



Concept des données liées au seins de l'Administration fédérale

Classification	Non classifié
Statut	Usage approuvé
Nom du programme	Stratégie Administration fédérale numérique Axe prioritaire 1 : Penser et concevoir l'administration numérique en réseau
Responsable de l'axe prioritaire	Karen Dijkstra, ChF
Version	1.0
Date	2 août 2024
Mandant	Transformation numérique et gouvernance de l'informatique (TNI)
Autrice	Karen Dijkstra, ChF
Conseil et révision	Conseil de l'architecture de la Confédération (ABB), Jean-Luc Cochard (AFS), Secrétariat OGD, Pasquale Di Donato (Swisstopo), Michael Luggen (ChF), Jörg De Bernardi (ChF), Elmar Ploskonka (ChF), Daniel Keller (ChF), Camilla Glagau (ChF),
Approuvé par	Andreas Spichiger (ChF)

Tableau des modifications

Date	Modification	Autrice / Auteur
05.07.2023	Création du document	Karen Dijkstra
07.07.2023	Description haut niveau des cas d'études sur la base des informations livrées par Michael Luggen	Karen Dijkstra
12.07.2023	Révision description ElCom	Karen Dijkstra
13.07.2023	Révision suite aux remarques et conseils de Michael Luggen	Karen Dijkstra
18.07.2023	Révision géodonnées (ch. 6.1.4)	Pasquale Di Donato
04.08.2023	Révision suite aux remarques et conseils de Jean-Luc Cochard	Karen Dijkstra
17.04.2023	Prise en compte des commentaires de Daniel Keller	Karen Dijkstra
04.09.2023	Fin de l'intégration des remarques de Jean-Luc Cochard	Karen Dijkstra
12.09.2023	Révision suite aux remarques du secrétariat OGD (Michèle Spichtig et Maik Roth)	Karen Dijkstra
15.02.2024	Allègement du document pour publication	Karen Dijkstra
16.05.2024	Rédaction des informations liées à la gouvernance (ch. 2)	Karen Dijkstra
05.06.2024	Révision et intégration des remarques suite à la consultation	Karen Dijkstra
27.06.2024	Soumission du document à l'ABB	Karen Dijkstra
02.08.2024	Version validée	Andreas Spichiger

Table des matières

MANAGEMENT SUMMARY	1
1 INTRODUCTION	2
1.1 PRÉSENTATION DU SUJET ET JUSTIFICATION DE SON IMPORTANCE.....	2
1.2 OBJECTIFS DU RAPPORT	2
2 GOUVERNANCE DES DONNÉES LIÉES	3
2.1 GOUVERNANCE ACTUELLE DANS L'ADMINISTRATION FÉDÉRALE	3
2.2 DÉFIS DE LA GOUVERNANCE FÉDÉRALE	4
2.3 CENTRALISATION VERSUS DÉCENTRALISATION	5
2.4 TRANSITION VERS L'UTILISATION DES DONNÉES LIÉES	6
3 IMPORTANCE DES DONNÉES LIÉES POUR L'ADMINISTRATION FÉDÉRALE	7
3.1 LA TRANSFORMATION NUMÉRIQUE DE L'ADMINISTRATION FÉDÉRALE PAR L'UTILISATION DES DONNÉES LIÉES.....	7
3.2 COMMENT LES DONNÉES LIÉES PEUVENT AIDER L'ADMINISTRATION FÉDÉRALE À ATTEINDRE SES OBJECTIFS	9
4 AVANTAGES DES DONNÉES LIÉES POUR L'ADMINISTRATION FÉDÉRALE	10
4.1 AMÉLIORATION DE L'EFFICACITÉ ET DE LA PRODUCTIVITÉ	10
4.2 AUGMENTATION DE LA TRANSPARENCE ET DE L'ACCESSIBILITÉ DES INFORMATIONS	10
4.3 FACILITATION DE LA COLLABORATION ENTRE LES DIFFÉRENTS DÉPARTEMENTS ET ENTITÉS	11
4.4 AUTRES AVANTAGES	12
5 DÉFIS ET LIMITATIONS	13
5.1 COMPLEXITÉ TECHNIQUE ET BESOIN DE COMPÉTENCES SPÉCIALISÉES	13
5.2 QUESTIONS DE SÉCURITÉ ET DE CONFIDENTIALITÉ.....	13
5.3 AUTRES DÉFIS	14
6 PRÉSENTATION DE CAS D'ÉTUDES OU D'EXEMPLES D'UTILISATION RÉUSSIE DES DONNÉES LIÉES	14
6.1 INFRASTRUCTURES ET PLATEFORMES PERTINENTE POUR L'ADMINISTRATION FÉDÉRALE	14
6.2 SUCCÈS RENCONTRÉS AVEC LES DONNÉES LIÉES.....	19
7 RECOMMANDATIONS POUR LA MISE EN ŒUVRE DES DONNÉES LIÉES	23
7.1 FACTEURS DE RÉUSSITE POUR UNE INTÉGRATION EFFICACE DES DONNÉES LIÉES	23
7.2 RÔLES ET RESPONSABILITÉS DANS LE PROCESSUS DE RÉUTILISATION DES DONNÉES	25
7.3 BASES LÉGALES	26
8 VISION FUTURE DE L'UTILISATION DES DONNÉES LIÉES DANS L'ADMINISTRATION FÉDÉRALE	26
8.1 COMMENT L'UTILISATION DES DONNÉES LIÉES POURRAIT ÉVOLUER À L'AVENIR.....	26
8.2 OPPORTUNITÉS POTENTIELLES POUR UNE UTILISATION PLUS LARGE DES DONNÉES LIÉES	27
9 CONCLUSION	27
ANNEXE A - L'ÉVOLUTION DU WEB AU FIL DU TEMPS	29
LE WEB 1.0	29
LE WEB 2.0	29
WEB 3.0 ET LE WEB SÉMANTIQUE	30

LE FUTUR WEB 4.0	30
ANNEXE B - COMPRÉHENSION DES DONNÉES LIÉES	32
DÉFINITION ET EXPLICATION DU CONCEPT DES DONNÉES LIÉES	32
MODÉLISATION, REPRÉSENTATION ET FONCTIONNEMENT DES DONNÉES LIÉES	32

Table des figures

FIGURE 1 : AVANTAGES DE L'ADOPTION DES DONNÉES LIÉES	8
FIGURE 2 : INTERACTION ENTRE LES OFFICES GRÂCE À UNE ARCHITECTURE DISTRIBUÉE.....	9
FIGURE 3 : SERVICES DE DONNÉES LIÉES.....	12
FIGURE 4 : VIDÉO DE PRÉSENTATION DU SERVICE LINDAS	15
FIGURE 5 : CONSULTATION DES DONNÉES RELATIVES AU PRIX DE L'ÉLECTRICITÉ.....	21
FIGURE 6 : EVOLUTION DU WEB DEPUIS 1991	29
FIGURE 7 : IDENTIFICATEUR DE RESSOURCE UNIFORME (URI).....	32
FIGURE 8 : GRAPH RDF BASIC	33
FIGURE 9 : TRIPLET RDF.....	34

Management summary

Le potentiel de la technologie des données liées pour l'Administration fédérale est immense et promet des avantages significatifs, y compris une amélioration de la collaboration entre les différents départements et offices fédéraux. Cependant, sa mise en œuvre n'est pas sans défis, tels que les investissements initiaux, l'alignement des systèmes administratifs existants et la nécessaire sensibilisation à cette nouvelle technologie. En dépit de ces défis, une approche d'adoption progressive pourrait se révéler être une voie fructueuse pour surmonter ces obstacles.

L'Office fédéral de topographie Swisstopo et la base de données LINDAS sont des exemples concrets de la façon dont les données liées peuvent être efficacement mises en œuvre. Swisstopo a créé un écosystème de données géospatiales liées grâce à des initiatives telles que Geo.admin.ch et Geocat.ch. Parallèlement, LINDAS a mis en place une infrastructure permettant de publier et de consulter les données RDF. Ces plateformes ont non seulement démontré l'efficacité et la viabilité de cette technologie, mais ont également révélé les avantages considérables en termes d'accès et de partage des données. Elles constituent un modèle à suivre pour d'autres unités administratives.

La transition vers les données liées nécessite des choix stratégiques importants concernant les données à transformer et les normes technologiques à adopter. Des standards, qui constituent la base de la modélisation des données liées, garantissent l'interopérabilité. Une fois les données transformées en ce format, elles peuvent être stockées de manière optimale pour une récupération et une interrogation aisée.

L'exploitation de ces données liées présente de nombreux avantages. Elle permet aux services d'extraire des informations consolidées provenant de différentes bases de données, réduisant ainsi la nécessité de maintenir leurs propres bases de données locales. Cette facilité d'accès à l'information peut stimuler l'innovation et améliorer la prise de décision au sein de l'administration. De plus, les données liées sont conçues pour être facilement navigables et accessibles, même pour les utilisateurs non spécialistes. Cela élargit la portée des données et améliore leur accessibilité pour un public plus large.

En plus de ces avantages techniques et opérationnels, l'intégration des données liées peut également améliorer la gouvernance au sein de l'Administration fédérale. La gouvernance actuelle, basée sur un modèle décentralisé, présente des défis significatifs. L'impression de centralisation offerte par les données liées permet de surmonter ces obstacles en offrant une vision unifiée tout en respectant la structure décentralisée des différentes unités administratives.

En somme, bien que l'adoption de données liées puisse représenter un défi initial significatif, tant en termes d'investissement que de stratégie, les avantages potentiels qu'elle offre sont considérables. L'efficacité opérationnelle, l'innovation, la collaboration et la transparence au sein de l'Administration fédérale peuvent être grandement améliorées, ce qui fait de cette technologie une initiative digne d'attention et d'investissement notamment en termes de formation.

1 Introduction

1.1 Présentation du sujet et justification de son importance

L'évolution rapide du numérique et des technologies de l'information a bouleversé la façon dont les informations sont créées et utilisées, non seulement dans le secteur privé, mais aussi dans le secteur public. Dans ce contexte, les données liées (ou Linked Data en anglais) émergent comme un outil permettant d'exploiter le plein potentiel des données dans le monde numérique. Les données liées permettent de partager des informations de manière plus efficace et plus interconnectée, ce qui peut contribuer à révolutionner la façon dont l'Administration fédérale fonctionne.

L'Administration fédérale est chargée de tâches complexes et variées qui nécessitent la gestion d'une grande quantité d'informations. Celles-ci sont souvent réparties entre différents départements et systèmes informatiques, ce qui peut rendre leur gestion difficile et inefficace. En outre, l'administration doit fournir des services publics de qualité, assurer la transparence de ses opérations et favoriser la participation citoyenne. Ces objectifs nécessitent un accès facile et rapide aux informations.

En permettant de lier et de partager des informations de manière plus fluide et interopérable, les données liées peuvent aider l'Administration fédérale à améliorer ses services et lui permettre de réaliser des économies en mettant en œuvre le principe « once-only » de la Stratégie administration fédérale numérique¹, en évitant la duplication des données et en facilitant leur gestion.

Cependant, malgré leur potentiel, l'adoption et l'utilisation des données liées restent à ce jour très limitées dans l'Administration fédérale. Cela peut être dû à divers facteurs, tels que le manque de connaissance ou de compétences techniques, les préoccupations en matière de sécurité et de confidentialité, ou le manque de normes et standards clairs au sein de l'administration.

Par conséquent, ce rapport se concentre sur l'examen des avantages de l'utilisation des données liées au sein de l'Administration fédérale, tout en abordant également les défis conceptuels associés. En explorant ce sujet, nous espérons contribuer à une meilleure compréhension de l'utilité des données liées pour encourager leur plus large utilisation.

1.2 Objectifs du rapport

Nous cherchons à examiner comment l'utilisation des données liées peut transformer l'administration fédérale. Ce rapport explore les avantages potentiels, tels que l'amélioration de l'efficacité, la transparence et la collaboration entre départements, tout en abordant les défis de mise en œuvre, comme la complexité technique et les questions de sécurité.

D'abord, nous analyserons la gouvernance actuelle des données et les défis posés par la décentralisation. Nous proposerons ensuite des solutions de centralisation pour surmonter ces obstacles. Ensuite, nous illustrerons l'importance des données liées pour la transformation numérique, en détaillant leurs avantages spécifiques et en proposant des recommandations pour une mise en œuvre réussie. Nous présenterons également des exemples d'utilisation réussie des données liées pour tirer des leçons pratiques. Enfin, nous explorerons l'avenir de ces technologies, envisageant leur évolution et les opportunités potentielles pour l'administration fédérale.

Pour une compréhension plus approfondie de l'évolution technologique du web et son impact sur l'administration publique, **veuillez-vous référer à l'Annexe A - l'évolution du web au fil du temps**. Cette annexe retrace l'évolution du web depuis ses débuts avec le Web 1.0 jusqu'à ses perspectives avec le Web 4.0, illustrant ainsi les transitions majeures et leurs implications pour les services publics.

Pour un aperçu détaillé des éléments fondamentaux de l'architecture des données liées et leur importance dans l'administration publique, **veuillez consulter l'Annexe B - Compréhension des données**

¹ <https://www.bk.admin.ch/bk/fr/home/digitale-transformation-ikt-lenkung/digitale-bundesverwaltung.html>

liées. Cette annexe présente une vue d'ensemble des concepts et technologies clés qui sous-tendent les données liées, ainsi que leur application pratique au sein des services publics.

2 Gouvernance des données liées

L'intégration des données liées dans l'administration fédérale offre une opportunité de moderniser la gouvernance et de renforcer l'efficacité des services publics. En exploitant les principes des données liées, nous pouvons établir une gouvernance équilibrée qui allie centralisation et décentralisation, favorisant ainsi une transparence et une meilleure prise de décision.

2.1 Gouvernance actuelle dans l'administration fédérale

La gouvernance actuelle de l'administration fédérale suisse repose sur les principes établis par le Nouveau Modèle de Gestion (NMG)². Ce modèle vise à aligner la gestion sur les objectifs et les résultats plutôt que sur les ressources utilisées. L'optimisation des ressources est centrale, avec un souci constant d'économie et d'efficacité, tout en garantissant la mise à disposition de données financières complètes et compréhensibles.

2.1.1 Principes du NMG

Le NMG repose sur trois principes fondamentaux : la gestion axée sur les objectifs et les résultats, l'optimisation des ressources, et la transparence financière.

- La gestion est orientée vers l'atteinte des objectifs fixés, plutôt que sur l'utilisation des ressources, permettant une meilleure évaluation de la performance. Cela signifie que les décisions de gestion et de financement sont basées sur les résultats attendus, avec un suivi des performances.
- Les ressources doivent être utilisées de manière optimale pour atteindre les objectifs avec un souci d'économie. Cela implique une gestion des fonds publics, visant à maximiser l'efficacité et à minimiser le gaspillage.
- Il est central de fournir des données financières complètes et compréhensibles, facilitant ainsi la prise de décision et la transparence. La transparence financière permet au Parlement, au public et à d'autres parties prenantes de suivre comment les fonds publics sont utilisés et d'évaluer l'efficacité des politiques publiques.

2.1.2 Finances

Le NMG a introduit des budgets globaux, offrant plus de flexibilité et de responsabilité dans la gestion des fonds publics. Cette structuration permet une meilleure adaptation aux besoins changeants et encourage une gestion plus dynamique des ressources.

- Les budgets globaux offrent une plus grande flexibilité aux entités administratives en leur permettant de réallouer les fonds en fonction des besoins émergents. Cela renforce également leur responsabilité dans la gestion des ressources financières.
- L'efficacité et l'efficacité des mesures sont évaluées régulièrement à l'aide d'indicateurs de performance. Cette évaluation continue aide à identifier les domaines nécessitant des améliorations et à ajuster les stratégies en conséquence.

2.1.3 Responsabilisation

La responsabilisation est un pilier essentiel de la gouvernance actuelle, encourageant une gestion axée sur la performance et la responsabilité des résultats.

² [Nouveau modèle de gestion de l'administration fédérale](#)

- Les entités administratives sont encouragées à se concentrer sur les résultats de leurs actions. Cette approche repose sur des indicateurs de performance mesurables, permettant de suivre les progrès réalisés et, si nécessaire, d'ajuster rapidement les actions.
- Un système de contrôle interne et des rapports de gestion réguliers permettent de surveiller et d'évaluer les performances, assurant ainsi la transparence et l'efficacité des opérations. Ces mécanismes de contrôle garantissent que les fonds publics sont utilisés conformément aux objectifs fixés et aux réglementations en vigueur.

Le Parlement et le public bénéficient de cette transparence, ce qui améliore le processus de prise de décision. Parallèlement, l'exécutif, représenté par le Conseil fédéral, joue un rôle clé dans la mise en œuvre des décisions prises.

2.1.4 Impact de la gouvernance actuelle

L'impact de cette gouvernance se manifeste à plusieurs niveaux.

- La mise à disposition de données financières claires et compréhensibles permet au Parlement et au public de mieux comprendre l'utilisation des fonds publics.
- En se concentrant sur les résultats et l'optimisation des ressources, l'administration fédérale peut fournir des services publics plus efficaces et efficaces.
- Les systèmes de contrôle interne et les rapports de gestion réguliers assurent que les entités administratives sont responsables de leurs performances.

2.2 Défis de la gouvernance fédérale

La gouvernance de l'administration fédérale fait face à plusieurs défis qui entravent son efficacité et son efficacité dans la mise en œuvre du NMG. Ces défis sont principalement liés à la fragmentation des données, aux problèmes de coordination, et à la transparence et l'accessibilité des informations.

2.2.1 Fragmentation des données

Les unités administratives indépendantes génèrent et conservent des données dans des silos, ce qui complique l'accès et l'intégration globale des informations. Cette fragmentation des données limite la capacité des différentes entités à collaborer efficacement et à partager des informations pertinentes. De plus, les différences dans les formats de données et les systèmes de gestion utilisés par ces unités administratives aggravent la situation, rendant la fluidité des échanges d'informations encore plus difficile.

Les silos de données sont une conséquence directe de la structure décentralisée de l'administration fédérale. Chaque unité administrative gère ses propres données de manière indépendante, ce qui peut entraîner des incohérences et des doublons. Par exemple, une information peut être stockée dans plusieurs systèmes différents, chacun ayant son propre format et ses propres protocoles de gestion. Cette situation non seulement complique l'intégration des données, mais aussi augmente les risques d'erreurs et de redondances, limitant ainsi l'efficacité globale du système administratif.

2.2.2 Problèmes de coordination

La collaboration entre les diverses entités administratives est complexe en raison de leurs structures et priorités divergentes. Chaque unité a ses propres objectifs et méthodes de travail, ce qui peut entraîner des conflits et des inefficacités. De plus, les canaux de communication sont souvent obstrués ou inadaptés, ce qui limite l'échange d'informations et complique la prise de décision collective.

Les problèmes de coordination sont exacerbés par le manque de standardisation dans les processus et les outils utilisés. Par exemple, certaines unités peuvent utiliser des plateformes numériques avancées pour la gestion des informations, tandis que d'autres peuvent encore dépendre de systèmes papier ou de logiciels obsolètes. Cette disparité complique la collaboration et l'échange de données en temps réel et rend difficile, pour la direction, le contrôle et le suivi des résultats conformément aux attentes du NMG.

2.2.3 Transparence et accessibilité

Pour les parties prenantes, la complexité et la non-uniformité des données limite parfois la compréhension des décisions de l'administration fédérale. Cette situation diminue la transparence et nuit à la confiance du public dans l'administration. De plus, l'accès restreint aux données en raison de limitations techniques ou réglementaires complique encore davantage la situation.

Pour illustrer, la diversité des formats et des structures de données signifie que les informations ne sont pas toujours présentées de manière claire et cohérente. Un parlementaire souhaitant comprendre les dépenses publiques pourrait être confronté à des documents financiers complexes et non standardisés, rendant l'analyse et la comparaison des données extrêmement difficiles. De plus, les restrictions techniques, comme les systèmes de gestion de bases de données incompatibles, et les restrictions réglementaires, comme les règles de confidentialité, peuvent limiter l'accès à des informations centrales pour une compréhension des activités administratives.

2.3 Centralisation versus décentralisation

2.3.1 Une décentralisation inhérente

L'administration fédérale est intrinsèquement décentralisée, chaque unité administrative disposant de ses propres processus et politiques. Cette structure décentralisée permet une gestion plus flexible, adaptée aux spécificités locales, mais pose également des défis en matière de coordination et de gestion des données. Les unités indépendantes génèrent et conservent des données dans des silos, ce qui complique l'accès et l'intégration globale des informations nécessaires pour une gouvernance efficace.

2.3.2 La gouvernance à distance

Selon David Giaouque dans son ouvrage *Une nouvelle économie politique des administrations publiques : la gouvernance à distance* (2023), la gouvernance à distance se nourrit des expériences de la Nouvelle Gestion Publique (NGP) correspondant à notre NMG, et utilise des indicateurs de performance et des dispositifs de management par objectifs. Ce système de pilotage par les nombres, les chiffres et les indicateurs, bien que pas totalement nouveau, connaît une accélération notable avec l'adoption des réformes de la NGP dans de nombreux pays de l'OCDE.

Les effets de la gouvernance à distance

- La gouvernance à distance génère une complexification institutionnelle en raison des besoins accrus en termes d'informations et de bureaucratization des politiques publiques.
- Elle contribue à une bureaucratization croissante des activités et favorise une myopie sur l'efficacité au détriment de l'efficacité dans la délivrance des prestations publiques.
- La focalisation sur les indicateurs de performance entraîne une perte de sens au travail pour les salariés publics, et une montée en puissance de la gestion de la performance individuelle au détriment de la collaboration et du bien-être des employés.

La gouvernance à distance représente une évolution de la gestion publique, construite sur les bases philosophiques et instrumentales de la NGP. Elle cherche à réintégrer l'information pour un pilotage efficace des politiques publiques tout en affrontant les défis posés par la fragmentation et la complexité des réseaux d'action publique. Toutefois, elle engendre également des dilemmes éthiques et des conflits de valeurs, soulignant la nécessité de trouver un équilibre entre performance et bien-être des salariés publics.

2.3.3 Les défis de la gouvernance décentralisée

La gouvernance décentralisée présente plusieurs défis pouvant être regroupé en trois catégories : la fragmentation des données, les problèmes de coordination et les questions de transparence et d'accessibilité.

Fragmentation des données

Dans une administration décentralisée, les unités administratives indépendantes génèrent et conservent des données de manière isolée, créant ainsi des silos d'information. Cette fragmentation entrave l'accès aux données et leur intégration globale, limitant ainsi la capacité de l'administration à avoir une vue d'ensemble cohérente. De plus, les différences dans les formats de données et les systèmes de gestion utilisés par ces unités compliquent davantage la fluidité des échanges d'informations. Ce manque d'harmonisation des données nuit à l'efficacité et à la rapidité des processus décisionnels.

Problèmes de coordination

La collaboration entre les diverses entités administratives est souvent complexe en raison des structures organisationnelles et des priorités qui divergent. Cette complexité est exacerbée par des canaux de communication parfois obstrués ou inadaptés, ce qui limite l'échange d'informations et entrave la prise de décision collective. L'insuffisance de mécanismes de coordination peut conduire à des redondances et une utilisation sous-optimale des ressources disponibles.

Transparence et accessibilité

La diversité des formats de données et la complexité des systèmes utilisés rendent difficile la compréhension des activités administratives par le public et les autres parties prenantes. Cette non-uniformité complique également la tâche des décideurs politiques qui doivent naviguer à travers des informations disparates pour évaluer les performances et l'impact des politiques publiques. De plus, l'accès aux données est souvent restreint par des limitations techniques ou réglementaires.

Pour répondre aux défis posés par la décentralisation, la nouvelle approche de gouvernance propose une centralisation permettant une vision unifiée pour des décisions cohérentes. Cette centralisation vise à combler l'écart entre les décideurs stratégiques et les exécutants sur le terrain, qui sont souvent confrontés à des réalités locales variées. Toutefois, cette approche centralisée peut également entraîner une dissonance opérationnelle. Les directives centralisées ne s'alignent pas toujours avec les contextes décentralisés, ce qui peut provoquer des frictions et des inefficacités. Par conséquent, il faut trouver un équilibre entre centralisation pour une coordination efficace et décentralisation pour la flexibilité et l'adaptabilité locales.

2.4 Transition vers l'utilisation des données liées

Face aux défis posés par la gouvernance décentralisée, l'utilisation des données liées semble apporter une solution de choix. Cette approche permet de concilier les avantages de la décentralisation avec les besoins de centralisation. Les données liées se fondent sur une architecture distribuée qui respecte l'autonomie des différentes unités administratives tout en offrant une interopérabilité sémantique et technique qui donne l'impression d'une centralisation des données.

2.4.1 Un pont entre décentralisation et centralisation

L'architecture des données liées est distribuée, ce qui signifie que les données restent physiquement réparties parmi les différentes unités administratives. Chaque unité conserve son autonomie et peut gérer ses données selon ses propres processus et politiques. Cependant, grâce à des standards communs et des vocabulaires contrôlés, ces données peuvent être interconnectées. Cela permet de surmonter les obstacles liés à la fragmentation des données et d'assurer la communication entre les différents systèmes de gestion.

Les données liées fonctionnent comme un pont entre les systèmes décentralisés et la nécessité de centralisation. Les données restent localisées au sein des unités, mais sont reliées par des mécanismes d'interopérabilité (sémantiques et techniques) qui permettent de les traiter comme un ensemble cohérent. Cette approche favorise une meilleure utilisation des ressources et une plus grande efficacité des processus décisionnels.

Avantages de l'approche des données liées :

- Les unités administratives conservent leur autonomie et leur capacité à gérer des politiques spécifiques, tout en contribuant à une base de connaissances collective. Cela garantit que chaque

unité peut fonctionner selon ses besoins particuliers tout en participant à un effort commun de partage de l'information.

- La centralisation des données n'est pas physique mais conceptuelle. Cela permet aux décideurs d'avoir une vision complète de l'ensemble des activités administratives, tout en respectant la structure décentralisée. Cette approche assure que les décisions stratégiques sont fondées sur des informations actualisées, sans pour autant compromettre l'indépendance des unités administratives.
- Les données liées facilitent la prise de décision en reliant les données transversales pour une meilleure compréhension des enjeux. En intégrant les informations provenant de diverses sources, les décideurs peuvent obtenir une vue d'ensemble plus précise et prendre de meilleures décisions.

2.4.2 Mise en oeuvre du Nouveau Modèle de Gestion (NMG)

L'utilisation des données liées s'aligne parfaitement avec les principes du Nouveau Modèle de Gestion (NMG). En facilitant l'accès et l'intégration des données, elle permet de surmonter les défis de la fragmentation des données et des problèmes de coordination identifiés dans la gouvernance décentralisée. De plus, elle améliore la transparence et l'accessibilité, répondant ainsi aux exigences de la nouvelle gouvernance publique. Les données liées constituent un outil idéal pour progresser dans la mise en place du NMG, en apportant une solution qui combine flexibilité locale et cohérence globale.

3 Importance des données liées pour l'Administration fédérale

3.1 La transformation numérique de l'Administration fédérale par l'utilisation des données liées

L'Administration fédérale, comme toute organisation complexe, génère et utilise une quantité importante d'informations. Ces informations, dispersées dans différents services et entités, peuvent poser des défis considérables en termes de gestion, d'accès, d'exploitation et de partage. Les données liées émergent comme une solution technique offrant une approche organisée et structurée pour la gestion des données.

Un avantage majeur des données liées réside dans leur capacité à soutenir une architecture de données distribuée. Dans un système de gestion de base de données traditionnel (DBMS), toute modification du modèle de données doit passer par l'autorité technique qui gère le DBMS. Cela signifie que la gestion des données est souvent centralisée. Cette approche centralisée peut limiter la flexibilité et nécessite une coordination pour toute modification du modèle de données.

Les données liées, en revanche, permettent une approche plus décentralisée de la gestion des données. Par exemple, supposons que l'Office fédéral de la statistique (OFS) définisse le modèle de données pour le registre des municipalités. Par la suite, Swisstopo, l'office fédéral de topographie, pourrait vouloir ajouter des informations sur les frontières des municipalités à ce modèle. Avec les données liées, cela est possible sans nécessiter de modifications de la part de l'OFS.

Chaque entité a la capacité d'étendre le modèle de données en fonction de ses propres besoins, tout en conservant la compatibilité avec les autres données basées sur RDF. Cela signifie que toutes les données peuvent être interrogées ensemble, malgré les ajouts indépendants au modèle de données. Une telle architecture soutient l'adoption des données liées, par leurs avantages en termes de :

- **Interopérabilité** : L'interopérabilité englobe l'interopérabilité légale, qui se réfère à l'harmonisation des cadres légaux pour permettre l'échange de données ; l'interopérabilité organisationnelle, qui vise à améliorer la coordination et la collaboration entre différentes entités administratives ; l'interopérabilité sémantique, qui s'assure que l'information est compréhensible et peut être utilisée de manière cohérente entre les systèmes ; et enfin l'interopérabilité technique, qui concerne l'utilisation de normes et formats communs pour permettre une

interaction fluide entre les systèmes³. Les données liées, grâce à l'utilisation de normes et formats communs, favorisent les aspects sémantiques et techniques de l'interopérabilité. Le partage et l'échange des données sont ainsi facilités.

- **Transparence et accès à l'information** : Les administrations publiques ont le devoir d'être transparentes envers leurs citoyens. Les Linked Open Data (LOD) jouent un rôle prédominant en rendant les informations plus accessibles et lisibles, soutenant ainsi cette obligation de transparence. Contrairement aux simples Linked Data (LD) qui peuvent ne pas être ouvertes au public, les LOD garantissent un accès libre à l'information pour les citoyens. Cela peut améliorer leur engagement et leur confiance dans le processus administratif. En outre, cela renforce également le droit à l'information.
- **Efficacité interne** : Les données liées offrent des mécanismes pour optimiser les processus internes, en permettant un accès plus rapide aux données pertinentes pour la prise de décisions et l'exécution des tâches. Cela se traduit par une amélioration de la productivité et de l'efficacité de l'administration.
- **Qualité des données** : La qualité des données ne dépend pas uniquement de leur nature liée, mais est également influencée par la manière dont elles sont générées. Les données liées facilitent une gestion des données plus précise, permettent une vérification, une validation et une mise à jour aisées, et offrent un contexte étendu. De plus, leur utilisation aide à prévenir les problèmes rencontrés avec la duplication des données. Sans elles, la tendance serait de copier localement les données produites par d'autres services, ce qui empêche de bénéficier des mises à jour automatiques des données d'origine. Opter pour les données liées assure un accès à des informations à jour. C'est essentiel pour maintenir l'intégrité des données et garantir de disposer d'informations actualisées.
- **Collaboration nationale et internationale** : Les données liées facilitent la collaboration, non seulement entre différentes branches de l'administration publique, mais aussi avec le secteur privé. Elles permettent une communication fluide et transparente, favorisant des synergies nationales et internationales. Cela ouvre également la porte à des collaborations transfrontalières plus efficaces, surmontant les barrières linguistiques et culturelles.

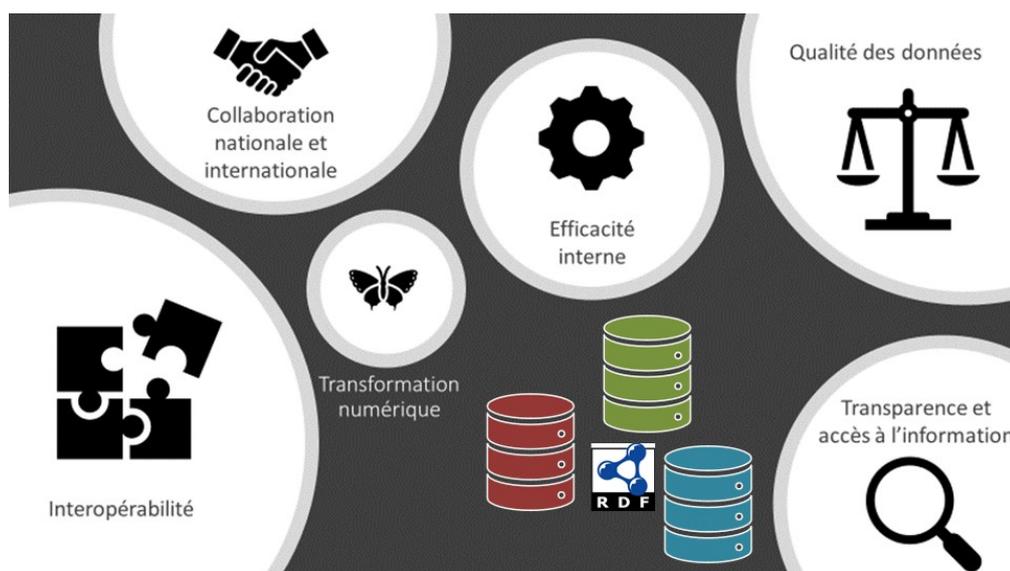


Figure 1 : Avantages de l'adoption des données liées

³ Cadre européen d'interopérabilité : https://ec.europa.eu/isa2/eif_en/

En somme, l'adoption des données liées offre à l'Administration fédérale un moyen de transformation numérique intéressant. Les avantages découlant d'une architecture distribuée se traduisent par une gestion et une exploitation des informations plus efficaces.

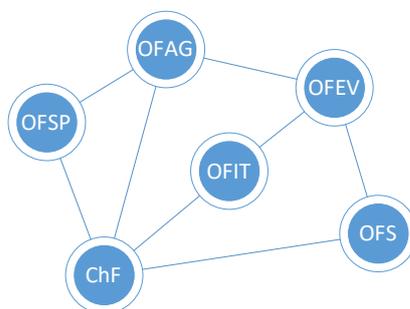


Figure 2 : Interaction entre les Offices grâce à une architecture distribuée

Ces caractéristiques contribuent à une plus grande fluidité dans le partage et l'échange de l'information, améliorant la cohésion et la coordination entre les différentes entités de l'administration. L'application des données liées a le potentiel de limiter considérablement les effets de « silos » au sein de l'Administration fédérale. En permettant à différentes entités de modifier et d'étendre indépendamment les modèles de données suivant les principes d'un modèle de gouvernance des données, les données liées favorisent une vision d'ensemble cohérente des informations. Cela élimine les barrières traditionnellement associées aux systèmes d'information isolés, favorisant un environnement de travail plus collaboratif.

En résumé, l'adoption des données liées par l'Administration fédérale pourrait marquer un tournant vers une gestion de l'information plus efficace et flexible. En brisant les silos, elle ouvre la voie à une collaboration plus dynamique.

3.2 Comment les données liées peuvent aider l'Administration fédérale à atteindre ses objectifs

L'adoption des données liées s'inscrit dans une démarche stratégique visant à optimiser la gestion des informations et à atteindre des objectifs spécifiques, tels que l'enrichissement et la publication de données, le soutien au principe Once-Only, et l'accès et la modélisation uniformes des données.

- Les données liées permettent une extension du modèle de données et une personnalisation de la perspective sur celui-ci. Cela joue un rôle dans l'enrichissement et la publication de divers registres publics, comme la liste des communes (Zefix⁴), le registre des entreprises (IDE⁵), registres des bâtiments et des logements (RegBL⁶), le service national des adresses (SNA⁷), etc. Toutefois, cela nécessite un « backlinking » cohérent, entraînant des efforts supplémentaires en matière de gestion des données. Cela favorise une utilisation uniforme des termes à travers les administrations.
- Les données liées sont un pilier du principe Once-Only, qui vise à minimiser les efforts administratifs en évitant la collecte redondante des mêmes données. La redondance peut aider à assurer la disponibilité des données, mais les données liées offrent une plus grande clarté et facilitent ainsi leur utilisation multiple. En utilisant le langage d'interrogation SPARQL, les données peuvent être extraites de leur source initiale, réduisant la nécessité de duplications locales. Cela diminue les risques d'incohérences et d'erreurs, et réduit le coût global de la gestion des données. En utilisant une URI stable, il est même possible d'éviter complètement les copies, élevant le principe Once-Only à un niveau supérieur.

⁴ <https://www.zefix.ch/fr/search/entity/welcome>

⁵ <https://www.uid.admin.ch/Search.aspx?lang=fr>

⁶ <https://www.housing-stat.ch/fr/madd/public.html>

⁷ <https://www.bfs.admin.ch/bfs/fr/home/registres/service-national-adresses.html>

- Les données liées offrent un point d'accès unifié, simplifiant ainsi l'implémentation des interfaces (intégrez SPARQL et communiquez avec tous). Cet accès uniforme facilite la navigation, le filtrage, et l'extraction des données, améliorant l'efficacité opérationnelle. Toutefois, cet avantage est limité si les données extraites ne suivent pas un modèle uniforme. La modélisation uniforme n'est pas un avantage en soi, mais résulte de l'accord d'utiliser le même langage standard, et certains acteurs peuvent ne pas décrire les données en RDF/OWL/RDFS. Malgré cela, l'adoption des données liées permet une interopérabilité plus fluide, éliminant les obstacles des modèles de données incohérents.

En conclusion, les données liées ne sont pas une simple technologie mais également un outil stratégique qui peut aider l'Administration fédérale à atteindre ses objectifs. Elles nécessitent un consensus sur l'utilisation d'un langage standard et peuvent impliquer des efforts en matière de gestion des données.

4 Avantages des données liées pour l'Administration fédérale

4.1 Amélioration de l'efficacité et de la productivité

La gestion des données s'appuie sur la collecte, l'analyse et la distribution de l'information pour fonctionner efficacement. Toutefois, sans une structuration et une accessibilité appropriée, les données peuvent se transformer en source de complexité et d'obstacles, entravant les processus de travail et diminuant l'efficacité et la productivité.

Dans ce contexte, les données liées, avec leur structure interconnectée et leurs normes d'interopérabilité, se présentent comme une solution permettant d'améliorer l'efficacité et la productivité. Elles offrent une manière plus intuitive et efficiente d'accéder aux informations.

Le langage d'interrogation SPARQL offre un accès rapide et précis aux données liées. Les utilisateurs peuvent utiliser SPARQL pour effectuer des requêtes sur n'importe quelle base de données RDF, que celle-ci soit hébergée par un office fédéral, un canton ou une commune. Toutefois, il convient de noter que l'avantage ici ne réside pas tant dans la nature intrinsèque de RDF, mais dans le fait que ces requêtes peuvent être automatisées sur plusieurs points de terminaison (« endpoints »), permettant ainsi des requêtes fédérées. Cette flexibilité dans l'accès aux données réduit le temps de recherche d'informations et facilite l'accomplissement de tâches dépendantes de l'accès à des données spécifiques.

L'utilisation de données liées supprime également le besoin de dupliquer les données. En éliminant le besoin de multiples copies des mêmes données, on économise non seulement des ressources de stockage, mais on réduit aussi les risques d'erreurs, d'incohérences ou de conflits entre différentes versions des mêmes données. À l'inverse, il est également possible de conserver délibérément les mêmes données selon le point de vue d'une organisation B et d'exprimer explicitement que ces données ne correspondent pas ou partiellement à celles de l'organisation A. Les données liées permettent d'exprimer des informations contradictoires au sein du même modèle de données.

La structure des données liées permet également aux machines d'interagir directement avec les données, ouvrant la voie à l'automatisation de nombreux processus. Non seulement les tâches répétitives peuvent être automatisées grâce à l'utilisation de machines capables de comprendre et d'interagir avec les données liées, mais ces données peuvent aussi être interrogées via API. Ceci souligne leur compatibilité avec la stratégie API First.

4.2 Augmentation de la transparence et de l'accessibilité des informations

La transparence est indispensable pour gagner et maintenir la confiance des citoyens et favorise une gouvernance responsable en rendant le processus décisionnel et les actions du gouvernement visibles. À cet égard, l'utilisation de données liées représente un moyen prometteur de renforcer la transparence de l'Administration fédérale.

Les données liées, grâce à leur structure interconnectée, bien décrite et standardisée, rendent l'information plus accessible pour les employés de l'administration et également pour le public. Elles sont organisées pour faciliter la navigation, la recherche et l'interprétation, ce qui permet à chacun d'accéder facilement aux

informations dont il a besoin. Par conséquent, les citoyens, journalistes, chercheurs et toute personne intéressée peuvent explorer les données liées ouvertes, comprendre leur signification et importance et les utiliser pour mieux saisir les activités de l'administration. En parallèle, les données sensibles et fermées au sein de l'administration doivent être gérées avec soin pour garantir la sécurité et la confidentialité. Cela favorise une participation plus active des citoyens en les aidant à comprendre comment les décisions sont prises et comment les ressources sont utilisées.

Par ailleurs, il est possible d'intégrer des informations internes non publiques, ou données fermées, au sein du même modèle de données, tout en publiant les parties qui peuvent être rendues publiques. Selon le contexte et le processus, certaines informations peuvent nécessiter une gestion et une validation multiples, plutôt que d'être gérées une seule fois selon le principe du « Once-Only ». Par exemple, des informations publiques sur des communes peuvent être référencées et utilisées à partir de systèmes fermés, sans qu'il soit nécessaire de les copier en interne. Des précautions doivent cependant être prises pour éviter que l'accès à partir de systèmes fermés ne puisse être interprété.

Une plus grande accessibilité des données peut également conduire à une plus grande responsabilité de l'administration. Avec des données plus transparentes, il est plus facile pour le public et les organismes de contrôle de suivre les actions de l'administration, de détecter d'éventuelles irrégularités et de demander des comptes en cas de besoin. Cela contribue à maintenir l'administration sous surveillance citoyenne, favorisant ainsi une gouvernance plus responsable.

De plus, une meilleure transparence et accessibilité de l'information peuvent conduire à une amélioration de la qualité des services publics. Les citoyens, en ayant accès à une information claire et à jour, peuvent mieux comprendre et utiliser les services proposés par l'administration. L'administration peut également bénéficier d'une meilleure compréhension des besoins et des attentes des citoyens, ce qui peut l'aider à améliorer ses services et à répondre plus efficacement à ces besoins.

4.3 Facilitation de la collaboration entre les différents départements et entités

L'Administration Fédérale se compose de nombreux départements et entités, chacun avec ses propres responsabilités et domaines de spécialisation. Dans ce contexte, la collaboration entre ces différentes entités est nécessaire pour une gouvernance cohérente. Un des principaux atouts des données liées est leur capacité à faciliter cette collaboration.

Les données liées se caractérisent par leur interconnexion et leur structure normalisée, ce qui permet un échange de données fluide entre différents départements. Les données sont stockées de manière uniforme et structurée, et peuvent être consultées et exploitées par différents départements sans nécessiter de conversions ou d'adaptations complexes. Cela facilite non seulement l'échange de données, mais également l'intégration de ces données dans différents systèmes ou processus. Par exemple, dans le cadre d'un projet collaboratif impliquant plusieurs départements, chacun d'entre eux pourrait avoir besoin d'accéder à des données provenant de plusieurs sources différentes. Avec les données liées, ces données disparates peuvent être facilement combinées et analysées de manière cohérente.

Prenons un exemple concret : un fournisseur de services propose une solution permettant à un utilisateur final (interne ou externe à l'Administration fédérale) d'accéder à des données de l'Office fédéral de l'agriculture (OFAG) et de l'Office fédéral de l'environnement (OFEV) afin de répondre à une demande spécifique nécessitant la consolidation des jeux de données des deux offices.

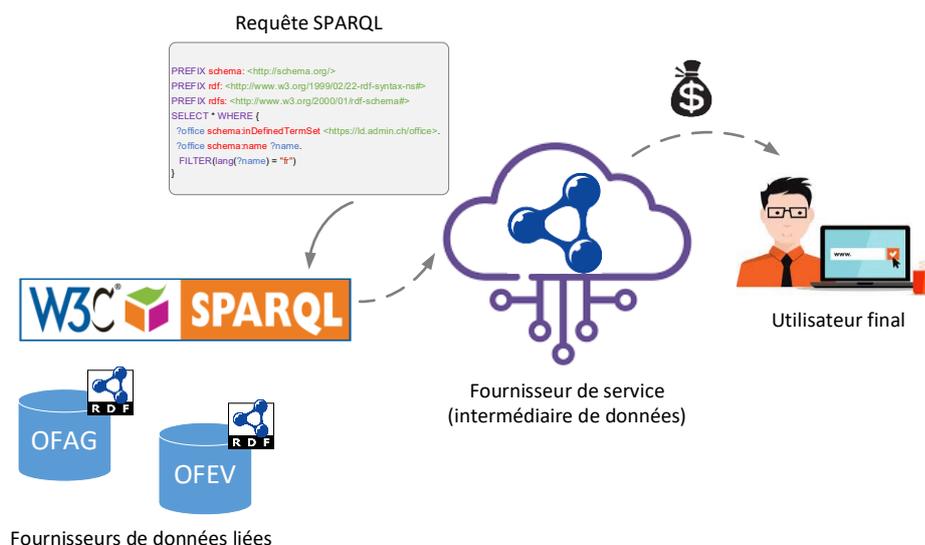


Figure 3 : Services de données liées

Les données liées peuvent également aider à éviter les doublons et les incohérences entre les données des différents départements. Grâce à la nature interconnectée des données liées, une même donnée peut être utilisée par plusieurs départements sans nécessité de duplication. Cela réduit non seulement les besoins de stockage, mais évite également les risques d'erreurs qui peuvent survenir lorsqu'il existe plusieurs copies des mêmes données.

4.4 Autres avantages

Les avantages des données liées ne se limitent pas à l'amélioration de l'efficacité et de la productivité ou à la promotion de la transparence et de la collaboration. Elles peuvent également apporter des avantages spécifiques à certaines activités ou projets au sein de l'Administration fédérale, stimuler l'innovation et aider à se conformer à certaines obligations légales.

L'un des atouts majeurs des données liées dans le contexte des activités internes de l'Administration fédérale réside dans la possibilité d'enrichir de manière décentralisée des listes communes. Ces listes peuvent servir à construire des vocabulaires communs utilisables par différents départements ou entités. En facilitant la collaboration sur des projets interdépartementaux ou la coordination de certaines activités, les données liées favorisent une gestion des données à la fois flexible et cohérente.

Un exemple représentatif est la gestion de la liste des noms de pays. Traditionnellement, cette liste pourrait être gérée de manière centralisée, ce qui pourrait entraîner des problèmes d'actualisation et de cohérence. Grâce aux données liées, chaque entité peut enrichir et utiliser la liste de manière décentralisée, garantissant ainsi que les informations sont toujours à jour et cohérentes à travers l'ensemble de l'administration. Cette opportunité illustre le potentiel d'amélioration de la qualité, de l'efficacité et de la cohérence de la gestion des données.

En termes d'innovation, les données liées peuvent avoir un impact majeur. En permettant aux développeurs, aux entreprises et aux chercheurs d'accéder facilement à des données publiques structurées et interconnectées, les données liées favorisent la création de nouveaux services. Cela peut générer de la croissance économique, stimuler l'innovation technologique et améliorer la qualité de vie des citoyens.

Enfin, l'adoption des données liées peut également aider l'Administration fédérale à se conformer à certaines obligations légales. Par exemple, la Loi fédérale sur l'utilisation de moyens électroniques pour l'exécution des tâches des autorités (LMETA⁸) stipule que les données ouvertes doivent être mises en ligne sous une forme lisible par machine. Les données liées, grâce à leur structure standardisée et à leur interconnectivité, répondent parfaitement à cette exigence.

⁸ [FF 2023 787 - Loi fédérale sur l'utilisation de moyens électroniques pour l'exécution des tâches des autorités \(LMETA\) \(admin.ch\)](#)

5 Défis et limitations

5.1 Complexité technique et besoin de compétences spécialisées

L'adoption et l'implémentation des données liées présentent une certaine complexité technique. Ce degré de complexité est inhérent à la nature même des données liées, qui requièrent un format spécifique et interconnecté pour être pleinement exploitables. Cette complexité s'intensifie lorsqu'il est nécessaire d'identifier et de convertir les données de base⁹ existantes en un format lié.

De plus, il est nécessaire de définir les vocabulaires de référence, souvent inspirés de ceux trouvés dans eCH¹⁰. Cependant, eCH ne se limite pas uniquement aux vocabulaires de référence, proposant également d'autres standards et normes essentiels pour le fonctionnement interne des administrations. Les contributions d'eCH peuvent varier en fonction du contexte ou du domaine spécifique de l'administration concernée. Cela souligne l'importance d'avoir des compétences spécialisées en interne, ainsi qu'une compréhension des normes, des outils et des protocoles associés à la gestion des données liées.

Il est également important de noter que pour beaucoup, la technologie des données liées peut sembler abstraite et difficile à comprendre, en particulier pour ceux qui ne sont pas déjà familiers avec les concepts de bases de données et d'informatique. Cela peut constituer un obstacle majeur, car les employés peuvent hésiter à adopter une technologie qu'ils perçoivent comme complexe ou inaccessible.

Pour surmonter ces défis, un investissement dans la formation et le développement des compétences de certains membres du personnel de la Confédération est nécessaire. Cette formation devrait être spécifiquement destinée à ceux qui travaillent directement avec les données liées, tels que les administrateurs de base de données, les développeurs web et les analystes de données. Pour ces spécialistes, l'accent doit être mis sur la pédagogie et l'accessibilité pour garantir qu'ils peuvent comprendre et utiliser efficacement les données liées dans leur travail quotidien.

Actuellement, bien que l'offre de formation spécifique sur les données liées reste limitée dans les universités et écoles supérieures, il est important de noter que les données liées font partie du programme de formation disponible sur le portail proposé par le centre de formation de l'Administration fédérale (CFAF¹¹).

En outre, il peut être utile d'adopter une approche progressive à l'adoption des données liées. Par exemple, il peut être bénéfique de commencer par les données ouvertes, un concept souvent associé aux Linked Open Data (LOD). Les données ouvertes sont généralement plus faciles à comprendre et à gérer, ce qui peut faciliter la transition vers l'utilisation de données liées plus complexes.

5.2 Questions de sécurité et de confidentialité

La sécurité et la confidentialité des données sont des enjeux de la numérisation et de la mondialisation de l'information. Ces préoccupations sont particulièrement élevées lorsqu'il s'agit de la publication de données liées, notamment celles qui sont sensibles ou personnelles. En effet, les données liées, en raison de leur nature interconnectée, sont susceptibles d'augmenter les risques de violations de la confidentialité si elles ne sont pas adéquatement protégées.

Les données personnelles, lorsqu'elles sont liées à d'autres sources, peuvent être exposées à des risques plus élevés de fuite ou d'utilisation abusive. Ce n'est pas nécessairement le système de données liées en lui-même qui est la cause, mais la potentialité de mettre en lumière des connexions entre les données. Cette liaison a l'avantage de nous faire prendre conscience des liens qui existent entre les données et de permettre une meilleure identification des risques. Cependant, cela peut entraîner des conséquences graves, non seulement pour les individus dont les informations sont compromises, mais aussi pour la réputation et la confiance envers l'Administration fédérale.

⁹ <https://www.bk.admin.ch/bk/fr/home/digitale-transformation-ikt-lenkung/bundesarchitektur/stammdatenvverwaltung.html>

¹⁰ <https://www.ech.ch/fr/standard-uebersicht>

¹¹ <https://www.epa.admin.ch/epa/fr/home/services/offre-de-formation.html>

Face à ces enjeux, il faut d'une part mettre en œuvre des mécanismes de sécurité appropriés pour assurer la protection des données et d'autre part se mettre en conformité avec les réglementations en vigueur sur la protection des données. Cela comprend non seulement les lois fédérales telles que la Loi fédérale sur la protection des données (LPD¹²) et la Loi fédérale sur la sécurité de l'information au sein de la Confédération (LSI¹³), mais aussi les réglementations internationales, en particulier si les données sont partagées ou utilisées à travers les frontières.

De plus, le personnel qui travaille avec des données liées doit être formé à comprendre et à mettre en œuvre les meilleures pratiques en matière de sécurité des données. Cette formation doit aborder des sujets tels que l'identification et la gestion des risques de sécurité, l'utilisation de protocoles de sécurité, la réponse aux incidents de sécurité, et le respect des lois et réglementations sur la protection des données.

5.3 Autres défis

Si tous les Offices fédéraux ne s'engagent pas à adopter cette technologie, les bénéfices potentiels en termes d'interopérabilité et de collaboration peuvent rester limités. Par conséquent, il faut encourager activement l'utilisation des données liées et fournir un soutien adéquat aux différentes organisations pour faciliter leur adoption.

L'adoption des données liées peut également nécessiter des investissements initiaux. Ces investissements peuvent concerner le temps, le personnel et les ressources financières. Par exemple, il peut être nécessaire d'établir de nouvelles infrastructures pour stocker et gérer les données liées, de convertir les données existantes au format RDF, de former le personnel à l'utilisation et à la gestion de ces données, et de maintenir régulièrement ces nouvelles infrastructures. Ce sont des coûts qui doivent être pris en compte lors de la planification de l'adoption des données liées.

Une stratégie d'adoption progressive pourrait être bénéfique. Par exemple, il pourrait être judicieux de commencer par des projets pilotes dans certains départements afin de démontrer les avantages des données liées. Ces projets pourraient être inclus dans la feuille de route pour la mise en œuvre de la stratégie de données liées, avec une description claire des objectifs, des délais et des indicateurs de performance. Il est à noter que, contrairement à de nombreuses technologies et plateformes, les données liées sont particulièrement adaptées à une introduction progressive basée sur des cas d'utilisation spécifiques. Cela est d'autant plus vrai que les données liées peuvent être exploitées en parallèle avec des systèmes RDBMS existants et des API comme REST, SOAP, etc.

6 Présentation de cas d'études ou d'exemples d'utilisation réussie des données liées

6.1 Infrastructures et plateformes pertinente pour l'Administration fédérale

6.1.1 LINDAS

L'initiative LINDAS¹⁴, pour Linked Data Service, est le fruit de l'ambition du SECO d'explorer le potentiel de la technologie des données liées. Le SECO a employé cette technologie pour connecter, améliorer et publier des annuaires d'autorités et de services, qui étaient alors créés et diffusés de manière disparate, ainsi que diverses autres données décentralisées. Un projet pilote fut mené entre 2014 et 2017, aboutissant à une version préliminaire de LINDAS.

De manière concomitante, les Archives fédérales (AFS) ont entrepris diverses expériences en collaboration avec d'autres archives dans le domaine des données liées. À partir de 2018, elles ont pris en charge la gestion de cette activité du SECO et ont déployé une version opérationnelle de LINDAS, mise en service en 2020.

¹² [RS 235.1 - Loi fédérale du 25 septembre 2020 sur la protection des données \(LPD\) \(admin.ch\)](#)

¹³ [RS 128 - Loi fédérale du 18 décembre 2020 sur la sécurité de l'information au sein de la Confédération \(Loi sur la sécurité de l'information, LSI\) \(admin.ch\)](#)

¹⁴ <https://lindas.admin.ch/>

LINDAS adhère au principe du « once-only », favorisant la publication de jeux de données issus de différentes organisations qui partagent des concepts communs, structurés sous forme de relations explicites entre les éléments de ces jeux de données. LINDAS permet ainsi l'utilisation directe d'une grande diversité de données, sans nécessité de copie locale, grâce à une interface programmable fondée sur le langage d'interrogation SPARQL. La présence d'un point d'accès unique réduit les coûts associés à la maintenance et à l'exploitation d'une multitude d'interfaces.

L'objectif de ce service est d'automatiser la création et la fourniture de données ouvertes et non sensibles sous forme de données liées. Cette approche facilite l'adoption de cette technologie par les nouveaux détenteurs de données et soutient les entités gouvernementales désireuses de partager leurs propres données dans ce format. En outre, LINDAS offre un accès simplifié à ces données pour les utilisateurs.



Figure 4 : Vidéo de présentation du service LINDAS¹⁵

LINDAS héberge une grande variété de jeux de données issus de différents domaines administratifs. Chaque ensemble de données est accompagné de métadonnées détaillées, fournissant des informations sur le contenu, l'origine, la qualité, le format, la licence et d'autres caractéristiques pertinentes des données. Ces métadonnées permettent aux utilisateurs de comprendre rapidement le contexte et la pertinence de chaque jeu de données. Toutes les données sur LINDAS sont au format RDF, assurant ainsi leur interopérabilité avec d'autres services de données liées. Les données sont facilement accessibles pour les systèmes automatisés grâce à l'utilisation du point d'accès SPARQL. De plus, un formulaire SPARQL est à la disposition des utilisateurs pour formuler des requêtes de données personnalisées.

LINDAS ne se limite pas à être un simple répertoire de données. Il s'agit d'un service avec plusieurs plateformes offrant un éventail d'outils et de services conçus pour aider les autorités à convertir leurs données en données liées. Ces outils facilitent la conversion des données, la gestion des métadonnées et l'application des normes d'interopérabilité. Parmi ces outils, citons le « Data Cube Creator » et le « Schema Manager », détaillés dans le document « Dienstleistung Linked Data Service LINDAS¹⁶ ».

En tant que composant essentiel de l'écosystème des données ouvertes, LINDAS favorise la collaboration entre diverses parties prenantes, y compris les fournisseurs de données, les développeurs de logiciels et les utilisateurs de données. En soutenant l'utilisation des standards ouverts du W3C, LINDAS facilite la création de solutions novatrices basées sur les données ouvertes.

6.1.2 Opendata Swiss

Le Secrétariat Open Government Data, rattaché à l'OFS (Office fédéral de statistique), joue un rôle dans la coordination de la mise en œuvre de la stratégie OGD Suisse pour la période 2019-2023. Ce travail de coordination est réalisé en collaboration étroite avec le Département fédéral de l'Intérieur (DFI). Cette stratégie vise principalement à rendre accessible un ensemble significatif de données de l'administration publique sur la plateforme numérique opendata.swiss, soutenant ainsi l'objectif d'une gouvernance ouverte et transparente. L'ancrage juridique du principe open by default est un objectif important de la mise en œuvre de la stratégie OGD. Un article correspondant à ce principe a été inséré dans la loi fédérale sur l'utilisation de moyens électroniques pour l'exécution des tâches des autorités (LMETA). L'ancrage juridique

¹⁵ <https://youtu.be/VWVsCMnh4BA>

¹⁶ https://www.bar.admin.ch/dam/bar/fr/dokumente/diverses/lindas_dienstleistung.pdf.download.pdf/LINDAS_Dienstleistung_de.pdf

est donc désormais acquis et l'élaboration d'une nouvelle stratégie n'est par conséquent pas nécessaire. Un Masterplan OGD 2024-2027 contenant les orientations, buts et mesures à appliquer sur le terrain est proposé en lieu et place d'une nouvelle stratégie.

Opendata.swiss se présente sous la forme d'un catalogue de métadonnées. Il ne publie pas des données elles-mêmes mais plutôt des références vers celles-ci. Les données publiques ouvertes (Open Government Data, OGD) sont des données administratives lisibles par machine, ouvertes et réutilisables. Ces données sont hébergées dans les centres de données des différentes organisations qui les ont produites. De cette manière, les données sources sont conservées à leur emplacement d'origine, assurant leur authenticité et leur intégrité. Elles sont référencées ou rendues accessibles par un lien via opendata.swiss, ce qui permet un accès aisé et centralisé à ces ressources.

En ce qui concerne la nature des données publiées dans le cadre de l'Open Government Data (OGD), il est important de souligner que sont exclus de la publication en tant qu'OGD les secrets commerciaux ou les données pour lesquelles il existe un intérêt de protection particulier ancré dans une loi spécifique (loi sur la protection des données, protection de l'information, droit d'auteur). En principe, les OGD doivent être publiées si possible sous forme de données brutes. Lors d'une publication en tant qu'OGD, s'il s'agit de données sensibles, une agrégation et une anonymisation sont nécessaires pour des raisons de protection des données. Une partie des jeux de données sur opendata.swiss n'est donc disponibles pour une réutilisation que sous forme agrégée et anonymisée. Ces données ont été traitées de manière à empêcher toute identification des personnes concernées, respectant ainsi les principes de confidentialité et de protection des données. Cette absence d'identification est une condition sine qua non pour la mise à disposition de ces données en libre accès.

L'approche décentralisée de la collecte de métadonnées provenant de différents catalogues de données nécessite un standard. Le portail opendata.swiss utilise DCAT-AP CH¹⁷. Le profil d'application DCAT pour les portails de données en Suisse (DCAT-AP CH) est un profil d'application du Data Catalog Vocabulary¹⁸ et sert à décrire les données répertoriées dans les portails de données suisses (p. ex. opendata.swiss). DCAT est un vocabulaire RDF qui doit faciliter l'interopérabilité entre les catalogues de données publiés sur le web. Il permet ainsi d'améliorer la facilité de recherche et la réutilisation des OGD. Les technologies et les spécifications du web sémantique soutiennent la publication de données sur le web. Afin de pouvoir utiliser tout le potentiel de ces technologies, opendata.swiss étudie l'utilisation d'un magasin RDF (« RDF store »). Cela a pour objectif de faciliter l'importation des catalogues de métadonnées, soutenir la conformité avec le standard européen DCAT-AP, améliorer la qualité des métadonnées et offrir la possibilité d'utiliser une autre interface avec un endpoint SPARQL. En outre, les métadonnées proposées en tant que Linked Data peuvent être reliées à d'autres données.

Les personnes responsables de publier les données s'assurent que leurs OGD sont disponibles dans un format ouvert, lisible par machine et standardisé non-propritaire selon les 3 étoiles du modèle 5 étoiles¹⁹ relatif aux données ouvertes. Une publication de Linked Open Data doit être examinée et mise en œuvre conformément à la directive OGD 18, en particulier pour les registres centraux et les vocabulaires contrôlés²⁰. L'utilisation du Resource Description Framework (RDF) et des standards qui l'emploient, tels que SPARQL et OWL, est la pratique idéale pour le codage et l'établissement de liens entre les données. Cette approche favorise la création de Linked Open Data (Données liées ouvertes), qui sont simultanément une composante du Web sémantique. Le Web sémantique permet aux machines de comprendre et de répondre aux données, tout comme les humains le feraient, ouvrant ainsi la voie à des applications plus intelligentes et automatisées.

¹⁷ <https://handbook.opendata.swiss/de/content/glossar/bibliothek/dcat-ap-ch.html>

¹⁸ <https://www.w3.org/TR/vocab-dcat-2/>

¹⁹ <https://handbook.opendata.swiss/de/content/glossar/bibliothek/ogd-richtlinien.html#maschinenlesbarkeit-der-daten-und-der-datenmodelle>

²⁰ <https://handbook.opendata.swiss/de/content/glossar/bibliothek/ogd-richtlinien.html#linked-open-data>

Les jeux de données Linked Data stockés sur LINDAS mais également sur d'autres portails tels que ceux de geo.admin²¹, de Bâle-Ville²² ou de la ville de Zurich²³, sont également référencés sur opendata.swiss. La visibilité des Linked Open Data sur opendata.swiss a été augmentée grâce à un filtre. Grâce à ce filtre, les utilisateurs de données peuvent limiter leur recherche aux différentes sérialisations RDF (Turtle, RDF/XML, etc.) et au service SPARQL.



Dans ce contexte, le service offert par LINDAS joue un rôle complémentaire en permettant le stockage de données sous forme de Linked Open Data. Opendata.swiss, d'autre part, assure le référencement de ces données. Ainsi, ces deux plateformes travaillent de concert pour favoriser l'accessibilité et l'utilisabilité des données publiques ouvertes.

Enfin, l'accès à toutes les métadonnées d'opendata.swiss est facilité par l'utilisation d'une API (Interface de Programmation d'Applications) dédiée. Cette API permet aux développeurs de logiciels d'intégrer facilement les données et fonctionnalités d'opendata.swiss dans leurs propres applications ou sites web, favorisant ainsi la réutilisation des données publiques ouvertes.

6.1.3 Data Europa EU

Data Europa EU, ou European Union Open Data Portal, est un acteur essentiel dans l'écosystème des données liées ouvertes (Linked Open Data - LOD) de l'Union européenne (UE). Le portail est un point central d'accès aux données ouvertes européennes provenant de portails internationaux, de l'Union européenne, nationaux, régionaux, locaux et de géodonnées. Il consolide l'ancien portail de données ouvertes de l'UE et le portail de données européen. Le portail se distingue par sa mise en œuvre technique des principes des LOD, favorisant l'interopérabilité des données et facilitant leur utilisation dans différents contextes.

Le portail est un catalogue de métadonnées. Pour favoriser la comparabilité des données publiées au-delà des frontières, il présente des références de métadonnées dans un format familier (profil d'application Data Catalogue Vocabulary (DCAT) pour les portails de données en Europe), en utilisant la technologie Resource Description Framework (RDF). Data.europa.eu utilise une approche innovante dans la structuration de ses données. Contrairement à la plupart des systèmes qui gèrent les données dans des bases de données relationnelles avant de les convertir en RDF pour la publication, data.europa.eu se base sur un triplestore. Les métadonnées sont enregistrées sous forme de LOD (basé sur la norme DCAT-AP) et mises à disposition pour la recherche et l'utilisation dans différentes sérialisations RDF. C'est-à-dire que le portail gère et rend ses (méta)données publiques directement dans le format RDF, ce qui est une pratique moins courante mais potentiellement plus efficace pour l'utilisation et le partage des données liées. En outre, l'utilisation d'URI persistants²⁴ permet de mettre en œuvre une autre composante importante des données liées.

Le portail propose des rapports lisibles par l'homme sur la qualité des métadonnées. Les données sont collectées à partir du Triplestore et transformées en un rapport visuellement attrayant. Pour ce faire, SHACL est utilisé en combinaison avec DQV. Pour assurer l'interopérabilité à travers les différents jeux de données

²¹ <https://www.geo.admin.ch/de/linked-data-dienst-geodaten-verlinken>

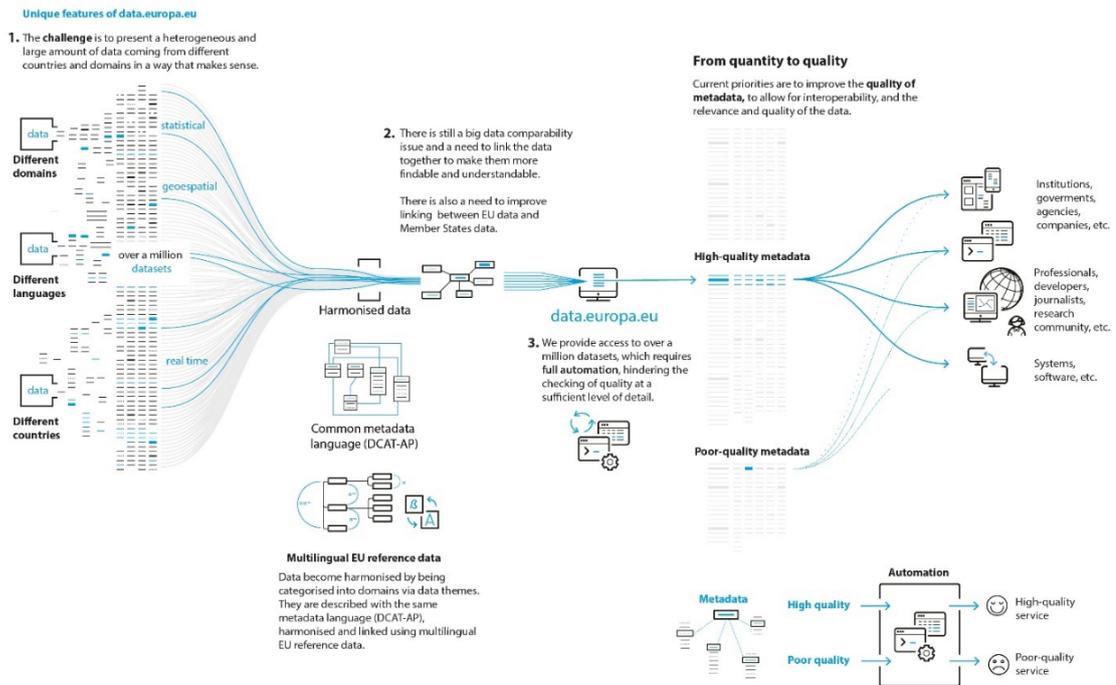
²² <https://www.opendata.bs.ch/ld.html>

²³ <https://www.stadt-zuerich.ch/prd/de/index/statistik/publikationen-angebote/linked-open-data.html>

²⁴ [The official portal for European data | data.europa.eu](https://data.europa.eu)

et les domaines, le portail adopte des vocabulaires contrôlés et des ontologies standardisées, comme SKOS (Simple Knowledge Organization System) ou Dublin Core. Ces standards facilitent l'échange d'informations et permettent aux machines de comprendre le sens des données.

En outre, la disponibilité d'outils de raisonnement peut favoriser la création de nouvelles idées et de données dérivées. Les capacités natives de RDF à gérer plusieurs langues soutiennent l'aspect transnational des données ouvertes.



En ce qui concerne l'accès aux données, le portail fournit des API dédiées qui permettent aux développeurs d'intégrer les données dans leurs propres applications ou services. Ces API utilisent également les standards du Web sémantique pour assurer que les données peuvent être facilement utilisées et réutilisées.

Data Europa EU favorise également la réutilisation des données grâce à l'adoption de licences ouvertes. En fournissant des métadonnées complètes pour chaque ensemble de données, y compris des informations sur la licence, le portail aide à assurer que les utilisateurs savent comment ils peuvent utiliser les données de manière légale et éthique.

Enfin, pour soutenir l'utilisation efficace des données liées, Data Europa EU propose diverses ressources éducatives et techniques, notamment des tutoriels, des guides et des exemples de requêtes SPARQL. Ces ressources peuvent aider les utilisateurs, qu'ils soient novices ou expérimentés, à exploiter au mieux les capacités offertes par les données liées ouvertes.

6.1.4 Intégration des données géospatiales

La richesse de l'offre de données géographiques en Suisse est grandement due à l'Office fédéral de topographie swisstopo qui a pris l'initiative il y a déjà plusieurs années de promouvoir l'accès, la standardisation et l'interconnectivité des géodonnées.

Swisstopo, par l'intermédiaire de l'organe fédéral de coordination de la géoinformation GCG et de son centre opérationnel COSIG, développe et exploite l'Infrastructure Fédérale de Données Géographiques IFDG en tant que base pour la conservation et la publication des géodonnées de l'Administration fédérale. L'IFDG se base juridiquement sur la législation suisse en matière de géoinformation (LGéo²⁵ / OGéo²⁶)

²⁵ <https://www.fedlex.admin.ch/eli/cc/2008/388/fr>

²⁶ <https://www.fedlex.admin.ch/eli/cc/2008/389/fr>

Le portail geo.admin.ch et son visualiseur de cartes map.geo.admin.ch ont été lancés pour offrir un accès centralisé à un grand nombre de données géographiques. En outre, une API et une série de services web géospatiaux²⁷ ont été mis à disposition pour encourager l'utilisation des géodonnées dans une multitude de contextes.

Ces outils tirent parti des normes du domaine de l'information géographique définies au niveau international par des organismes tels que le comité technique 211 de l'ISO et l'Open Geospatial Consortium, ainsi que des normes définies au niveau suisse.

Parallèlement, swisstopo supervise également geocat.ch, le catalogue géographique de la Suisse²⁸. [Geocat.ch](http://geocat.ch) fournit un répertoire de métadonnées géographiques normalisées. En tant que catalogue national, geocat.ch facilite l'accès, l'échange et la réutilisation des géométadonnées.

L'information est structurée conformément à la norme internationale pour les géométadonnées définie par l'ISO. Parallèlement, les géométadonnées de geocat.ch sont reprises par la plateforme opendata.swiss et sont mappées dans le format DCAT-AP. Toutefois, il convient de noter que cette conversion ne signifie pas nécessairement que les géométadonnées sont disponibles sous forme de données liées. En outre, les géométadonnées de geocat.ch sont également disponibles dans le portail européen data.europa.eu²⁹ par le biais d'opendata.swiss.

En mars 2017, swisstopo a lancé le service de données liées geo.ld.admin.ch comme nouveau canal de publication des géodonnées de l'IFDG. Le service est basé sur le standard international GeoSPARQL de l'Open Geospatial Consortium. GeoSPARQL définit un vocabulaire pour l'encodage RDF des géodonnées et fournit une extension du langage SPARQL pour supporter les requêtes géospatiales.

L'offre actuelle de géodonnées de geo.ld.admin.ch se compose d'un certain nombre de jeux de géodonnées d'intérêt général (adresses, routes, arrêts de transports publics, limites administratives) ainsi que d'autres jeux de géodonnées soutenant les activités de divers offices fédéraux dans le cadre de la plateforme LINDAS. Cette offre évolue constamment en fonction des besoins exprimés par les différents offices fédéraux (notamment dans le cadre de la collaboration avec LINDAS) et par la communauté suisse des Linked Data en général.

Les données liées sont actuellement l'un des thèmes de la stratégie swisstopo 2025³⁰ et de la stratégie suisse pour la géoinformation³¹.

Grâce à l'effort coordonné de swisstopo à travers geo.admin.ch et geocat.ch, la Suisse a réussi à créer un écosystème solide de données géospatiales. Ces ressources, non seulement largement accessibles, sont aussi interconnectées et interopérables, encourageant leur utilisation innovante et créative.

6.2 Succès rencontrés avec les données liées

6.2.1 Visualisation des données ouvertes de l'administration publique suisse

Au travers de « [visualize](http://visualize.admin.ch)³² », l'outil de visualisation des données ouvertes de l'Administration publique suisse, développé par l'Office fédéral de l'environnement, nous sommes en mesure de constater le potentiel fascinant de la visualisation des données ouvertes.

Au cœur de ce service se trouve une application innovante qui rend possible la visualisation de données à caractère unidimensionnel ou multidimensionnel. Cette solution tire sa force des données brutes et intrinsèques, dont la représentation peut également être obtenue directement par la visualisation. Une particularité notable de cette application réside dans sa capacité à refléter en temps réel les modifications apportées aux données. Pour illustrer cela, considérons un exemple : lorsqu'une nouvelle année est ajoutée

²⁷ <https://api.geo.admin.ch/>

²⁸ <https://www.geocat.ch/geonetwork/srv/fre/catalog.search#/home>

²⁹ Par exemple <https://data.europa.eu/data/datasets/86cb844f-296b-40cb-b972-5b1ae8028f7c-bundesamt-fur-landestopografie-swisstopo?locale=fr>

³⁰ <https://www.swisstopo.admin.ch/fr/swisstopo/vision.html>

³¹ <https://www.geo.admin.ch/fr/strategie-et-mise-en-oeuvre/>

³² <https://visualize.admin.ch/fr?dataSource=Prod>

à un ensemble de données, cette information est immédiatement visible au sein de la visualisation, sous réserve que cette fonctionnalité ait été configurée.

La structure fondamentale de l'application est basée sur l'ontologie ou le schéma « Cube³³ ». Ce choix n'est pas anodin car ce schéma permet non seulement d'organiser les données de manière efficace, mais facilite aussi leur interprétation et leur utilisation.

En outre, pour assurer une cohérence et une précision optimales, l'application fait appel à l'utilisation de vocabulaires contrôlés. Ce recours aux vocabulaires contrôlés garantit que les termes employés sont homogènes et universellement compris, ce qui constitue un facteur important dans le domaine de la gestion des données.

Enfin, nous voulons souligner l'aspect intégratif de cette application. En effet, elle permet l'incorporation des données provenant de divers offices tels que l'OFS, Swisstopo, l'OFEV et l'OFEN. Cette fonctionnalité favorise non seulement une vue d'ensemble plus complète, mais aussi une compréhension plus profonde des différents contextes dans lesquels ces données sont générées et utilisées.

En somme, de tels outils de visualisation des données peuvent révolutionner la façon dont nous interagissons avec les informations, en rendant les données ouvertes de l'Administration fédérale non seulement plus accessibles, mais aussi plus compréhensibles et exploitables. Pour plus d'information, rendez-vous sur YouTube³⁴ afin de visualiser les tutoriaux disponibles en libre accès.

6.2.2 Gestion avancée des données juridiques

La gestion optimisée des données juridiques est une priorité pour toutes les administrations publiques. À cet égard, Fedlex, la plate-forme de publication du droit fédéral opérationnelle mais en constante évolution, pilotée par la Chancellerie fédérale, apporte des solutions novatrices à ces défis. Fedlex ne se limite pas à une simple mise à jour des systèmes préexistants, c'est un véritable renouvellement destiné à rendre les lois, règlements et traités internationaux plus accessibles.

Dans sa phase initiale, Fedlex a mis l'accent sur la publication d'une série de documents essentiels : la Feuille fédérale (FF), le Recueil officiel du droit fédéral (RO) et le Recueil systématique (RS). Toutefois, son ambition va plus loin. Pour la première fois, la base de données des traités internationaux applicables à la Suisse, auparavant gérée par le Département fédéral des affaires étrangères, a été intégrée à la plate-forme du droit fédéral. Ce processus a permis de créer une ressource unifiée et exhaustive pour tous les textes juridiques pertinents.

L'innovation majeure de Fedlex réside peut-être dans son engagement envers les principes du Web sémantique et des données ouvertes liées. Cela signifie que toutes les informations publiées sur la plate-forme sont structurées de manière à faciliter leur navigation, évaluation, interconnexion et utilisation par les humains et les machines. Pour l'instant, seules les métadonnées des lois sont présentes dans le triplestore, mais Fedlex s'efforce de fournir une bibliothèque numérique interconnectée qui favorise la transparence, la compréhension et l'efficacité juridiques.

Afin de faciliter cette interconnectivité, Fedlex utilise une combinaison de langages et de formats de données pour publier des métadonnées. La structure de ces métadonnées est définie par des modèles tels que JOLux³⁵ et ELI³⁶ (European Legislation Identifier). Grâce à l'utilisation d'éléments largement reconnus tels que le framework RDF, les langages de balisage comme HTML, les vocabulaires basés sur RDF comme SKOS, ainsi que les formats tels que PDF et DOCX, Fedlex s'assure que ses données peuvent être aisément intégrées et utilisées dans une multitude de contextes.

De plus, Fedlex ne se limite pas à la simple publication de textes juridiques. Il génère un véritable graphe de connaissance, une structure de données interconnectée qui englobe toutes les informations et les relations

³³ <https://cube.link>

³⁴ <https://www.youtube.com/@visualizetutorials/videos>

³⁵ <https://fedlex.data.admin.ch/fr-CH/home/intro>

³⁶ <https://eur-lex.europa.eu/content/help/eurlex-content/eli.html>

entre les textes juridiques. Ce graphe est mis à disposition en open data, fournissant une ressource précieuse pour la recherche, l'analyse et le développement de services basés sur les données juridiques.

En outre, la collaboration de la Chancellerie fédérale avec le Service central de législation du Grand-Duché du Luxembourg souligne l'aspect international et collaboratif de Fedlex. En utilisant la suite logicielle « Casemates » du Luxembourg, Fedlex s'appuie sur l'expertise et l'expérience internationales pour améliorer ses services.

En somme, Fedlex marque le début d'une nouvelle ère pour la gestion et la publication des données juridiques en Suisse. Grâce à son engagement envers la transparence, l'accessibilité et l'interopérabilité, Fedlex constitue un prérequis pour la digitalisation des décisions politiques, et sert d'exemple pour les administrations publiques du monde entier.

6.2.3 Transparence des prix de l'électricité

L'Autorité fédérale de régulation de l'électricité en Suisse, ElCom, est déjà en avance dans le domaine de l'ouverture des données. ElCom s'engage activement à rendre les informations accessibles et transparentes, jouant un rôle dans l'avancement des données ouvertes au sein du secteur énergétique.

Actuellement, ElCom gère une application qui offre un accès public aux données relatives aux prix de l'électricité. Cette application facilite la prise de décisions éclairées par les consommateurs en ce qui concerne leur utilisation de l'électricité. En outre, elle encourage une transparence accrue et une compétitivité plus saine au sein du secteur. L'application tire ses données de l'OFS et elles sont conservées dans une base de données relationnelle classique.

ElCom a déjà mis en œuvre la publication des données sur les prix de l'électricité sur la plateforme LINDAS. Cette initiative facilite une analyse approfondie de la structure tarifaire de l'électricité en Suisse, améliorant ainsi la transparence dans le secteur. La conversion de ces données en un format de données liées a simplifié leur utilisation par les développeurs et les analystes de données, et a garanti une meilleure interopérabilité avec d'autres bases de données. L'application « Prix de l'électricité en Suisse »³⁷, par exemple, tire directement ses données de LINDAS.

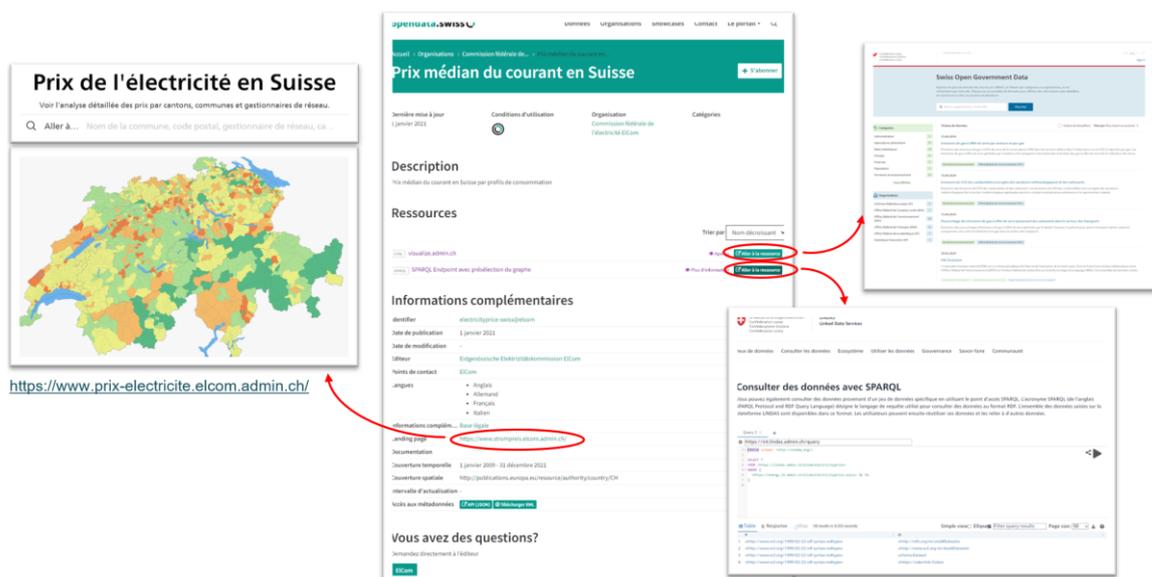


Figure 5 : Consultation des données relatives au prix de l'électricité

L'application Web d'ElCom, destinée aux utilisateurs finaux, offre une visualisation interactive et facile à comprendre des prix de l'électricité. Les utilisateurs peuvent y trouver des informations détaillées et actualisées sur les prix de l'électricité dans différentes régions de la Suisse. Dans le cadre des améliorations

³⁷ <https://www.prix-electricite.elcom.admin.ch/>

futures, l'utilisation des données de Swisstopo et de l'outil "Visualize"³⁸ de l'Office fédéral de l'environnement est envisagée.

Globalement, l'initiative d'ElCom pour l'ouverture des données est un exemple de la façon dont la publication des données liées peut promouvoir la transparence, l'accessibilité et l'innovation au sein des secteurs publics clés. Il met également en évidence l'importance de proposer des interfaces utilisateur attrayantes et intuitives pour faciliter l'accès du public aux données ouvertes.

Enfin, il est important de noter l'intérêt potentiel de ces informations pour Monsieur Prix et le SECO. De plus, bien que l'application reçoive d'ordinaire environ 1'000 clics par an, ce nombre a grimpé à environ 173'000 en septembre 2022, lors des annonces d'augmentation des prix de l'électricité.

6.2.4 Amélioration de la collaboration dans le secteur de la santé

Dans le contexte de la santé, les avantages des données liées sont bien réels. Le Swiss Personalized Health Network (SPHN) est une illustration éloquent de l'exploitation réussie des données liées pour améliorer la collaboration et l'interopérabilité dans le secteur de la santé.

Le SPHN, en tant qu'initiative nationale, est dédié à la promotion de la médecine personnalisée en Suisse. Pour atteindre cet objectif, le SPHN a recours aux données liées afin de faciliter la coopération entre les différents intervenants du secteur de la santé.

En ce sens, le SPHN a mis en place le service web « SPHN Schema Forge » et l'outil « SPHN Dataset2RDF Tool »³⁹. Le premier permet aux chercheurs et aux professionnels de la santé de concevoir des schémas de données harmonisés pour la collecte, le stockage et le partage des données de santé, assurant une meilleure comparabilité et intégrité des données. Le deuxième, quant à lui, facilite la conversion des ensembles de données de santé en un format compatible avec les technologies de données liées. Cela permet une meilleure accessibilité et exploitabilité des données de santé, contribuant ainsi à une collaboration, et une recherche, plus efficaces.

Récemment, le rôle du SPHN en tant que Data Coordination Center (DCC) a été confirmé et renforcé pour 2025. Ce renforcement témoigne de la reconnaissance croissante de l'importance des données liées dans le secteur de la santé et de leur potentiel pour améliorer la collaboration et la coordination dans le domaine de la santé à l'échelle nationale.

6.2.5 Renforcement de la collaboration internationale

Les avancées technologiques et l'interconnexion globale ont changé la façon dont les nations collaborent. Dans ce contexte, l'utilisation des données liées s'est avérée être un outil essentiel pour faciliter et améliorer la collaboration internationale. Les données liées permettent une communication plus fluide, une coopération plus efficace et des interactions plus transparentes entre les différentes organisations, qu'elles soient publiques ou privées.

Un exemple éminent de cette application des données liées est le SEMIC Support Centre⁴⁰ de l'Union Européenne. Le SEMIC (Semantic Interoperability Community) est une initiative de l'Union Européenne (UE) qui vise à améliorer l'interopérabilité sémantique entre les systèmes informatiques européens. Grâce à l'utilisation des données liées, SEMIC permet d'harmoniser la terminologie à l'échelle de l'Union, rendant ainsi les informations plus accessibles et compréhensibles, indépendamment de la langue ou du pays d'origine. Leur travail sur la clarification terminologique est disponible sur SEMIC.

De plus, le projet « Authority tables » de l'Office des publications de l'UE illustre comment les données liées peuvent être utilisées pour créer une base de données plus complète et plus efficace. Ces tables d'autorité fournissent un ensemble de valeurs standardisées pour divers concepts utilisés dans les publications de l'UE, garantissant ainsi une compréhension uniforme et une utilisation cohérente des termes⁴¹.

³⁸ <https://visualize.admin.ch/en/browse?includeDrafts=true&search=elcom&order=SCORE&dataSource=Prod>

³⁹ <https://sphn.ch/2023/05/02/sphn-schema-forge-release/>

⁴⁰ <https://joinup.ec.europa.eu/collection/semic-support-centre>

⁴¹ <https://semiceu.github.io/style-guide/1.0.0/terminological-clarifications.html>

7 Recommandations pour la mise en œuvre des données liées

7.1 Facteurs de réussite pour une intégration efficace des données liées

Une intégration progressive des données liées au sein de l'Administration fédérale pourrait se faire en plusieurs étapes, suivant l'ordre des facteurs de réussite définis pour chaque cas spécifique. Chaque étape mettrait en pratique les facteurs de réussite appropriés, depuis la sensibilisation et la formation initiales jusqu'à la garantie de la qualité, la promotion des données liées, et la garantie de la sécurité et de la confidentialité. Un suivi régulier de la progression de la mise en œuvre selon la stratégie établie serait également nécessaire.

La mise en œuvre réussie des données liées au sein de l'Administration fédérale dépend de plusieurs facteurs clés qui sont décrits dans les sous-chapitres suivants.

7.1.1 Sensibilisation et formation

La première étape est la sensibilisation et la formation. Cela signifie que les décideurs, les développeurs, les architectes de solutions et le personnel technique spécialisé doivent comprendre ce que sont les données liées, comment elles fonctionnent, quels en sont les avantages et comment elles peuvent être utilisées.

La formation devrait couvrir une gamme de sujets, allant des principes de base des données liées (comme le RDF, le SPARQL et les URIs) à des sujets plus avancés comme la modélisation des données, l'interopérabilité et la transformation des données. Les formations pourraient également inclure des ateliers pratiques, où les participants peuvent s'exercer à utiliser les outils et les techniques des données liées.

7.1.2 Elaboration d'une tactique

Le secteur Transformation numérique et gouvernance de l'informatique de la Chancellerie fédérale (TNI)⁴² peut jouer un rôle en coordonnant l'élaboration d'une tactique, en collectant et en publiant les stratégies existantes et les meilleures pratiques dans ce domaine.

En parallèle de la sensibilisation et de la formation, une tactique claire pour l'adoption et la mise en œuvre des données liées peut être élaborée. Cette tactique devrait définir les objectifs à atteindre, les ressources nécessaires, un calendrier réaliste pour la mise en œuvre, et les méthodes pour évaluer le succès.

Il est important de noter que cette tactique doit également prendre en compte la disponibilité et la gestion des données. En effet, l'adoption réussie des données liées nécessite un inventaire précis des jeux de données dont dispose un office, un prérequis qui n'est actuellement pas toujours atteint. L'inventaire des données est également une tâche clé dans la mise en œuvre d'un Système de Gestion de la Sécurité de l'Information (ISMS), avec l'inventaire des objets de protection primaires (les données) et secondaires (les applications qui manipulent ces données).

La tactique doit également identifier les ensembles de données qui bénéficieraient le plus de la transformation en données liées. Ces ensembles de données pourraient être ceux qui sont souvent utilisés, qui sont essentiels pour la prise de décisions, ou qui présentent un potentiel d'interopérabilité avec d'autres ensembles de données.

7.1.3 Choix des normes et identification des données

Une fois la tactique définie, il est essentiel de sélectionner les normes pour la mise en œuvre des données liées. Ce choix doit être guidé par la prévalence et la reconnaissance dans le domaine des données liées. À cet égard, l'administration fédérale devrait établir un cadre qui privilégie l'utilisation de RDF (Resource Description Framework) et de SPARQL (SPARQL Protocol and RDF Query Language) comme normes de base.

En outre, le cadre pourrait également recommander l'utilisation de certaines ontologies, comme RDFS (RDF Schema), OWL (Web Ontology Language), SKOS (Simple Knowledge Organization System), Dublin Core,

⁴² <https://www.bk.admin.ch/bk/fr/home/digitale-transformation-ikt-lenkung/bereichdti.html>

DCAT (Data Catalog Vocabulary), et schema.org. Bien que le choix des ontologies soit moins critique que le choix de RDF et SPARQL, leur utilisation peut améliorer la compatibilité et l'interopérabilité des données.

Au-delà de ces normes de base, il existe également des modèles spécifiques qui peuvent être appliqués pour structurer les données. Par exemple, SKOS (Simple Knowledge Organization System), Cube.link, DCAT (Data Catalog Vocabulary), PROV-O (Provenance Ontology) et Version.link sont utilisés pour définir des modèles de données génériques. Pour des domaines spécifiques, des modèles tels que JOLux, ESCO, Schema.org, et d'autres peuvent être employés.

L'identification des ensembles de données à transformer en données liées est une autre étape importante. Cette tâche implique généralement une évaluation approfondie des ensembles de données existants pour déterminer leur pertinence et leur potentiel pour la transformation en données liées. L'évaluation peut impliquer une analyse des critères tels que la fréquence d'utilisation des données, leur importance pour la prise de décisions, et leur potentiel d'interopérabilité avec d'autres systèmes ou ensembles de données.

En fonction des domaines spécifiques des données, il est à noter que certaines ontologies, qui s'appuient sur RDF, ont été développées au niveau international. Il serait donc souhaitable de s'appuyer sur ces modèles thématiques une fois les jeux de données identifiés.

L'objectif est de sélectionner des ensembles de données qui, une fois transformés en données liées, apporteront le plus de valeur ajoutée à l'organisation. Cela peut nécessiter une connaissance approfondie des données disponibles, ainsi qu'une compréhension claire des objectifs et des besoins de l'organisation en matière de données.

7.1.4 Modélisation, transformation et publication des données

Une fois les ensembles de données identifiés, ils doivent être modélisés en utilisant des ontologies et des vocabulaires standardisés. Les données modélisées sont ensuite transformées en données liées, ce qui nécessite des compétences techniques spécialisées et l'utilisation d'outils dédiés.

Après la transformation, les données liées doivent être publiées de manière à être facilement accessibles. Les plateformes de publication, y compris les plateformes de référencement comme I14Y ou opendata.swiss, doivent être choisies en fonction de leur facilité d'utilisation, de leur capacité à gérer des données liées, de leur compatibilité avec les normes choisies et de leur capacité à garantir la sécurité et la confidentialité des données.

Des technologies comme les triplestores (des bases de données conçues pour stocker des données RDF) et des interfaces de requêtes SPARQL sont utilisées pour ces fins. Ce n'est plus une possibilité, mais une pratique courante dans la gestion et la publication des données liées.

La publication des données doit respecter les réglementations en vigueur concernant la confidentialité et la protection des données. Des mécanismes d'accès contrôlé, de cryptage et d'audit pourraient être nécessaires pour garantir la sécurité des données.

7.1.5 Garantie de qualité et promotion

Après la publication des données liées, l'implémentation de processus de contrôle de qualité permet de s'assurer la fiabilité et la précision de ces données. Ce processus devrait intégrer des vérifications régulières, des mécanismes de feedback pour signaler les erreurs et des mises à jour fréquentes pour garantir que les données restent actuelles et pertinentes. À ce sujet, il est possible de recourir à des outils tels que SHACL-Checker, qui permettent d'automatiser en grande partie ce processus lors de l'utilisation de RDF. Il est important de noter que ce contrôle de qualité devrait également être effectué en amont, avant la conversion des données en RDF, s'inscrivant ainsi dans une gestion de qualité globale et non spécifique aux données liées.

Parallèlement à ces actions de contrôle de qualité, l'utilisation des données liées doit être promue. Des initiatives comme des présentations, ateliers ou démonstrations peuvent être mises en place pour sensibiliser les utilisateurs potentiels. L'accompagnement technique pour faciliter l'utilisation des données liées peut aussi être envisagé. De plus, il est essentiel de s'associer à des institutions éducatives, comme les

Hautes Ecoles Spécialisées (HES), afin de renforcer la formation des développeurs d'applications dans l'utilisation de ces données liées.

7.1.6 Sécurité et confidentialité

Dans le cadre de l'intégration des données liées, la sécurité et la confidentialité des données doivent être garanties. Il est nécessaire de distinguer deux contextes d'utilisation : les données ouvertes et les données fermées. Dans le cas des données ouvertes, l'objectif principal est la publication de données, la confidentialité n'est donc pas un problème critique. Cependant, pour les données fermées, qui sont destinées à un public restreint, des mesures spécifiques de sécurité et de confidentialité doivent être prises en compte. Ces mesures peuvent inclure le chiffrement des données, l'authentification des utilisateurs, l'autorisation d'accès aux données, l'audit des accès aux données et la conformité avec les réglementations en matière de protection des données.

Au-delà de la mise en place de ces mesures, il convient d'éduquer les utilisateurs sur les bonnes pratiques en matière de sécurité des données. Il convient de les tenir informés des politiques et procédures en vigueur. Toutefois, cette éducation à la sécurité des données est une question générale qui dépasse le cadre spécifique des données liées.

7.1.7 Suivi régulier

Enfin, une fois que l'intégration des données liées a commencé, un suivi régulier est nécessaire pour évaluer les progrès et ajuster la stratégie si nécessaire. Cela pourrait impliquer la collecte de métriques sur l'utilisation des données liées, le suivi des problèmes signalés par les utilisateurs, l'évaluation de la satisfaction des utilisateurs et la révision régulière de la stratégie d'implémentation. Ce suivi permettrait d'assurer que les objectifs sont atteints et que les données liées sont utilisées de manière efficace. L'importance de ce suivi régulier souligne la nécessité d'une stratégie transversale plutôt que d'une stratégie spécifique à chaque office. Cette approche garantit une vision plus cohérente et plus globale, optimisant ainsi l'efficacité et l'utilisation des données liées.

7.2 Rôles et responsabilités dans le processus de réutilisation des données

La mise en œuvre des données liées au sein de l'Administration fédérale est un effort collectif qui nécessite la participation de divers acteurs clés, chacun jouant un rôle unique. Il est judicieux de se référer au catalogue existant des fonctions IT afin d'éviter de créer une nouvelle structure. Les groupes d'acteurs clés incluent :

- **Les décideurs** : Ils jouent un rôle clé dans le soutien et l'orientation de la transition vers l'utilisation des données liées. Leur rôle consiste à comprendre les avantages et les principes des données liées, à prendre des décisions stratégiques éclairées et à allouer les ressources nécessaires à la mise en œuvre. Il est essentiel que les décideurs soient bien informés et convaincus des avantages que les données liées peuvent apporter à l'Administration fédérale.
- **Les responsables de la gestion des données** : Ils supervisent le processus de sélection, de transformation et de publication des données. Ils doivent s'assurer que les ensembles de données appropriés sont identifiés, que ces données sont correctement transformées en données liées conformément aux normes et formats appropriés, et qu'elles sont ensuite publiées de manière efficace et sécurisée. Ils sont également responsables de la mise en place de processus pour maintenir la qualité des données, en veillant à leur validation et à leur mise à jour régulière.
- **Les développeurs et les spécialistes des technologies de l'information** : Ces experts techniques sont chargés de la mise en œuvre pratique des normes et des outils nécessaires à l'utilisation des données liées. Ils ont la responsabilité de développer et de maintenir les API et les points d'accès SPARQL, qui permettent l'échange et l'accès aux données liées. En outre, ils jouent un rôle dans la garantie de la sécurité des données. La formation des fournisseurs de services de l'office fédéral

de l'informatique et de la télécommunication (OFIT)⁴³ est une initiative importante pour assurer leur maîtrise de ces technologies et leur proposition pour le développement de solutions web.

Chacun de ces acteurs joue un rôle important dans le processus de mise en œuvre des données liées, mais aussi dans leur réutilisation dans des applications internes ou des portails web. Leur collaboration et leur engagement sont essentiels pour assurer une adoption réussie des données liées au sein de l'Administration fédérale.

7.3 Bases légales

Sur le plan juridique, l'utilisation des données liées est soutenue par plusieurs lois et ordonnances. En vertu de l'art. 3 de l'Ordonnance sur la coordination de la transformation numérique et la gouvernance de l'informatique dans l'administration fédérale (OTNI⁴⁴), la Chancellerie fédérale a la charge de la transformation numérique dans ses domaines de compétences et assure la gouvernance de l'informatique. En conséquence, le secteur TNI de la Chancellerie fédérale est chargé de définir et de mettre en œuvre des modèles de données pour l'administration fédérale (art. 4 al. 2 OTNI) et de gérer des instruments aidant à la gouvernance des données (art. 4 al. 3 OTNI). Selon l'art. 17 al. 1 let. d, le secteur TNI est également chargé d'établir les normes découlant de l'architecture d'entreprise, nécessaires pour assurer l'interopérabilité.

D'autre part, la Loi fédérale sur l'utilisation de moyens électroniques pour l'exécution des tâches des autorités (LMETA⁴⁵), qui est entrée en vigueur en janvier 2024, prévoit la mise en ligne des données ouvertes sous une forme lisible par machine (art. 10 al. 4). Cela rend l'utilisation des données liées parfaitement admissible et favorable à la mise en œuvre de cet article. Selon l'art. 12 al. 1, la Chancellerie fédérale peut imposer aux autorités fédérales des normes techniques destinées à favoriser l'interopérabilité de différents systèmes. En outre, la régulation de la forme des métadonnées, qui doivent être publiées sur la plateforme d'interopérabilité, peut être réalisée par l'Office fédéral de la statistique (OFS) en accord avec la Chancellerie fédérale (art. 14 al. 2).

Ces dispositions légales établissent une base pour l'adoption des données liées au sein de l'Administration fédérale, en favorisant leur utilisation pour améliorer l'interopérabilité, la gouvernance des données et l'exécution des tâches publiques.

8 Vision future de l'utilisation des données liées dans l'Administration fédérale

8.1 Comment l'utilisation des données liées pourrait évoluer à l'avenir

Au fur et à mesure de son adoption non seulement au sein de l'Administration fédérale mais aussi auprès des prestataires informatiques, l'usage des données liées pourrait s'étendre. Cela ouvrirait la voie à une plus grande interopérabilité entre les différentes entités administratives, permettant une intégration approfondie de données issues de diverses sources. En conséquence, les analyses pourraient être plus profondes et précises, offrant une vision complète et claire des divers enjeux administratifs et politiques.

Avec la progression de la technologie et une adoption croissante des données liées tant en interne qu'auprès des prestataires, de nouveaux outils et techniques pourraient voir le jour. L'intelligence artificielle (IA) et l'apprentissage automatique (« Machine Learning ») pourraient offrir des perspectives inédites pour l'exploitation de ces données. Par exemple, des algorithmes d'apprentissage pourraient détecter des connexions entre les données de manière optimale, tandis que l'IA pourrait faciliter l'analyse et l'interprétation de ces connexions, permettant ainsi une prise de décision basée sur des informations tangibles.

Il est essentiel de noter que ces données, étant stockées dans une base de données, sont interrogeables par des applications, en particulier web. Elles suivent ainsi le modèle de toute autre base de données relationnelle associée à une application. Avec ces données structurées dans un triplestore, une séparation

⁴³ <https://www.bit.admin.ch/bit/fr/home.html>

⁴⁴ RS 172.010.58 - Ordonnance du 25 novembre 2020 sur la coordination de la transformation numérique et la gouvernance de l'informatique dans l'administration fédérale (Ordonnance sur la transformation numérique et l'informatique, OTNI)

⁴⁵ FF 2023 787 - <https://www.fedlex.admin.ch/eli/fga/2023/787/fr>

vitale entre les données et les applications est établie. Cela permettrait à l'administration de reprendre le contrôle sur ses données, laissant aux développeurs la mission de construire des applications basées sur nos données, contrairement à la situation actuelle où les développeurs maîtrisent à la fois les données et le code.

L'adoption des données liées pourrait également mener à la création d'applications centrées sur l'utilisateur, offrant une interaction et une personnalisation accrues. En effet, la capacité de personnaliser les services, notamment gouvernementaux, pourrait être décuplée grâce à l'utilisation des données liées. Par exemple, des services gouvernementaux pourraient, tout en respectant les lois sur la protection des données ainsi que sur protection et le respect de la sphère privée, utiliser des données liées pour fournir des informations pertinentes adaptées à des groupes généraux d'utilisateurs. De plus, les données liées pourraient faciliter la création d'interfaces utilisateurs interactives, rendant la navigation et l'interaction avec les données plus instinctives pour les citoyens.

Enfin, il est nécessaire de souligner l'importance des métadonnées comme structures descriptives. L'utilisation de données liées dans l'inventaire, la classification, et l'enrichissement d'un catalogue de métadonnées local pour une entité administrative est nécessaire. Cela permettrait à l'entité de mieux maîtriser son patrimoine numérique, une tâche qui deviendra de plus en plus centrale dans les années à venir.

8.2 Opportunités potentielles pour une utilisation plus large des données liées

L'utilisation des données liées possède un potentiel allant bien au-delà de l'Administration fédérale. Elles pourraient influencer divers secteurs tels que l'éducation, la recherche, la santé, et l'industrie.

Dans l'éducation, ces données pourraient faciliter des parcours d'apprentissage sur mesure, adaptés aux compétences, intérêts et antécédents des étudiants. Dans le domaine de la santé, et en écho à des initiatives telles que le SPHN Semantic Framework⁴⁶, elles pourraient soutenir des soins personnalisés, associant les dossiers médicaux des patients aux dernières découvertes cliniques, accentuant ainsi l'efficacité des traitements. L'industrie, quant à elle, bénéficie déjà de ces technologies comme le montre l'exemple de l'entreprise NXP avec son « Enterprise Data Hub »⁴⁷. L'utilisation de données liées permet une analyse approfondie des performances de production et des chaînes d'approvisionnement, offrant des insights précieux.

L'application de ces données pourrait mener à l'émergence de nouveaux services, applications, et innovations, redéfinissant les horizons de l'économie numérique.

Dans le secteur public, l'utilisation des données liées pourrait instaurer une transparence inégalée. En interconnectant les informations gouvernementales, elles pourraient offrir au public une compréhension claire des actions gouvernementales et de l'utilisation des ressources publiques. De plus, l'intégration des données liées pourrait booster la participation citoyenne, en offrant à chacun la possibilité de contribuer aux bases de données publiques. Pour illustrer, les citoyens pourraient enrichir les bases de données environnementales par leurs observations locales.

En conclusion, les données liées se présentent comme un pilier de la transformation numérique de notre société. En fusionnant des informations issues de différentes sources et en optimisant leur analyse, ces données ont le potentiel d'aborder et de surmonter les défis majeurs de notre époque, tels que le changement climatique, les enjeux de santé publique ou les inégalités socio-économiques. Elles incarnent la transition de l'information en connaissance, et de la connaissance en action concrète.

9 Conclusion

Au travers de notre exploration de la technologie des données liées, il est devenu évident que son incorporation est une étape essentielle pour la modernisation et l'optimisation des services de l'Administration fédérale. Avec leurs capacités à faciliter l'interopérabilité, la transparence, l'efficacité et l'innovation, les données liées ont le potentiel non seulement d'améliorer la manière dont l'information est

⁴⁶ <https://sphn-semantic-framework.readthedocs.io/en/latest/index.html>

⁴⁷ <https://2017.semantics.cc/nxp-enterprise-data-hub>

gérée et exploitée, mais aussi de révolutionner la collaboration et la prise de décision au sein de l'administration.

Au long de ce rapport, nous avons examiné les principes sous-jacents des données liées, leur fonctionnement, les normes techniques associées et surtout, l'impact réel et potentiel de leur mise en œuvre, comme en témoigne le succès de l'Office fédéral de topographie Swisstopo ou du service LINDAS proposé par les Office fédéral des archives fédérales. Nous avons également discuté des avantages substantiels qu'elles pourraient apporter à l'Administration fédérale, en termes d'interopérabilité améliorée, d'efficacité opérationnelle et de transparence.

Annexe A - l'évolution du web au fil du temps

L'histoire du web, en partant du web 1.0 jusqu'à l'arrivée prochaine du web 4.0, illustre une transition majeure de la diffusion unidirectionnelle d'informations à l'immersion complète dans un monde numérique omniprésent. Pour l'administration publique, ces transitions ont redéfini les relations avec les citoyens et la manière dont les services sont délivrés.

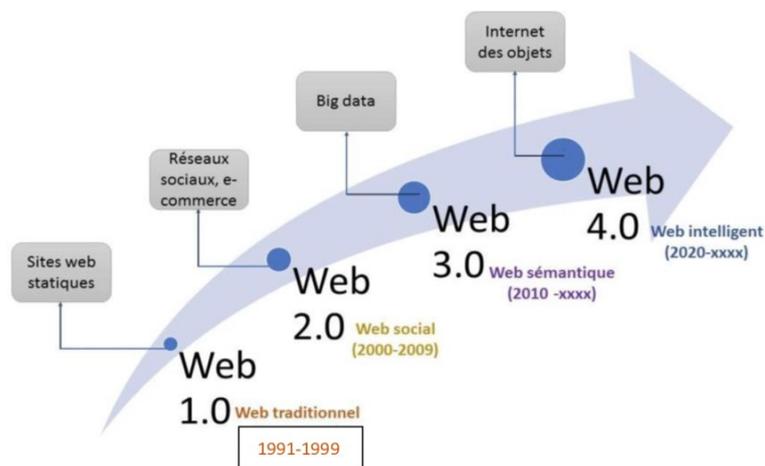


Figure 6 : Evolution du web depuis 1991⁴⁸

Pour l'administration publique, ces transitions soulignent la nécessité de développer des compétences en gestion des données liées. La capacité à exploiter efficacement ces données peut non seulement améliorer la prestation de services, mais aussi renforcer la confiance des citoyens et augmenter leur engagement.

Le Web 1.0

Le web 1.0, souvent qualifié de « web statique », est la première version du World Wide Web qui a vu le jour dans les années 1990. Dans cette phase du développement du web, les sites web étaient principalement des documents statiques hébergés sur un serveur spécifique. Ils étaient principalement utilisés pour diffuser de l'information et ne proposaient pas d'interactivité ou d'échange d'information entre l'utilisateur et le site web.

Le contenu du web 1.0 était créé et mis à jour par des webmasters, et les utilisateurs étaient essentiellement des lecteurs ou des consommateurs de ces contenus. Le web 1.0 est souvent comparé à une bibliothèque ou à un catalogue en ligne où les utilisateurs peuvent chercher et lire des informations, mais ils ne peuvent pas interagir avec ces informations, les modifier ou les commenter.

Parmi les exemples typiques de sites web 1.0, on peut citer les sites web d'entreprises qui se limitent à présenter les produits et services de l'entreprise, les sites web d'information ou les premiers sites d'e-commerce qui permettaient simplement d'acheter des produits ou services mais sans interactivité ou possibilité de donner son avis. Cependant, malgré leurs limitations, dans le web 1.0, le contenu des pages web était facilement analysable car les attributs html de structuration servaient à définir la mise en page du contenu.

Le Web 2.0

Le Web 2.0, souvent appelé « Web participatif » ou « Web social », est la deuxième génération du World Wide Web qui a émergé au début des années 2000. Par rapport à la première version du Web, le Web 2.0 est caractérisé par une interaction plus grande entre les utilisateurs et le contenu en ligne, ainsi que par une plus grande interaction entre les utilisateurs eux-mêmes.

⁴⁸ Source : https://www.researchgate.net/figure/du-Web-10-au-Web-40-son-evolution-depuis-1991-Source-Sika-Technologie-mars-2017_fig2_363135944

Dans le Web 2.0, les sites web ne sont plus simplement des points de diffusion d'information, mais deviennent des plateformes permettant aux utilisateurs de créer et de partager leur propre contenu. Le Web 2.0 comprend une large gamme de formats, tels que les blogs, les wikis, les réseaux sociaux, les plateformes de partage de vidéos ou de photos, et autres applications centrées sur l'utilisateur.

Ces caractéristiques du Web 2.0 permettent une plus grande participation des utilisateurs, qui deviennent des créateurs de contenu et non plus simplement des consommateurs. Cela a conduit à l'essor de la « culture participative », où tout un chacun peut contribuer à la création de contenu et influencer le Web. Le Web 2.0 a également facilité l'émergence de nouvelles formes de collaboration et de partage d'informations, grâce à des technologies telles que le RSS, l'Ajax et les API Web.

Cependant, avec ces avancées sont également apparus de nouveaux défis. Non seulement il a fallu gérer une surcharge d'information, mais également mettre en place de nouvelles approches en matière de confidentialité et de sécurité des données. De plus, des questions concernant la qualité et la fiabilité du contenu généré par les utilisateurs se sont posées. Un autre corollaire important de ces développements a été le passage d'un « langage de marquage » des pages à un « langage de maquillage » des pages, marquant une transition de la structure à l'esthétique. Cela a augmenté la complexité des pages web, rendant l'interprétation automatique du contenu plus difficile. Par conséquent, les données qui sont publiées dans ces pages sont souvent inutilisables en tant que telles, ce qui représente un défi pour l'exploitation des données à grande échelle.

Web 3.0 et le web sémantique

Le Web 3.0, également connu sous le nom de « Web sémantique » ou « Web des données », est la prochaine étape de l'évolution du World Wide Web⁴⁹. Alors que le Web 2.0 se concentrait sur l'interaction sociale et la participation des utilisateurs, le Web 3.0 vise à apporter une plus grande intelligence et une meilleure compréhension des données en ligne. Dans un sens, c'est un retour aux sources : tout comme le Web 1.0, le Web 3.0 met l'accent sur la structure et l'organisation des informations.

Il vise à organiser les informations de manière plus sémantique, c'est-à-dire en leur donnant du sens et en les reliant de manière logique. Il cherche à permettre aux ordinateurs de comprendre le contenu des pages web de manière similaire à la façon dont les humains le font. Cette ambition est rendue possible grâce à l'utilisation de technologies et de normes spécifiques. Parmi ces outils, on trouve le modèle de données RDF et ses différentes sérialisations (RDF/XML, Turtle, N3, etc.) qui permettent une mise en œuvre dans les pages HTML à travers RDFa, Microdata, JSON-LD, etc. Sur ce modèle de données, des ontologies et des bases de connaissances peuvent être construites et utilisées, enrichissant le potentiel d'interprétation et de liaison des données.

Grâce au Web 3.0, les machines peuvent interpréter et analyser les informations de manière plus sophistiquée, permettant une recherche plus précise, des recommandations personnalisées et des services plus intelligents. Par exemple, un moteur de recherche Web 3.0 pourrait comprendre la signification d'une demande de l'utilisateur, et fournir des résultats plus pertinents en prenant en compte son contexte. C'est dans cette perspective que Google, par exemple, a amélioré la qualité de ses résultats de recherche. En effet, lors du processus d'analyse des sites web, Google prend en compte les attributs RDFa qui utilisent l'ontologie schema.org, offrant ainsi une meilleure contextualisation et pertinence des résultats pour l'utilisateur.

Cette version du Web vise également à améliorer l'interopérabilité des données en permettant à différentes sources de données de se connecter et d'échanger des informations de manière transparente. Cela ouvre la voie à des applications plus avancées, telles que l'Internet des objets (IoT) et les villes intelligentes, où les dispositifs et les systèmes sont interconnectés et échangent des données en temps réel.

Le futur Web 4.0

Le concept de Web 4.0, aussi souvent désigné sous les termes de « Symbiotic Web » ou « Pervasive Web », est encore en cours de développement et suscite de nombreuses discussions et spéculations. Le Web 4.0 représente une vision de l'avenir dans laquelle le web est profondément enraciné dans tous les aspects de notre vie quotidienne, créant une symbiose entre les mondes physique et numérique. Cela permettrait par exemple à

⁴⁹ https://fr.wikipedia.org/wiki/Web_s%C3%A9mantique

chaque véhicule ou individu d'être directement informé, rendant potentiellement obsolètes les feux de signalisation ou les passages piétons.

Toutefois, il est généralement considéré que le Web 4.0 représente une évolution ultérieure du Web qui explore les possibilités de l'intelligence artificielle (IA) et de l'Internet des objets (IoT) pour créer un environnement numérique plus avancé. Il vise à intégrer les technologies intelligentes dans notre quotidien, connectant les objets physiques et virtuels pour permettre des interactions plus naturelles et personnalisées.

Dans cet écosystème, les objets connectés peuvent communiquer entre eux de manière autonome, prenant des décisions basées sur les données qu'ils collectent et les modèles qu'ils apprennent grâce à des technologies telles que l'apprentissage automatique, le traitement du langage naturel et la vision par ordinateur. Cette intégration profonde permet des applications innovantes telles que les maisons intelligentes, les véhicules autonomes, les soins de santé intelligents, et les villes intelligentes, entre autres.

Annexe B - Compréhension des données liées

Définition et explication du concept des données liées

A la base du concept de données liées, nous avons l'idée que les données peuvent être mises en relation sans être limitées par des contraintes techniques ou organisationnelles. Un peu de la même manière que l'Internet relie les pages HTML.

Chaque élément dans un réseau de données liées, qui peut comprendre des éléments variés tels que des organisations, des thèmes, des lieux, des personnes, etc., est attribué à un identificateur de ressource uniforme (URI). Ces URI sont univoques et lisibles par machine, créant ainsi un maillage global d'informations interconnectées. Grâce à cette structure, il est possible de faire des requêtes sur des données hétérogènes et isolées les unes des autres. Les résultats de ces requêtes peuvent révéler des relations et insights à la fois nouveaux et innovants.

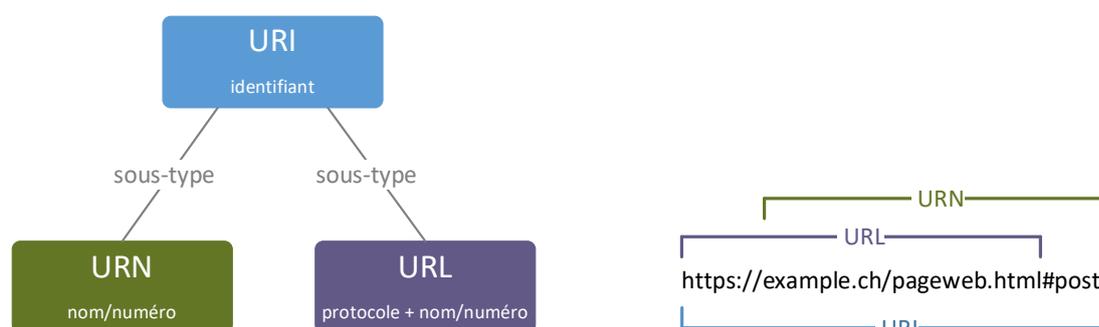


Figure 7 : identificateur de ressource uniforme (URI)

Un élément essentiel du concept des données liées est le Resource Description Framework (RDF). RDF est un modèle de données conçu à l'origine pour décrire les métadonnées des ressources informatiques. Aujourd'hui, RDF est utilisé dans un contexte plus large, y compris en relation avec le web sémantique. Ce modèle représente les informations sous la forme sujet-prédicat-objet, permettant ainsi de construire des structures de graphes orientés et étiquetés. Les systèmes capables de stocker et de traiter ces triplets sont appelés triplestores, une sous-classe spécialisée des systèmes de gestion de bases de données (SGBD) de graphes, spécifiquement conçus pour gérer les données RDF dans le cadre du web sémantique. Il existe d'autres types de bases de données de graphes, mais les triplestores sont particulièrement adaptés à la modélisation et à la gestion des données liées.

Modélisation, représentation et fonctionnement des données liées

Normes et langages pour les données liées

La modélisation et la représentation des données liées dépendent fortement de l'usage de certaines normes et langages spécifiques. En particulier, RDF (Resource Description Framework), RDFS (RDF Schema), et OWL (Web Ontology Language) jouent un rôle dans la structuration et la sémantique des données liées. Ces normes sont souvent représentées sous la forme d'une pyramide de complexité, où RDF constitue la base. Dans cette représentation, l'axe vertical (y) correspond à la complexité croissante des technologies et des normes, tandis que l'axe horizontal (x) peut représenter les différentes fonctionnalités ou composants des données liées. RDF, en tant que langage standard pour coder des informations en format de graphe, sert de fondement, utilisé pour décrire des ressources sur le Web. À mesure que l'on monte dans la pyramide, on trouve des normes et des technologies de plus en plus sophistiquées, construites sur la base solide que fournit RDF.

Comprendre RDF et les bénéfices des données liées

Au-dessus de RDF, RDFS et OWL définissent un modèle de classes qui ajoute plus de structure et de sémantique aux données. RDFS est une extension de RDF qui offre des capacités de modélisation de données supplémentaires, permettant de décrire des hiérarchies de classes et de propriétés.

OWL, d'autre part, fournit des capacités encore plus riches pour exprimer la sémantique des données, en allant au-delà des simples relations sujet-prédicat-objet fournies par RDF. OWL permet de définir des classes, des propriétés et leurs instances, ainsi que des relations entre elles, facilitant l'expression de concepts plus complexes et de logique d'inférence. L'utilisation de ces normes et langages, y compris OWL, permet une représentation plus précise des données, en ajoutant des couches de signification et de contexte qui facilitent leur interprétation et leur utilisation dans divers contextes.

Les données liées reposent sur plusieurs principes et technologies clés. L'un des plus importants est le Resource Description Framework (RDF), qui est le socle du concept de données liées. Comprendre RDF et son utilité dans la modélisation des données est un préalable indispensable pour appréhender le fonctionnement des données liées.

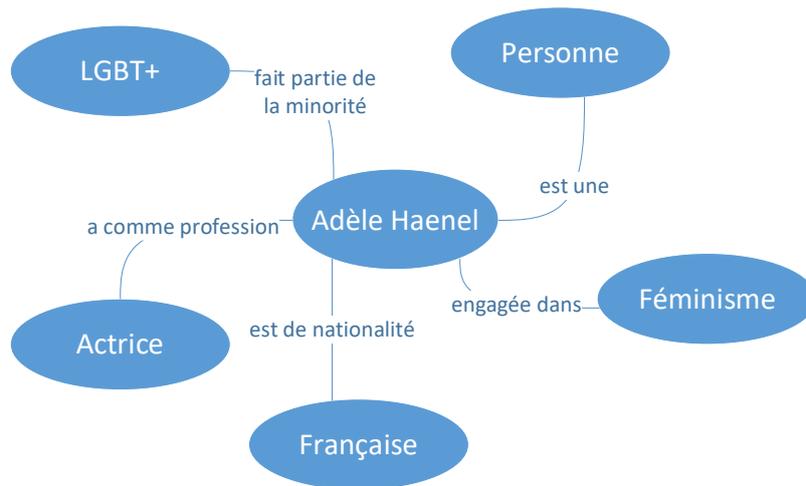


Figure 8 : Graph RDF basic

L'adoption des données liées, soutenues par des technologies comme RDF et SPARQL (SPARQL Protocol and RDF Query Language - une langue de requête utilisée pour récupérer et manipuler les données stockées en RDF), offre une approche plus flexible de la gestion des données. SPARQL permet d'interroger et de naviguer dans les données RDF, facilitant l'accès à des informations spécifiques et leur manipulation. Cela permet non seulement une meilleure interopérabilité entre différentes sources de données, mais également une meilleure accessibilité et une plus grande facilité d'utilisation pour tous les utilisateurs.

La modélisation des données avec RDF

Le Resource Description Framework (RDF) est un modèle standardisé conçu pour un échange de données flexible et adaptable à de nombreux scénarios. Son utilisation dépasse le cadre des systèmes de gestion de bases de données (DBMS) traditionnels et les restrictions syntaxiques. Un lien important avec le Web est établi par la syntaxe des identificateurs universels de ressources (URI) dans RDF, qui reprend celle des URL des pages web. Cependant, il est important de noter que l'usage de RDF n'est pas directement lié au web mais plutôt à son approche universelle de la description des ressources.

RDF propose une représentation des données basée sur des triplets, formés d'un sujet, d'un prédicat et d'un objet (S-P-O). Chaque triplet forme une déclaration factuelle, comme "Paris a pour pays la France", où "Paris" est le sujet, "a pour pays" est le prédicat, et "la France" est l'objet. En rassemblant plusieurs de ces triplets, on peut créer un graphe de données, qui offre une visualisation riche et connectée des relations entre divers ensembles de données.

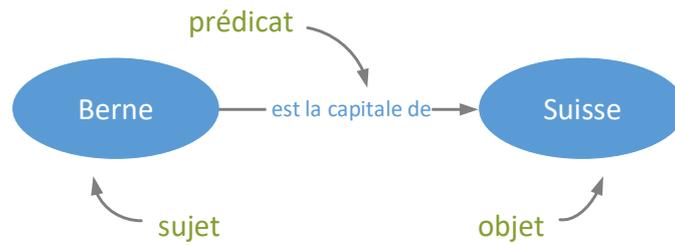


Figure 9 : Triplet RDF

RDF offre un avantage en termes d'interopérabilité, car il permet de représenter des données de manière indépendante de la syntaxe. Plusieurs syntaxes, comme RDF/XML, Turtle, JSON-LD et RDFa, peuvent être utilisées pour exprimer des données RDF, offrant une flexibilité sans précédent pour représenter et échanger des données sur différents systèmes et plateformes.

Les exemples ci-dessous montrent comment le même graphe RDF peut être exprimé en différentes syntaxes tout en conservant la même signification sémantique.

```
/* Exemple avec JSON-LD */
{
  "@context" : {
    "ex" : "http://example.ch/"
  },
  "@id" : "ex:AdèleHaenel",
  "@type" : "ex:Person",
  "ex:partOfMinority" : "ex:LGBT+",
  "ex:engagedIn" : "ex:Feminism",
  "ex:nationality" : "ex:French",
  "ex:profession" : "ex:Actress"
}
```

```
# Exemple de code avec turtle
@prefix ex: <http://example.org/> .
ex:AdèleHaenel a ex:Person ;
ex:partOfMinority ex:LGBT+ ;
ex:engagedIn ex:Feminism ;
ex:nationality ex:French ;
ex:profession ex:Actress .
```

Notez que ces exemples sont une simplification et dans un scénario réel, les URI pour les individus et les propriétés devraient être bien définis pour correspondre à des entités réelles et uniques dans votre domaine de connaissance.

```
/* Exemple basé sur le Conseil fédéral */
@prefix : <https://example.com>.
@prefix schema : <http://schema.org/>.
@prefix xsd : <http://www.w3.org/2001/XMLSchema#>.

:KarinKeller-Sutter schema:hasOccupation :Conseil fédéral ;
:parti :PLR ;
:nombrede frères et sœurs 3.
:ViolaAmherd schema:hasOccupation :Conseil fédéral ;
:parti :LeCentre ;
:nom complet "Viola Patricia Amherd".

:NadineMasshardt schema:hasOccupation :Conseiller national ;
:parti :PS ;
:nédans :AffolternAmAlbis ;
schema:birthDate "1984-10-04"^^xsd:date.

:MayaGraf schema:hasOccupation :Staenderat ;
:parti :Verts ;
schema:parent :FritzGraf.

:FritzGraf schema:hasOccupation :BaselbieterLandrat ;
:parti :UDC ;
schema:children :MayaGraf.
```

Le stockage des données RDF : Les triplestores

Après avoir modélisé les données en RDF, ces dernières doivent être stockées pour être accessibles. C'est là qu'interviennent les triplestores, des bases de données spécialement conçues pour stocker et récupérer des triplets RDF.

Les triplestores sont optimisés pour le stockage et la gestion des données RDF. Ils sont capables d'effectuer des requêtes complexes sur ces données, grâce à leur capacité à comprendre la structure de graphe des données RDF et à naviguer efficacement dans cette structure.

L'interrogation des données RDF avec SPARQL

Une fois les données RDF stockées dans un triplestore, il est nécessaire de pouvoir les interroger et les récupérer de manière efficace. C'est ici qu'intervient SPARQL, le langage d'interrogation standard pour les données RDF.

SPARQL permet d'extraire des données d'un triplestore, grâce à sa capacité à comprendre la structure de graphe des données RDF et à naviguer efficacement dans cette structure. SPARQL permet également de faire des requêtes fédérées qui relient plusieurs sources de données, ce qui rend possible l'interrogation de plusieurs triplestores en une seule requête. Il est important de noter que, lors de l'interrogation de plusieurs sources, la gestion de l'autorisation peut devenir un enjeu complexe. Selon la configuration et les politiques de sécurité des sources de données, il peut être nécessaire d'harmoniser les autorisations ou de gérer différentes identités pour chaque source, ce qui exige une attention particulière aux détails de l'implémentation et de la sécurité.

Trouvez ci-dessous un exemple permettant de consulter les données provenant d'un jeu de données disponible sur LINDAS⁵⁰. La liste des offices fédéraux en français résulte de cette requête SPARQL :

```
PREFIX schema: <http://schema.org/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
SELECT * WHERE {
  ?office schema:inDefinedTermSet <https://ld.admin.ch/office>.
  ?office schema:name ?name.
  FILTER(lang(?name) = "fr")
}
LIMIT 20
```

L'exploitation des données liées

L'utilisation de SPARQL pour interroger et mettre à jour des triplestores offre de nombreuses possibilités pour l'exploitation des données liées. Un fournisseur de services peut accéder aux données disponibles, formuler une requête SPARQL pour lire les données, et utiliser SPARQL UPDATE pour insérer, modifier ou supprimer des triplets, consolidant ainsi des informations provenant de différentes bases de données.

Les résultats de lecture peuvent être fournis sous une forme directement utilisable dans leurs applications, ce qui réduit le besoin de maintenir une base de données locale et simplifie la gestion des données. La capacité de mettre à jour les données à travers SPARQL favorise également une utilisation plus innovante et créative des données, car les fournisseurs de services peuvent facilement accéder à des informations provenant de sources diverses et variées, et interagir avec elles de manière plus dynamique.

L'accessibilité des données liées

Les données liées sont conçues pour être facilement navigables. Grâce à leur structure de graphe et à l'utilisation de standards comme RDF et SPARQL, elles peuvent être parcourues comme on naviguerait sur des pages web. Par exemple, des outils tels que le Graph Explorer⁵¹ permettent d'explorer visuellement ces structures de données comme si l'on naviguait entre les liens d'un site web.

⁵⁰ <https://ld.admin.ch/sparql/>

⁵¹ <https://test.lindas.admin.ch/graph-explorer/?lang=fr>

Cela rend les données liées accessibles à un public plus large, allant des spécialistes des données aux utilisateurs non experts. En effet, un utilisateur peut suivre les liens entre les données pour explorer le graphe de connaissances, même s'il ne possède pas une connaissance technique approfondie de la structure sous-jacente des données.

Cette accessibilité des données liées renforce leur potentiel pour améliorer non seulement l'interopérabilité des données, mais aussi leur accessibilité et leur utilisabilité pour une gamme plus large d'utilisateurs. Elle souligne l'importance des initiatives de données ouvertes et de données liées pour favoriser l'innovation, améliorer la prise de décision, et promouvoir la transparence.