.2 Autres critères . . . 22
4.3 Identifier les offres du marché . . . 23

4.4 Évaluer les offres selon les critères établis 25
4.4.1 Notation 25
4.4.2 Pondération 26
4.4.3 Présentation des résultats 27
4.4.4 Importance relative des différents critères 29

5 Liste comparative d'outils **31**
5.1 Choix des outils étudiés 31
5.2 Résultats 33
5.3 Commentaires 33

6 Mener à bien un projet d'implantation d'un ELN **36**
6.1 Introduction 36
6.2 Parties prenantes et pilotage du projet 36
6.3 Avant-projet 36
6.4 Phase pilote 37
6.5 Appel d'offres 37
6.6 Déploiement 38
6.7 Changement d'outils 38
6.8 Accompagnement du changement 38
6.9 Profils et dimensionnement en RH 40
6.10 Situation de co-tutelle 41

7 Conclusion **42**

A Annexes **43**
A.1 Contributeurs hors GT 43
A.2 Lettre de mission du GT 43
A.3 Vision détaillés des usages - approche observationnelle 48
A.3.1 Démarche ethnographique et analyse « qualitative » (sociologie, anthropologie, design, recherche participative, etc.) . . 48
A.3.2 Collecte et enregistrement de données dans des environnement contraints (archéologie, sciences de l'environnement) . 49
A.4 Description des critères 51
A.5 Liste des outils analysés et non analysés 59

B Glossaire et abréviations **63**

1 Contexte et objectifs du document

1.1 Contexte

Le cahier de laboratoire est un journal de bord utilisé pour détailler au quotidien les activités liées aux projets de recherche. Il assure le suivi du descriptif des expériences et des protocoles utilisés ainsi que les contributions de chacune et chacun. Son utilisation permet de répondre aux exigences de la démarche qualité et de l'intégrité scientifique en garantissant la traçabilité de la méthode scientifique, la reproductibilité des données et des résultats de recherche. Il permet, ainsi devant les offices de brevets, d'établir la preuve de l'invention, de ses inventrices ou inventeurs et donc de ses ayants-droit.

Le cahier de laboratoire électronique est la version numérique dématérialisée du cahier de laboratoire. Le terme « électronique », souvent remplacé par « numérique » pour d'autres domaines est l'adjectif le plus fréquemment employé pour les cahiers de laboratoires.

De par sa position au cœur de l'activité de recherche et son rôle dans la gestion et la protection de la connaissance scientifique, le cahier de laboratoire électronique est un outil stratégique essentiel qui s'inscrit pleinement dans une démarche de science ouverte. En outre, il doit répondre aux mêmes objectifs et enjeux scientifiques que le cahier papier, c'est pourquoi le choix de l'outil, son paramétrage et son utilisation, sont ainsi d'une importance capitale.

Afin d'éclairer les choix qui sont faits par les équipes scientifiques parmi les différentes solutions actuellement disponibles, un groupe de travail a été créé dans le cadre collège des données de la recherche du comité pour la science ouverte.

Terminologie : Dans la suite de ce document ainsi que dans l'ensemble des travaux produits par le groupe de travail, le cahier de laboratoire électronique est désigné par l'acronyme ELN (*Electronic Lab Notebook*). Le choix de l'anglais a été proposé afin d'éviter toute confusion, l'acronyme CLE étant déjà utilisé pour désigner un déploiement spécifique de cahier de laboratoire électronique.

1.2 Mission du groupe de travail

La lettre de mission du groupe (en annexes) précise l'objectif du groupe de travail qui vise à proposer une méthodologie d'analyse aux équipes scientifiques ayant à faire le choix d'un cahier de laboratoire électronique, en fonction de leurs besoins spécifiques. Pour réaliser cette étude, le groupe de travail s'est ainsi fixé

plusieurs étapes :

- Établir une vision partagée sur la définition, le cadrage, les usages et le périmètre fonctionnel du "cahier de laboratoire électronique" (voir section 2).
- Produire un ensemble de recommandations relatives à l'interopérabilité entre les différents outils de cahiers de laboratoire électroniques et d'autres outils ou systèmes d'information déjà utilisés. En effet, les cahiers de laboratoires doivent pouvoir s'intégrer, sans difficulté, dans les environnements informatiques déjà existants (y compris les entrepôts de données), en particulier dans un contexte académique français multi-tutelles (voir section 3).
- Produire un ensemble de recommandations sur les critères de choix d'un outil, en fonction des besoins métiers, des disciplines et domaines de recherche, ainsi que des contraintes institutionnelles (voir section 4).
- Établir une liste comparative de certains outils existants selon les critères recommandés pour mesurer leurs performances et/ou pertinence suivant les contextes d'usages (voir section 5).
- Produire un ensemble de recommandations permettant de mener à bien le projet d'implantation d'un ELN (voir section 6).

Il est rappelé que l'objectif du groupe est de proposer des critères d'aide à la décision et non de fournir des éléments absolus de comparaison entre différents outils, ni de préconiser le choix d'un outil unique pour l'ensemble de la communauté de l'enseignement supérieur et de la recherche.

1.3 Constitution du groupe

Le groupe de travail est composé de 16 personnes issues de différents établissements, représentant la diversité scientifique de l'enseignement supérieur et de la recherche. Chaque membre, selon sa discipline et son expérience, a su apporter son expertise, d'un point de vue :

- **Scientifique/expérimentateur** : avec une vision de terrain, en tant qu'utilisateur
- **Juridique et stratégique**, en considérant les aspects liés à l'intégrité scientifique, à la propriété intellectuelle et à la valorisation du contenu.
- **Technique**, tant en matière d'infrastructures et de l'hébergement d'outil, qu'en matière de problématique de gestion de données, d'interopérabilité et de sémantique.

1.3.1 Pilotes

Les personnes suivantes ont été missionnées par le Comité pour la Science Ouverte pour le pilotage de l'activité du groupe :

- Gilles MATHIEU, Ingénieur de Recherche en informatique au Département du Système d'Information (DSI) de l'Inserm
- Dominique PIGEON, Ingénieur de Recherche en informatique au Département du Système d'Information (DSI) de l'Inserm
- Tovo RABEMANANTSOA, Membre de la direction pour la science ouverte (DipSO) d'INRAE

1.3.2 Membres

Les personnes suivantes ont été membres du groupe de travail :

- Christophe CHIPEAUX (ICOS/INRAE)
- Simon DUVILLARD (Réseau CURIE)
- Célya GRUSON-DANIEL (Inno^3/UTC)
- Marie-Emilia HERBET (Université Lyon I)
- Arnaud LEGRAND (INRIA/CNRS)
- Nathalie LEON (CNRS)
- Domenico LIBRI (CNRS/INSB)
- Jean-Baptiste LILY (ICOS/INRAE)
- Jean-François PEYRAT (Université Paris Saclay)
- Agnès PINET (CEA)
- François SABOT (IRD)
- Moussa SEYDI (IRD)
- Véronique THEISEN (CEA)

1.4 Fonctionnement du GT

Le groupe de travail a réalisé ses travaux entre novembre 2020 et juillet 2021. Les échanges ont eu lieu uniquement par visioconférence sous la forme :

- de réunions plénières qui se sont tenues mensuellement pour une durée d'une heure.
- d'ateliers qui ont été proposés aux membres du GT, pour travailler sur des sujets particuliers.
- De façon exceptionnelle, des personnes extérieures ont participé à certaines sessions, ou ont été sollicitées pour des réflexions et échanges sur des points particuliers. La liste des personnes concernées est fournie en annexe.

Par ailleurs, les travaux ont été coordonnés avec ceux du groupe de travail « Réussir l'appropriation de la science ouverte » [1] piloté par Anne Vanet au sein du Comité pour la science ouverte et avec le projet « Cahiers de laboratoire électroniques », en cours au CNRS, piloté par Nathalie Léon et Domenico Libri. Pour cela, le groupe de travail a compté dans sa composition un membre nommé au sein du collège de données, un membre du groupe de travail « Réussir l'appropriation de la science ouverte », ainsi que les deux co-pilotes du projet du CNRS.

1.5 Finalité du rapport

Le groupe de travail a formulé des recommandations pour le choix des cahiers de laboratoire électroniques sous forme, d'une part, de ce rapport, qui reprend le cheminement du groupe de travail, présente les outils et des exemples de leur utilisation, et, d'autres part, d'outils méthodologiques :

- une grille de critères pour l'aide au choix d'un cahier de laboratoire électronique (tableau éditable, avec formules de calcul incluses) ; les critères sont accompagnés de propositions de pondération en fonction des usages ;
- des recommandations pour la conduite d'un projet de choix et de mise en place d'un cahier de laboratoire électronique ;
- une fiche synthétique présentant de manière brève et concise les travaux du groupe.

1. https ://www.ouvrirlascience.fr/reussir-lappropriation-de-la-science-ouverte/

2 Définition et périmètre : une vision partagée du cahier de laboratoire électronique

2.1 Définition

Un cahier de laboratoire électronique (*Electronic Laboratory Notebook*, ou ELN) est un logiciel qui permet de recueillir et centraliser toutes les informations essentielles à la reproduction d'une expérience scientifique. Ses fonctions documentaires et sa force probatoire lors du dépôt de brevet recoupent celles du cahier papier, mais elles les dépassent largement pour constituer un cahier « enrichi » et « augmenté » assimilable à un véritable outil de gestion des connaissances du laboratoire. Les différences les plus significatives reposent sur l'approche collaborative (travail en réseau), la traçabilité renforcée (adoption de protocoles-type ou modèles de données) et le gain de temps (requête des expériences par moteur de recherche). Au-delà du recueil des modes opératoires, le cahier électronique offre une approche intégrée des travaux scientifiques. Il établit une continuité entre l'expérience menée et les données brutes associées, sans avoir vocation à assurer le stockage de celles-ci. Il facilite le traitement des données grâce à la possibilité d'interfaçage avec les logiciels disciplinaires, les instruments de laboratoire et les bases de données. Enfin, lorsque l'ELN est couplé à une solution de type LIMS (cf. section 2.2.3 « outils connexes » ci-dessous), son rôle scientifique est complété par des fonctionnalités logistiques de nature à améliorer l'organisation du laboratoire en matière de gestion quotidienne des stocks (substances, échantillons, animaux, etc.) ou d'équipements (réservation, maintenance).

2.2 Périmètre

2.2.1 Personnes concernées

L'ELN fait intervenir les actrices et acteurs suivants :

- **Les actrices et acteurs principaux de la recherche**, à qui l'ELN est principalement dédié. Il s'agit des utilisatrices et utilisateurs principaux de l'ELN, dans le cadre du traçage et de l'enregistrement de leur activité ;
- **La ou le responsable d'équipe**, qui aura un rôle de référent-e voire d'administratrice ou administrateur. On pourra également parler d'investigatrice ou investigateur principal-e, ou *principal investigator* (PI) ;
- **La ou le signataire**, qui aura la responsabilité finale de valider les travaux de recherche consignés dans l'ELN ;
- **Les services juridiques et valorisation**, qui pourront s'appuyer sur l'ELN pour définir ou confirmer la date et la paternité des travaux de recherche consignés dans l'ELN, notamment dans les processus de dépôt de

brevet ;

- **Les différents partenaires** des actrices et acteurs principaux de la recherche (communautés, industriels, hospitaliers, etc.), qui pourront utiliser l'ELN comme outil collaboratif et de partage d'information ;
- **Les services d'archives**, qui seront amenés à assurer la conservation de l'information contenue dans l'ELN sur le long terme.

2.2.2 Usages

Les usages de l'ELN sont très différents selon la discipline scientifique, mais aussi – et surtout – selon les pratiques de recherche. Les appellations de carnet de terrain, cahier de manipulation, journal de bord, cahier de fouilles sont alors préférées à celles de « cahiers de laboratoire » avec une diversité de formes (notamment numériques) mais aussi de finalités en fonction des besoins des communautés. En dehors du « laboratoire », les enquêtes de terrain s'accompagnent d'un certain nombre de supports et d'outils dont les fonctions diffèrent en fonction des démarches méthodologies de recherche. Il peut s'agir ainsi :

- d'une aide au recueil d'une variété d'observations de terrain ou de notes d'analyse tout au long d'une recherche « qualitative »de manière individuelle ou collective dans des démarches ethnographiques, sociologiques, ou anthropologiques
- de la collecte et de l'organisation systématisée de données en archéologie, en sciences de l'environnement donnant lieu à des bases de données et systèmes d'information

Qu'il s'agisse de prises de notes sur un éditeur de texte ou de la mise en œuvre de systèmes d'information géographique, la nature des outils numériques employés varient selon les pratiques.

Dans le domaine des SHS (sociologie, anthropologie, etc.) ou des sciences du vivant (écologie, etc.), des outils informatiques ont commencé à remplacer sur le terrain les supports papier pour les prises de notes d'observation, d'analyse, etc. Sans pour autant avoir les mêmes besoins que la recherche expérimentale, ces disciplines peuvent tirer un avantage certain de l'usage des ELN, pour ce qui est de la sécurisation des données, de leur indexation, des liens avec d'autres outils numériques, etc. Un panorama bien documenté de ces questions, sans objectif d'exhaustivité, est présenté en annexes.

En termes de pratiques de recherche, de manière transversale aux différents domaines scientifiques, le groupe de travail a identifié trois modalités principales caractérisant des usages différents de l'ELN :

- La recherche à dominante **expérimentale**. Le cœur de la recherche s'ar-

ticule autour de manipulations contrôlées et volontaires, qu'elles soient en laboratoire ou en milieu naturel.
- La recherche à dominante **observationnelle**, où l'activité de recherche a pour principal objet l'observation d'objets, phénomènes, comportements en dehors d'un cadre manipulatoire.
- La recherche à dominante **analytique**, principalement centrée sur l'analyse informatique et le traitement de données, quelle que soit leur origine.

On notera que la plupart des disciplines scientifiques présentent des domaines et sujets de recherche se rapportant à ces trois modalités et qu'il n'y a pas de frontière stricte entre elles.

2.2.3 Outils connexes

Avec l'avancée des technologies numériques, les différents outils pouvant être mis en place sont de moins en moins considérés comme des objets séparés : à différentes échelles, des environnements numériques de travail se développent mettant en avant l'interopérabilité entre différents logiciels. Le cahier de laboratoire en contexte numérique est donc bien loin d'être isolé comme un cahier papier et s'articule avec différents outils et infrastructure de recherche permettant une gestion fluide et sécurisée des données.

Parmi les outils les plus à même d'interagir avec l'ELN, nous pouvons noter :
- Les **LIMS (*Laboratory Information Management System*)**. Il s'agit de progiciels de gestion intégrés qui gèrent par exemple la saisie sur le terrain et la traçabilité des échantillons, les utilisatrices et utilisateurs, les instruments (principalement des automates), les stocks et approvisionnements, et le suivi des produits et des équipements. Pour les entités de recherche tels que les laboratoires, travaillant principalement selon une modalité expérimentale, il est particulièrement pertinent de disposer de fonctionnalités propres à la gestion de ces informations. On notera que de nombreuses solutions d'ELN intègrent des fonctionnalités de type LIMS.
- Les ***notebooks***. Il s'agit d'interfaces de programmation interactive permettant de combiner des sections en langage naturel et des sections en langage informatique. Ces outils sont particulièrement utilisés par les communautés de recherche travaillant principalement en mode analytique.
- Les **forges logicielles**. Il s'agit de systèmes de gestion et de maintenance collaborative de code, texte et documents souvent liés à un projet informatique ou un produit logiciel. Ces outils sont principalement utilisés en recherche analytique.
- Les **systèmes d'information disciplinaires**, comme par exemple les SIG

(Systèmes d'Information Géographiques), utilisés en archéologie.
- Les **entrepôts de données**, permettant de faire le lien entre la démarche de recherche présentée dans l'ELN et les données finales, potentiellement publiées, issues de ces travaux.

Les liens entre le suivi d'un travail de recherche, les données traitées, les équipements utilisés ou encore le code employé sont souvent nécessaires pour répondre aux exigences de traçabilité et reproductibilité. En fonction du contexte, l'interopérabilité avec ces différents outils peut s'avérer nécessaire pour l'utilisation correcte d'un ELN.

3 Recommandations relatives à l'interopérabilité entre différents outils

3.1 Interopérabilité

Dans sa définition communément admise [2], l'interopérabilité est la capacité que possède un produit ou un système à fonctionner avec d'autres produits ou systèmes existants ou futurs et ce sans restriction d'accès ou de mise en œuvre. L'interopérabilité concerne uniquement le comportement externe des produits et n'influe pas leur conception interne. Cette capacité est assurée par des normes et des formats connus. En France, ces normes peuvent être décrites dans le référentiel général de l'interopérabilité (RGI) [3] dont la version actuelle a été approuvée par l'arrêté du 20 avril 2016 [4]. Le RGI est un document décrivant un ensemble de normes et bonnes pratiques communes aux administrations publiques françaises dans le domaine informatique.

Les enjeux de l'interopérabilité se situent sur les formats des données échangées entre les différents produits ainsi que sur les interfaces et protocoles de communication.

Un laboratoire de recherche a, généralement, un écosystème numérique riche que l'on qualifie souvent de système d'information (SI). Il peut être composé d'infrastructures de gestion de données (bases de données, gestion documentaire, entrepôt, etc.), d'infrastructures de calcul (HPC, *Notebook*, chaînes de traitements, etc.), d'environnement virtuels de recherche (VRE) ou tout simplement d'outils bureautique. Le personnel de recherche a également accès à des ressources se situant au-delà du périmètre du laboratoire tels que les entrepôts de données publiques, des catalogues et des espaces de stockage et/ou de calcul. Pour que l'adoption d'un ELN apporte une valeur ajoutée, celui-ci devra s'interfacer au maximum avec les éléments du SI existant. L'interopérabilité de l'ELN est l'élément majeur de son intégration dans le SI.

3.2 Recommandations

3.2.1 Co-tutelles

Quand chaque tutelle préconise ou utilise déjà un ELN, il est important que chaque ELN puisse exporter son contenu dans un format homogène, ouvert et do-

2. https ://fr.wikipedia.org/wiki/Interop%C3%A9rabilit%C3%A9
3. https ://www.numerique.gouv.fr/publications/interoperabilite/
4. https ://www.legifrance.gouv.fr/jorf/id/JORFTEXT000032438896

cumenté. Cela permettra aux différents scientifiques de faire circuler et partager les contenus et, ainsi, de décloisonner les travaux de recherche.

Étant donné qu'actuellement, aucun format standard n'existe pour les contenus d'ELN, l'accessibilité et de l'utilisabilité de ces formats d'export sont primordiales afin de créer les passerelles entre les ELN existants.

Il est vital que l'ELN puisse attribuer, *a minima*, un identifiant pérenne (*Persistent IDentifier* ou PID) et un horodatage à chacune de ses entrées. Dans l'idéal, il faudrait que l'ELN puisse adresser directement ces identifiants à travers des référentiels (DOI, ORCID, etc.)

3.2.2 Intégration

Privilégier les ELN qui s'intègrent de manière transparente avec les outils déjà utilisés dans l'établissement.

Cette intégration peut se faire nativement sans nécessité de programmation mais peut également se réaliser à travers l'usage d'interfaces de programmation (API). Dans ce cas, privilégier un ELN dont les API sont bien documentés.

L'ELN doit pouvoir s'interfacer avec les systèmes déjà présents tels que la gestion des identités, les GED, les VRE, entrepôts (Dataverse, forge logicielle, etc.) ou encore les environnements de calculs (*notebooks* tels que Jupyter et Scilab, R-studio etc.)

3.2.3 Accès aux données

Considérer la capacité de l'ELN à accéder et à lire les données stockées en dehors du système et plus particulièrement les grands jeux de données qui ne peuvent y être importés.

Il est important que l'ELN puisse accéder aux données et métadonnées stockées dans les dépôts (Dataverse, CKAN, etc.) afin de permettre la réutilisation des données de la recherche.

3.2.4 Sauvegarde et archivage

Dans le cas où l'ELN embarque un mécanisme de sauvegarde, il doit pouvoir être paramétrable et opérable sans limitation aux responsables de l'établissement utilisateur. Les éléments sauvegardés doivent être accessibles et restaurables à gré.

Dans le cas contraire, la structure de l'ELN doit le permettre. La sauvegarde des données de l'ELN doit pouvoir s'effectuer de l'extérieur à travers des images de l'arborescence du système de fichiers et des bases de données de l'ELN. Ces images doivent être restaurables, a minima, sur la même version de l'ELN et, dans l'idéal, avec une compatibilité ascendante et descendante.

Dans tous les cas, l'établissement utilisateur a tout intérêt à pouvoir accéder à un historique des contenus de l'ELN qu'il utilise. En cas de corruption ou de perte de données, il doit pouvoir faire face à travers une restauration de cet historique.

L'ELN doit pouvoir exporter tout ou partie de ses entrées dans un format ouvert afin de devenir une archive pérenne. Ces exportations doivent inclure toutes les métadonnées, médias associés et références nécessaires afin qu'ils puissent être exploitables indépendamment de l'ELN.

3.3 Résumé

Veiller à ce que l'ELN puisse stocker et manipuler des données aux formats ouverts en entrée ET en sortie.

Veiller à ce que l'ELN offre une API bien documentée pour faciliter son intégration dans l'écosystème informatique s'il ne le fait pas nativement.

Veiller à ce que le contenu de l'ELN puisse être sauvegardé et archivé.

4 Recommandations sur la procédure de choix d'un outil

4.1 Besoins

4.1.1 Recueillir les souhaits des différents acteurs

L'objectif étant de répondre aux besoins des agents souhaitant une solution électronique adaptée aux pratiques de la recherche, l'une des premières phases est l'établissement, de façon macroscopique, des exigences fonctionnelles. Celui-ci repose sur une étude permettant de recueillir les besoins des utilisateurs et définir les spécifications techniques découlant des contributions des différentes parties prenantes (direction, scientifiques, juridique, systèmes d'information, sécurité, etc.). Afin d'assurer la traçabilité et l'intégrité scientifique des résultats de la recherche, la réflexion doit porter sur le fait d'offrir toute la sécurité, la confidentialité et la traçabilité nécessaires à la protection des résultats et du patrimoine scientifique de l'établissement et aussi de ses partenaires de recherche. Aussi, la collaboration entre établissements partenaires et le plus souvent co-tutelles d'unités est essentielle afin d'émettre des recommandations partagées facilitant l'accès aux outils pour la cible première que sont les utilisateurs.

La direction se doit d'apporter les lignes directrices du ou des objectifs attendus, du périmètre sur lequel porte l'étude ou le besoin et de l'organisation globale à mettre en œuvre en réponse à l'atteinte de ce ou ces objectifs selon des indicateurs qui seront fixés en amont. Y seront définies les étapes d'arbitrage et de validation permettant l'avancement cohérent de l'étude au regard des résultats intermédiaires et permettant une certaine agilité pour adapter les jalons au cours du temps si besoin.

4.1.2 Prendre en compte la discipline et les modalités de pratique de la recherche

Habituellement utilisé dans le cadre d'expérimentations à la paillasse ou sur le terrain, le cahier de laboratoire voit ses applications démultipliées dans sa version électronique, comme montré dans la section « usages » (2.2.2) du présent rapport. Il existe un socle commun de fonctionnalités propres à la plupart des outils (création d'expériences, recherche avancée, modèles d'expérience, etc.), complété par des fonctionnalités métier parfois très spécifiques à une discipline donnée.

La sélection d'un outil peut ainsi s'avérer délicate en fonction des contraintes métier dues aux domaines et sous-domaines de recherche. Dans une même disci-

pline, les attentes peuvent être très différentes : un-e chimiste de synthèse n'aura pas la même utilisation de l'ELN qu'un-e chimiste qui fait de l'analytique avec du code. Or ces deux profils peuvent très bien se rencontrer au sein du même laboratoire. L'angle de choix doit donc plutôt s'effectuer au niveau des pratiques de recherche plus qu'au niveau de la discipline.

En restant sur l'exemple de la chimie et plus particulièrement de la chimie de synthèse, la plus-value de l'ELN réside dans des fonctionnalités telles qu'une table stœchiométrique réalisant les calculs de masses et de quantités automatiquement, une recherche par structure et sous-structure chimique efficace, un éditeur de molécule et un dispositif de gestion des stocks de réactifs et de produits utilisés durant les manipulations. C'est cet ensemble cohérent d'outils disciplinaires qui peut faire de l'ELN un atout dans le quotidien du chercheur.

La gestion des stocks citée précédemment est une fonctionnalité se rapportant plus à des outils de type LIMS (*Laboratory Information Management System*). Cependant, on trouve aujourd'hui dans la plupart des solutions de cahier de laboratoire électronique une partie gestion des stocks, plus ou moins aboutie. Dans le cas où le laboratoire est amené à gérer quotidiennement des produits, des équipements ou des appareils comme peuvent l'être des laboratoires de chimie ou de biologie, le choix de l'ELN peut être en partie déterminé par cette fonctionnalité.

Pour les chercheuses et chercheurs issus de l'analytique et du développement logiciel, les outils répondant à la définition stricte du cahier électronique sont peu adaptés : calcul avancé, édition et exécution de code, gestion de version, etc., sont des fonctionnalités indispensables inexistantes à l'heure actuelle. De plus, de nombreux outils très performants réalisant la plupart de ces fonctions existent et sont déjà utilisés par les différentes communautés. On peut par exemple citer Git pour la gestion des versions, Gitlab pour le développement collaboratif, Emacs pour l'édition de code et Jupyter Notebook pour la mise en forme de code et l'exécution de celui-ci. L'ELN ne peut pas se substituer à ces outils, mais peut intervenir en complément. Certains éditeurs d'ELN mettent à disposition une API, permettant d'interconnecter l'outil avec d'autres logiciels. Il peut être intéressant pour les chercheuses et chercheurs en analytique de voir s'il est possible de « greffer » leurs outils à l'ELN grâce à l'API.

Ainsi, l'importance à accorder à telle ou telle fonctionnalité variera-t-elle selon la discipline visée et les modalités de recherche prépondérantes auprès du public cible.

4.1.3 Réfléchir au niveau de sécurité attendu

L'une des problématiques les plus importantes à prendre en compte au moment du choix concerne le niveau de sécurité attendu pour la conservation, le traitement et la mise à disposition des données. Selon le contexte technique, scientifique, et humain propre à chaque étude il sera nécessaire de définir l'échelle des besoins en la matière. Il est recommandé de s'appuyer pour cela sur une démarche d'étude de risques comme la méthodologie EBIOS [5] de l'Agence Nationale de la Sécurité des Systèmes d'Information (ANSSI), en se référant à des critères normés comme par exemple ceux de la norme ISO/CEI 27001 [6]. Le contexte éventuel de laboratoire classé en ZRR (zone à régime restrictif) devra également être pris en compte lors du choix. Le périmètre des ELN les associe généralement à un besoin fort en matière de sécurité, notamment selon les quatre critères d'évaluation communément admis :

- La **confidentialité** des données traitées doit être assurée par le biais d'un système d'**autorisation** : seules les personnes autorisées peuvent avoir accès aux informations qui leur sont destinées (notions de droits ou permissions) et tout accès indésirable doit être empêché.
- L'**authenticité** doit être assurée *via* un système d'**authentification** : les utilisatrices et utilisateurs doivent prouver leur identité par l'usage de code d'accès. Cela permet de gérer les droits d'accès aux ressources concernées et maintenir la confiance dans les relations d'échange.
- L'**intégrité** : les données doivent être celles que l'on attend, et ne doivent pas être altérées de façon fortuite, illicite ou malveillante. En clair, les éléments considérés doivent être exacts et complets. Cet objectif utilise généralement des méthodes de calculs de *checksum* ou de hachage.
- La **disponibilité** : l'accès aux ressources du système d'information doit être permanent et sans faille durant les plages d'utilisation prévues. Les services et ressources sont accessibles rapidement et régulièrement.

Ces besoins peuvent également être majorés selon la nature des données traitées, notamment pour les données concernées par le secret professionnel, les secrets industriels et commerciaux, les données personnelles et la protection du potentiel scientifique. Le pays et les conditions d'hébergement des données de ce cahier devront faire l'objet de la plus grande vigilance, ainsi que les conditions de leur préservation et de l'accès par des tiers.

Au-delà de ces quatre critères principaux, d'autres aspects doivent aussi être considérés dans l'étude de choix d'un ELN, notamment :

- La **traçabilité** (ou « preuve ») : garantie que les accès et tentatives d'accès

5. https ://www.ssi.gouv.fr/administration/management-du-risque/la-methode-ebios-risk-manager/

6. https ://www.iso.org/fr/isoiec-27001-information-security.html

aux éléments considérés sont tracés et que ces traces sont conservées et exploitables.
- La **non-répudiation** et l'**imputation** : aucune utilisatrice ou utilisateur ne doit pouvoir contester les opérations qu'elle/il a réalisées dans le cadre des actions auxquelles il est autorisé et aucun tiers ne doit pouvoir s'attribuer les actions d'un autre utilisateur.

4.1.4 Réfléchir au niveau d'implémentation de l'outil

Il faut distinguer le niveau de déploiement de l'ELN et celui de son hébergement :

- On peut créer un ELN par contributeur d'un projet de recherche ou, plus généralement, par équipe de recherche ; plus rarement pour une structure de recherche multi-équipes, voire pour tout un centre de recherche.
- Ces cahiers peuvent être hébergés sur le poste de la chercheuse ou du chercheur, dans le bâtiment abritant la structure de recherche ou dans un centre de données régional ou national, voire dans un cloud.

L'implantation de l'ELN au niveau du poste de travail peut conférer à son possesseur de l'autonomie en matière de sauvegarde et de mise en qualité des données (horodatage des actions) ; en revanche, elle rend difficile le partage de celles-ci. Or, l'un des intérêts de l'ELN par rapport à sa version papier est la possibilité de partager l'information au sein d'une équipe de recherche. Un avantage de l'implantation au niveau du poste de travail est la possibilité de travailler en mode hors ligne ou lors d'une panne de réseau. L'implantation des ELN sur un serveur central facilite le partage mais rend dépendant de la qualité du réseau entre les postes de travail et le serveur. Les différentes combinaisons étant possibles, il conviendra de les étudier notamment sous les angles suivants :

- Partage de l'information
- Sécurité des données
- Possibilité d'interconnexion avec les autres outils numériques
- Performance, continuité de service

4.1.5 Identifier les systèmes qui seront en interaction avec l'ELN

Selon le contexte de recherche, les modalités principales et les disciplines considérées, l'ELN choisi devra être en mesure d'interagir avec différents systèmes tiers. Ce sera notamment le cas pour les outils connexes décrits en section 2.2.3. Les recommandations relatives à l'interopérabilité de la solution, décrits en section 3, devront donc s'appliquer pleinement.

4.2 Établir les critères de choix

4.2.1 Caractéristiques techniques et fonctionnelles

Le choix d'un ELN passe d'abord par l'élaboration de la liste des fonctionnalités souhaitées, ainsi que des contraintes et caractéristiques techniques nécessaires. Concernant les fonctionnalités, la liste suivante a été établie par le groupe de travail :

TABLE 1: liste des fonctionnalités retenues.

Fonctionnalité
Description des expériences et résultats
Possibilité de dupliquer des expériences (sans les résultats de recherche), comme un modèle/template pour une nouvelle expérience
Description des protocoles
Possibilité de dupliquer des protocoles, comme un modèle/template pour un nouveau protocole
Gestion de pièces jointes : bureautiques (PDF, audios, vidéos, etc.) ou scientifiques et techniques (git, chromatogrammes, spectres, images, etc.)
Possibilité d'exporter des éléments de l'ELN à un format imprimable
Accès à des données externes (c'est-à-dire non stockées dans l'ELN), brutes ou non, pour lecture et référencement
Structuration avec moteur et filtre de recherche interne ; intégration de mots-clés (tags)
Contenu accessible (pour fouille/mining)
Possibilité d'associer des métadonnées aux jeux de données manipulés, qu'il s'agisse de métadonnées génériques (dates, auteurs, etc.) ou spécifiques (ontologies, taxonomies, etc.)
Utilisation et maintenance des équipements (y compris remontées d'alertes)
Gestion des stocks (réactifs, lignées biologiques, etc.)
Outils pour des calculs simples (fonctions mathématiques de base, conversion d'unité, etc.)
Outils pour des calculs complexes (statistiques ou analyses nécessitant un langage de programmation)
Outil de construction/dessin de molécules
Notes sur les expériences ou le protocole (pendant l'expérimentation), comptes rendus, etc.
Possibilité de citer des contenus issus de bibliographies (type labtex.bib Text)
Accès aux outils bureautiques du poste de travail

Fonctionnalité (suite)
Gestion des outils logiciels et de codes informatiques (intégration continue, déploiement continu, gestion des versions, etc.)
Gestion du partage par les utilisatrices et utilisateurs
Agenda partagé
Edition partagée de documents (en ligne et en simultané par plusieurs utilisatrices ou utilisateurs)
Annotation de contenu et ajout de commentaires

Sur le plan des caractéristiques techniques, les travaux du groupe ont mené à la liste suivante :

TABLE 2: liste des caractéristiques techniques retenues.

Catégorie	**Caractéristique**
Protection des données	Respecter le RGPD pour les données personnelles des utilisatrices et utilisateurs
	Respecter le RGPD pour les données de recherche : données personnelles des participant-e-s à la recherche si elles sont contenues dans les ELN
	Avoir un stockage maîtrisé, redondant et sauvegardé (sites distants)
Intégrité	Respecter des contraintes réglementaires ou législatives concernant la localisation de l'hébergement des données de recherche
	Permettre la signature des documents (hachage)
	Permettre l'horodatage des actions et des documents
	Identifier de manière unique les entrées (protocole, expérience, note, etc.) - UUID
	Assurer un versionnement/suivi des modifications
	Permettre d'implémenter des workflows prédéfinis (par exemple, suite d'actions obligatoires dans la description d'une expérience ou d'un résultat)
Authentification, sécurité	Permettre une signature électronique des actrices et acteurs
	Utiliser un système d'authentification sécurisé
	Permettre de définir les règles de gestion des utilisatrices et utilisateurs : niveau d'autorisation d'accès selon rôles et projets et cycle de vie (départ/arrivée agent, ouverture/fermeture projet, etc.)

Sauvegarde des données	Chiffrer le transport des données (TLS, algorithmes conformes RGS)
	Chiffrer les données
	Intégrer des sauvegardes régulières et des restaurations
	Définir les durées de conservation et modalités d'archivage ou destruction des données
Confidentialité	Assurer la confidentialité des données exportées pour sauvegarde ou archivage.
	Préserver la confidentialité des informations pouvant constituer un savoir-faire valorisable ou relever du secret des affaires (préserver la nouveauté en cas d'invention brevetable)
	Préserver la confidentialité des informations échangées avec un-e partenaire dans le cadre d'une collaboration
Compatibilité, interopérabilité	Être compatible avec les différents systèmes informatiques (Windows, Linux, etc.) pour la partie serveur
	Permettre la gestion (*Import/Export*) des données sous un format structuré (par exemple en cas de changement de logiciel ou d'évolution administrative)
	Permettre l'intégration avec un système d'information local (bases de données, annuaires, etc.)
	Permettre les liens avec des entrepôts de données
	Utiliser, quand ils existent, les standards ouverts des données en entrée ou en sortie pour permettre la réexploitation des informations de façon pérenne
	Permettre les échanges de données en proposant des services d'intégration et d'exposition, si possible via des *Web-Service* ou API REST
Usage, utilisation	Pouvoir installer facilement les mises à jour et évolutions, en garantissant l'accessibilité, dans le cas de l'utilisation à partir d'un navigateur Web
	Être ergonomique et intuitif
	Permettre la modularité/personnalisation/flexibilité
	Disposer des références documentaires (guide, tutoriels, consignes d'utilisation, etc.)
	Avoir un code source accessible pour contrôler, si nécessaire, l'évolution de l'outil, sa maintenance et surtout les données qu'il contient
	S'appuyer sur une communauté d'utilisatrices et d'utilisateurs

	Proposer l'aspect multilingue (a minima l'anglais)
	Permettre une utilisation hors-ligne avec synchronisation a posteriori
	Être utilisable en mobilité (sur tablette et smartphone)
	Prendre en compte l'empreinte environnementale : consommation électrique des serveurs, etc.

Bien entendu, les différentes fonctionnalités et caractéristiques techniques listées n'interviendront pas avec le même niveau d'importance dans le choix final. Elles doivent principalement servir de point d'entrée pour l'établissement d'une grille d'évaluation, dont le principe est présenté plus loin. Des explications sur certaines de ces fonctionnalités et caractéristiques sont présentées en annexe.

4.2.2 Autres critères

Les fonctionnalités et contraintes techniques ne sont pas les seuls critères à prendre en compte. Pour une étude complète, il convient aussi de considérer :

- Le **coût global de la solution**, qu'il s'agisse de coûts financiers ou de ressources nécessaires, notamment en termes de ressources humaines. Ce coût est à estimer pour l'acquisition, le déploiement et l'opération sur le long terme de la solution retenue. Outre les éventuels coûts de licence d'une solution propriétaire, l'étude doit également inclure les coûts de maintenance liés à l'hébergement, à l'environnement technique nécessaire, à l'accompagnement et la conduite du changement.
- La **facilité d'adoption et l'aide disponible**. Ceci peut concerner la qualité du support éditeur le cas échéant, la présence de documentation, mais aussi la taille et le dynamisme de la communauté d'utilisatrices et d'utilisateurs.
- Les **critères subjectifs d'adéquation avec le public cible**. Bien que difficile à quantifier, il s'agit pourtant d'un élément clé dans le succès du choix d'un outil. Pour certaines communautés, cela peut se traduire par la capacité de l'ELN à correspondre au mode de travail habituel des utilisatrices et utilisateurs, qui conduira à faciliter le lien entre leur cheminement mental de recherche et la saisie d'informations dans l'ELN. L'étude de cet aspect peut difficilement reposer sur des critères objectifs et doit souvent faire l'objet d'une étude de cas d'utilisation auprès d'utilisatrices et utilisateurs pilotes, suivi d'un retour d'expérience et recueil d'avis.

4.3 Identifier les offres du marché

Le marché des ELN est en pleine expansion. Entre 2014 et 2020, le volume de revenus générés a plus que doublé passant de 222 millions à 576 millions de dollars aux États-Unis [7]. Ce dynamisme a été largement porté par la conversion technologique du secteur de l'industrie pharmaceutique, qui a très tôt perçu le potentiel des ELN en tant que levier d'amélioration du processus qualité, d'innovation et de compétitivité. Certains géants du secteur, comme BMS, Astra Zeneca, Johnson & Johnson, GSK, Eli Lilly and Company ou encore Roche ont ainsi opté pour la dématérialisation des cahiers de laboratoire dès le début des années 2000.

Cette volonté de transformation s'est traduite par une explosion des solutions disponibles. En 2006, un rapport américain recensait moins de 30 fournisseurs de cahiers de laboratoire électroniques [8]. En 2017, une étude conduite par l'Université de Southampton dénombre 103 solutions différentes [9].

L'atomicité de l'offre, qui peut complexifier le processus décisionnel quant à la solution à retenir, se reflète dans les différentes tentatives d'analyses comparatives (= *benchmarking*) existantes. Entre 2018 et 2021, la Faculté de médecine de Harvard et le campus médical de Boston ont mené un travail de *benchmarking* portant sur 33 solutions afin « d'aider les chercheuses et chercheurs [...] à identifier concrètement les outils de type ELN répondant aux besoins spécifiques de leur recherche. » [10] La matrice, toujours disponible en ligne [11], a été réalisée grâce à l'envoi de questionnaires auprès des éditeurs concernés. D'autres initiatives sont accessibles librement, à l'instar du recensement, plus synthétique, proposé par l'Université de Cambridge [12], assorti de courtes études de cas centrées sur quatre solutions (Onenote, Hivebench, Benchling et LabArchives [13]). Cette dernière application, déployée notamment par l'Université de Monash à Melbourne (Australie), est largement décrite sur le site de la bibliothèque universitaire de l'établissement [14]. De manière analogue, les bibliothèques universitaires de Lyon 1

7. https ://www.grandviewresearch.com/industry-analysis/electronic-lab-notebook-eln-market
8. https ://journals.sagepub.com/doi/10.1016/j.jala.2009.01.002
9. https ://jcheminf.biomedcentral.com/articles/10.1186/s13321-017-0221-3
10. https ://zenodo.org/record/4723753#.YORvu0lLiUk
11. Le groupe de travail porteur du document a décidé de cesser toute mise à jour au-delà d'avril 2021 : https ://docs.google.com/spreadsheets/d/1ar8fgwagOh30E31EAPL-Gorwn_g6XNf81g3VDQnQ_I8/
12. https ://www.data.cam.ac.uk/data-management-guide/electronic-research-notebooks/electronic-research-notebook-products
13. https ://www.data.cam.ac.uk/data-management-guide/electronic-research-notebooks/electronic-notebook-case-studies
14. https ://www.monash.edu/library/researchers/researchdata/eln/about/who-is-using

et de Grenoble Alpes ont élaboré une typologie détaillée de huit solutions de cahiers de laboratoire électroniques, dont certaines se démarquent par leurs fonctions adaptées à la chimie [15].

Une première appréhension des offres disponibles peut consister à distinguer les différents ELN selon le modèle d'usage, le modèle économique et le modèle d'hébergement :

Modèle d'usage	Générique
	Disciplinaire
	Fonctions LIMS intégrées
Modèle économique	Open Source
	Propriétaire
Modèle d'hébergement	Cloud du fournisseur exclusif
	In-house exclusif
	Cloud ou in-house au choix

Cette grille sommaire permet ainsi d'opérer un premier tri en fonction des offres actuellement disponibles et des priorités des établissements, selon un modèle proposé ci-dessous à partir de 4 exemples emblématiques de la diversité du paysage des ELN :

critère / solution		**Chemotion**	**elabFTW**	**Mbook**	**Labstep**
Modèle d'usage	Générique		X		X
	Disciplinaire	X		X	
	Fonctions LIMS intégrées			X	X
Modèle économique	Open Source	X	X		
	Propriétaire			X	X
Modèle d'hébergement	Cloud du fournisseur exclusif				X
	In-house exclusif	X			
	Cloud ou in-house au choix		X	X	

15. https ://www.datacc.org/bonnes-pratiques/utiliser-un-cahier-de-laboratoire-numerique/typologie-de-8-cahiers-de-laboratoire-a-destination-des-chimistes/

Les déploiements opérés ces dernières années dans le secteur de la recherche et de l'enseignement supérieur ne permettent pas de dégager une tendance de fond au profit d'un outil, tant l'hétérogénéité des choix prédomine. Certains établissements ont privilégié des solutions *open source* (elabFTW à l'Inria, Institut Curie, Institut Jacques Monod, Université de Düsseldorf, Université de technologie de Graz, etc.), quand d'autres optent pour des solutions propriétaires (Labguru à l'Inserm, eLabJournal à l'Institut Pasteur et à l'Université de Genève, où Rspace a également été retenu [16], etc). Parmi les logiciels ayant volontairement une approche disciplinaire, l'offre se concentre essentiellement sur les besoins de la communauté des biologistes et des chimistes.

Le marché des ELN est également marqué par une incursion des grands groupes éditoriaux, dont certains rachètent des solutions initialement développées par des start-up. En témoignent l'acquisition de Hivebench par Elsevier en 2016, ou encore l'achat de parts au sein de Biodata (éditeur de Labguru) par Digital Science, dont la maison-mère est Holtzbrinck. Ils font par ailleurs partie des stratégies Commerciales de groupes d'envergure mondiale à l'image de Dassault (Biovia), PerkinElmer (Signals) ou encore Agilent (Slims).

Ces solutions contrastent avec les stratégies portées par le secteur académique, qui tendent à se développer ces dernières années : elabFTW (Institut Curie), Chemotion (Institut de Karlsruhe) ou encore OpenBIS (Université de Zurich).

4.4 Évaluer les offres selon les critères établis

Dans la mesure du possible, la comparaison d'outils sur le plan technique (fonctionnalités et caractéristiques) doit s'appuyer sur une liste de critères objectifs permettant de différencier les solutions selon l'importance accordée, dans le contexte de l'étude, à tel ou tel critère.
Cela nécessite une réflexion à deux niveaux, d'une part sur l'échelle et la granularité souhaitée pour évaluer ces critères (notation), et d'autre part sur l'importance relative d'un critère par rapport à un autre (pondération).

4.4.1 Notation

Le niveau minimum pour la notation d'un outil selon les critères établis est binaire : l'outil évalué propose la fonctionnalité ou il ne la propose pas ; il remplit la contrainte ou ne la remplit pas. Les différentes solutions peuvent ainsi être évaluées fonctionnalité par fonctionnalité sur une évaluation objective qui n'inclut

16. https ://www.unige.ch/researchdata/fr/collecter-organiser/eln/

pas de jugement de valeur.

Pour autant, selon le contexte de l'étude, le profil de l'évaluatrice ou évaluateur, ou encore la finesse nécessaire, différentes échelles de notation peuvent être utilisées. Le modèle de grille fourni par le groupe de travail propose une notation à 4 niveaux :

- **0** = ne fait pas ou information non disponible
- **1** = répond partiellement au besoin
- **2** = répond au besoin
- **3** = fait au-delà de l'attente, ou dispose d'atouts particuliers

Si l'étude menée fait appel à de nombreuses évaluatrices ou évaluateurs, on peut imaginer une échelle bornée continue (de 0 à 10 par exemple) permettant d'obtenir, pour chaque fonctionnalité, la moyenne des notes attribuées par les différentes évaluatrices ou évaluateurs.

Dans tous les cas, le choix de l'échelle de notation doit être fait et expliqué aux évaluatrices ou évaluateurs afin d'éviter toute confusion et mauvaise interprétation.

4.4.2 Pondération

Si la notation permet de comparer différents outils sur un critère donné, il est nécessaire de différencier dans le choix final les critères d'importance de ceux plus optionnels. En guise d'exemple, il semble logique dans le cas d'un ELN d'accorder plus d'importance à la fonctionnalité « permet de décrire les expériences » (qui est l'une des fonctions de bases sans laquelle un ELN n'est pas un ELN) qu'à la fonctionnalité d'exporter le texte en version imprimable.

Il convient donc, pour obtenir une vision équilibrée de la qualité d'une solution, de pondérer les différentes notes de manière à considérer les résultats de l'évaluation dans leur ensemble.

Deux échelles de pondération ont été proposées par le groupe. Dans un premier temps, selon une échelle à 4 niveaux :

- **1** = inutile
- **2** = secondaire
- **3** = important
- **4** = indispensable

Puis, dans un second temps et pour arriver à une analyse plus fine, une échelle non linéaire a été proposée. Son principe est d'arriver à une somme des valeurs de pondérations égale à 100, et répartir les valeurs selon leur importance respective. C'est cette échelle qui est appliquée dans la grille présentée au paragraphe suivant.

4.4.3 Présentation des résultats

L'utilisation de la grille à bon escient doit permettre de présenter les résultats de l'étude comparative technique. Cette représentation peut, selon le besoin et le contexte, prendre plusieurs formes :

- Un tableau binaire « fait/ne fait pas » : cette représentation peut s'avérer efficace et utile pour effectuer un premier tri dans une large liste d'outils (voir section 5)
- Une notation globale par outil, obtenue en multipliant pour chaque critère la note obtenue par la valeur de pondération, puis en ajoutant les résultats.
- Une représentation graphique, pour chaque outil, des points forts et points faibles. Cette représentation peut prendre la forme d'un diagramme de Kiviat.

Les figures 1 et 2 montrent un exemple de grille de notation, pris parmi les outils analysés dans la section 5.

La représentation graphique sous la forme d'un diagramme de Kiviat peut être créé ainsi : pour chaque approche, les sous-totaux pour l'ensemble des fonctionnalités d'une part et pour chaque catégorie de caractéristiques d'autre part sont comparés à la note maximale de 2 par critère (le 3 permettant de dépasser les 100% d'adéquation de la solution évaluée par rapport au besoin) puis rapportés à 5. Les 8 valeurs obtenues sont représentées sur un graphique en toile d'araignée, qui s'avère pertinent pour comparer plusieurs solutions entre elles. On se méfiera du biais que représente le fait que l'importance des groupes de caractéristiques les uns par rapport aux autres n'a pas été pris en compte. Un exemple de diagramme et de la table de valeurs permettant sa génération est présenté sur les figures 3 et 4.

Commentaires :

- Dans l'exemple donné ci-dessus (Figure 1), on voit apparaître, pour l'outil considéré, une potentielle meilleure adéquation des fonctionnalités à l'approche expérimentale alors que les caractéristiques techniques ne démontrent aucune différence significative.
- On remarque quelques items avec une note à 0 (Figure 2). – En effet, cet ELN est uniquement accessible en mode hébergé (SaaS) aux Etats-Unis,

Catégorie	Fonctionnalité	Note attribuée	Note pondérée Exp.	Note pondérée Analyt.	Note pondérée Obs.	expérimentale	analytique	observationnelle
Fonctionnalités	Description des expériences et résultats	3	30	30	30	10	10	10
	Possibilité de dupliquer des expériences (sans les résultats de recherche), comme un modèle/template pour une nouvelle expérience	3	15	21	12	5	7	4
	Description des protocoles	1	10	7	10	10	7	10
	Possibilité de dupliquer des protocoles, comme un modèle/template pour un nouveau protocole	2	10	14	10	5	7	5
	Gestion de pièces jointes : bureautiques (PDF, audios, vidéos, etc.) ou scientifiques et techniques (git, chromatogrammes, spectres, images,	3	21	15	21	7	5	7
	Possibilité d'exporter des éléments de l'ELN à un format imprimable	2	4	4	4	2	2	2
	Accès à des données externes (c'est-à-dire non stockées dans l'ELN), brutes ou non, pour lecture et référencement	2	20	20	20	10	10	10
	Structuration avec moteur et filtre de recherche interne ; intégration de mots-clés (tags)	2	14	12	16	7	6	8
	Contenu accessible (pour fouille/mining)	0	0	0	0	2	2	2
	Possibilité d'associer des métadonnées aux jeux de données manipulés, qu'il s'agisse de métadonnées génériques (dates, auteurs, etc.) ou spécifiques (ontologies, taxonomies, etc.)	2	16	12	16	8	6	8
	Utilisation et maintenance des équipements (y compris remontées	1	3	1	1	3	1	1
	Gestion des stocks (réactifs, lignées biologiques, etc.)	2	6	0	4	3	0	2
	Outils pour des calculs simples (fonctions mathématiques de base, conversion d'unité, etc.)	3	9	9	9	3	3	3
	Outils pour des calculs complexes (statistiques ou analyses nécessitant un langage de programmation)	0	0	0	0	0	10	5
	Outil de construction/dessin de molécules	3	12	3	0	4	1	0
	Notes sur les expériences ou le protocole (pendant l'expérimentation), comptes-rendus, etc.	1	2	1	3	2	1	3
	Possibilité de citer des contenus issus de bibliographies (type labtex.bib	1	3	2	4	3	3	4
	Accès aux outils bureautiques du poste de travail	0	0	0	0	3	3	4
	Gestion des outils logiciels et de codes informatiques (intégration continue, déploiement continu, gestion des versions, etc.)	0	0	0	0	2	10	0
	Gestion du partage par les utilisateurs	2	10	10	10	5	5	5
	Agenda partagé	0	0	0	0	2	1	2
	Édition partagée de documents (en ligne et en simultané par plusieurs utilisateurs)	0	0	0	0	2	1	2
	Annotation de contenu et ajout de commentaires	1	2	1	2	2	1	2
	Sous-Total fonctionnalités	34	187	162	172	100	100	100

FIGURE 1 – Exemple de notation des fonctionnalités

son éditeur étant américain. Bien que les conditions d'utilisation indiquent que l'éditeur est conforme au RGPD sur les données personnelles des utilisatrices et utilisateurs, il ne fait aucune mention concernant la protection des données de recherches contenues dans l'ELN. Nous pouvons donc raisonnablement penser que des données de recherche à caractère personnel ne sont pas protégées. Par contre, ni la documentation, ni le site de l'éditeur n'indiquent si le contenu est chiffré ou pas. L'information est donc manquante.

- Dans les notes à 1, si on considère la fonctionnalité « utilisation et maintenance des équipements », cet ELN permet effectivement d'accéder à l'inventaire des équipements du laboratoire, de déterminer s'ils sont libres ou utilisés, cependant, nous ne sommes pas informés si un équipement a dépassé sa date de contrôle ou pas.
- Pour les cas où les notes sont à 3, c'est-à-dire que l'ELN va au-delà des attentes sur une fonctionnalité ou une caractéristique, si nous prenons en exemple la fonctionnalité « outil de dessin/construction de molécules », cet ELN va beaucoup plus loin que le simple dessin ou construction puisqu'il permet également de vérifier/valider les liaisons atomiques, calculer les masses molaires, etc.

Caractéristiques techniques						Pondération par approche		
Catégorie	Caractéristique	Note attribuée	Note pondérée Exp.	Note pondérée Analyt.	Note pondérée Obs.	expérimentale	analytique	observationnelle
Protection des données	Respecter le RGPD pour les données personnelles des utilisateurs	2	8	8	8	4	4	4
	Respecter le RGPD pour les données de recherche : données personnelles des participants à la recherche si elles sont contenues dans	0	0	0	0	4	4	4
	Avoir un stockage maîtrisé, redondant et sauvegardé (sites distants)	2	8	8	8	4	4	4
Intégrité	Respecter des contraintes réglementaires ou législatives concernant la localisation de l'hébergement des données de recherche	2	8	8	8	4	4	4
	Permettre la signature des documents (hashage)	0	0	0	0	4	4	4
	Permettre l'horodatage des actions et time-stamping des documents	1	4	4	4	4	4	4
	Identifier de manière unique les entrées (protocole, expérience, note, etc.) - UUID	0	0	0	0	4	4	4
	Assurer un versioning/suivi des modifications	1	3	3	3	3	3	3
	Permettre d'implémenter des workflows prédéfinis (par exemple, suite d'actions obligatoires dans la description d'une expérience ou d'un	1	3	3	3	3	3	3
Authentification/sécurité	Permettre une signature électronique des acteurs	1	4	4	4	4	4	4
	Utiliser un système d'authentification sécurisé	2	6	6	6	3	3	3
	Permettre de définir les règles de gestion des utilisateurs : niveau d'autorisation d'accès selon rôles et projets et cycle de vie (départ/arrivée agent, ouverture/fermeture projet, etc.)	1	4	4	4	4	4	4
Sauvegarde des données	Chiffrer le transport des données (TLS, algorithmes conformes RGS)	2	8	8	8	4	4	4
	Chiffrer les données	0	0	0	0	3	3	3
	Intégrer des sauvegardes régulières et des restaurations	2	8	8	8	4	4	4
	Définir les durées de conservation et modalités d'archivage ou destruction des données	2	4	4	4	2	2	2
Confidentialité	Assurer la confidentialité des données exportées pour sauvegarde ou archivage	0	0	0	0	4	4	4
	Préserver la confidentialité des informations pouvant constituer un savoir-faire valorisable ou relever du secret des affaires (préserver la nouveauté en cas d'invention brevetable)	0	0	0	0	4	4	4
	Préserver la confidentialité des informations échangées avec un partenaire dans le cadre d'une collaboration	1	2	2	2	2	2	2
Compatibilité/interop	Etre compatible avec les différents systèmes informatiques (Windows, Linux, etc.) pour la partie serveur	1	3	3	3	3	3	3
	Permettre la gestion (Import/Export) des données sous un format structuré (par exemple en cas de changement de logiciel ou d'évolution	1	4	4	4	4	4	4
	Permettre l'intégration avec un système d'information local (bases de données, annuaires...)	2	8	8	8	4	4	4
	Permettre les liens avec des entrepôts de données	0	0	0	0	3	3	3
	Utiliser, quand ils existent, les standards ouverts des données en entrée ou en sortie pour permettre la réexploitation des informations de façon	1	4	4	4	4	4	4
	Permettre les échanges de données en proposant des services d'intégration et d'exposition, si possible via des WebService ou API REST	1	4	4	4	4	4	4
Usage/Utilisation	Pouvoir installer facilement les mises à jour et évolutions, en garantissant l'accessibilité, dans le cas de l'utilisation à partir d'un navigateur Web	2	6	6	6	3	3	3
	Etre ergonomique et intuitif	2	8	8	8	4	4	4
	Permettre la modularité/personnalisation/flexibilité	1	2	2	2	2	2	2
	Disposer des références documentaires (guide, tutoriels, consignes d'utilisation, etc.)	3	12	12	12	4	4	4
	Avoir un code source accessible pour contrôler, si nécessaire, l'évolution de l'outil, sa maintenance et surtout les données qu'il contient	0	0	0	0	3	3	3
	S'appuyer sur une communauté d'utilisateurs	1	3	3	3	3	3	3
	Proposer l'aspect multilingue (a minima l'anglais)	1	4	4	4	4	4	4
	Permettre une utilisation hors-ligne avec synchronisation a posteriori	0	0	0	0	4	4	4
	Etre utilisable en mobilité (sur tablette et smartphone)	2	8	6	8	4	3	4
	Prendre en compte l'empreinte environnementale : conso électrique des serveurs, etc.	0	0	0	0	2	2	2
	Sous-Total caractéristiques	38	136	134	136	123	122	123
		Note attribuée	expérimentale	analytique	observationnelle			
	Grand Total	72	323	296	308			
	Note ramenée sur 100		**72**	**67**	**69**			

FIGURE 2 – Exemple de notation des caractéristiques techniques

4.4.4 Importance relative des différents critères

Si une grille d'évaluation permet d'objectiver au maximum les critères fonctionnels et techniques, rappelons qu'elle ne doit pas constituer la seule base de choix d'un ELN, mais être mise en regard des autres critères évoqués (coût, facilité d'adoption, etc.) L'importance relative de ces différents aspects ne peut être définie qu'avec la connaissance précise du contexte politique et économique du projet d'implantation.

approche expérimentale			
Catégorie	**note par catég**	**note max**	**note rapportée à 5**
Fonctionnalités	187	200	4,7
Protection des données	2	24	0,4
Intégrité	1	44	0,1
Authentification/sécurité	1	22	0,2
Sauvegarde des données	2	26	0,4
Confidentialité	1	20	0,3
Compatibilité/interop	1	44	0,1
Usage/Utilisation	0	66	0,0

approche analytique			
Catégorie	**note par catég**	**note max**	**note rapportée à 5**
Fonctionnalités	162	200	4,1
Protection des données	16	24	3,3
Intégrité	18	44	2,0
Authentification/sécurité	14	22	3,2
Sauvegarde des données	20	26	3,8
Confidentialité	2	20	0,5
Compatibilité/interop	23	44	2,6
Usage/Utilisation	41	64	3,2

approche observationnelle			
Catégorie	**note par catég**	**note max**	**note rapportée à 5**
Fonctionnalités	172	200	4,3
Protection des données	16	24	3,3
Intégrité	18	44	2,0
Authentification/sécurité	14	22	3,2
Sauvegarde des données	20	26	3,8
Confidentialité	2	20	0,5
Compatibilité/interop	23	44	2,6
Usage/Utilisation	43	66	3,3

	Note sur 5 par approche		
Catégorie	**expérimentale**	**analytique**	**observationnelle**
Fonctionnalités	4,7	4,1	4,3
Protection des données	0,4	3,3	3,3
Intégrité	0,1	2,0	2,0
Authentification/sécurité	0,2	3,2	3,2
Sauvegarde des données	0,4	3,8	3,8
Confidentialité	0,3	0,5	0,5
Compatibilité/interop	0,1	2,6	2,6
Usage/Utilisation	0,0	3,2	3,3

FIGURE 3 – Exemple de calculs en vue d'une représentation graphique

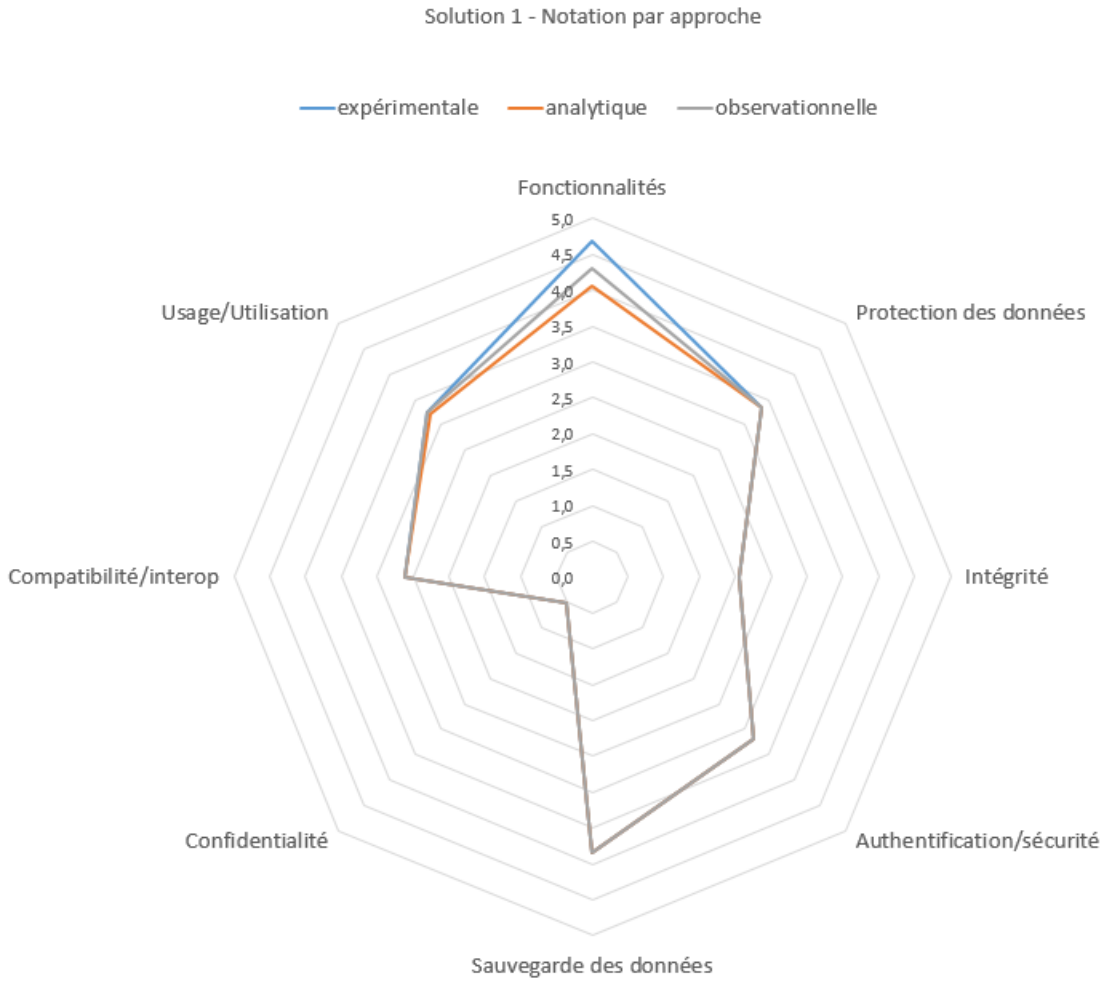


FIGURE 4 – Exemple de représentation graphique par diagramme de Kiviat

5 Liste comparative d'outils

Sur la base des méthodes décrites en section 4, le groupe de travail a comparé un ensemble de solutions du marché.

5.1 Choix des outils étudiés

La liste des outils retenus est présentée dans le tableau ci-dessous (les analyses ont porté sur la version courante de l'outil en mai-juin 2021, sauf indication contraire) :

TABLE 3: liste des outils évalués.

Nom	**Editeur**	**Version**	**Type**
BIOVIA	Dassault Systèmes		Commercial
Chemotion	Karlsruhe Institute of Technology	0.5.0	Open source
ElabFTW_dev	Deltablot	4.0	Open source
ElabFTW_QeR	Deltablot	3.6.6	Open source
Elog			Open source
FindMolecule	FindMolecule, inc.		Commercial
Jogl	Just One Giant Lab MVAC		Open source
Jupyter notebook			Open source
LabArchives	Labarchives LLC		Commercial
LabBook	eNovalys		Commercial
LabCollector	Agilebio	6.032 avec add-on ELN 4.13	Commercial
Labforward	LabForward gmbh		Commercial
LabGuru	Biodata		Commercial
Mbook	mestrelab	2.1.1	Commercial
OSF.io	Center of Open Science		nonprofit organization
RSpace	ResearchSpace		Commercial

Le choix des outils que nous avons étudiés s'est fait en veillant à un équilibre de répartition sur plusieurs critères :

- Entre ceux sous licences libres et commerciaux
- Entre ceux spécialisés dans certains domaines scientifiques (biologie, chimie,etc.) et les génériques
- Entre ceux disponibles uniquement en ligne (*Software as a Service* - SaaS) et ceux que l'on peut héberger dans les établissements (on-premise).

Nous avons également privilégié les outils qui sont déjà utilisés ou déployés dans nos établissements afin d'avoir un regard le plus pertinent possible lors de leur évaluation.

Enfin, nous avons choisi quelques *outsiders* dont la vocation première n'était pas d'être un ELN mais dont l'usage a été détourné.

On trouvera en annexe 5, d'une part, des liens vers les éditeurs des solutions évaluées et, d'autre part, la liste des autres outils identifiés mais qui n'ont pas pu

être analysés faute de temps.

5.2 Résultats

Une vision globale des résultats est présentée sur les figures 5 et 6. Nous avons choisi le mode le plus simple de notation, avec un code de couleur :

- vert = la solution remplit la fonction, plus ou moins bien
- rouge = la solution ne remplit pas la fonction, ou l'information n'est pas disponible

Note : Nous n'avons pas eu les moyens d'évaluer entièrement la solution eLog, d'où une colonne laissée vide au niveau de ses caractéristiques

5.3 Commentaires

Un tableau de ce type, qui ne tient pas compte du poids relatif des critères entre eux ni des exigences propres aux différentes approches de la recherche, peut seulement permettre d'éliminer des solutions qui ne répondraient pas à un critère, ou un ensemble de critères, qui seraient jugés indispensables. Il n'était pas attendu du groupe de travail qu'il donne des recommandations hors contexte sur tel ou tel outil.

La liste des solutions n'est donc pas exhaustive et leur notation imparfaite, d'une part car elle est décorrélée d'une réalité de terrain, et d'autre part car les évaluations ont été conduites en un temps restreint, par des groupes hétérogènes.

Ces évaluations permettent néanmoins de mettre en perspective l'importance de définir correctement les critères de choix et leur poids dans un réel processus d'évaluation.

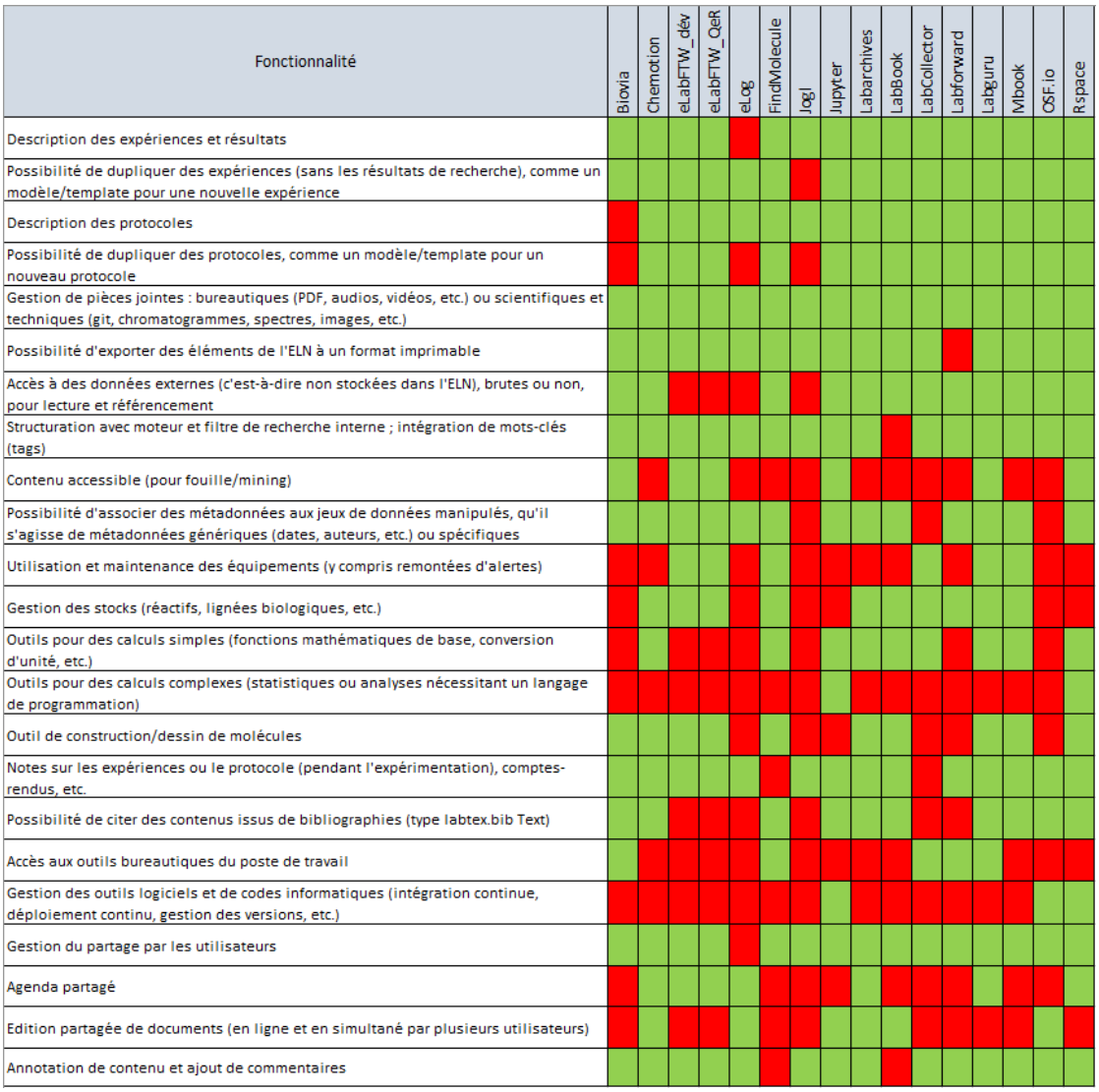


FIGURE 5 – Notation des fonctionnalités pour les 16 solutions évaluées



FIGURE 6 – Notation des caractéristiques techniques pour les 16 solutions évaluées

6 Mener à bien un projet d'implantation d'un ELN

6.1 Introduction

Ce chapitre n'a pas pour but de décrire comment doit se dérouler un projet en général mais plutôt de faire ressortir les particularités et les points de vigilance d'un projet de cette nature.

6.2 Parties prenantes et pilotage du projet

En plus des parties prenantes classiques d'un projet d'implantation d'une solution logicielle : le service chargé des achats, le service juridique, le service des systèmes d'information, il faut impliquer très tôt les services qui bénéficieront en premier lieu des retombées de l'utilisation de l'ELN, notamment :

- Qualité
- Science ouverte
- Valorisation
- Intégrité scientifique
- Archivage
- Sécurité des données
- Gestion des données

Les parties prenantes et les pilotes du projet doivent pouvoir s'appuyer sur des représentant-e-s des futur-e-s utilisatrices et utilisateurs. L'appui de la direction, que ce soit celle d'une structure de recherche ou d'un établissement de niveau national, est un gage de succès de l'aboutissement du projet.

Il peut être difficile de concilier les exigences des juristes, qui doivent veiller à l'application du RGPD, et celles de la Science Ouverte.

6.3 Avant-projet

En phase d'avant-projet, il convient de :

- S'enquérir des recommandations du ou des ministères de tutelle : préconisent-ils ou imposent-ils une solution, une méthode de choix ?
- S'enquérir de la situation dans les autres établissements des éventuelles cotutelles : préconisent-ils ou imposent-ils une solution ?
- Bien caractériser les populations ciblées : par exemple sur la base de la discipline mais également sur la base de l'approche de la recherche (expérimentale, analytique ou observationnelle) ; cf. paragraphe 2.2.2

- Réfléchir au nombre de solutions à retenir : la simplicité voudrait qu'on n'en retienne qu'une mais plusieurs solutions permettent de mieux couvrir les besoins des différentes communautés ; on peut également parvenir à cela en faisant le choix d'une solution simple mais possédant des modules additionnels.
- Réfléchir au caractère obligatoire ou non de l'utilisation du ou des outils retenus.
- Vérifier les contraintes en matière de mise en sécurité des données (hébergement en UE ou en France, par exemple).
- Réfléchir aux règles d'affectation des cahiers : à l'échelle d'un projet de recherche, d'une équipe, d'un sous-groupe, d'une unité ou pour un équipement.

6.4 Phase pilote

Le déploiement d'un ELN est facilité si le choix de la solution a été fait par un panel d'utilisatrices et utilisateurs, durant une phase pilote. Celle-ci peut démarrer par l'utilisation d'un nombre restreint (2 à 4) d'outils sélectionnés lors d'une phase de pré-étude (cf. chapitre 4), par un ensemble de futures utilisatrices et utilisateurs, représentant au mieux la diversité des futurs usages (organisation des équipes, disciplines, approches de la recherche, localisation, etc.) Cette phase peut durer entre un trimestre et une année avec des évaluations périodiques et la réalisation d'enquêtes de satisfaction auprès des utilisateurs.

6.5 Appel d'offres

Les établissements ou laboratoires peuvent s'appuyer sur la grille de fonctionnalités et caractéristiques techniques du chapitre 4 pour alimenter leur cahier des charges, afin d'analyser des outils pouvant répondre à leurs besoins et de s'assurer de ne pas omettre certains éléments. Dans le cas d'une recherche d'outil hébergé par un tiers, le cahier des charges devra exiger que les lieux, conditions d'hébergement, d'accès... soient détaillées.

Il peut être intéressant de demander :

- que la réponse à l'appel d'offres comporte une feuille de route des évolutions prévues pour la solution présentée. Ainsi, on pourra évaluer, par exemple, si la solution va s'enrichir de composants pouvant l'apparenter à un LIMS ou si son intégration dans l'environnement numérique du chercheur sera facilitée.
- qu'on puisse continuer à exploiter l'outil en lecture seule, une fois le contrat terminé.

6.6 Déploiement

Le déploiement de la solution retenue gagne à être effectuée progressivement, pour évaluer la qualité de l'architecture technique et du support et pouvoir ainsi anticiper la montée en charge.

6.7 Changement d'outils

Un point de vigilance d'importance : les contraintes des marchés publics peuvent faire qu'un changement d'outils doive intervenir une fois les possibilités de reconduction épuisées [17] ; cela peut être très fastidieux si la reprise de données ne peut pas être automatisée et/ou si le nombre de cahiers est important. La préservation de l'intégrité des données doit être garantie. Devoir former l'ensemble des utilisateurs à un nouvel ELN peut également avoir un coût, financier et humain, important.

En résumé, un changement d'ELN constituerait un frein à la recherche pendant la phase de transition.

6.8 Accompagnement du changement

Cette activité est un facteur très important de réussite du projet. Pour nombre d'actrices et d'acteurs de la recherche, le cahier de laboratoire est un outil quotidien. Il importe donc que celles/ceux-ci puissent prendre en main rapidement sa version électronique et trouver rapidement des solutions à leurs questions. De plus, certaines caractéristiques de l'ELN peuvent apporter des changements importants dans l'organisation du travail : travail collaboratif (le cahier est le plus souvent partagé par une équipe, visible en permanence par la ou le *Principal Investigator* (PI)), accès à distance (facilite le recours au télétravail), etc. D'où la nécessité de prévoir, pour l'utilisatrice ou utilisateur final-e comme pour l'administratrice ou administrateur du cahier (chef-fe d'équipe, PI), un accompagnement approprié.

Pour commencer, il semble pertinent de (re) sensibiliser les utilisatrices et utilisateurs aux bonnes pratiques de rédaction d'un cahier de laboratoire, qu'il soit papier ou électronique. Un frein tout particulier concerne les équipes qui ne sont pas au départ utilisatrices de cahiers de laboratoire papier car elles considèrent que c'est un outil totalement inadapté à leur activité et n'appréhendent pas l'intérêt que peut avoir un cahier de laboratoire électronique pour la protection et la

17. C'est le cas quand l'objet du marché est l'acquisition d'une solution d'ELN et non la mise en oeuvre d'une solution donnée.

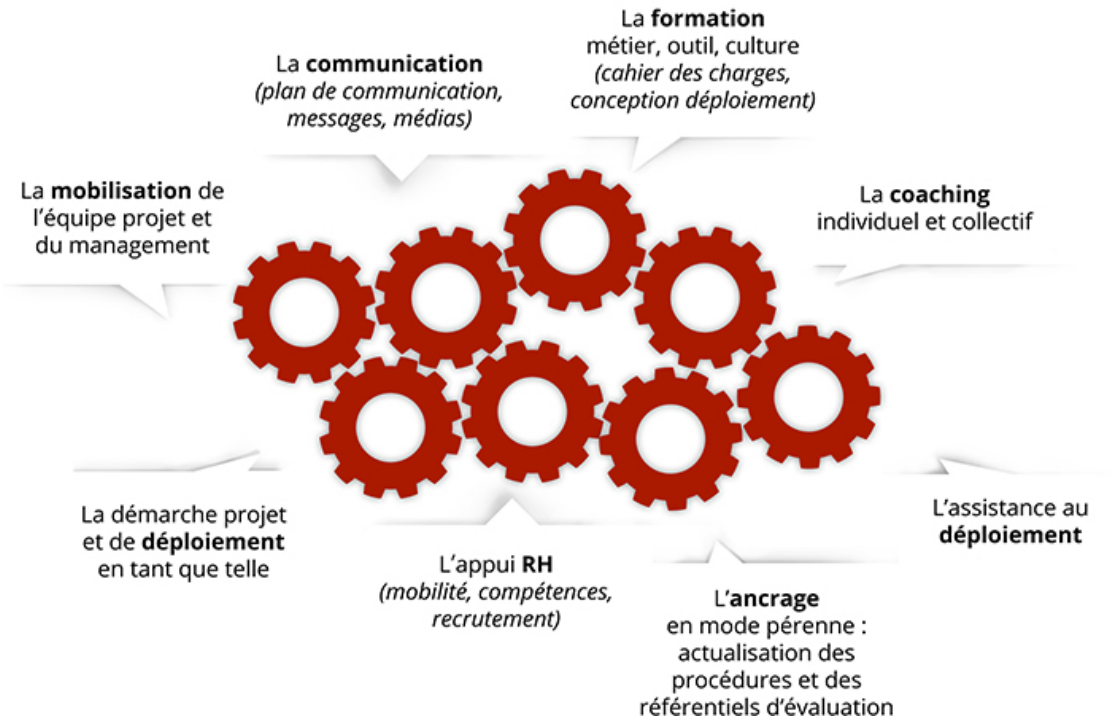


FIGURE 7 – Les facteurs de succès d'un déploiement de solution logicielle *(repris du livre "Pro en Conduite du changement", Julie Ricou et Valérie Moissonnier, 2018, éd. Vuibert)*

valorisation de leurs résultats.

Les craintes principales à l'utilisation d'un tel outil concernent les aspects informatiques au sens large : ergonomie des outils, difficultés de mise en œuvre notamment lors de missions de terrain, les coûts associés à la fois financiers (licences, équipements) et humains (compétences techniques, accompagnement, support) ainsi qu'un manque de confiance dans la fiabilité des outils informatiques et la pérennité des résultats de la recherche.

Cet accompagnement, qui se doit d'être structuré et planifié, concerne deux grandes phases du projet, qui sont :

- l'accompagnement et le dépannage en phase d'installation et de problématiques techniques ;
- l'accompagnement et le service de support en phase d'utilisation en cas de difficultés lors de l'exploitation de l'outil.

Dans le premier cas, plusieurs issues peuvent être envisagées selon l'implémen-

tation de l'outil choisie : ce support peut être proposé en externe (prestataire dédié, solution intégrée avec l'acquisition de la solution) ou prévu en interne au travers de l'assistance des services de systèmes d'informations locaux, nationaux, etc.

Dans le second cas, l'accompagnement peut être proposé en interne par le biais de l'organisation de formations, informations, échanges avec une communauté d'utilisatrices et utilisateurs. L'éditeur de la solution retenue peut, quant à lui, fournir un service support en cas de blocage technique (tchat, service d'aide en ligne, ligne téléphonique, etc.)

Aussi, un accompagnement au changement est à proposer sous différentes formes :

- des actions de communication, de présentations de l'étude et des choix réalisés, de formations individuelles et collectives d'utilisation de l'outil,
- la mise en place de communautés d'utilisatrices et utilisateurs,
- la mise à disposition de documentation et de tutoriels vidéos (dans l'environnement de l'ESRI, une réflexion pourrait être entreprise quant à la mise en place de MOOC (Massive Open Online Course : type ouvert de formation à distance capable d'accueillir un grand nombre de participant-e-s) pour validation du savoir-faire des étudiant-e-s par exemple.

En parallèle, un service de support permettant de répondre aux difficultés et besoins de dépannage des utilisatrices et utilisateurs, que ce soit lors de la phase d'installation ou de l'utilisation de l'outil, se doit d'être formalisé.

6.9 Profils et dimensionnement en RH

Le renouvellement relativement rapide des doctorant-e-s et post-doctorant-e-s dans les laboratoires impose que soit maintenue en place une offre d'accompagnement.

L'établissement se doit de se doter d'une (ou de plusieurs) interlocutrices ou interlocuteurs identifié-e-s en capacité de répondre aux demandes émises autour de la question de la traçabilité et de la gestion des résultats de la recherche (direction ou service, comité spécifique, référent-e technique, etc.), de la sécurité des données, etc.

Une fonction de *Lab Data Manager* (qui dépasse la problématique de l'ELN) peut être encouragée dans les laboratoires, pour gérer les données, paramétrer l'ELN et assurer la continuité de la connaissance de celui-ci par les personnes nouvellement arrivées.

6.10 Situation de co-tutelle

Comment gérer les situations où les co-tutelles d'un laboratoire promeuvent des solutions différentes ? Il s'agirait alors d'inventorier les politiques et stratégies mises en œuvre par les établissements co-tutelles de la structure, si elles existent ; puis de vérifier si une co-tutelle propose une solution électronique, si son utilisation serait pertinente dans la structure, en termes de fonctionnalités techniques et de déploiement (modalités d'accès, modalités de stockage et de sécurité, etc.) Le cas échéant, une analyse de solutions peut être initiée au regard des besoins et attentes identifiées et conformément aux règles (de mise en concurrence ou d'achat par exemple) de la structure.

7 Conclusion

Le développement récent de nombreuses offres d'ELN d'une part, et la volonté des organismes de recherche de se doter d'une solution d'autre part, montrent clairement l'importance grandissante de l'ELN dans le contexte du développement des outils numériques des équipes de recherche, de la mise en qualité des données et des pratiques de recherche. En découle l'importance d'un choix réfléchi du fait de la croissance de l'offre, de la diversité des approches de la recherche, de la nécessité d'interopérer l'ELN avec d'autres outils voire avec d'autres ELN.

Dans cette optique, le groupe de travail a produit :

- Des listes de fonctionnalités et caractéristiques techniques attendues des ELN
- Un exemple de pondération de ces critères en fonction des pratiques : expérimentale, analytique ou observationnelle.
- Un exemple de notation et de représentation graphique pour un ELN, dont la répétition pourrait permettre des comparatifs.
- Le comparatif d'une quinzaine d'outils propriétaires ou open source, en fonction de l'existence ou non des critères recherchés.
- Des recommandations sur la manière de conduire un projet d'implantation d'ELN, au-delà du choix de l'outil.

Ces productions sont disponibles et réutilisables, et constituent les livrables principaux du groupe dont la mission s'est achevée en juillet 2021.

Le présent rapport rend compte de l'ensemble des travaux réalisés, et a été fourni dans sa présente version en septembre 2021.

A Annexes

A.1 Contributeurs hors GT

En plus des membres du groupe de travail, ses copilotes tiennent à remercier pour leur contribution les personnes suivantes :

- Jérémie LEONARD, Chargé de mission traitement des données, Université Lyon 1
- Nicolas ARGENTO, chef d'équipe applications, Ecole polytechnique fédérale de Lausanne
- Philippe RIZAND, directeur de la DSI, Institut Curie
- Nicolas CARPI, concepteur de eLabFTW, UMR144, CNRS et Institut Curie
- Nicolas RENARD, UMR6521, CNRS
- Alain RIVET, UPR530 - Henri VALEINS, UMR5536 - Yaël HERSANT– UMR6296, membres du réseau Qualité en Recherche, CNRS
- Ambra MARI, responsable d'équipe, CNRS et Université Lyon 1
- Frédéric LEROUX, directeur, UMR7042, CNRS
- Auriane DENIS MEYERE, Céline CHARAVAY, Eric FAUDRY - Equipe administratrice de l'ELN, UMR 5075 - IRIG, CEA
- Fanny BRIZZI, cheffe de projet ELN, DSI, Inserm

Les copilotes du GT remercient également le GT du CNRS sur le cahier de laboratoire électronique, qui a partagé ses résultats avant leur finalisation.

A.2 Lettre de mission du GT

La lettre de mission adressée aux pilotes, accompagnée de la feuille de route du groupe de travail, est jointe dans son intégralité sur les 4 pages suivantes.

MINISTÈRE
DE L'ENSEIGNEMENT
SUPÉRIEUR,
DE LA RECHERCHE
ET DE L'INNOVATION

Liberté
Égalité
Fraternité

Direction générale
de la recherche et de l'innovation

Affaire suivie par :
Claire LEYMONERIE
Tél : 01 55 55 80 51
Mél : claire.leymonerie@recherche.gouv.fr

1 rue Descartes
75231 Paris SP 05

Paris, le 17 novembre 2020

Messieurs,

Les cahiers de laboratoire sont des outils scientifiques et juridiques qui permettent de garantir la traçabilité des expériences de laboratoire, de maintenir la continuité du cheminement intellectuel aboutissant aux résultats scientifiques, d'identifier les contributions de chacun et d'établir les dates d'acquisition des résultats. Ils contribuent à la reproductibilité et au caractère cumulatif des connaissances scientifiques et s'inscrivent donc pleinement dans une démarche de science ouverte.

Il importe donc d'éclairer les choix qui sont faits par les équipes scientifiques parmi les différentes solutions actuellement disponibles.

Nous vous confions à cet effet le pilotage d'un groupe de travail qui aura pour objectifs d'établir une vision partagée sur la définition, le périmètre fonctionnel et les usages du cahier de laboratoire électronique, de produire des recommandations sur les critères de choix d'un outil et son interopérabilité avec d'autres outils, dont notamment les entrepôts de données. Vous établirez ainsi une analyse comparative d'outils existants (sans viser à l'exhaustivité) et en vous appuyant sur les études déjà réalisées par certains établissements ou le réseau CURIE.

Vos travaux s'inscriront dans le cadre du Plan national pour la science ouverte et seront menés en lien avec la Direction générale de la recherche et de l'innovation du ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation, et le collège des données de la recherche du Comité national pour la science ouverte. Vous présenterez les résultats des travaux menés devant le Secrétariat permanent de la science ouverte.

Gilles Mathieu
INSERM, Département Systèmes d'Information

Dominique Pigeon
INSERM, Département Systèmes d'information

Tovo Rabemanantsoa,
INRAE, Direction pour la science ouverte

Copies :
Véronique Stoll et Pierre-Yves Arnould, pilotes du collège données de la recherche du Comité pour la science ouverte
Anne Vanet, responsable du groupe de travail « réussir l'appropriation de la science ouverte »

Pièce jointe : Feuille de route du groupe de travail « Cahiers de laboratoire »

Vous veillerez à coordonner vos travaux avec ceux du groupe de travail « réussir l'appropriation de la science ouverte » piloté par Anne Vanet au sein du Comité pour la science ouverte et avec le projet « cahiers de laboratoire électronique » en cours au CNRS. Pour cela, votre groupe de travail comptera dans sa composition un membre nommé au sein du collège de données et du groupe de travail « réussir l'appropriation de la science ouverte ». Plus globalement, la composition du groupe devra autant que possible représenter la diversité disciplinaire et structurelle de l'ESR.

Vous trouverez joint à ce courrier une feuille de route précisant les objectifs et les modalités de fonctionnement du groupe de travail.

Nous vous remercions pour votre engagement dans cette mission et vous prions de recevoir, Messieurs, nos cordiales salutations.



Isabelle Blanc
P/O le bureau de la Science Ouverte

Comité pour la science ouverte
Groupe de travail « Cahiers de laboratoire »
Feuille de route

1. Contexte

Le cahier de laboratoire est un journal de bord utilisé pour détailler au quotidien les activités liées aux projets de recherche, les expériences et les protocoles utilisés. Il permet, notamment devant les offices de brevets, d'établir la preuve de l'invention, de ses inventeurs et donc de ses ayants-droit. Le cahier facilite les démarches qualité, contribue à la reproductibilité des données et des résultats de recherche et répond aux obligations légales et contractuelles.

De par sa position au cœur de l'activité de recherche, le cahier de laboratoire électronique est un outil stratégique, tant au niveau de l'équipe et de la structure, qu'au niveau des tutelles institutionnelles. Il fait partie des outils socles des projets de recherche qui s'inscrivent dans une dynamique de science ouverte. Le choix de l'outil, son paramétrage et surtout son utilisation, sont ainsi d'une importance capitale

2. Objectifs et productions

Le groupe de travail "cahiers de laboratoire électroniques" aura pour objectifs principaux :

- **Établir une vision partagée sur la définition, le périmètre fonctionnel et les objets du "cahier de laboratoire électronique"**. Il s'agira en particulier de clarifier ce qu'un tel outil fait ou ne fait pas, ce à quoi il est adapté, et les outils complémentaires selon le contexte.
- **Produire un ensemble de recommandations sur les critères de choix d'un outil.** Les critères seront établis de manière objective en fonction des besoins métiers, des disciplines et domaines de recherche, ainsi que des contraintes institutionnelles. Ces critères serviront également pour des études comparatives entre les outils et pour mesurer leurs performances et/ou pertinence suivant les contextes d'usages.
- **Produire un ensemble de recommandations relatives à l'interopérabilité entre différents outils.** Dans le contexte académique français multi-tutelles, le choix d'outils différents ne doit pas mener à une étanchéification des travaux de recherche. Les cahiers de laboratoire devront pouvoir s'intégrer dans les écosystèmes informatiques sans difficulté (acquisition de licences spécifiques, actions de rétro-ingénierie...) et permettre la circulation des données (exposition des données) vers des entrepôts de données notamment
- **Établir une liste comparative (non exhaustive) d'outils selon les critères recommandés.** Cette liste pourra être utilisée à titre d'exemple pour la réalisation d'étude de choix.

3. Méthodologie

3.1. Constitution du groupe

Il est proposé de constituer un groupe d'entre 8 et 12 personnes, chacune représentant un ou plusieurs des angles de vue suivants :

- **Vue de terrain** : utilisateurs de cahiers ou potentiels utilisateurs, si possible provenant de communautés scientifiques différentes
- **Vue intégrité scientifique, juridique, propriété intellectuelle** : juristes, spécialistes de la donnée, responsables qualité...
- **Vue technique/infrastructures** : opérateurs de services, familiers avec les notions de performances, sauvegardes, sécurité...

- **Vue technique/données** : data managers, responsables de plateformes, familiers avec les notions de formats, d'interopérabilité, d'archivage...
- **Vue stratégique/valorisation du contenu** : représentants institutionnels ou scientifiques, responsables valorisation...

La composition du groupe devra autant que possible représenter la diversité disciplinaire et structurelle de l'ESR. La répartition par établissement d'origine et discipline scientifique constituera donc un critère de choix pour la constitution finale. Le groupe devra aussi accueillir un membre du collège des données.

3.2. Fonctionnement

Le groupe réalisera ses travaux entre octobre 2020 et juin 2021.
Du fait de la nature distribuée du groupe, il est prévu de tenir la majorité des réunions et ateliers de travail en visioconférence, selon une fréquence qui sera définie conjointement une fois le groupe constitué.
Aux réunions régulières s'ajouteront des points d'étapes et de restitution trimestriels, en janvier 2021 et avril 2021 auprès de l'équipe science ouverte du MESRI et du collège de données. Les conclusions et le livrable final seront produits à la fin des travaux du groupe, en juin 2021.

A.3 Vision détaillés des usages - approche observationnelle

A.3.1 Démarche ethnographique et analyse « qualitative » (sociologie, anthropologie, design, recherche participative, etc.)

Dans des démarches ethnographiques (par exemple en anthropologie ou en sociologie dite « qualitative »), le « journal de bord », le « carnet de terrain » constitue un fil directeur à la fois pour garder trace des informations recueillies tout autant que pour expliciter les raisons de tels ou tels choix effectués lors du travail de recherche. Ces documents jouent un rôle important dans la structuration de la pensée au cours de la recherche. Expliciter ce travail de recueil d'informations tout autant que l'analyse qui en est faite est associé à un critère de qualité et de validité pour les communautés de recherche développant ces démarches (Vanlint, 2019).

Lors du travail d'enquête, les ethnographes emploient par exemple plusieurs cahiers ou des feuilles plus ou moins structurées sous format papier (Weber, 1991) mêlant différentes fonctions (notes d'observation, d'analyse, etc.). Sous format numérique, cela prend souvent la forme d'un ensemble de fichiers de prise de notes de natures diverses comportant des compte-rendus/mémos d'analyse, des notes d'entretiens, des fiches d'observations. Ces prises de notes textuelles (dans un éditeur de texte) peuvent s'accompagner de croquis, de graphiques, d'enregistrement audio ou vidéo. Aujourd'hui, dans un format numérique, peu d'outils « clefs en main » existent pour assembler des notes et fichiers parfois éparses. Des outils génériques de carnet de note en ligne (type Evernote) peuvent être employés. Dans le champ du design, certaines plateformes se développent également pour faciliter une prise de note collective sous divers formats (audio, texte, vidéo) et une « narration collective » d'évènements dans le cadre de projets de recherche participative.

Dans la phase d'analyse des observations obtenues et des informations recueillies, des outils d'analyse qualitative (QDA) sont employés afin d'expliciter les interprétations faites sur le codage des informations provenant d'un corpus de texte (verbatims, etc.)

Quelques exemples :

- Un exemple de carnet papier d'enquête en sociologie dans le cadre d'une recherche participative :
  https ://reflexivites.hypotheses.org/11932
- Le projet « Les carnets nambikwara de Lévi-Strauss – NAMBIKWARA » projet ANR visant le déchiffrage, la numérisation et l'édition critique

des carnets de terrain de Claude Lévi-Strauss rédigés lors de sa seconde expédition chez les Nambikwara (mai 1938-janvier 1939) :
https ://anr.fr/Projet-ANR-18-CE27-0017

- De nombreuses enquêtes numérisées et anonymisées dans le cadre du projet BeQuali (banque d'enquête qualitative en sciences humaines et sociales » :
https ://bequali.fr/fr/
Par exemple, l'enquête qualitative sur « les conditions de vie professionnelle des femmes enceintes » par Anne Marie Devreux (acccompagnée de l'équipe de BeQuali pour l'anonymisation) :
https ://cdsp.sciences-po.fr/fr/ressources-en-ligne/ressource/cdsp_bq_s12/
- L'atelier des chercheurs propose plusieurs prototypes open source utilisés dans le cadre de pratiques d'observation variées et collectives dans des projets de design :
https ://latelier-des-chercheurs.fr/infos
- Recension d'outils visant à assister l'analyse de contenu, l'interprétation de récits de vie ou d'entretiens en sociologie qualitative :
http ://www.squash.ulg.ac.be/logiciels/

A.3.2 Collecte et enregistrement de données dans des environnement contraints (archéologie, sciences de l'environnement)

En archéologie, les carnets de fouille prenaient auparavant une forme permettant le partage d'observations au fil des recherches. Dans les années 1980, ces pratiques observationnelles se sont systématisées et standardisées et ont donné lieu à la mise en place de fiches « papier » proposant déjà une lecture analytique du terrain avec un enregistrement d'informations dans des rubriques définies (Desachy, 2016). Ces informations font aujourd'hui l'objet d'un enregistrement numérique direct par le biais d'application/tablettes de terrain donnant lieu à la création de bases de données relationnelles puis au développement de systèmes d'information géographique de terrain (SIG).

Les pratiques de collecte d'information sur le terrain dans des environnements contraints sont aussi centrales en écologie, en sciences de l'environnement. Ce que l'on appelle aujourd'hui « carnets de terrain électroniques » représentent ici différents supports et outils associés à ces démarches. Il s'agit par exemple de supports matériels mobiles (téléphone, tablette, etc.) développés pour être résistants à des environnements extrêmes. Ces derniers sont associés à des logiciels (*software*) de collecte et d'organisation de données. Plusieurs fonctionnalités sont proposées telles qu'un mode « hors ligne » et de synchronisation des données a posteriori (en cas d'absence de connexion internet sur le terrain), tout autant que des modalités de travail collaboratif afin de remplir collectivement les bases de données et les SIG

du projet de recherche.

Quelques exemples :

- OpenData Kit est une application *open source* pour collecter, gérer et utiliser des données dans le cadre d'environnements contraints.
- Pi4X4un projet de tablette tout-terrain sur la base de l'utilisation d'un Raspberry Py.
  https ://odk-x.org/
- Présentations de différents projets en sciences de l'environnement et en écologie lors d'un atelier « carnet de terrain électronique » en mars 2018.
  https ://rbdd.cnrs.fr/spip.php ?article270
- Un exemple de collecte de données dans le cadre d'un projet de santé publique en Afrique employant des téléphones portables et open data kit.
  https ://www.youtube.com/watch ?v=vB0jHQgVkEs

Les outils au plus proche du travail de recherche d'observation, d'analyse, d'interprétation de terrain et d'enquête prennent des formes variées en fonction des modalités de production de savoirs et de leur validité. L'idée même d'une reproduction d'expérience (présente dans la notion de « cahier de laboratoire ») est antinomique de démarches qualitatives visant à observer et à apporter une interprétation de phénomènes sociaux que ce soit par l'étude de communautés/populations données ou par l'analyse critique de corpus de texte.

On retrouve néanmoins plusieurs besoins que des outils numériques viennent décliner en fonction des usages et des pratiques : utilisation sur des supports mobiles avec des modes « hors ligne », sauvegarde et confidentialité des informations recueillies, moteur de recherche interne, lien avec des entrepôts de données, dimension collective et gestion des droits d'accès.

Pour en savoir plus :

- Baribeau, Colette. 2005. "L'instrumentation dans la collecte de données - Le journal de bord du chercheur" Hors-Série (2) : 17.
- Desachy Bruno. 2016. Du Carnet de Fouilles Aux Systèmes d'information Archéologiques de Terrain : Quelques Remarques Sur l'évolution de l'enregistrement et l'impact de l'informatisation · Médiathèque de La MSH Mondes · Médiathèque MSH Mondes.
  http ://mediatheque.mae.cnrs.fr/s/fr/item/883.
- Gruson-Daniel, Célya. 2019. "Les outils numériques d'enquête." In Guide

décolonisé et pluriversel de formation à la recherche en sciences sociales et humaines. Éditions science et bien commun.
https ://scienceetbiencommun.pressbooks.pub/projetthese/chapter/les-outils-numeriques-denquete/.

- Olivier de Sardan, Jean-Pierre. 1995. "La politique du terrain. Sur la production des données en anthropologie." Enquête. Archives de la revue Enquête, no. 1 (Octobre) : 71–109.
https ://doi.org/10.4000/enquete.263.
- Pecqueux, Anthony. n.d. "L'écoute comme politique de l'enquête : 5/ Écoute acousmatique, ou : Le nez dans le carnet ou dans le guidon ?" Billet. Espaces réflexifs (blog). Consulté le 27 mai 2020.
https ://reflexivites.hypotheses.org/11932.
- Vanlint, Alice. 2019. "Le journal de bord ou de terrain." dans Guide décolonisé et pluriversel de formation à la recherche en sciences sociales et humaines. Éditions science et bien commun.
https ://scienceetbiencommun.pressbooks.pub/projetthese/chapter/le-journal-de-bord-ou-de-terrain/.
- Weber, Florence. 1991. "L'enquête, la recherche et l'intime ou : pourquoi censurer son journal de terrain ?" Espace Temps 47 (1) : 71–81.
https ://doi.org/10.3406/espat.1991.3788.

A.4 Description des critères

Les fonctionnalités et caractéristiques citées en section 4 sont décrites ci-dessous quand cela semble nécessaire :

Fonctionnalité	**Description**
Description des expériences et résultats	
Possibilité de dupliquer des expériences (sans les résultats de recherche), comme un modèle/template pour une nouvelle expérience Description des protocoles	
Possibilité de dupliquer des protocoles, comme un modèle/template pour un nouveau protocole	
Gestion de pièces jointes : bureautiques (PDF, audios, vidéos, etc.) ou scientifiques et techniques (git, chromatogrammes, spectres, images, etc.)	

Fonctionnalité	**Description**
Possibilité d'exporter des éléments de l'ELN à un format imprimable	La conception du logiciel doit permettre l'exportation de l'ensemble de son contenu afin qu'il puisse être imprimé. Cela implique une mise en forme de tous types d'informations (images, schémas, textes...).
Accès à des données externes (c'est-à-dire non stockées dans l'ELN), brutes ou non, pour lecture et référencement	Pouvoir enregistrer des liens vers des jeux de données, idéalement sous forme de DOI (indentifiant unique d'objet numérique)
Structuration avec moteur et filtre de recherche interne ; intégration de mots-clés (tags)	Pouvoir effectuer des recherches simples sur des éléments contenus dans les cahiers, principalement au travers de l'interface utilisateur
Contenu accessible (pour fouille/mining)	Un outil de TDM (text and data mining) peut accéder aux données dans l'ELN pour y faire des recherches.
Possibilité d'associer des métadonnées aux jeux de données manipulés, qu'il s'agisse de métadonnées génériques (dates, auteurs, etc.) ou spécifiques (ontologies, taxonomies, etc.)	
Utilisation et maintenance des équipements (y compris remontées d'alertes)	Une partie GMAO (Gestion de Maintenance Assistée par Ordinateur) doit être proposée à l'utilisateur. Il pourra ainsi gérer les plannings de réservation et d'utilisation des équipements, mais également de la maintenance de ces derniers. Un service d'alertes (équipement non rendu, maintenance préventive ou étalonnage à faire, équipement non revenu de chez le fournisseur...) doit également être présent.

Fonctionnalité	**Description**
Gestion des stocks (réactifs, lignées biologiques, etc.)	Une partie gestion des stocks doit être proposée dans le cahier de laboratoire électronique. Elle permet aux utilisateurs d'avoir une vue d'ensemble et d'anticiper l'approvisionnement de consommables (réactifs, lignées biologiques, produits chimiques...).
Outils pour des calculs simples (fonctions mathématiques de base, conversion d'unité, etc.)	
Outils pour des calculs complexes (statistiques ou analyses nécessitant un langage de programmation)	
Outil de construction/dessin de molécules	
Notes sur les expériences ou le protocole (pendant l'expérimentation), comptes-rendus, etc.	Pouvoir créer des documents simples, associés à d'autres.
Possibilité de citer des contenus issus de bibliographies (type labtex.bib Text)	Idéalement, il faudrait pouvoir établir une connexion avec un référentiel bibliographique, (par exemple PubMed ou HAL).
Accès aux outils bureautiques du poste de travail	Ceci permet de consulter ou modifier des fichiers sans quitter l'ELN. En cliquant sur le fichier présent dans l'ELN, celui-ci se trouve ouvert par l'application correspondante, si elle est présente sur le poste de travail.
Gestion des outils logiciels et de codes informatiques (intégration continue, déploiement continu, gestion des versions, etc.)	Fonctionnalité rarement trouvée dans un ELN, dont l'alternative est la connexion à un serveur spécialisé, du type GitLab ou SourceSup.
Gestion du partage par les utilisatrices et utilisateurs	Chaque utilisateur doit pouvoir en autoriser un autre à accéder aux informations relatives à un projet, une expérimentation. Cela permet une diffusion rapide des informations aux ayants droit, mais en toute connaissance de cause.

Fonctionnalité	**Description**
Agenda partagé	Si un cahier de laboratoire électronique fait référence à par exemple un projet et que cela implique des points d'avancements ou des expérimentations (donc faisant intervenir plusieurs personnes), il est indispensable qu'un agenda partagé soit proposé aux ayants droits.
Edition partagée de documents (en ligne et en simultané par plusieurs utilisatrices ou utilisateurs)	Le cahier de laboratoire électronique doit permettre à plusieurs utilisateurs d'éditer en ligne et en simultané des documents. De cette façon, toutes les modifications sont prises en compte lors d'une édition simultanée et cela évite qu'un document soit bloqué car déjà en cours d'édition.
Annotation de contenu et ajout de commentaires	Il s'agit de faire des ajouts à des descriptions d'expérience ou de protocole, sans créer de nouveau document.

La liste des caractéristiques techniques citée en section 4 est détaillée ci-dessous :

Caractéristique	**Description**
Respecter le RGPD pour les données personnelles des utilisatrices et utilisateurs	
Respecter le RGPD pour les données de recherche : données personnelles des participant-e-s à la recherche si elles sont contenues dans les ELN	
Avoir un stockage maîtrisé, redondant et sauvegardé (sites distants)	
Respecter des contraintes réglementaires ou législatives concernant la localisation de l'hébergement des données de recherche	

Caractéristique	Description
Permettre la signature des documents (hashage)	Il ne s'agit pas de la signature par un individu (cf. critère "Authentification/sécurité/Permettre une signature électronique des acteurs") mais par un programme, qui génère une suite de caractères propres au fichier (son "hash"). Si le fichier est fourni à un tiers, celui-ci doit pouvoir, avec le même mécanisme de signature, retrouver la même suite de caractères (même s'il y a eu chiffrement et déchiffrement), ce qui est la preuve que l'intégrité du fichier a été préservée. cf. https ://www.cnil.fr/fr/securite-chiffrer-garantir-lintegrite-ou-signe
Permettre l'horodatage des actions et time-stamping des documents	
Identifier de manière unique les entrées (protocole, expérience, note, etc.) - UUID	
Assurer un versionnement/suivi des modifications	Chaque action effectuée doit être notifiée par exemple par l'incrémentation d'un numéro de version du document et/ou d'une date de dernière modification. Le versionning garantit qu'aucune action ne puisse être effectuer sans laisser de trace
Permettre d'implémenter des workflows prédéfinis (par exemple, suite d'actions obligatoires dans la description d'une expérience ou d'un résultat)	
Permettre une signature électronique des actrices et acteurs	Chaque acteur du cahier de laboratoire électronique doit pouvoir le signer électroniquement. Comme le veut l'usage, un témoin (responsable d'unité, collègue ne travaillant pas sur le projet ou d'autres équipes...) doit signer régulièrement le cahier afin d'attester de ce qui est écrit et ainsi de garantir la valeur juridique.
Utiliser un système d'authentification sécurisé	

Caractéristique	**Description**
Permettre de définir les règles de gestion des utilisatrices et utilisateurs : niveau d'autorisation d'accès selon rôles et projets et cycle de vie (départ/arrivée agent, ouverture/fermeture projet, etc.)	
Chiffrer le transport des données (TLS, algorithmes conformes RGS)	
Chiffrer les données	Chaque sauvegarde des données doit être protégée par chiffrement afin de garantir la confidentialité de celles-ci. Il est impensable qu'une sauvegarde puisse être utilisée sans aucune restriction, au même titre que le cahier de laboratoire électronique.
Intégrer des sauvegardes régulières et des restaurations	Afin de préserver le contenu du cahier de laboratoire électronique, des sauvegardes régulières doivent être paramétrables par l'utilisateur (fréquence et éventuellement son emplacement) puis automatiquement exécutées. Il doit à l'inverse être possible d'effectuer une restauration de contenu en cas d'erreur.
Définir les durées de conservation et modalités d'archivage ou destruction des données	
Assurer la confidentialité des données exportées pour sauvegarde ou archivage.	
Préserver la confidentialité des informations pouvant constituer un savoir-faire valorisable ou relever du secret des affaires (préserver la nouveauté en cas d'invention brevetable)	
Préserver la confidentialité des informations échangées avec un-e partenaire dans le cadre d'une collaboration	
Être compatible avec les différents systèmes informatiques (Windows, Linux, etc.) pour la partie serveur	

Caractéristique	**Description**
Permettre la gestion (Import/Export) des données sous un format structuré (par exemple en cas de changement de logiciel ou d'évolution administrative)	
Permettre l'intégration avec un système d'information local (bases de données, annuaires, etc.) Permettre les liens avec des entrepôts de données Utiliser, quand ils existent, les standards ouverts des données en entrée ou en sortie pour permettre la ré-exploitation des informations de façon pérenne	
Permettre les échanges de données en proposant des services d'intégration et d'exposition, si possible via des WebService ou API REST	
Pouvoir installer facilement les mises à jour et évolutions, en garantissant l'accessibilité, dans le cas de l'utilisation à partir d'un navigateur Web	
Être ergonomique et intuitif	
Permettre la modularité/personnalisation/flexibilité	
Disposer des références documentaires (guide, tutoriels, consignes d'utilisation, etc.)	

Caractéristique	**Description**
Avoir un code source accessible pour contrôler, si nécessaire, l'évolution de l'outil, sa maintenance et surtout les données qu'il contient	La solution proposée est en open source, c'est à dire que le code source du logiciel développé est mis à disposition sous licence libre et ouverte. Idéalement le code source est disponible sur une forge logicielle et est documenté. La dimension open source peut aussi être associée à une dynamique de développement communautaire. Concernant les données, les CGU et la « politique de données » précisent les conditions d'utilisation, de conservation des données. Si il y a des API, ces dernières sont aussi documentées et également ouvertes.
S'appuyer sur une communauté d'utilisatrices et d'utilisateurs	Il est important que l'utilisateur ait la possibilité en cas d'interrogation ou de problème de contacter d'autres utilisateurs (Par exemple au travers d'un forum en ligne) afin de trouver une solution. A l'inverse, l'utilisateur doit également pouvoir mettre en ligne toute solution trouvée face à une situation bloquante.
Proposer l'aspect multilingue (a minima l'anglais)	Comme dans la plupart des communautés scientifiques, l'aspect international est essentiel. En effet, tout le monde ne parle pas (ou pas suffisamment bien) la langue du pays de rattachement de son laboratoire et il est donc indispensable que tout utilisateur puisse sélectionner la langue qu'il pratique. Ce paramètre ne s'appliquera bien entendu qu'à la structure du cahier de laboratoire électronique et non à son contenu. On peut donc en convenir que la langue anglaise doit être à minima proposée.

Caractéristique	Description
Permettre une utilisation hors-ligne avec synchronisation a posteriori	Lorsqu'un accès au réseau n'est pas possible (par exemple lorsque l'on se trouve en expérimentation en extérieur ou dans un bâtiment non connecté), l'utilisateur doit pouvoir utiliser (de façon hors-ligne) le cahier de laboratoire électronique. Une synchronisation des nouvelles données apportées et/ou des modifications apportées doit ensuite être réalisée une fois sur le réseau.
Être utilisable en mobilité (sur tablette et smartphone)	Le cahier de laboratoire électronique doit pouvoir être utilisé sur des supports mobiles comme une tablette ou un smartphone. Cela implique que l'ergonomie du logiciel soit compatible avec les écrans tactiles et des différentes résolutions d'écran (icônes pas trop petites, clavier numérique...). Il est aussi indispensable que le logiciel puisse fonctionner en mode hors-ligne et pourvu d'une fonction de synchronisation.
Prendre en compte l'empreinte environnementale : consommation électrique des serveurs, etc.	

A.5 Liste des outils analysés et non analysés

La liste des outils évalués en section 5 est rappelée ci-dessous, accompagnée de notes de travail pour chaque outil :

Nom	Editeur	Type	Note
BIOVIA [18]	Dassault Systèmes	Commercial	stockage USA par défaut.

18. https ://www.3ds.com/products-services/biovia/products/laboratory-informatics/electronic-lab-notebooks/

Nom	Editeur	Type	Note
Chemotion [19]	Karlsruhe Institute of Technology	Open source	Développé à l'origine pour la chimie organique, mais est également étendu à d'autres domaines. Le développement de nouvelles solutions spécifiques à une discipline est en cours.
ElabFTW_dev [20]	Deltablot	Open source	Version standard. Des exemples de pratiques sont disponibles [21]
ElabFTW_QeR	Deltablot	Open source	Version du réseau qualité en recherche, du CNRS.
Elog [22]		Open source	outil Open source développé et maintenu par l'ensemble de la communauté des physicien-ne-s des particules et nucléaires.
FindMolecule [23]	FindMolecule, inc.	Commercial	ELN et gestionnaire d'inventaire de produits chimiques. Sous législation Québéquoise, gratuit pour le monde académique
Jogl [24]	Just One Giant Lab MVAC	Open source	Solution de gestion de projet en SaaS ; ne peut pas remplacer un ELN mais peut y amener.
Jupyter notebook [25]		Open source	Permet de développer, documenter et exécuter le code, puis communiquer les résultats.
LabArchives [26]	Labarchives LLC	Commercial	
LabBook [27]	eNovalys	Commercial	Outil utilisé par des unités à la DR10 du CNRS et des universités (Polytechnique, Paris Sud, Strasbourg).

19. https ://www.chemotion.net/ chemotionsaurus/index.html
20. https ://www.elabftw.net/
21. https ://fosdem.org/2021/ schedule/event/open_research_using_elabftw/
22. https ://elog.psi.ch/elog/
23. https ://findmolecule.com/
24. https ://jogl.io/
25. https ://jupyter-notebook.readthedocs.io/en/stable/
26. https ://www.labarchives.com/eln-for-research/
27. http ://www.enovalys.com/book

Nom	Editeur	Type	Note
LabCollector [28]	Agilebio	Commercial	Utilisé au CEA, domaine Biologie essentiellement. https ://labcollector.com/
Labforward [29]	LabForward gmbh	commercial	
LabGuru [30]	Biodata	Commercial	Déployé à l'Inserm. ELN avec quelques propriétés de LIMS.
Mbook [31]	mestrelab	Commercial	pour les chimistes de synthèse.
OSF.io [32]	Center of Open Science	nonprofit organization	
RSpace [33]	ResearchSpace	Commercial	Interfaçage avec entrepôts déjà implémenté - Sous législation britannique.

La liste suivante présente les outils sélectionnés mais qui n'ont pas été évalués par le groupe :

Nom	Editeur	Type	Note
Scinote [34]	scinote llc	Commercial	n'existe qu'en mode cloud
Benchling [35]	benchling	Commercial	utilisé plutôt dans le privé (dont Sanofi, Syngenta, BP)
Labstep [36]	Labstep Ltd.	Commercial	N'existe qu'en mode cloud, gratuit pour le monde académique
ISIA			consortium AnaEE France
Cassandre [37], NVivo [38]	Université de Liège	Open source	Cassandre et NVivo sont les logiciels Open source pour l'analyse de données qualitatives, utilisés en SHS.
Hivebench [39]	Elsevier	Commercial	

28. https ://labcollector.com/
29. https ://www.labforward.io/
30. https ://www.labguru.com/
31. https ://mestrelab.com/software/mbook/
32. https ://osf.io/
33. https ://www.researchspace.com/enterprise/
34. https ://www.scinote.net/
35. https ://www.benchling.com/notebook/
36. https ://www.labstep.com/notebook
37. https ://github.com/Hypertopic/Cassandre
38. https ://qsrinternational.com/nvivo
39. https ://www.hivebench.com/ ?anchor=apps

Nom	Editeur	Type	Note
Bookitlab [40]	Prog4biz Software Ltd	Commercial	
OpenDataKit [41]		Open source	
Elabjournal [42]	Bio-itech - Ependorf	Commercial	Utilisé par l'Institut Pasteur
Laby [43]	KYLI SAS	commercial	Possède certaines caractéristiques d'un LIMS. Développé en France.
Evernote [44]	Evernote corporation	Commercial	

40. https ://bookit-lab.com/
41. https ://getodk.org/#features
42. https ://www.elabjournal.com
43. https ://laby.io/
44. https ://evernote.com/

B Glossaire et abréviations

Terme ou Notion	**Traduction**	**Sens**
ANSSI	Agence Nationale de la Sécurité des Systèmes d'Information	Autorité nationale en matière de sécurité et de défense des systèmes d'information.
API	Application Programming Interface (= interface de programmation applicative)	Ensemble normalisé de classes, de méthodes, de fonctions et de constantes qui sert de façade par laquelle un logiciel offre des services à d'autres logiciels
Checksum, hachage, hash	s.o.	Fonction particulière qui, à partir d'une donnée fournie en entrée, calcule une empreinte numérique servant à identifier rapidement la donnée initiale, au même titre qu'une signature pour identifier une personne.
CKAN	Comprehensive Knowledge Archive Network	Application web permettant le stockage et la distribution de données telles que des tableurs ou le contenu de bases de données.
CoSO	Comité pour la Science Ouverte	Rattaché à la Direction générale de la recherche et de l'innovation du ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation
Dataverse	Comprehensive Knowledge Archive Network	Application web à code source ouvert permettant de préserver, partager, citer, rechercher et analyser des données de recherche.
DMP	Data management plan (= PGD : plan de gestion de données)	Document évolutif qui aide le chercheur ou le chargé de projet de la recherche à définir un plan pour gérer les données utilisées et générées dans le cadre de son activité ou de son projet de recherche.

Terme ou Notion	**Traduction**	**Sens**
DOI	Digital object identifier (= identifiant numérique d'objet)	Mécanisme d'identification de ressources, qui peuvent être des ressources numériques, comme un film, un rapport, des articles scientifiques, mais également des personnes ou tout autre type d'objets.
EBIOS	Expression des Besoins et Identification des Objectifs de Sécurité	Outil complet de gestion des risques de sécurité des systèmes d'information, conforme au RGS et aux dernières normes ISO ; elle permet de les apprécier et de les traiter.
ELN	Electronic Lab Notebook (= cahier de laboratoire électronique)	cf. chapitre 1 du rapport
Emacs	s.o.	Famille d'éditeurs de texte disposant d'un ensemble extensible de fonctionnalités. Il est populaire parmi les programmeurs et plus généralement les personnes ayant des compétences techniques sur les ordinateurs.
ESRI	Enseignement supérieur, Recherche et Innovation	
GED	Gestion électronique de documents	Procédé informatisé visant à définir, concevoir, produire, organiser, diffuser et gérer des informations et des documents papier ou des documents électroniques au sein d'une organisation.
Git	« quand on lui a demandé pourquoi il avait appelé son logiciel "git", qui est à peu près l'équivalent de "connard" en argot britannique, Linus Torvalds a répondu "je ne suis qu'un sale égocentrique, donc j'appelle tous mes projets d'après ma propre personne. D'abord Linux, puis Git."	Logiciel libre de gestion de versions décentralisé.

Terme ou Notion	**Traduction**	**Sens**
GitLab	s.o.	Logiciel libre de forge basé sur Git proposant les fonctionnalités de wiki, un système de suivi des bugs, l'intégration continue et la livraison continue.
GMAO	Gestion de maintenance assistée par ordinateur	Méthode de gestion assistée d'un logiciel destiné aux services de maintenance d'une entreprise afin de l'aider dans ses activités.
HAL	Hyper Article en Ligne	Archive ouverte pluridisciplinaire destinée au dépôt et à la diffusion d'articles scientifiques de niveau recherche, publiés ou non, et de thèses, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
HPC	High performance computing (= calcul de haute performance)	Science des superordinateurs.
ICOS	Integrated Carbon Observation System	Infrastructure de recherche européenne dédiée à l'observation des flux de gaz à effet de serre.
INSB	Institut des sciences biologiques	l'un de ses dix instituts thématiques du CNRS
LIMS	Laboratory Information Management System (= système de gestion de l'information du laboratoire)	cf. Paragraphe 2.2.3 du rapport
ORCID	Open Researcher and Contributor ID (= identifiant ouvert pour chercheur et contributeur)	Code alphanumérique non propriétaire, qui permet d'identifier de façon pérenne et univoque les chercheurs et auteurs de contributions académiques et scientifiques.

Terme ou Notion	**Traduction**	**Sens**
Parangonnage	(en anglais : benchmarking)	Méthode consistant à trouver les entreprises qui réalisent de la manière la plus performante un processus ou une tâche donnée, d'aller les étudier et d'adapter ensuite ce processus à sa propre entreprise.
PI	principal investigator (= chercheur principal)	Désigne, en Amérique du Nord, le titulaire d'une subvention indépendante et le chercheur principal d'un projet de subvention, généralement dans le domaine des sciences ; souvent synonyme de « chef de laboratoire » ou de « chef de groupe de recherche ».
PID	Persistent IDentifier (= identifiant pérenne)	Chaîne de caractères alphanumériques qui désigne de façon pérenne une ressource indépendamment de son emplacement. Cette ressource peut être tout objet réel ou conceptuel (une personne, une structure, une production de recherche...).
PubMed	s.o.	Principal moteur de recherche de données bibliographiques de l'ensemble des domaines de spécialisation de la biologie et de la médecine. PubMed est gratuit et donne accès à la base de données bibliographique MEDLINE, rassemblant des citations et des résumés d'articles de recherche biomédicale. Développé et hébergé par les Instituts Nationaux de la Santé (NIH, Etats-Unis).

Terme ou Notion	**Traduction**	**Sens**
REST	Representational state transfer	Style d'architecture logicielle définissant un ensemble de contraintes à utiliser pour créer des services web. Les services web conformes au style d'architecture REST, aussi appelés services web RESTful, établissent une interopérabilité entre les ordinateurs sur Internet.
RGI	Référentiel général de l'interopérabilité	Cadre de recommandations référençant des normes et standards qui favorisent l'interopérabilité au sein des systèmes d'information de l'administration.
RGPD	Règlement général sur la protection des données (en anglais : GDPR)	Règlement du Parlement européen et du Conseil européen relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données.
RGS	Référentiel général de sécurité	Cadre règlementaire permettant d'instaurer la confiance dans les échanges au sein de l'administration et avec les citoyens.
RStudio	s.o.	Environnement de développement gratuit, libre et multiplateforme pour R, un langage de programmation utilisé pour le traitement de données et l'analyse statistique.
SHS	Sciences humaines et sociales	Ensemble de disciplines étudiant divers aspects de la réalité humaine sur le plan de l'individu et sur le plan collectif

Terme ou Notion	**Traduction**	**Sens**
SourceSup	s.o.	Plateforme de gestion de logiciels destinée aux établissements de recherche et d'enseignement supérieur ; gérée par le groupement d'intérêt public RENATER, elle permet de rendre publics des logiciels dont la licence est libre et d'héberger des projets sans les rendre publics.
TDM	Text and data mining (= fouille de textes et de données)	Processus d'obtention d'informations à partir de matériel lu par une machine ; fonctionne en copiant de grandes quantités de matériel, en en extrayant les données et en les recombinant pour identifier des modèles.
TLS	Transport Layer Security (= Sécurité de la couche de transport)	Protocole de sécurisation des échanges par réseau informatique, notamment par Internet.
UUID	Universally unique identifier (= identifiant unique universel)	Système permettant à des systèmes distribués d'identifier de façon unique une information sans coordination centrale importante.
VRE	Virtual research environment (= environnement virtuel de recherche)	Système en ligne qui aide les chercheurs à collaborer.