



Innovation et R&D

Projet LORIA

dvc DataValue®
CONSULTING

Lab'INNO

Introduction



Rostand Nya Djiki

Directeur BU Innovation & R&D
chez DataValue Consulting

Le Lab'INNO. est un centre d'excellence des DEEP TECH, créé par DataValue Consulting, qui vise à allier toutes nos expertises technologiques et celles d'un réseau d'organisations créatives, afin de construire de nouvelles solutions disruptives.

Depuis bientôt 3 ans, notre cellule participe à de nombreux programmes de recherche et d'innovation français et européens. Nos travaux et notre veille technologique nous ont permis, entre autres, de renouveler pour la 3ème fois le statut JEI (Jeune Entreprise Innovante), marque de reconnaissance et de soutien de l'Etat français. Nous sommes également très optimistes sur la validation prochaine de notre agrément CIR.

En tant que centre de R&D, nous intervenons sur de nombreux domaines d'activité tels que l'e-santé (recherche médicale, santé mentale), l'IoT (logiciels embarqués) et la Cybersécurité (IA pour SD-WAN), pour ne citer que quelques exemples.





Innovation et R&D



DataValue Consulting, société innovante soutenue par Bpifrance, mène de nombreux projets en R&D, porteurs de valeurs et d'innovations autour de l'utilisation du Big Data pour de nombreux secteurs.

Les missions de notre Lab'INNO

- Aider les organisations dans le déploiement de l'ingénierie basée modèles pour le renforcement de l'excellence opérationnelle.
- Contribuer aux initiatives d'interopérabilité et coopération pour la consolidation de l'écosystème français et européen en matière d'outils technologiques.
- Positionner le groupe DataValue Consulting sur le devant de la scène internationale de la production de services numériques multi-domaines.

Mentions et reconnaissances

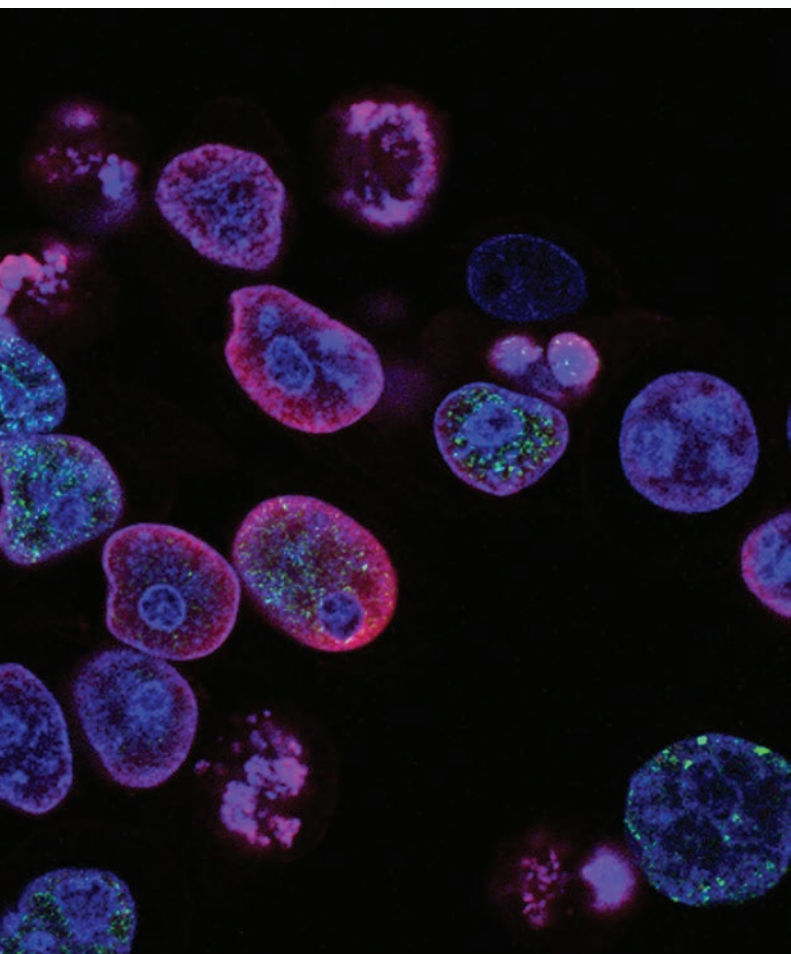


Article

L'état de l'art biomédical augmenté : industrialisation d'un graphe de connaissance sur l'oncologie et le Sars-cov-2

DataValue Recherche & Innovation (DRI Lab'Inno), centre R&D du groupe DataValue Consulting mène de nombreuses expérimentations technologiques alliant l'expertise de la valorisation de la donnée et les enjeux d'excellence opérationnelle de métiers relevant des domaines de l'e-santé, la cybersécurité et la GPEC ⁽¹⁾.

Avec de nombreux programmes R&D à son actif, la DRI Lab'Inno entend contribuer activement aux initiatives d'interopérabilité pour la consolidation de l'écosystème français et européen en matière d'outils stratégiques.



Compte tenu de la situation de pandémie COVID-19 et de sa brutale résurgence, nous avons souhaité prendre l'initiative de fournir un dispositif technologique dont la vocation première est d'aider à l'accélération de la recherche biomédicale. **Ce programme R&D nommé LORIA a pour ambition de proposer une ontologie de connaissance pour l'état de l'art biomédical augmenté**, exposée via une interface interopérable et capable de fournir une situation actualisée des travaux de recherche opérés sur des thèmes médicaux prédéfinis, en l'occurrence, l'oncologie et le Sars-cov-2.

Plus précisément, la promesse incubée est celle d'exposer sous la forme d'un modèle sémiotique (ou graphe de connaissance), **des informations pertinentes et exploitables par l'expert métier en recherche clinique**. Ces informations intelligibles sont extraites d'un grand ensemble d'articles scientifiques traitant d'un thème biomédical prédéfini, tous issus de sources bibliographiques de référence internationale (bioRxiv.org ; pubMed.ncbi.nlm.nih.gov...) fréquemment actualisées par la communauté de pairs.

La spécification formelle sous-jacente à ces informations tire profit d'une **conceptualisation partagée des besoins d'usage pratique par une communauté d'experts cliniciens**. La valeur perçue dans nos opérations de recherche, réside ainsi sur la prise en compte de la connaissance métier de la recherche biomédicale, à chaque niveau de notre démarche de spécification formelle de la conceptualisation partagée des articles scientifiques.

⁽¹⁾GPEC (Gestion Prévisionnelle de l'Emploi et des Compétences) – Nos initiatives portent sur la construction de solutions activables pour une GPEC élargie à l'échelle territoriale, tout en s'accommodant des prérogatives du plan de relance transformation numérique de l'Etat et des Territoires <https://www.economie.gouv.fr/plan-de-relance/appels-projets>.

Plateforme de création et d'enrichissement
d'ontologies pour la recherche biomédicale

PROJET LORIA

Univers de la
connaissance

bioRxiv
THE PREPRINT SERVER FOR BIOLOGY

PubMed

Bibliovid

NCBI
National
Center for
Biotechnology
Information

Experts R&D
cliniciens

Catalogue d'agents
algorithmiques
(offre de services)



API-01

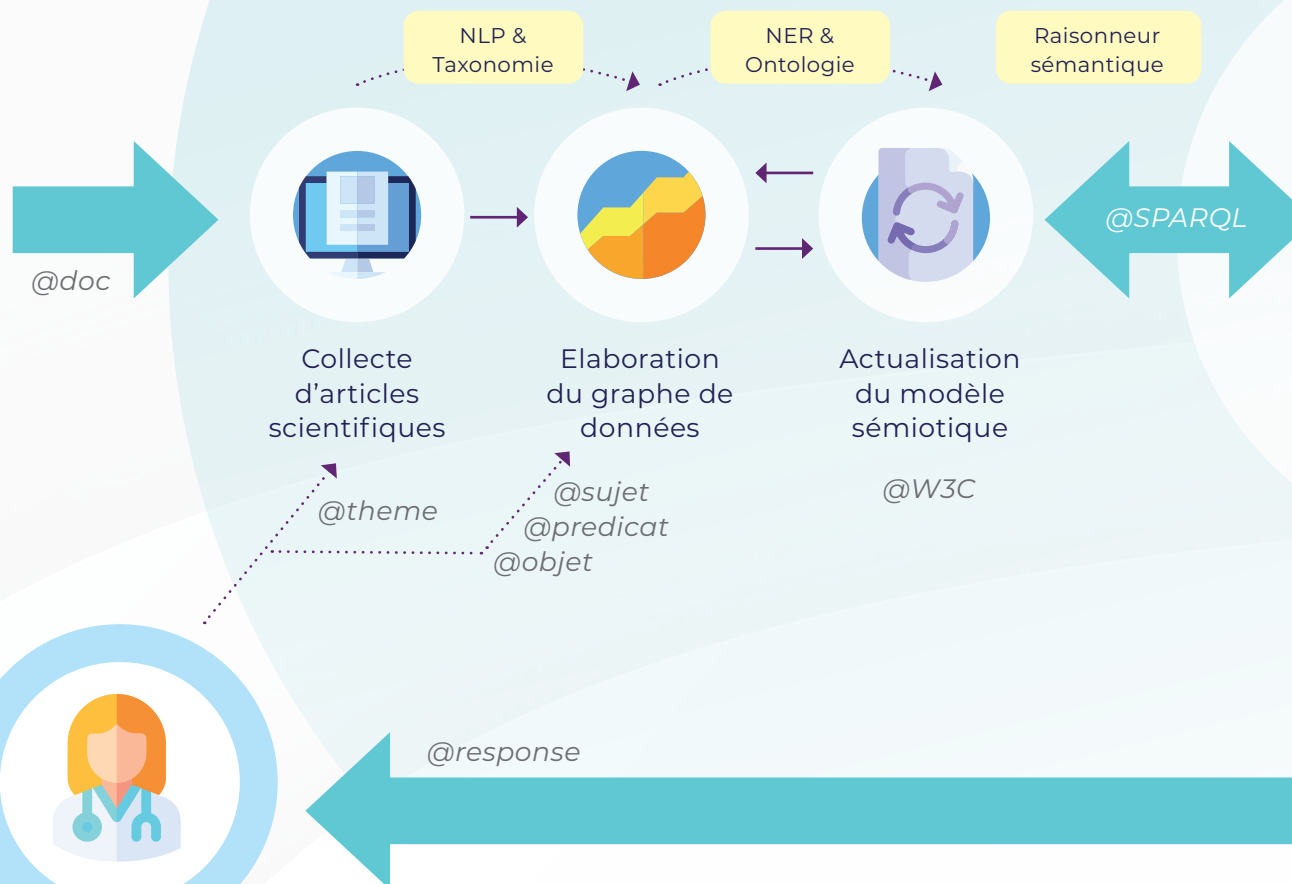


API-02



API-03

IA hybrides
et explicables

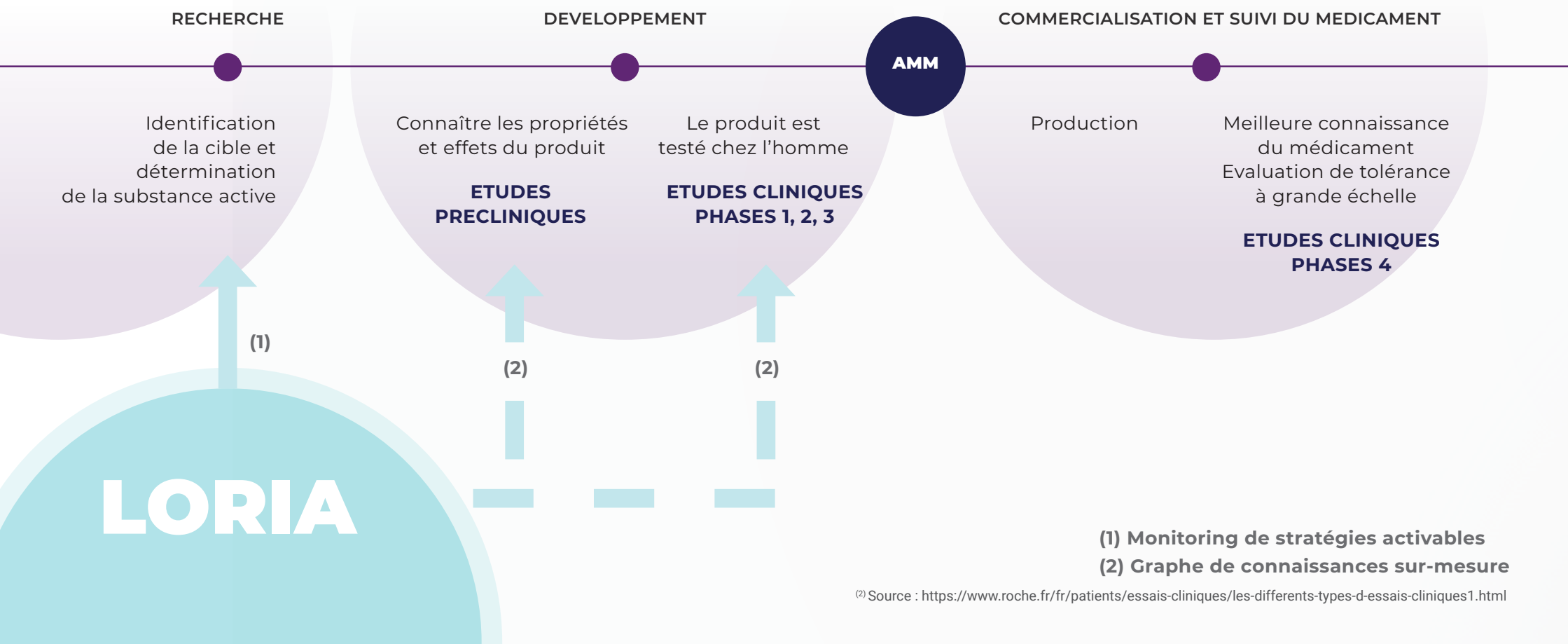




LORIA ambitionne de renforcer l'outillage stratégique pour l'accélération de la recherche clinique en suivant les trois étapes ⁽²⁾ distinctes du développement d'un médicament :

- La recherche,
- Le développement, comprenant les essais pré-cliniques et les essais cliniques
- La commercialisation et le suivi.

Le projet vise la création de valeur sur la chaîne de développement d'un médicament

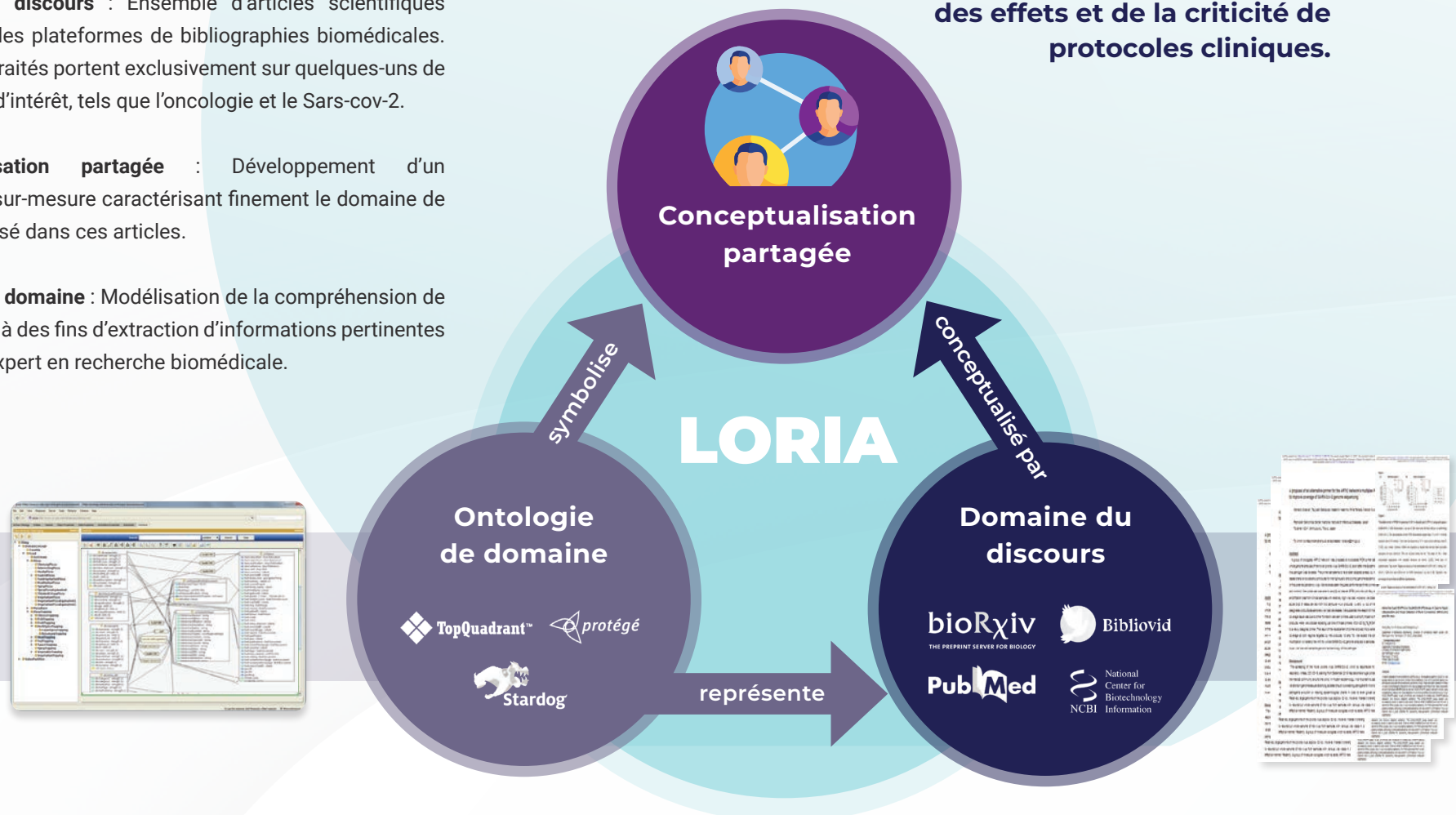


⁽²⁾ Source : <https://www.roche.fr/fr/patients/essais-cliniques/les-differents-types-d-essais-cliniques1.html>

Forts d'une expertise dans la construction d'ontologie de connaissance, nous tentons de construire **un modèle sémiotique** capable d'interopérer avec le web grâce au formalisme d'ontologie sémantique promu par le W3C. Ce modèle se compose de trois principales entités :

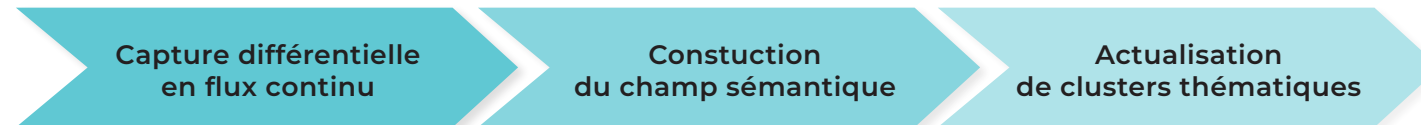
- **Domaine du discours** : Ensemble d'articles scientifiques publiés sur des plateformes de bibliographies biomédicales. Les articles traités portent exclusivement sur quelques-uns de nos thèmes d'intérêt, tels que l'oncologie et le Sars-cov-2.
- **Conceptualisation partagée** : Développement d'un vocabulaire sur-mesure caractérisant finement le domaine de discours utilisé dans ces articles.
- **Ontologie de domaine** : Modélisation de la compréhension de ces discours à des fins d'extraction d'informations pertinentes et utiles à l'expert en recherche biomédicale.

La plateforme LORIA expose des marqueurs rentrant dans l'évaluation des effets et de la criticité de protocoles cliniques.



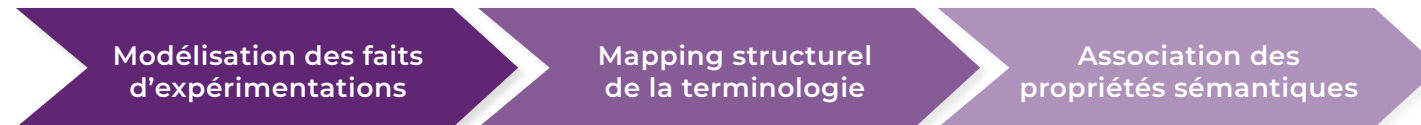
Nous tentons de construire un graphe de connaissances à partir de ces articles scientifiques, sous la forme d'une liste d'énoncés (sujet/prédicat/objet) agglomérés.
Ce graphe de connaissances autoenrichi formalise deux niveaux d'abstraction :

(1) Niveau 1 - Graphe de données : A partir d'une thématique biomédicale définie, par exemple «Sars-cov-2», nous construisons un pipeline de traitements des articles scientifiques correspondants. Ce pipeline a pour but d'élaborer un modèle de liens entre les termes fréquents. **Il s'agit d'une formalisation sémantique des faits scientifiques relevant de la dite thématique.** Soit le pipeline ci-dessous :



Pipeline LORIA pour la modélisation du protocole d'exploitation de la bibliographie scientifique selon les besoins d'usage opérationnel de la recherche clinique.

(2) Niveau 2 - Réseau sémantique : Il s'agit de définir une modélisation des classes des termes à partir de la connaissance métier de la recherche biomédicale. Nous tentons ainsi d'attribuer un sens « métier » aux termes stockés dans le graphe de données. Un expert R&D clinicien nous accompagne dans la formalisation des sujets d'intérêt et leurs attributs escomptés d'une exploitation fine de la bibliographie scientifique de son domaine métier. Le réseau sémantique formel définit donc l'ensemble des classes sémantiques dont les liens sont inspirés de règles régies par la connaissance du métier de la R&D biomédicale.



Pipeline LORIA pour l'enrichissement de la base de connaissance spécifique à une pratique méthodologique d'un quelconque laboratoire de recherche biomédicale.

Ainsi, la combinaison de ces deux niveaux d'abstraction produit les faits et la sémantique extraits d'une base bibliographique centrée sur une unique thématique de recherche clinique.

Notre ontologie sémantique s'appuie sur une norme ouverte à tous les niveaux d'abstraction, ce qui lui confère le caractère interopérable dans le web.

Nos modèles sémantiques sont formels, évolutifs et assurent un bon équilibre entre expressivité et décidabilité. Nous leur associons nos algorithmes d'IA (Intelligence Artificielle) combinant les modèles connexionnistes, symboliques et multi-agents – afin de renforcer l'expérience utilisateur par **la profusion de cas d'usage potentiels** :

- Production de synthèse automatique,
- Annotation instantanée de cooccurrences de faits cliniques,
- Génération d'opportunités d'expérimentations,
- Recherche de facteurs exogènes,
- Génération de séquences combinatoires.

A l'heure où les attentes de la recherche biomédicale sont de plus en plus marquées, le Lab'Inno⁽³⁾ entend contribuer activement, par son projet LORIA, d'une part, à l'effort de relance ; et d'autre part, à la capacité de résilience humanitaire.

Le dispositif projet LORIA se tient ouvert à toute collaboration pour des partenariats de recherche et aussi pour contribuer à l'excellence opérationnelle d'unités scientifiques de laboratoires de recherche clinique.

Fort d'un capital humain qualifié et expérimenté, le centre de recherche et innovation du groupe DataValue Consulting incube depuis plus de cinq années de nombreux programmes de R&D traitant des thèmes du Cloud sémantique, de l'ontologie de connaissance et des méthodes formelles en informatique.

⁽³⁾Lab'Inno : Centre de Recherche, Développement et Innovation du groupe DataValue Consulting, Data Specialist par passion.



Rostand Nya Djiki - Directeur BU Innovation & R&D

Rostand NYA DJIKI est Ingénieur de Recherche spécialisé dans la construction d'algorithmes et de stratégies Data multi-domaines activables.

Spécialiste de la sûreté des systèmes complexes, il rejoint en 2008 un leader mondial de la construction de systèmes pour l'industrie aéronautique et spatiale, avec la responsabilité de développer des solutions de surveillance de la santé de moteurs à réaction. Fort de plusieurs publications scientifiques et de brevets d'invention, fervent passionné de la Data, il rejoint en 2019 un cabinet d'Experts en valorisation de la donnée, où il conçoit des ontologies et méthodologies neuronales.

rostand.nya-djiki@datavalue-consulting.com



DataValue Consulting

5 Boulevard du Général de Gaulle - 92120 - Montrouge

Tél : 01 81 89 41 70 - www.datavalue-consulting.com

contact@datavalue-consulting.com