

Chapitre 3 : Linked Open Data (Données Liées Ouvertes)

1. Introduction

Le Linked Open Data (LOD) est un mouvement qui vise à publier et à relier des données ouvertes sur le Web, en utilisant les standards du Web sémantique. Le LOD permet de créer un réseau global de données interopérables, accessibles et réutilisables, qui peuvent être exploitées par des applications et par utilisateurs humains. Le LOD a de nombreuses applications potentielles dans des domaines variés, tels que l'éducation, la culture, la science, la santé, le journalisme, le gouvernement ou l'entreprise.

2. Linked Open Data

Les Linked Open Data ou LOD est une combinaison de deux concepts : le Linked Data et l'Open Data.

Les **données liées**, ou **Linked Data**, sont un ensemble de bonnes pratiques pour partager des données structurées sur le Web, en utilisant les technologies du Web sémantique. Le Web sémantique est une vision du Web où les données sont liées entre elles par des relations sémantiques, et où les machines peuvent traiter ces données de manière intelligente. Les données liées reposent sur quatre principes énoncés par Tim Berners-Lee :

1. Utilisez des **URI** comme noms pour les éléments.
2. Utilisez des **URI HTTP** afin que les personnes puissent rechercher ces noms.
3. Lorsque quelqu'un recherche une **URI**, fournir des informations utiles en utilisant les normes (**RDF**, **SPARQL**).
4. Incluez des liens vers d'autres **URI**, afin plus d'éléments puissent être découverts [1].

L'**Open Data**, ou **données ouvertes**, est un mouvement qui promeut la publication et la diffusion de données libres de droits, qui peuvent être utilisées et réutilisées sans restriction. L'Open Data repose sur des principes tels que la transparence, la participation et la collaboration. L'Open Data vise à favoriser l'accès à l'information, à stimuler la création de valeur et à résoudre des problèmes sociaux. L'Open Data concerne des domaines variés, tels que les données publiques, les données scientifiques, les données culturelles ou les données personnelles.

L'Open Data peut être publié sous différents formats, tels que des fichiers CSV (Comma-Separated Values), des documents PDF (Portable Document Format) ou des pages HTML (HyperText Markup Language). Cependant, ces formats ne sont pas toujours adaptés pour représenter la structure et la sémantique des données, ni pour permettre leur interconnexion avec d'autres sources de données. C'est pourquoi il est recommandé d'utiliser les données liées pour publier de les données ouvertes, afin de créer des **Linked Open Data**.

Historique et Évolution

Le linked open data est intimement lié à l'évolution du web sémantique. Cette idée trouve ses racines dans le concept initial du World Wide Web, imaginé par Tim Berners-Lee, le pionnier du web, qui a introduit l'idée du web sémantique dans un article publié en 2001. Toutefois, à ses débuts, cette idée a eu du mal à se concrétiser, principalement en raison de la complexité technique sous-jacente. En 2006, Berners-Lee a publié un second article intitulé simplement "Linked Data," ce qui a marqué le début de la popularisation de ce concept et a accru sa visibilité.

Initialement, la majorité des données "ouvertes et liées" sur le Web étaient principalement en anglais, mais il existait déjà des données interconnectées au-delà des barrières linguistiques, notamment grâce aux liens interlangue dans des projets comme Wikipédia, désormais pris en charge par Wikidata.

Progressivement, des liens ont été établis dans d'autres langues, ouvrant ainsi la voie à l'adoption de cette approche dans le reste du monde. Il est important de noter que Wikipédia a été conçu comme un réseau ouvert et interconnecté au niveau international, facilitant la diffusion et l'échange de connaissances dans différentes langues, contribuant ainsi à l'essor du linked open data au-delà des frontières linguistiques [2].

Les avantages des Linked Open Data

Les Linked Open Data constitue ainsi une source de données riche, diversifiée et évolutive, qui peut être utilisée pour alimenter des applications, des services ou des analyses. Le Linked Open Data peut également être considéré comme une forme d'intelligence collective, où les données sont produites, partagées et enrichies par une multitude d'acteurs.

L'adoption des Linked Open Data présente de nombreux avantages, comme l'interopérabilité, la réutilisation, la découvrabilité, la visibilité, l'enrichissement, l'innovation et la transparence des données ouvertes.

- **Interopérabilité** : Les LOD sont conçus pour être interopérables, ce qui signifie que les données peuvent être intégrées et combinées de manière transparente, quel que soit leur origine. Cela facilite l'agrégation de données provenant de sources diverses, améliorant ainsi la qualité de l'analyse et de la prise de décision.
- **Réutilisation** : Les LOD sont conçues pour être réutilisables. Les données sont accessibles via des URI et peuvent être utilisées dans diverses applications sans nécessiter une transformation majeure. Cela réduit le temps et les efforts nécessaires pour exploiter ces données à des fins diverses. Les utilisateurs peuvent tirer parti des données déjà disponibles pour répondre à de nouveaux besoins.
- **Interconnexion** : Les données ouvertes liées sont interconnectées, ce qui signifie qu'elles sont liées à d'autres ressources pertinentes. Cela facilite la découverte de nouvelles sources de données et de liens entre les données, ce qui peut conduire à de nouvelles

perspectives et à une meilleure compréhension. L'interconnexion permet de créer un tissu de connaissances mondial, où les données peuvent être explorées et naviguées de manière plus intuitive.

- **Visibilité** : En publiant des données sous forme de LOD, les organisations peuvent augmenter leur visibilité en ligne. Les données deviennent accessibles à un public mondial, ce qui peut renforcer la notoriété de l'organisation et encourager la collaboration.
- **Enrichissement** : Les LOD permettent d'enrichir les données existantes en les reliant à d'autres sources de données. Cela peut améliorer la qualité et la pertinence des informations en fournissant des contextes supplémentaires.
- **Innovation** : Les LOD offrent un terrain propice à l'innovation. Les développeurs, les chercheurs et les entreprises peuvent utiliser ces données pour créer de nouvelles applications, services et produits. Cela favorise l'innovation dans de nombreux secteurs.
- **Transparence des données ouvertes** : L'utilisation de LOD pour publier des données ouvertes améliore la transparence des données gouvernementales, des données de recherche et d'autres types de données publiques. Cela renforce la responsabilité et la confiance du public.

Utilisation des Linked Open Data

Les Linked Open Data présentent une multitude de bénéfices pour les producteurs, les consommateurs et la société en général [2]:

Pour les producteurs de données :

Les producteurs peuvent être les gouvernements, les entreprises, institutions académiques, organisations à but non lucratif, plateformes en ligne.

- **Amélioration de la visibilité** : En publiant des données sous forme de LOD, les producteurs augmentent leur visibilité et leur accessibilité sur le web, ce qui peut renforcer leur présence en ligne et accroître leur notoriété.
- **Interconnexion des données** : Les producteurs peuvent lier leurs données à d'autres ensembles de données, ce qui permet des analyses plus riches et une meilleure compréhension des informations contextuelles.
- **Innovation et collaboration** : Le LOD favorise l'innovation en permettant aux producteurs de collaborer avec d'autres acteurs, cela peut provoquer des collaborations et de nouvelles possibilités d'affaires.

Pour les consommateurs de données :

- **Accès à une mine d'informations** : Les consommateurs ont accès à une vaste quantité de données provenant de sources diverses, ce qui peut les aider dans leurs recherches, prises de décision et analyses.

- **Données actualisées et fiables** : Le LOD est souvent mis à jour régulièrement, offrant des données fiables et à jour pour des analyses précises.
- **Interconnectivité** : Les consommateurs peuvent explorer des données provenant de sources multiples, ce qui leur permet de croiser des informations pour obtenir une vision plus complète.

Pour la société en général :

- **Innovation et développement** : Les données ouvertes et liées peuvent stimuler l'innovation dans divers domaines, de la recherche scientifique à la gestion des entreprises, favorisant ainsi le développement économique et technologique.
- **Transparence et responsabilité** : Les données ouvertes et liées peuvent renforcer la transparence des gouvernements et des organisations, ce qui peut améliorer la responsabilité et la confiance du public.
- **Solutions aux problèmes complexes** : En combinant et en analysant des ensembles de données divers, le LOD peut aider à résoudre des problèmes sociétaux complexes, tels que la gestion de l'environnement, la santé publique et l'éducation.
- **Facilitation de l'apprentissage automatique et de l'intelligence artificielle** : Les données ouvertes et liées sont essentielles pour former et alimenter des systèmes d'intelligence artificielle, ce qui peut conduire à des avancées significatives dans de nombreux domaines.

Exemples de cas d'utilisation courants :

Les Linked Open Data offrent de nombreux cas d'utilisation courants qui profitent de sa capacité à relier des ensembles de données hétérogènes, à enrichir le contexte sémantique et à faciliter l'accès aux informations. Voici quelques exemples de cas d'utilisation :

- **Recherche d'informations améliorée** : Les moteurs de recherche et les applications de recherche peuvent utiliser le LOD pour améliorer la pertinence des résultats. Par exemple, en intégrant des données liées sur les produits, les entreprises ou les sujets de recherche, les moteurs de recherche peuvent offrir des résultats plus précis et contextuels.
- **Intégration de données** : Les entreprises et les organisations utilisent le LOD pour intégrer des données provenant de sources diverses. Par exemple, une entreprise peut combiner des données de ventes avec des données de clients provenant de sources externes, créant ainsi une vision holistique de ses activités.
- **Enrichissement sémantique** : Les éditeurs de contenu, tels que les sites web d'actualités ou les encyclopédies en ligne, peuvent enrichir leur contenu en utilisant des données liées. Par exemple, un article sur une personnalité publique peut inclure automatiquement des informations sémantiquement liées, comme des liens vers sa biographie, des articles connexes ou des données démographiques.

- **Recommandations personnalisées** : Les plateformes de recommandations, telles que Netflix ou Amazon, utilisent le LOD pour comprendre les préférences des utilisateurs et leur recommander des produits, des films, des livres ou d'autres contenus similaires.
- **Analyse de données avancée** : Les chercheurs et les analystes utilisent le LOD pour effectuer des analyses de données avancées en intégrant des ensembles de données divers. Par exemple, en reliant des données économiques avec des données environnementales, ils peuvent étudier l'impact des politiques sur le changement climatique.
- **Navigation contextuelle** : Les applications de cartographie et de navigation peuvent intégrer le LOD pour fournir des informations contextuelles sur les lieux, les monuments, les attractions, etc. Cela permet aux utilisateurs de mieux comprendre leur environnement.
- **Gestion des connaissances** : Les entreprises utilisent le LOD pour créer des bases de connaissances internes qui relient des informations sur les produits, les processus, les employés et les clients, facilitant ainsi la prise de décision et l'apprentissage organisationnel.
- **Transparence et responsabilité** : Les gouvernements et les organisations publiques utilisent le LOD pour publier des données ouvertes et transparentes sur les finances, les dépenses publiques, la gouvernance, etc., renforçant ainsi la responsabilité et la confiance du public.
- **Assistance à la recherche médicale** : Les chercheurs en médecine peuvent utiliser le LOD pour relier des informations sur les maladies, les médicaments, les essais cliniques et d'autres données pertinentes, accélérant ainsi la recherche médicale.
- **Éducation** : Le LOD est utilisé pour créer des ressources éducatives en ligne interconnectées, offrant aux étudiants et aux enseignants un accès à des informations riches et contextuelles.

Les données ouvertes et liées 5 étoiles

Les données ouvertes liées à cinq étoiles sont un terme inventé par Tim Berners-Lee pour décrire un ensemble de bonnes pratiques pour publier et relier des données sur le Web [3]. Les cinq étoiles sont :

★ : Rendez vos données disponibles sur le Web sous une licence ouverte. Cela permet à quiconque d'accéder et de réutiliser vos données sans barrières juridiques.

★★ : Rendez vos données disponibles sous forme de données structurées, telles que des fichiers Excel ou CSV. Cela permet aux machines de traiter et d'analyser vos données plus facilement.

★★★ : Rendez vos données disponibles dans un format ouvert non propriétaire, tel que CSV ou JSON. Cela garantit que vos données peuvent être accessibles par quiconque, quel que soit le logiciel qu'il utilise.

★★★★ : Utilisez des URI (Uniform Resource Identifiers) pour identifier vos données et les choses qu'elles décrivent. Cela permet à vos données d'être liées à d'autres données sur le Web, créant ainsi un réseau d'informations.

★★★★★ : Liez vos données à d'autres sources de données pertinentes sur le Web, en fournissant du contexte et des informations supplémentaires. Cela enrichit vos données et les rend plus utiles pour les utilisateurs.

3. Recommandations pour la publication de données liées.

Le W3C décrit une série de meilleures pratiques conçues pour faciliter le développement et la diffusion de données ouvertes sous forme de données liées ouvertes. Les données liées ouvertes transforment le World Wide Web en une base de données mondiale (le Web de données). En utilisant les principes des données liées, les développeurs peuvent interroger des données liées à partir de plusieurs sources en même temps et les combiner sans avoir besoin d'un schéma commun unique que toutes les données partagent [4].

Étape 1 : préparer les parties prenantes

Le premier pas consiste à identifier les données que nous souhaitons publier en tant que données liées ouvertes. Cela peut inclure des données gouvernementales, des données de recherche, des données d'entreprise, etc. Il est essentiel d'examiner de manière approfondie les modèles de données, les métadonnées et les données elles-mêmes.

Étape 2 : sélectionner des données

Avant la publication, il est important de nettoyer, structurer et normaliser les données. Cela peut inclure l'élimination des doublons, la résolution de problèmes de qualité des données et la transformation des données dans un format compatible avec les standards RDF (Resource Description Framework). Choisissez des données qui apportent des avantages à d'autres pour la réutilisation. Il faut une vue d'ensemble et préparer une sélection de données et d'informations qui sont utiles pour la publication.

Étape 3 : modéliser les données

La modélisation de données liées implique de représenter des objets de données et leurs relations de manière indépendante de l'application. Modéliser les données consiste à structurer et organiser les informations que l'on souhaite publier sous forme de graphe de données liées ce qui facilite l'interopérabilité, la réutilisation et l'analyse des données par d'autres parties intéressées

Étape 4 : spécifier une licence appropriée

Spécifiez une licence de données ouvertes appropriée. La réutilisation des données est plus probable lorsqu'il existe une déclaration claire sur l'origine, la propriété et les conditions liées à

l'utilisation des données publiées. Il est courant d'utiliser des licences ouvertes, telles que la licence Creative Commons, qui permettent une réutilisation facile des données.

Étape 5 : bons URIs pour les données liées

Le cœur des données liées est une stratégie de nommage URI bien réfléchie et un plan de mise en œuvre, basé sur les URIs http. La prise en compte de la dénomination des objets, le support multilingue, les modifications des données au fil du temps et la stratégie de persistance sont les éléments clés pour des données liées utiles. Attribuez des URI (Uniform Resource Identifiers) aux ressources de vos données, de manière cohérente et significative, pour les identifier de manière unique.

Étape 6 : utiliser des vocabulaires standards

Décrivez les objets avec des vocabulaires précédemment définis chaque fois que possible. Étendez les vocabulaires standard lorsque cela est nécessaire et créez des vocabulaires (uniquement lorsque cela est requis) qui suivent les meilleures pratiques autant que possible.

Étape 7 : convertir les données

Convertissez les données en une représentation de données liées. Cela est généralement fait par script ou d'autres processus automatisés. Sélectionnez les formats de données RDF appropriés pour représenter vos données. Les formats couramment utilisés incluent Turtle, JSON-LD, RDF/XML, N-Triples, etc. Modélisez vos données en utilisant des ontologies et des schémas RDF. Cela implique de définir des classes, des propriétés et des relations pour structurer vos données de manière sémantique.

Étape 8 : fournir un accès machine aux données

Mettez les données RDF en ligne sur un serveur web accessible au public. Assurez-vous que les données sont accessibles en utilisant des URI valides.

Étape 9 : annoncer les nouveaux ensembles de données

Annoncez la publication de vos données liées ouvertes sur des portails de données ouverts, des réseaux sociaux, des listes de diffusion et d'autres canaux pour attirer l'attention des utilisateurs potentiels. Il est important de se rappeler qu'en tant qu'éditeur de données liées ouvertes, un contrat social implicite est en vigueur.

Étape 10 : reconnaître le contrat social

Reconnaissez votre responsabilité dans la maintenance des données une fois qu'elles sont publiées. Assurez-vous que le ou les ensembles de données restent disponibles là où votre organisation dit qu'ils le seront et qu'ils sont entretenus dans le temps.

4. Les Principales Initiatives et Sources de Données Liées

DBpedia¹

DBpedia est un projet majeur du Web sémantique qui vise à extraire des données structurées et sémantiques à partir de Wikipédia. DBpedia a pour principal objectif de transformer le contenu textuel de Wikipédia en une forme de données structurées, de manière à rendre l'énorme quantité d'informations contenue dans Wikipédia accessible de manière plus structurée et interconnectée. Concrètement, DBpedia extrait des informations sur les entités, les concepts, les catégories, les relations et les attributs contenus dans les articles de Wikipédia. En combinant toutes les versions linguistiques, DBpedia (2016-10) contient 6,6 millions d'entités.[5].

Wikidata²

Wikidata est une base de connaissances libre et ouverte qui peut être lue et éditée à la fois par des êtres humains et des machines. Elle sert de stockage central pour les données structurées des projets Wikimedia, y compris Wikipédia.

Wikidata est un graphe de connaissances pour stocker les données sous forme d'énoncés structurés selon un format de triplet (sujet-prédicat-objet). Les éléments de Wikidata représentent des sujets, des concepts, des objets, des événements, des personnes, etc. Chaque élément possède un identifiant unique, des libellés dans plusieurs langues, des descriptions, des alias et des énoncés.

Par exemple, l'élément Wikidata pour Berlin a l'identifiant Q64 et contient des énoncés sur la population de Berlin, son emplacement, son maire, etc. Le contenu de Wikidata est disponible sous une licence Creative Commons CC0 qui permet la réutilisation, la modification et le partage libres des données. La base de données sous-jacente et l'interface d'édition sont en open source.

Wikidata a été lancée en 2012 et est soutenue par la Wikimedia Foundation. En 2022, elle contenait plus de 92 millions d'éléments dans son graphe de connaissances. Wikidata sert de source importante de données structurées pour de nombreuses autres plateformes en ligne.

OpenCyc³

OpenCyc est un projet d'intelligence artificielle open source visant à rassembler une ontologie complète et une base de connaissances de la connaissance du sens commun de la vie quotidienne. Le projet a été lancé en 1984 par Cycorp dans le but de développer une machine dotée de capacités de raisonnement de niveau humain. La base de connaissances Cyc a été gardée privée jusqu'en 2001, date à laquelle une version open source appelée OpenCyc a été publiée.

¹ <https://www.dbpedia.org/>

² <https://www.wikidata.org/>

³ <https://cyc.com/>

OpenCyc contient plus de 239 000 concepts, 2 093 000 faits et 4 301 relations. Il possède des connaissances dans des domaines tels que la science, la médecine, la géographie, et bien d'autres.

Google Knowledge Graph :

Google Knowledge Graph a été lancé en 2012 pour améliorer la recherche Google avec des informations de recherche sémantique sur les personnes, les lieux et les choses.

Le graphe de connaissances contient plus de 500 millions d'entités et 3,5 milliards de faits à leur sujet. Les données sont collectées à partir de diverses sources publiques telles que Wikipedia, le CIA World Factbook, Wikidata, ainsi que des données structurées provenant de pages web.

Il alimente les panneaux d'informations qui apparaissent sur le côté droit des pages de résultats de recherche Google. Ils fournissent un résumé de l'entité et des attributs tels que la date de naissance, l'emplacement, etc.

Le Knowledge Graph est constamment mis à jour à mesure que de nouvelles informations deviennent disponibles à partir de sources publiques et que de nouveaux schémas sont ajoutés.

Freebase :

Freebase était un grand graphe de connaissances collaboratives lancée en 2007 et développée par Metaweb. Elle contenait plus de 40 millions de sujets et 2,4 milliards de faits.

Le but était de construire une base de connaissances sur le monde de façon collaborative, couvrant de nombreux domaines. Freebase était alimentée par des contributions provenant à la fois d'utilisateurs humains et de sources de données structurées comme Wikipedia.

Freebase a été largement utilisée par les chercheurs en IA pour entraîner des systèmes d'apprentissage automatique. En 2010, Metaweb a été racheté par Google, qui a par la suite intégré Freebase dans sa base de connaissances Knowledge Graph en 2014. Freebase n'est plus activement développée depuis son intégration à Google, mais elle a joué un rôle important en démontrant la faisabilité des bases de connaissances collaboratives à grande échelle.

YAGO :

YAGO (Yet Another Great Ontology) a été développée à l'Université de Leipzig en Allemagne à partir de 2007. Le graphe contient plus de 10 millions d'entités et 120 millions de faits reliant ces entités.

YAGO exploite les infoboxes et catégories de Wikipedia pour en extraire des informations structurées. Elle utilise également WordNet pour typer sémantiquement les entités et GeoNames pour ajouter des informations géographiques.

YAGO est conçue pour être une source de connaissances précises et de haute qualité, plutôt que de viser l'exhaustivité.

Microsoft Academic Graph ⁴:

Microsoft Academic Graph est un graphe de connaissances contenant des informations sur plus de 200 millions de publications académiques et scientifiques. Le graphe a été développé par Microsoft Research en 2015 et couvre des publications de diverses disciplines : sciences, sciences sociales, arts, humanités, etc.

Le graphe contient des métadonnées détaillées sur les publications : auteurs, affiliations, citations, domaines de recherche, conférences/journaux. Ces métadonnées permettent d'établir des liens entre chercheurs, publications, institutions, domaines de recherche. Microsoft Academic Graph fournit donc un graphe de connaissances riche du monde académique. Les données proviennent de multiples sources en open access : PubMed, arXiv, ACM, IEEE, Web of Science, etc.

Facebook Entity Graph:

Le Graph Social de Facebook, développé depuis 2010, constitue une base de connaissances inestimable qui capture les intérêts variés des utilisateurs. Avec plus de 3 milliards d'entités et leurs connexions, ce graph couvre un large éventail de centres d'intérêt, allant des célébrités aux marques, produits et lieux. L'objectif de cette ressource est de mieux comprendre les intérêts personnels de chaque utilisateur et les relations entre ces entités. Grâce à cela, il améliore le ciblage publicitaire en permettant des annonces plus pertinentes et les recommandations personnalisées sur Facebook. Les données proviennent des profils et des activités des utilisateurs sur Facebook et Instagram, alimentant ainsi constamment le graphe en élargissant ses entités. Le graphe de Facebook est un exemple puissant de la manière dont les entreprises utilisent la modélisation des données pour offrir une expérience utilisateur plus riche et plus personnalisée.

GeoNames

GeoNames est une base de données géographiques fournissant des informations sur plus de 25 millions de lieux dans le monde. Elle a été lancée en 2006 et est gérée par l'organisation à but non lucratif GeoNames. Elle couvre une large gamme de types de lieux : pays, régions administratives, villes, villages, montagnes, rivières, monuments, etc. Pour chaque lieu nommé, GeoNames fournit des informations comme les coordonnées géographiques, la population, l'élévation, etc.

Les données proviennent de sources officielles gouvernementales et autres sources faisant autorité dans chaque pays. Elle peut être utilisée pour la géocodification, l'enrichissement de contenus, des applications cartographiques, etc.

⁴ <https://www.microsoft.com/en-us/research/project/microsoft-academic-graph/>

Elle est disponible en licence libre, permettant une large réutilisation des données. GeoNames fournit l'une des sources de données géographiques les plus complètes et faisant autorité au niveau mondial.

Linked Open Data Cloud⁵

Le Linked Open Data Cloud, est une initiative mondiale visant à interconnecter des ensembles de données structurées et sémantiques sur le web de manière à créer un réseau mondial de données liées entre elles. Le Linked Data Cloud est conçu pour être une toile mondiale de données, similaire à la manière dont le World Wide Web (WWW) est une toile mondiale de documents. Cependant, au lieu de simplement relier des pages web, il relie des ensembles de données structurées et sémantiques. Cela signifie que chaque donnée liée peut être connectée à d'autres données liées, créant ainsi un réseau mondial d'informations interconnectées. Il est composé de milliers d'ensembles de données provenant de divers domaines, y compris la culture, les sciences, la santé, la géographie, les transports, et bien d'autres. L'objectif ultime du Linked Data Cloud est de permettre aux chercheurs, aux développeurs et au grand public de tirer parti de cette toile mondiale de données pour des applications telles que la recherche avancée, l'analyse de données, la création de connaissances et le développement de services innovants. Cette initiative vise à rendre les données plus accessibles, interconnectées et utiles pour un large éventail d'utilisateurs et de cas d'utilisation.

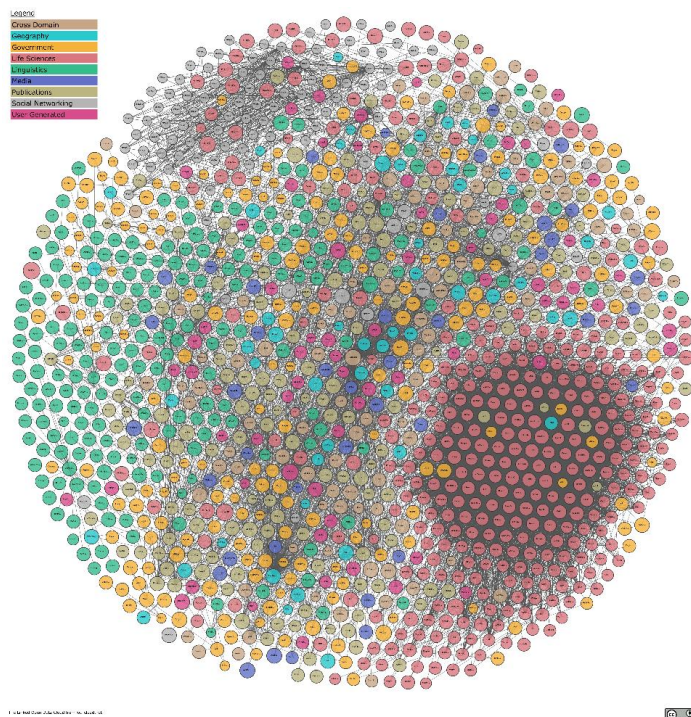


Figure 1 Le Linked Open Data Cloud

⁵ <https://lod-cloud.net/>

5. L'Écosystème des Données Liées Ouvertes

L'économie des données liées ouvertes représente un potentiel de création de valeur importante grâce à l'exploitation des données ouvertes et interconnectées. L'économie des données liées ouvertes offre des opportunités économiques significatives en favorisant l'innovation, en réduisant les coûts, en créant des marchés de données et en améliorant la productivité. De nombreux secteurs d'activité bénéficient des LOD en exploitant les avantages de l'accessibilité, de l'interopérabilité et de la réutilisation des données.

- Santé : Les données liées ouvertes dans le domaine de la santé permettent la recherche médicale, la surveillance épidémiologique et l'amélioration des soins de santé.
- Transport et Logistique : Les LOD aident à optimiser les chaînes d'approvisionnement, à améliorer la gestion du trafic, et à fournir des informations en temps réel aux voyageurs.
- Éducation : Les LOD soutiennent la création de ressources pédagogiques ouvertes, la personnalisation de l'apprentissage et la recherche académique.
- Gouvernement : Les gouvernements utilisent les LOD pour accroître la transparence, améliorer la prestation de services publics et faciliter l'engagement citoyen.
- Finance : Les données liées ouvertes sont utilisées pour l'analyse financière, la détection de fraudes et la création de produits financiers innovants.
- Tourisme : Les informations touristiques interconnectées permettent de créer des guides de voyage personnalisés et de faciliter la planification de voyages.
- Recherche Scientifique : Les LOD favorisent la collaboration entre chercheurs, la publication ouverte des données et la découverte de nouvelles connaissances.

Défis et Questions Éthiques

L'adoption et l'utilisation des données liées ouvertes (LOD) présentent de nombreux défis, notamment en ce qui concerne la qualité des données, la gouvernance, la confidentialité et la sécurité. De plus, des questions éthiques importantes sont soulevées par la collecte et l'utilisation de ces données. Voici une exploration de ces aspects :

Qualité des Données :

- Les données liées et ouvertes proviennent de sources diverses, ce qui peut entraîner des variations de qualité, de précision et de cohérence des données.

Gouvernance :

- Déterminer la propriété des données et les droits d'accès peut être complexe, en particulier lorsque les données sont collectées et partagées par de multiples parties.
- La gouvernance des LOD exige des mécanismes pour gérer les mises à jour et l'attribution de responsabilités.

Confidentialité et Sécurité :

- Les LOD peuvent contenir des informations sensibles, ce qui nécessite des mesures de sécurité pour prévenir l'accès non autorisé.

- La mise à disposition de données ouvertes peut créer des vulnérabilités potentielles, telles que la divulgation involontaire d'informations personnelles.

Éthiques :

- La collecte de données LOD peut soulever des questions de consentement des individus, en particulier si les données personnelles sont impliquées.
- La protection de la vie privée et le respect des droits des individus sont des préoccupations éthiques majeures.
- L'utilisation des données LOD à des fins éthiques, notamment en évitant la discrimination ou les préjugés, est essentielle.

6. Résumé

- ✓ Le linked open data (LOD) vise à publier et interconnecter des données ouvertes sur le web en utilisant les principes des données liées.
- ✓ Initié par Tim Berners-Lee, le mouvement a pris son essor dans les années 2000, notamment grâce à des initiatives comme Wikipédia/Wikidata.
- ✓ Les LOD présentent de nombreux avantages : interopérabilité, réutilisation, enrichissement, innovation, transparence.
- ✓ Les LOD bénéficient aux producteurs et consommateurs de données ainsi qu'à la société dans son ensemble.
- ✓ De nombreux cas d'usage existent : recherche, intégration de données, recommandations, navigation contextuelle, IA.
- ✓ Les principales sources de LOD sont Wikidata, DBpedia, Geonames, etc.
- ✓ La publication de LOD implique des étapes comme la préparation, la modélisation, l'utilisation de standards et de bons URIs.
- ✓ Des défis subsistent concernant la gouvernance, la qualité, la confidentialité et l'éthique des LOD.

Bibliography

- [1] W3C, "LinkedData," W3C, [Online]. Available: <https://www.w3.org/DesignIssues/LinkedData.html>.
- [2] "Benefits of the Linked Data Approach," W3C, 20 September 2011. [Online]. Available: <https://www.w3.org/2005/Incubator/lld/wiki/Benefits>. [Accessed 20 October 2023].
- [3] "5 ★ OPEN DATA," [Online]. Available: <https://5stardata.info/en/>. [Accessed 15 October 2023].
- [4] "Best Practices for Publishing Linked Data," W3C, 09 January 2014. [Online]. Available: <https://www.w3.org/TR/ld-bp/>. [Accessed 20 October 2023].
- [5] "DBpedia," DBpedia, 2023. [Online]. Available: <https://www.dbpedia.org/>. [Accessed 20 October 2023].