



Methode zur Identifizierung und Validierung von Anwendungsfällen für verknüpfte Daten (Linked Data)

Digitale Verwaltung vernetzt denken und gestalten

Klassifizierung	nicht klassifiziert
Status	in Bearbeitung
Name des Programms	Strategie Digitale Bundesverwaltung Schwerpunkt 1: Digitale Verwaltung vernetzt denken und gestalten
Projektverantwortliche	Karen Dijkstra (BK)
Version	1.0
Datum	21.10.2024
Auftraggeber	Digitale Transformation und IKT-Lenkung (DTI)
Verfasserin	Karen Dijkstra
Fachgruppen	Architekturboard Bund (ABB)
Genehmigt von	Andreas Spichiger (BK)



Table des matières

1	INTRODUCTION	5
1.1	CONTEXTE ET OBJECTIFS	5
1.2	PORTÉE ET LIMITATIONS	5
2	PHASE 1 : ANALYSE DES EXIGENCES MÉTIER ET DES OBJECTIFS STRATÉGIQUES	6
2.1	TRAVAUX PRÉPARATOIRES	6
2.2	IDENTIFICATION DES BESOINS MÉTIER PRIORITAIRES	6
3	VALIDATION DES BESOINS AVEC LES PARTIES PRENANTES	7
4	PHASE 2 : IDENTIFICATION DES TECHNOLOGIES ET COMPÉTENCES REQUISES	8
4.1	SERVICES ET TECHNOLOGIES PROPRES AUX DONNÉES LIÉES	8
4.2	COMPÉTENCES REQUISES POUR LA GESTION DES DONNÉES LIÉES	11
5	PHASE 3 : SÉLECTION DES CAS D'USAGE	14
5.1	CONSOLIDATION DES DONNÉES RECUEILLIES	14
5.2	VALIDATION DES CAS D'USAGES	18
6	PHASE 4 : TESTS ET VALIDATION	19
6.1	DÉFINITION DES INDICATEURS DE PERFORMANCE (KPI)	19
6.2	EVALUATION STATISTIQUE	19
6.3	COMPARAISON AVEC LES OBJECTIFS INITIAUX	21
6.4	RECOMMANDATIONS ET AJUSTEMENTS	22
7	PHASE 5 : RÉFÉRENCEMENT	23

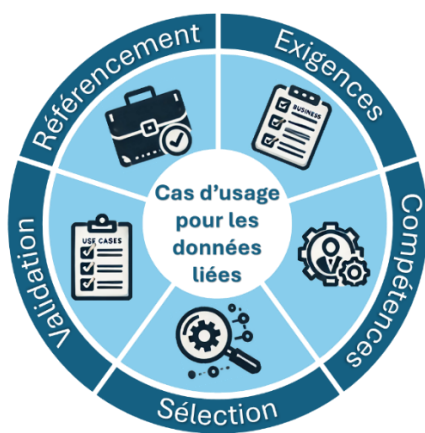


Management Summary

Diese Anleitung knüpft an den Bericht «Concept des données liées au sein de l'Administration fédérale» (Konzept der verknüpften Daten in der Bundesverwaltung) an¹, der das Potenzial der verknüpften Daten für die verbesserte Effizienz der Verwaltung unterstreicht. Der Erstbericht dokumentiert die bedeutenden Vorteile der Interoperabilität der öffentlichen Dienste; sie vereinfacht die departementsübergreifende Zusammenarbeit und die Prozessharmonisierung.

Diese Anleitung soll die strategische Vision der Interoperabilität in die Praxis übertragen. Sie gibt den Unternehmensarchitektinnen und Projektmanagern, die das Konzept der verknüpften Daten anwenden, den praktischen Rahmen vor. Die Anleitung beinhaltet konzeptuelle Empfehlungen und liefert kapazitäten- und ressourcengerechte, konkrete Instrumente, sodass der Nutzen für die Verwaltung möglichst gross wird.

Der in dieser Anleitung verfolgte Ansatz ist konsequent auf die Operationalisierung ausgerichtet und verfolgt das klare Ziel, messbare Ergebnisse zu erzielen. Das Dokument ist in fünf aufeinanderfolgende Phasen gegliedert, in denen jeweils spezifische methodische Instrumente eingesetzt werden, welche die praktische Umsetzung erleichtern:



Das nebenstehende Schema stellt die fünf Schlüsselphasen zur Identifizierung und Validierung von Anwendungsfällen für verknüpfte Daten dar:

1. Anforderungen: Bedürfnisse der Interessengruppen evaluieren.
2. Kompetenzen: Notwendige Fähigkeiten evaluieren.
3. Auswahl: Relevante Anwendungsfälle auswählen.
4. Validierung: Durchführbarkeit und Wirksamkeit prüfen.
5. Katalogisierung: Validierte Fälle für die Wiederverwendung katalogisieren.

Der Fünf-Phasen-Prozess gewährleistet die effiziente Einbindung verknüpfter Daten in strategische Projekte.

1. Analyse der Geschäftsanforderungen und der strategischen Ziele: In der ersten Phase sollen die Geschäftsanforderungen der verschiedenen Verwaltungseinheiten identifiziert und priorisiert werden – im Einklang mit den langfristigen strategischen Zielen der Bundesverwaltung. Es soll sichergestellt werden, dass jedes Transformationsprojekt auf den Zielen der jeweiligen Verwaltungseinheit aufbaut.
2. Identifizierung der notwendigen Technologien und Kompetenzen: In dieser Phase werden die Technologien ermittelt, die den identifizierten Geschäftsanforderungen am besten Rechnung tragen. Gleichzeitig werden die für die effiziente Umsetzung erforderlichen technischen Kompetenzen evaluiert.
3. Auswahl der relevanten Anwendungsfälle: Sobald die Bedürfnisse und die Technologien identifiziert sind, werden die Anwendungsfälle nach ihrer strategischen Wirksamkeit und ihrer technischen Machbarkeit priorisiert. Dabei werden Instrumente wie Entscheidungshilfen für die Auswahl genutzt.
4. Prüfung und Validierung der Anwendungsfälle: Die Wirksamkeit der ausgewählten Anwendungsfälle kann mittels Leistungskennzahlen (Key Performance Indicators; KPI) gemessen werden. Diese Phase umfasst strenge Prüf- und Validierungsprozesse, sodass sichergestellt werden kann, dass die Ergebnisse mit den Anfangszielen übereinstimmen.
5. Katalogisierung und Dokumentation: Schliesslich werden die validierten Anwendungsfälle zusammen mit Empfehlungen für ihre Wiederverwendung in künftigen Projekten dokumentiert und katalogisiert. So können Erfolge und erprobte Lösungen innerhalb der Bundesverwaltung geteilt werden.

¹ <https://www.bk.admin.ch/bk/fr/home/digitale-transformation-ikt-lenkung/bundesarchitektur/daten-und-informationsmanagement/linked-data.html>



Die Methode zur Identifizierung und Validierung von Anwendungsfällen für verknüpfte Daten ist flexibel und ausbaufähig; der pragmatische Ansatz fördert geprüfte und bewährte Lösungen. Im Gegensatz zu den früheren theoretischen Arbeiten liegt der Schwerpunkt in diesem Dokument auf der konkreten Umsetzung und der Optimierung von digitalen Projekten in der Bundesverwaltung.

Zusammenfassend liefert das Dokument eine strukturierte Methode und praktische Instrumente, um verknüpfte Daten als Schlüsselement für die digitale Transformation der Bundesverwaltung zu nutzen.



1 Introduction

1.1 Contexte et objectifs

L'administration fédérale s'engage dans une démarche stratégique visant à optimiser l'utilisation des données au sein de ses différentes entités. L'exploitation des données liées se présente comme une opportunité stratégique pour améliorer l'interopérabilité des services publics.

Ce travail s'ancre dans la volonté d'adopter une architecture d'information interconnectée, répondant mieux aux besoins des citoyens et facilitant la collaboration entre les départements. Il s'appuie sur une analyse intitulée « Concept des données liées au sein de l'Administration fédérale »², qui a révélé les bénéfices significatifs des données liées pour l'administration, notamment en améliorant la prise de décision et en optimisant les processus.

Objectifs spécifiques :

- Définir une approche structurée pour l'identification des cas d'usage, alignant les besoins métier avec les opportunités et les défis posés par les technologies des données liées. Cette approche vise à distinguer clairement entre les contextes où les données liées offrent une plus-value et ceux où leur application pourrait s'avérer moins pertinente.
- Fournir des critères objectifs et un processus d'évaluation pour classer les cas d'usage en fonction de leur impact stratégique et de leur apport potentiel aux utilisateurs finaux. L'accent est mis sur l'identification de critères qui permettent de reconnaître non seulement les opportunités mais aussi les limitations inhérentes à l'utilisation des données liées.
- Elaborer et tester une méthodologie de validation des cas d'usage dans des contextes réels pour confirmer leur pertinence. Cette méthodologie comprend des étapes pour tester les hypothèses et ajuster les critères d'évaluation en fonction des retours d'expérience des différents offices fédéraux.

Ce manuel, fondé sur les principes d'architecture d'entreprise, vise à assurer que la sélection et l'évaluation des cas d'usage sont systématiquement alignées avec les stratégies de l'Administration. Il est conçu pour les architectes d'entreprise, les chefs de projet, et les responsables stratégiques, mais également pour les développeurs et les spécialistes des données qui souhaitent adopter les meilleures outils et techniques en fonction de l'état de leur compétence en matière de données liées.

1.2 Portée et limitations

Ce manuel vise à établir un moyen de déterminer où et comment les données liées peuvent être appliquées pour maximiser les bénéfices. Il s'efforce de fournir un cadre méthodologique pour naviguer au travers des données liées en tenant compte des ressources disponibles ainsi que de leurs compétences en la matière.

1.2.1 Portée

- Le manuel couvre la phase d'identification et d'évaluation préalable des cas d'usage des données liées.
- Il est destiné à un public varié au sein de l'Administration fédérale, incluant les architectes d'entreprise, les chefs de projet, les décideurs, les développeurs, les spécialistes de données.
- Il propose des critères d'évaluation pour aider à décider de la pertinence des données liées pour des cas d'usage spécifiques.

1.2.2 Limitations

- Les recommandations fournies sont conçues pour être adaptées et ajustées en fonction des spécificités de chaque unité administrative ou projet au sein de l'Administration. Une évaluation individuelle permet de déterminer l'applicabilité des données liées à chaque cas.
- Les considérations légales et éthiques autour de l'utilisation des données, notamment en ce qui concerne la protection de la vie privée et la sécurité des données ne sont pas l'objet de ce manuel.

² <https://www.bk.admin.ch/dam/bk/fr/dokumente/dti/themen/linked-data/concept-linked-data-bund.pdf.download.pdf/Concept%20des%20donn%C3%A9es%20li%C3%A9es%20au%20sein%20de%20l%E2%80%99Administration%20f%C3%A9d%C3%A9rale%20.pdf>



Les utilisateurs sont encouragés à consulter les directives légales et éthiques pertinentes en vigueur au sein de l'Administration fédérale.

2 Phase 1 : Analyse des exigences métier et des objectifs stratégiques

La première étape dans l'identification des cas d'usage consiste à effectuer un recueil des besoins métier. Cette démarche s'inscrit dans une logique de diagnostic stratégique visant à aligner les capacités technologiques avec les exigences et objectifs métier des parties prenantes. Cette approche permet de garantir que les cas d'usage identifiés dans les étapes suivantes sont ancrés dans la réalité des besoins métier et alignés avec les orientations stratégiques de l'organisation.

2.1 Travaux préparatoires

2.1.1 Revue des objectifs stratégiques

Cette étape consiste à analyser les documents stratégiques tels que les stratégies, tactiques ou plans d'actions pour s'assurer de l'alignement avec les objectifs à long terme de l'administration fédérale. L'objectif est d'identifier les priorités stratégiques, les axes de développement et les synergies entre les différentes initiatives. Cette revue permet notamment de repérer les éventuels écarts ou redondances entre les différents plans et de prioriser les actions à mener pour maximiser l'efficacité organisationnelle et l'impact des initiatives liées aux données.

2.1.2 Cartographie des acteurs

La première étape consiste à établir une cartographie des acteurs. Cette cartographie permet d'identifier les parties prenantes essentielles, les responsabilités, ainsi que les flux d'informations entre les différents acteurs. Idéalement, une analyse des processus métier doit être réalisée. Cette analyse vise à comprendre les points où les données liées peuvent apporter une réelle valeur ajoutée, en facilitant l'automatisation, en simplifiant les flux de travail, ou en améliorant la prise de décision.

Il est important de cartographier à la fois les processus actuels et ceux qui pourraient bénéficier de la transformation numérique, en mettant l'accent sur les interactions entre les services et sur les flux de données entre les systèmes. Cette étape sert de fondation pour définir les besoins spécifiques qui seront adressés dans les étapes suivantes.

2.1.3 Entretiens avec les parties prenantes

Dans cette phase, des entretiens doivent être menés avec les parties prenantes clés pour comprendre les préoccupations de chaque acteur et identifier les opportunités d'amélioration. Les informations recueillies doivent ensuite être synthétisées sous forme de diagrammes de motivation qui illustrent les exigences stratégiques. Le but est de transformer ces perspectives en exigences concrètes et alignées avec les objectifs à long terme de l'organisation.

2.2 Identification des besoins métier prioritaires

Notre ambition est de concevoir une infrastructure où les exigences sont intégrées de manière holistique au sein de notre système d'architecture d'entreprise. Cette infrastructure visera à capturer de manière exhaustive les attentes des parties prenantes, les transformant en une série de spécifications qui guideront la mise en œuvre de solutions alignées sur les objectifs stratégiques de l'Administration.

Composants clés	Description
Cadre de gestion des exigences	Un cadre unifié qui permet une gestion systématique des exigences, depuis leur capture initiale jusqu'à leur réalisation, assurant que chaque spécification est clairement définie et correctement priorisée.
Processus de gouvernance	Des mécanismes de gouvernance qui garantissent que les exigences sont évaluées et approuvées par les bonnes autorités, et que les changements sont gérés de manière contrôlée.



Composants clés	Description
Interconnectivité avec les outils d'architecture d'entreprise	Une interconnexion avec des outils d'architecture d'entreprise existants et futurs, fournissant une vue d'ensemble des exigences, des architectures, des processus et des technologies impliquées.
Support à la décision basé sur des données	Des capacités d'analyse de données permettant aux architectes et analystes métier de formuler des stratégies basées sur des informations actualisées.

Spécifications	Description des fonctionnalités de l'outil
Traçabilité	Il doit permettre une traçabilité des exigences, de leur origine à leur concrétisation, en passant par toutes les étapes intermédiaires.
Flexibilité	Il doit s'adapter aux changements des besoins métier et être capable d'intégrer de nouvelles exigences en cours de projet.
Collaboration	Il doit faciliter une collaboration interdépartementale, soutenant une synergie entre les unités administratives.
Sécurité et conformité	La solution doit respecter les normes de sécurité et s'assurer que la gestion des exigences est conforme aux réglementations en vigueur.
Intégration	Il doit se connecter avec d'autres systèmes existants pour former un écosystème cohérent.

Nous imaginons un système où la gestion des exigences est une partie intégrante de la stratégie. Un outil qui non seulement répond aux besoins actuels, mais qui est également évolutif pour s'adapter aux exigences futures. Cela nécessite une plateforme qui non seulement recueille les exigences, mais qui les aligne avec les objectifs à long terme de l'Administration.

Chaque fois que le symbole du couteau suisse apparaît dans ce document, il renvoie à un outil simple mis à disposition pour faciliter l'identification ou l'évaluation des cas d'usage. Ils sont décrits en détail dans le Guide d'utilisation des outils pour l'identification et l'évaluation des cas d'usage des données liées, disponible sur Internet³.



Catalogue des exigences

3 Validation des besoins avec les parties prenantes

Présentez une synthèse des besoins métier identifiés aux parties prenantes pour validation et ainsi vous assurez que la compréhension des besoins est partagée et acceptée avant de procéder à l'identification des cas d'usage potentiels. Il est également essentiel pour l'ensemble des acteurs de comprendre l'impact sur les stratégies de la Confédération.

Pour simplifier, chaque besoin métier exprimé doit pouvoir se justifier sur la base de la mise en œuvre des différentes stratégies pertinente pour la politique publique à mettre en place ou développer dans le cadre des activités du ou des offices impliqués.

Notre ambition est de concevoir une infrastructure technologique qui intègre les besoins identifiés et les aligne avec les stratégies de la Confédération. Cet outil vise à capturer, de manière exhaustive et structurée, les attentes des parties prenantes et les transformer en un ensemble de spécifications précises.

³ <https://www.bk.admin.ch/bk/fr/home/digitale-transformation-ikt-lenkung/bundesarchitektur/daten-und-informationsmanagement/linked-data.html>



Composants clés	Description
Plateforme de collaboration pour la validation des besoins	Un système centralisé pour collecter, organiser et présenter les besoins métier identifiés aux parties prenantes, facilitant leur validation et leur approbation.
Outils d'alignement Stratégique	Des fonctionnalités qui permettent de relier chaque besoin métier aux objectifs stratégiques de l'organisation, montrant clairement comment ils soutiennent la vision globale.
Système de feedback et de révision	Un mécanisme recueillant les retours des parties prenantes et permettant des ajustements des besoins métier proposés.
Tableau de bord de suivi des validations	Tableaux de bord fournissant une vue d'ensemble des statuts de validation, facilitant le suivi des modifications requises.

Spécifications	Description des fonctionnalités de l'outil
Traçabilité des besoins	Capacité à suivre chaque besoin depuis sa proposition initiale jusqu'à sa validation finale, incluant toutes les interactions et modifications.
Adaptabilité aux stratégies de la Confédération	L'outil doit permettre une mise à jour des critères d'alignement stratégique pour refléter les changements dans les orientations politiques ou les priorités de la Confédération.
Facilitation de la collaboration	Support d'une collaboration interdépartementale avec des fonctionnalités de partage, de discussion et de notification.
Conformité et sécurité	Assurer que la plateforme respecte les normes de sécurité et est conforme aux réglementations en vigueur concernant la gestion et le partage des informations.
Intégration avec l'outil d'architecture d'entreprise	L'outil doit s'intégrer avec les systèmes d'architecture d'entreprise existants, permettant un échange de données et une vue cohérente des informations stratégiques.
Analyse et support à la décision	Fonctionnalités d'analyse pour évaluer l'impact des besoins métier sur les stratégies globales.



Validation des exigences

4 Phase 2 : Identification des technologies et compétences requises

Après le recueil des besoins métier, il convient d'explorer les possibilités technologiques liées à l'utilisation des données liées. Cette phase permet d'assurer l'adéquation entre les besoins métier identifiés et les solutions technologiques disponibles ou développables par les équipes à disposition.

4.1 Services et technologies propres aux données liées

4.1.1 LINDAS

Le service de données liées LINDAS⁴ (Linked Data Service) de l'Administration fédérale offre un cadre pour la gestion et la publication de données liées. Conçu pour améliorer l'interopérabilité et la réutilisation des données, LINDAS s'appuie sur les principes et les technologies ci-dessous.

⁴ <https://lindas.admin.ch/>



LINDAS repose sur plusieurs modèles de données spécifiques, chacun adapté à des cas d'utilisation définis. Parmi les principaux modèles utilisés, on trouve :

- DCAT (Data Catalog Vocabulary) : Vocabulaire RDF permettant de structurer et standardiser les métadonnées des catalogues de données, facilitant leur interopérabilité et leur accès via des plateformes comme Opendata.swiss.
- SKOS (Simple Knowledge Organization System) : Ontologie RDF utilisée pour organiser des vocabulaires contrôlés en thésaurus et taxonomies, garantissant une cohérence sémantique et facilitant la recherche et l'interopérabilité des données.
- cube.link (voir chapitre 4.1.2) : Utilisé pour transformer les données brutes en données liées, cube.link est adapté pour l'intégration des données structurées en tables. Cette ontologie est supportée par l'outil Cube Creator, facilitant la conversion et la publication des données brutes sous forme de données RDF (Resource Description Framework).
- version.link : Ce modèle est essentiel pour la gestion des données versionnées, où l'évolution temporelle des ressources est critique. Il permet de mapper les changements des ressources dans le temps, assurant la traçabilité et l'historisation des données.

L'un des fondements de LINDAS est l'utilisation rigoureuse de modèles d'URI⁵ (Uniform Resource Identifier) et d'espaces de noms (namespaces⁶) pour assurer une identification unique et cohérente des ressources. Les modèles d'URI définissent des schémas standardisés pour générer des URI cohérentes dans le but de faciliter l'interopérabilité entre différents systèmes et datasets. Par exemple, les espaces de noms comme ceux définis sur schema.ld.admin.ch jouent un rôle clé dans la normalisation et la structuration des données.

LINDAS utilise un ensemble de concepts⁷ et d'entités de base⁸ définis pour structurer les données de manière sémantique. Ces concepts sont basés sur des vocabulaires standardisés tels que SKOS qui permet d'organiser les concepts en thésaurus, facilitant ainsi la recherche et la navigation au sein des données. Les entités de base incluent des éléments essentiels comme les organisations, les lieux, les événements, et les personnes, chacun défini de manière à maximiser la réutilisation et l'interopérabilité des données.

Le langage de requête SPARQL⁹ (SPARQL Protocol and RDF Query Language) permet d'interroger les bases de données¹⁰ RDF spécialement conçues pour stocker et gérer des triples RDF (triplestores). Les utilisateurs peuvent formuler des requêtes complexes pour extraire des données liées de manière efficace, ce qui est essentiel pour les applications analytiques avancées et la génération de rapports dynamiques. Les triplestores assurent la séparation des données et des applications, permettant ainsi une gestion plus flexible et évolutive des données.

La gouvernance des données dans LINDAS est assurée par un ensemble de directives et de bonnes pratiques. La sécurité des données est assurée grâce à des mécanismes en place pour garantir la confidentialité et l'intégrité des données. Les contrôles d'accès stricts, le chiffrement et l'audit régulier des accès sont des mesures couramment mises en œuvre.

L'écosystème LINDAS comprend plusieurs outils et services destinés à faciliter la publication et l'utilisation des données liées. Cela inclut des plateformes de publication, des interfaces de requêtes SPARQL, et des applications spécifiques qui démontrent l'usage pratique des données liées dans divers contextes administratifs.

En somme, LINDAS offre un cadre technologique de gestion des données liées, reposant sur des standards et des pratiques largement connus. Son adoption et sa mise en œuvre au sein de l'Administration fédérale permettent d'améliorer la collaboration inter-organisationnelle.

⁵ <https://lindas.admin.ch/governance/URI-templates/>

⁶ <https://lindas.admin.ch/governance/namespaces/>

⁷ <https://lindas.admin.ch/governance/concepts/>

⁸ <https://lindas.admin.ch/governance/core-entities/>

⁹ <https://lindas.admin.ch/sparql/>

¹⁰ <https://lindas.admin.ch/datasets/>



4.1.2 Cube.link

Le projet cube.link permet de gérer et de transformer des données brutes en données liées grâce à des technologies comme RDF (Resource Description Framework). Cube.link est une spécification et un cadre pour la publication de données sous forme de cubes de données RDF. Les cubes de données sont particulièrement utiles pour représenter des données multidimensionnelles.

Principales caractéristiques de cube.link:

- Cube.link utilise le vocabulaire RDF Data Cube Vocabulary (QB) pour modéliser les données. Ce vocabulaire permet de définir des cubes de données composés de dimensions, de mesures et d'attributs, chacun représenté par des URI uniques.
- Grâce à l'utilisation de RDF et des vocabulaires standardisés, les cubes de données peuvent être facilement intégrés avec d'autres sources de données RDF.
- Le modèle cube.link permet de définir des relations complexes entre les données, notamment via des propriétés de relation¹¹ spécifiques.

Les métadonnées¹² fournissent des informations contextuelles pour comprendre et utiliser correctement les données. Les points clés sont :

- Les métadonnées incluent des descriptions des dimensions, mesures et attributs, facilitant leur compréhension et réutilisation.
- Les informations sur la provenance des données et leur qualité permettent une évaluation de leur fiabilité et de leur pertinence.

Ces relations permettent de créer des liens explicites entre différents cubes de données ou entre les éléments au sein d'un même cube. Cela enrichit les capacités analytiques et permet des explorations de données plus sophistiquées.

4.1.3 Cube Creator

Cube Creator est un outil développé par l'entreprise Zazuko¹³, conçu pour automatiser la transformation des données brutes en cubes de données RDF. Disponible sur GitHub¹⁴. Ses fonctionnalités premières sont les suivantes :

- Cube Creator permet de convertir des données CSV et d'autres formats tabulaires en cubes de données RDF avec une intervention humaine minimale. Le processus de transformation est guidé par des configurations définies dans des fichiers YAML¹⁵.
- Le fichier releases.yaml montre comment Cube Creator peut être intégré dans des pipelines d'intégration continue et de déploiement continu (CI/CD), assurant ainsi que les transformations de données sont reproductibles et automatisées.
- Cube Creator supporte diverses options de configuration, permettant de définir des mappings spécifiques pour les colonnes de données, les URI, et les vocabulaires utilisés.

Ci-dessous, un aperçu de cas d'utilisation¹⁶ réalisé avec cube.link et Cube Creator pour transformer et publier des données statistiques complexes.

- Cube Creator permet de transformer des données statistiques en datasets multidimensionnels.
- Les datasets publiés peuvent être interconnectés avec d'autres sources de données RDF.

Le projet cube.link et Cube Creator s'appuient sur plusieurs vocabulaires standardisés pour assurer l'interopérabilité et la compatibilité des données. Parmi les vocabulaires couramment utilisés, on trouve ceux listés sur Zazuko Vocabularies¹⁷.

- SKOS -> gestion des vocabulaires contrôlés et des thésaurus.

¹¹ <https://cube.link/relation/>

¹² <https://cube.link/meta/>

¹³ <https://zazuko.com/>

¹⁴ <https://github.com/zazuko/cube-creator>

¹⁵ <https://github.com/zazuko/cube-creator/blob/master/.github/workflows/releases.yaml>

¹⁶ <https://zazuko.com/blog/data-cube-creating-datasets/>

¹⁷ <https://zazuko.com/category/Vocabularies.html>



- DCAT -> description des catalogues de données.

4.1.4 Visualize

Visualize¹⁸ est une plateforme qui permet de visualiser des données unidimensionnelles ou multidimensionnelles en temps réel. Elle est conçue pour offrir une visualisation interactive des données, permettant aux utilisateurs de créer des graphiques et des tableaux de manière intuitive. Cette application est centralisée autour de la gestion et la visualisation des Linked Open Government Data (LOGD). Ses principales caractéristiques sont les suivantes :

- Toute mise à jour des données se reflète immédiatement dans les visualisations.
- La plateforme permet de manipuler des données provenant de diverses sources.
- Les utilisateurs peuvent configurer des visualisations à l'aide de différents types de graphiques et filtres personnalisés.

L'architecture de Visualize repose sur une ontologie ou un schéma de type « Cube »¹⁹. Les points clé de cette architecture sont d'une part l'utilisation d'une ontologie Cube pour structurer les données en cubes multidimensionnels et d'autre part l'utilisation de vocabulaires contrôlés garantissant l'uniformité et la précision des termes employés.

Visualize utilise RDF (Resource Description Framework) pour modéliser les données et SPARQL pour les interroger, ce qui permet une flexibilité dans la gestion des données liées. L'usage de SKOS²⁰ (Simple Knowledge Organization System) pour la gestion des vocabulaires permet de standardiser les termes et d'assurer une cohérence sémantique des données visualisées. Parmi les outils de développement, comme nous l'avons vu précédemment, Cube Creator est utilisé pour transformer les données brutes en cubes de données RDF, facilitant ainsi l'intégration avec Visualize. Les codes sources et la gestion de projet pour Visualize sont disponibles sur GitHub²¹, et la coordination du projet est assurée via GitLab.



- ① Les données sont importées à partir de diverses sources de données ouvertes.
- ② Les données sont transformées en cubes RDF à l'aide de Cube Creator.
- ③ Les utilisateurs peuvent configurer des visualisations personnalisées en choisissant parmi différents types de graphiques et en appliquant des filtres.
- ④ Les visualisations peuvent être partagées via des liens ou intégrées dans des sites web.

4.2 Compétences requises pour la gestion des données liées

Pour réaliser la chaîne complète de transformation des données brute en données liées, il est nécessaire de posséder un ensemble de compétences techniques spécifiques.



Arbre décisionnel

4.2.1 Ontologies, schéma et structure de données

L'utilisation d'ontologies pertinentes structure les données de manière cohérente et sémantiquement riche. Les équipes doivent être en mesure d'identifier les situations où une ontologie comme SKOS est nécessaire pour structurer des vocabulaires et organiser des données dans des cubes multidimensionnels via des outils tels que Cube Creator ou Cube.link. L'application des ontologies garantit que les données peuvent être reliées à d'autres sources dans un cadre standardisé.

¹⁸ <https://visualize.admin.ch/fr?dataSource=Prod>

¹⁹ <https://visualize.admin.ch/fr/browse?dataSource=Prod>

²⁰ <https://www.w3.org/TR/skos-primer/>

²¹ <https://github.com/visualize-admin>



Par exemple, lors de la modélisation de données géographiques ou temporelles, les cubes de données multidimensionnels permettent d'associer des mesures précises à des dimensions hiérarchiques. Dans ces cas, il est essentiel d'utiliser des « Shared Dimensions » pour éviter les doublons de données et améliorer la gestion des informations réutilisées dans plusieurs cubes. Cela permet une gestion des relations sémantiques tout en assurant la standardisation et la pérennité des données. L'intégration de ces « Shared Dimensions », soutient une meilleure agrégation et visualisation des données à grande échelle.

4.2.2 Identification de la source de données

Les équipes doivent être compétentes dans l'extraction et la manipulation de données provenant de diverses sources, qu'il s'agisse de fichiers PDF, Excel, JSON, ou encore de bases de données relationnelles. Utiliser des outils comme Tabula ou PDFMiner permet de convertir des fichiers non structurés, tels que des rapports PDF, en formats exploitables comme CSV, qui pourront ensuite être transformés en RDF via Cube Creator.

Lorsque les données sont déjà structurées sous forme hiérarchique ou relationnelle, comme avec des fichiers JSON ou SQL, une compréhension des formats de données permet d'assurer leur intégration dans des triplestores et leur conversion en cubes RDF. La maîtrise de la manipulation de données brutes et structurées permet ainsi de simplifier l'extraction, la transformation, et l'utilisation de ces données dans les systèmes de données liées. Enfin, les équipes doivent être capables de sélectionner et structurer les données avant transformation en RDF, garantissant ainsi leur cohérence sémantique dans les cubes et triplestores.

4.2.3 Vocabulaires

Utiliser des vocabulaires contrôlés et des ontologies standardisées, telles que SKOS, permet de maintenir une cohérence sémantique tout au long du projet. Les équipes doivent être capables de créer, gérer et appliquer ces vocabulaires de manière à assurer l'interopérabilité des données, tant à l'intérieur qu'à l'extérieur des unités administratives.

Dans un contexte où les mêmes données ou termes sont utilisés dans plusieurs cubes, il est souvent plus efficace de créer des « Shared Dimensions » plutôt que de dupliquer les informations ou de répéter les traductions dans chaque cube. Les « Shared Dimensions » permettent de centraliser certaines informations communes à plusieurs cubes. Cela garantit la réutilisation et l'amélioration de l'interopérabilité entre différents ensembles de données.

4.2.4 Transformation en RDF

La transformation des données en RDF (Resource Description Framework) permet de structurer et publier des données liées de manière interopérable. Ce processus repose sur une bonne compréhension des principes fondamentaux du RDF, notamment la structure en triplets sujet-prédicat-objet, qui permet de modéliser des relations sémantiques entre différentes entités de données.

Les équipes doivent maîtriser des outils comme Cube Creator pour transformer les données brutes en triplets RDF. Ce processus implique de structurer les fichiers CSV dans un format long, ce qui facilite la gestion des dimensions hiérarchiques et permet une meilleure manipulation des données lors de leur conversion en cubes RDF. Cube Creator est particulièrement efficace pour organiser les données multidimensionnelles et les publier ultérieurement sur des plateformes comme Visualize.

Par ailleurs, la maîtrise des ontologies telles que SKOS assure la cohérence sémantique des données transformées. L'utilisation d'ontologies permet de définir les relations entre les termes et concepts utilisés dans les jeux de données. Cela assure l'interopérabilité avec d'autres sources de données liées. Les équipes doivent être capables de gérer des vocabulaires contrôlés, en créant, modifiant et maintenant des ensembles de termes standardisés.

4.2.5 Programmation

La programmation joue un rôle central dans l'automatisation du traitement et de la transformation des données. Les spécialistes doivent être capables de rédiger des scripts en Python, R ou autres langages, pour nettoyer, normaliser et convertir les données brutes en formats RDF compatibles avec les triplestores. Des bibliothèques comme Pandas pour Python permettent de manipuler les DataFrames et gérer des fichiers CSV.



Par ailleurs, l'utilisation d'outils ETL (Extract, Transform, Load) permet d'automatiser les flux de travail, afin de garantir une transformation cohérente des données.

La création de pipelines automatisés de traitement des données permet d'assurer un flux constant d'informations entre les différentes étapes du projet. Ces pipelines transforment les données brutes en RDF prêtes à être publiées. Le but visé est l'efficacité et la réduction des risques d'erreurs manuelles.

Outre la manipulation des données, les spécialistes doivent maîtriser l'écriture de requêtes SPARQL pour interroger les triplestores. Cela implique la capacité de concevoir des requêtes complexes pour extraire des données spécifiques et gérer des jointures entre plusieurs sources RDF. L'utilisation de filtres et d'agrégations permet d'affiner les résultats des requêtes, tandis que des fonctions de navigation dans les graphes de données améliorent la précision des analyses.

Pour garantir des performances optimales, les équipes doivent également savoir optimiser l'exécution des requêtes en réduisant le nombre de triplets examinés. La connaissance des techniques d'indexation et de partitionnement des données au sein des triplestores aide à améliorer l'efficacité des requêtes.

4.2.6 Publication des Linked Data

La publication des données liées sur des plateformes comme LINDAS ou Cube.link assure l'accessibilité et l'interopérabilité des informations. Les équipes doivent être capables d'utiliser LINDAS pour publier et gérer les données liées, en maîtrisant les mécanismes de chargement des données RDF et l'utilisation des APIs pour automatiser le processus de publication. La gestion des accès et des permissions sur la plateforme assure la sécurité et la traçabilité des données publiées.

En outre, la maîtrise de Cube.link est nécessaire pour gérer efficacement les métadonnées et les relations entre les différentes sources de données. Les équipes doivent comprendre les principes de modélisation des données RDF et savoir utiliser des vocabulaires standardisés, tels que DCAT²² (Data Catalog Vocabulary), afin d'assurer l'interopérabilité entre les jeux de données.

Afin de garantir la pérennité et l'interopérabilité des données publiées, les spécialistes doivent être capables de créer des URI persistants pour les entités de données. Cela implique de configurer des templates URI selon des schémas de nommage cohérents, en suivant les recommandations du W3C²³. L'intégration des données sur des plateformes publiques comme opendata.swiss permet de rendre ces informations accessibles à une large communauté d'utilisateurs.

Une attention particulière doit être portée à la documentation des jeux de données. La bonne documentation facilite la réutilisation des données. Enfin, il est important de référencer les métadonnées sur des plateformes d'interopérabilité comme I14Y²⁴ pour garantir l'accès centralisé et l'intégration des jeux de données avec d'autres systèmes.

4.2.7 Visualisation des données

Les équipes doivent être capables de configurer des visualisations personnalisées en sélectionnant parmi divers types de graphiques et en appliquant des filtres adaptés pour affiner l'affichage des données. Cette compétence implique une compréhension des options de configuration disponibles dans des outils tels que Visualize, permettant aux utilisateurs d'explorer les données de manière interactive.

En plus de la personnalisation, les spécialistes doivent avoir connaissance des mécanismes d'importation et de liaison des données au sein des outils de visualisation. Cela inclut la gestion des métadonnées et l'utilisation des URI persistants pour garantir la traçabilité des informations.

L'objectif est de créer des visualisations dynamiques basées sur des données RDF, qui permettent de naviguer facilement à travers les différentes dimensions des données publiées. Ce type de visualisation interactive offre une meilleure compréhension des relations entre les données et aide à la prise de décision. La capacité à adapter ces visualisations aux besoins spécifiques des utilisateurs finaux assure une qualité d'analyse et d'utilisation des données.

²² <https://www.w3.org/TR/vocab-dcat-3/>

²³ <https://www.w3.org/policies/uri-persistence/>

²⁴ <https://www.i14y.admin.ch/fr/home>



5 Phase 3 : Sélection des cas d'usage

La troisième étape consiste en une identification préliminaire des cas d'usage potentiels. Cette phase soutient la conversion des besoins métier et des capacités technologiques en applications concrètes. Une fois ce processus terminé, les unités administratives en charge de l'implémentation se réfèrent à leur processus internes et procèdent à leur catégorisation.

5.1 Consolidation des données recueillies

5.1.1 Objectifs et importance

Dans le cadre de l'identification des cas d'usage, la consolidation des données permet de développer une vue d'ensemble cohérente, assurant que toutes les parties prenantes partagent une compréhension commune des objectifs. Elle garantit également que les initiatives technologiques sont alignées avec les objectifs stratégiques de l'organisation, priorisant ainsi les projets ayant le plus grand impact. En optimisant l'allocation des ressources et en concentrant les efforts sur les projets les plus prometteurs, la consolidation des données minimise les redondances et inefficacités. De plus, elle fournit une base factuelle d'aide à la décision en centralisant les informations pertinentes. Enfin, cette étape permet d'identifier les opportunités et les risques.

Voici pourquoi cette consolidation est fondamentale :

- En regroupant les informations recueillies lors des étapes précédentes, les décideurs peuvent s'assurer que les initiatives technologiques sont en adéquation avec les objectifs stratégiques globaux de l'organisation. Cette approche évite les silos d'information et favorise une meilleure coordination entre les différentes unités administratives.
- En synthétisant les données, il devient possible d'identifier et de prioriser les besoins les plus critiques. Cette hiérarchisation permet d'allouer les ressources de manière à concentrer les efforts sur les initiatives qui auront le plus grand impact. La consolidation aide d'une part à identifier les redondances et les inefficacités potentielles et d'autre part à rationaliser les processus.
- La consolidation des données fournit une base factuelle pour soutenir les décideurs dans l'évaluation des implications des différents cas d'usage. Cela réduit les risques associés à l'implémentation de nouvelles technologies.
- La consolidation des données permet d'identifier les écarts entre les capacités actuelles et les exigences futures, et ainsi élaborer des stratégies pour combler ces lacunes. Cela inclut l'optimisation des infrastructures existantes et la planification des investissements technologiques nécessaires.



Consolidation des données (dans le même fichier que la gestion des exigences)

Base de réflexion pour l'analyse SWOT relative à **LINDAS** :

Forces

- LINDAS améliore l'interopérabilité des données entre différentes entités administratives grâce à l'utilisation de modèles standardisés comme RDF et SKOS.
- LINDAS s'appuie sur des standards reconnus tels que RDF, SPARQL, et les ontologies SKOS, ce qui facilite l'intégration avec d'autres systèmes.
- La plateforme permet une publication centralisée des données, ce qui améliore la visibilité et l'accès aux données liées pour les utilisateurs internes et externes.

Faiblesses

- La mise en œuvre de LINDAS requiert une compréhension approfondie des modèles RDF, des URI et des namespaces, ce qui peut nécessiter des compétences techniques spécifiques et un investissement en formation.
- La transformation et la modélisation des données en RDF, ainsi que l'utilisation d'ontologies standardisées, peuvent être complexes et requérir des compétences spécialisées qui ne sont pas toujours disponibles en interne.



- Utilisation de version.link pour la gestion des données versionnées, assurant ainsi la traçabilité et l'historisation des ressources dans le temps.

- L'utilisation de Cube Creator et d'autres outils nécessite des configurations YAML et une intégration dans les pipelines CI/CD, ce qui peut ajouter une couche de complexité opérationnelle.

Opportunités

- Avec l'augmentation de l'intérêt pour les données ouvertes et la transparence gouvernementale, LINDAS peut jouer un rôle clé dans la publication et la gestion des données ouvertes, facilitant ainsi l'accès public et la réutilisation des données.
- L'intégration et l'interopérabilité des données peuvent conduire à des analyses plus précises et à de meilleure prise de décision dans l'administration publique.
- La plateforme favorise la collaboration entre différents départements en permettant le partage et l'utilisation commune des données liées, ce qui peut améliorer la coordination des politiques publiques.

Menaces

- Bien que des mécanismes de sécurité soient en place, la gestion des données sensibles et la conformité aux réglementations sur la protection des données restent des défis.
- LINDAS devra constamment s'adapter aux nouvelles normes et outils pour rester pertinent et efficace, ce qui peut nécessiter des investissements en temps et en ressources.
- L'adoption de nouvelles technologies et de nouveaux processus peut rencontrer de la résistance de la part des utilisateurs internes, ce qui peut ralentir l'implémentation et l'adoption de LINDAS.

Base de réflexion pour l'analyse SWOT relative à **Visualize** :

Forces

- Visualize permet la création de visualisations interactives qui reflètent les mises à jour des données en temps réel.
- Conçu pour la visualisation des données gouvernementales ouvertes, Visualize améliore la transparence et l'accessibilité des données publique.
- Les utilisateurs peuvent configurer des visualisations personnalisées, en choisissant parmi divers types de graphiques et en appliquant des filtres.
- Visualize peut manipuler des données provenant de diverses sources, facilitant ainsi une vue d'ensemble cohérente des informations disponibles.

Faiblesses

- L'utilisation de Visualize requiert une ontologie Cube et des vocabulaires contrôlés, ce qui peut ajouter de la complexité pour les utilisateurs non familiers avec ces concepts.
- L'intégration des données et la configuration des visualisations peuvent être complexes et nécessiter des compétences techniques avancées.
- La performance et l'efficacité de Visualize dépendent de la qualité et de la structure des données RDF sous-jacentes, ce qui peut être un défi en cas de données mal structurées.

Opportunités

- En facilitant l'accès et la visualisation des données publiques, Visualize peut améliorer la transparence gouvernementale.
- Visualize peut devenir un outil clé pour les administrations publiques souhaitant publier et visualiser leurs données de manière interactive et accessible.
- Le développement de Visualize et l'intégration de nouvelles fonctionnalités de visualisation peuvent offrir des opportunités pour des analyses de données plus sophistiquées.

Menaces

- Bien que Visualize améliore l'accessibilité des données, il faut encore gérer les questions de sécurité et de confidentialité des données sensibles.
- La gestion et la mise à jour continue des données RDF pour garantir des visualisations précises peuvent être laborieuses et nécessiter des ressources dédiées.
- Visualize devra s'adapter aux nouvelles technologies et aux attentes des utilisateurs pour rester pertinent.



- Visualize pourrait être adapté pour d'autres secteurs nécessitant des visualisations de données interactives, comme l'éducation, la recherche, ou l'industrie.

- Le maintien et le support de la plateforme peuvent nécessiter des investissements en temps et en ressources humaines.

Base de réflexion pour l'analyse SWOT relative à I14Y :

Forces

- La plateforme I14Y permet de centraliser et de référencer les métadonnées, facilitant l'interopérabilité des données entre différentes entités administratives et systèmes.
- I14Y assure une accessibilité centralisée des métadonnées, ce qui simplifie la recherche et l'utilisation des données.
- I14Y utilise des standards pour la gestion des métadonnées, garantissant une cohérence dans la manière dont les données sont décrites et référencées.
- La plateforme est conçue pour renforcer la transparence et l'accessibilité des informations publiques.

Faiblesses

- Le succès de I14Y dépend de la qualité et de la cohérence des métadonnées fournies par les différents contributeurs, ce qui peut varier considérablement.
- Intégrer I14Y avec des systèmes existants peut être complexe et nécessite des compétences techniques avancées.
- La gestion continue des métadonnées et la nécessité de maintenir la plateforme à jour peuvent nécessiter des ressources importantes.

Opportunités

- Avec une adoption croissante, I14Y peut devenir un standard pour l'interopérabilité des données.
- En facilitant l'accès aux métadonnées, I14Y peut jouer un rôle clé dans l'amélioration de la responsabilisation de l'administration fédérale.
- I14Y pourrait être adopté par d'autres secteurs nécessitant une gestion centralisée et standardisée des métadonnées.
- L'évolution des technologies de gestion des métadonnées et de l'interopérabilité pourrait offrir de nouvelles fonctionnalités à la plateforme.

Menaces

- La centralisation des métadonnées pose des défis en termes de sécurité et de confidentialité, notamment en cas de violation de données.
- L'adoption de nouvelles technologies et de nouveaux standards peut rencontrer une résistance, ralentissant l'intégration et l'utilisation de I14Y.
- L'émergence de nouvelles plateformes de gestion des métadonnées ou l'amélioration des plateformes existantes pourrait poser des défis concurrentiels à I14Y.

Base de réflexion pour l'analyse SWOT relative à Opendata.swiss :

Forces

- Opendata.swiss est une plateforme centralisée qui permet d'accéder à une collection de données ouvertes.
- La plateforme renforce la transparence en rendant les données gouvernementales accessibles au public.
- Opendata.swiss offre des jeux de données provenant de différentes unités administratives, couvrant des domaines tels que la santé, l'éducation, les transports, ect.
- En fournissant des données ouvertes, la plateforme encourage l'innovation et le

Faiblesses

- La qualité des jeux de données peut varier, certains jeux de données pouvant être incomplets ou mal documentés, ce qui limite leur utilité.
- Certains jeux de données ne sont pas mis à jour régulièrement, ce qui peut rendre les informations obsolètes.
- Pour les utilisateurs non techniques, la manipulation et l'analyse des données brutes peuvent être complexes.



développement de nouvelles applications et services basés sur les données.

Opportunités

- Opendata.swiss peut augmenter sa valeur et son utilité en encourageant plus d'institutions à publier leurs données.
- La plateforme peut renforcer la collaboration avec des entreprises privées pour développer des applications basées sur les données ouvertes.
- Le développement de meilleures interfaces de programmation d'application (API) et d'outils de visualisation peut rendre les données plus accessibles et exploitables.

Menaces

- Le succès de la plateforme dépend de l'engagement de l'administration fédérale à publier et à maintenir leurs jeux de données.
- L'émergence d'autres plateformes de données ouvertes, tant au niveau national qu'international, peut poser des défis concurrentiels à Opendata.swiss.

Base de réflexion pour l'analyse SWOT relative à Cube Creator :

Forces

- Cube Creator permet de convertir automatiquement des données CSV et autres formats tabulaires en cubes de données RDF.
- L'outil utilise des fichiers YAML pour définir les configurations, permettant une personnalisation facile des mappings spécifiques pour les colonnes de données, les URI et les vocabulaires utilisés.
- Cube Creator peut être intégré dans des pipelines CI/CD, assurant la reproductibilité et l'automatisation des transformations de données.
- L'outil est conçu pour fonctionner de manière transparente avec cube.link, facilitant ainsi l'intégration des données multidimensionnelles.

Faiblesses

- La nécessité de configurer des fichiers YAML peut être complexe pour les utilisateurs non techniques.
- L'utilisation de Cube Creator peut nécessiter une compréhension approfondie des concepts RDF et des vocabulaires standardisés, ce qui peut limiter son adoption par des utilisateurs non spécialisés.
- Cube Creator peut avoir des limitations lorsqu'il s'agit de transformer des données à partir de formats non tabulaires ou très spécifiques.

Opportunités

- En ajoutant des fonctionnalités pour supporter plus de formats de données et en simplifiant les processus de configuration, Cube Creator pourrait élargir sa base d'utilisateurs.
- Collaborer avec d'autres outils et plateformes de données liées pourrait améliorer l'interopérabilité et offrir des solutions plus complètes aux utilisateurs.
- Offrir des ressources de formation et des tutoriels détaillés pourrait aider à surmonter les barrières techniques et encourager une adoption plus large.
- Développer une interface utilisateur plus intuitive pourrait rendre l'outil accessible à un public plus large.

Menaces

- L'émergence de nouveaux outils et plateformes offrant des fonctionnalités similaires pourrait réduire l'attractivité de Cube Creator.
- La complexité technique et les besoins en formation pourraient limiter l'adoption de Cube Creator par des organisations ou des utilisateurs qui recherchent des solutions plus simples et intuitives.



Base de réflexion pour l'analyse SWOT relative à Cube.link :

Forces

- Cube.link utilise le vocabulaire RDF Data Cube Vocabulary (QB) pour modéliser des données sous forme de cubes, permettant une représentation structurée des données.
- Grâce à l'utilisation de RDF et des vocabulaires standardisés, Cube.link facilite l'intégration avec d'autres sources de données RDF.
- Le modèle Cube.link permet de définir des relations complexes entre les données, notamment via des propriétés de relation spécifiques.
- Les métadonnées fournissent des informations contextuelles, comme les descriptions des dimensions, mesures et attributs.

Faiblesses

- La mise en œuvre de Cube.link peut nécessiter une compréhension approfondie des concepts RDF et des vocabulaires standardisés.
- L'utilisation de Cube.link peut être limitée par la nécessité de recourir à des vocabulaires standardisés spécifiques, réduisant sa flexibilité dans certains contextes.
- La gestion et la mise à jour des cubes de données et des vocabulaires associés peuvent nécessiter des ressources et des compétences spécialisées.

Opportunités

- En développant des fonctionnalités supplémentaires pour supporter d'autres formats de données et en simplifiant l'utilisation, Cube.link pourrait attirer un plus large éventail d'utilisateurs.
- Fournir des formations et des tutoriels détaillés pourrait aider à surmonter les barrières techniques, favorisant une adoption plus large.

Menaces

- La complexité technique et les besoins en formation pourraient limiter l'adoption de Cube.link par des organisations recherchant des solutions plus simples et intuitives.
- Les incompatibilités potentielles avec des systèmes ou des vocabulaires non standardisés pourraient poser des défis pour l'intégration et l'utilisation de Cube.link.

5.2 Validation des cas d'usages

La validation des cas d'usage vise à identifier s'ils répondent aux besoins métier et s'intègrent de façon réaliste dans l'environnement technologique existant. Cette validation doit déterminer la pertinence de transformer les données existantes en données liées (Linked Data). Ceci implique l'évaluation de divers critères afin d'identifier les cas d'usage les plus prometteurs et alignés avec les objectifs stratégiques et opérationnels de l'organisation.

L'objectif principal est de s'assurer que les projets de transformation des données apportent une réelle valeur ajoutée et sont réalisables dans le cadre des ressources et des compétences disponibles. Pour ce faire, trois critères principaux sont pris en compte.

L'utilisation de l'arbre décisionnel (checkliste) réalisé sur la base du tableau des compétences proposé au chapitre 4.2 est conseillé.

Alignement stratégique

Vérifier que les cas d'usage proposés sont en phase avec les objectifs stratégiques. Chaque projet doit contribuer de manière significative à la réalisation des buts et des missions de l'entité administrative concernée.

Faisabilité technique

Évaluer la compatibilité des cas d'usage avec les technologies existantes et identifier les besoins en nouvelles infrastructures. Les solutions techniques nécessaires doivent être mises en place sans difficulté majeure.

Priorisation des initiatives

Évaluer l'impact de chaque projet, les prioriser en fonction de leur valeur ajoutée et de leur faisabilité. Les projets les plus prometteurs, offrant le plus grand bénéfice avec des ressources raisonnables, seront mis en avant.



6 Phase 4 : Tests et validation

L'évaluation des résultats des cas d'usage est essentielle pour garantir que les objectifs initiaux sont atteints et pour fournir des retours d'expérience pour les projets futurs. Cette étape permet de vérifier si les bénéfices attendus se concrétisent et d'identifier les points d'amélioration.

6.1 Définition des indicateurs de performance (KPI)

Avant de commencer l'évaluation, il est conseillé de définir des indicateurs de performance mesurables qui permettront de juger du succès de chaque cas d'usage. Les KPIs doivent être alignés avec les objectifs stratégiques et opérationnels identifiés lors de la phase de sélection des cas d'usage. Voici quelques exemples de KPIs potentiels :

Processus	Indicateur	Variable
Inputs	Ressources disponibles pour les linked data	<ul style="list-style-type: none">• Nombre de spécialistes RDF et SPARQL• Budget alloué aux projets de linked data• Formation continue des spécialistes• Partenariats technologiques
Activities	Efficacité des activités de transformation des données	<ul style="list-style-type: none">• Nombre de jeux de données convertis en RDF• Nombre d'ontologies créées ou adaptées• Fréquence des mises à jour des datasets RDF• Elaboration de guides pour les utilisateurs• Processus de tests et de validation des données• Création de documentation et matériel de support
Outputs	Performance des données liées	<ul style="list-style-type: none">• Quantité de données disponibles en RDF• Qualité des données (cohérence et précision des ontologies)• Accessibilité des données (temps de réponse des requêtes SPARQL)• Nombre d'offices utilisant les données liées
Outcomes	Adoption et satisfaction des linked data	<ul style="list-style-type: none">• Fréquence d'accès et d'utilisation des données liées• Niveau de satisfaction des utilisateurs finaux• Nombre de nouvelles applications ou analyses basées sur les données liées
Goals	Impact des linked data sur l'administration publique	<ul style="list-style-type: none">• Amélioration de l'interopérabilité des systèmes• Facilitation de la prise de décision grâce à des données interconnectées• Augmentation de la transparence et de l'accès aux données publiques• Satisfaction des parties prenantes• Impact économique global des données liées

6.2 Evaluation statistique

Pour évaluer l'efficacité des cas d'usage, il est utile de procéder à une analyse statistique au moyen d'outils standards (Excel peut tout à fait être suffisant). Quelques exemples sont décrits dans le présent chapitre. Tout autre analyse statistique plus sophistiquée est encouragée si elles aident à terme les offices à augmenter leur efficacité opérationnelle.

6.2.1 Régression linéaire

La régression linéaire peut être utilisée pour modéliser la relation entre une ou plusieurs variables explicatives (X - indépendantes) et une variable de réponse (Y - dépendante). Dans notre contexte, il pourrait être intéressant d'examiner des relations telles que :



Relation	Variable X	Variable Y
Impact des ressources sur la performance des données liées.	<ul style="list-style-type: none">• Nombre de spécialistes• Budget alloué• Fréquence de formation continue	<ul style="list-style-type: none">• Nombre de jeux de données transformés en Linked Data• Performance des requêtes SPARQL
Impact des données liées sur l'utilisation des services	<ul style="list-style-type: none">• Nombre de jeux de données liés publiés• Fréquence de mise à jour des données• Qualité des données (évaluée par des mesures de complétude et d'exactitude)	<ul style="list-style-type: none">• Fréquence d'accès aux données• Satisfaction des utilisateurs

6.2.2 Analyse de l'endogénéité

L'endogénéité se produit lorsqu'une ou plusieurs variables explicatives (X) sont corrélées avec l'erreur de régression. Pour détecter et corriger l'endogénéité, et ainsi tester si les variables explicatives sont corrélées avec les erreurs du modèle, vous pouvez par exemple utiliser le test de Hausman. En utilisant cette approche, il est possible de minimiser les biais causés par l'endogénéité et obtenir des estimations plus fiables de l'impact des ressources sur les performances des initiatives de Linked Data.

Pour corriger l'endogénéité, il est d'usage d'utiliser des variables qui sont corrélées avec les variables explicatives mais non corrélées avec l'erreur. Pour illustrer ceci, supposons que vous souhaitiez étudier l'impact des ressources (comme le budget alloué et le nombre de spécialistes) sur la performance des requêtes SPARQL. La variable de réponse est la performance des requêtes SPARQL (mesurée en temps de réponse, taux de succès, etc.), et les variables explicatives sont le budget alloué et le nombre de spécialistes travaillant sur les données liées.

Variable Y	Variable X	Problème d'endogénéité
Performance des requêtes SPARQL	<ul style="list-style-type: none">• X1 : Budget alloué aux initiatives de Linked Data• X2 : Nombre de spécialistes travaillant sur les initiatives de Linked Data	L'endogénéité peut survenir si le budget alloué (X1) est corrélé avec des facteurs non observés qui affectent également la performance des requêtes SPARQL (Y). Par exemple, une organisation pourrait allouer un budget plus important aux initiatives de Linked Data précisément parce qu'elle a identifié des besoins spécifiques ou des problèmes de performance antérieurs. Ces besoins spécifiques ou problèmes antérieurs sont des facteurs non observés qui peuvent influencer à la fois le budget et la performance des requêtes.

Pour corriger l'endogénéité, nous pouvons utiliser une variable instrumentale (VI). Cette variable est corrélée avec la variable explicative endogène (le budget alloué), mais n'est pas corrélée avec les erreurs du modèle de régression. Dans le cas présent, vous pourriez utiliser le budget global de digitalisation de l'organisation comme une variable instrumentale. Il serait ainsi possible d'évaluer de manière plus précise l'impact du budget alloué et du nombre de spécialistes sur la performance des requêtes SPARQL.

6.2.3 Analyse de la corrélation

L'analyse de la corrélation permet d'identifier la force et la direction de la relation entre deux variables. Une corrélation positive indique que lorsque l'une des variables augmente, l'autre tend à augmenter également. À l'inverse, une corrélation négative signifie que lorsque l'une des variables augmente, l'autre tend à diminuer.

Concrètement, il semble pertinent d'examiner la corrélation entre la fréquence de mise à jour des données et leur utilisation. Les variables considérées dans cette analyse sont la fréquence de mise à jour des données (mesure à quelle fréquence les données sont actualisées) et la fréquence d'accès aux données (indique à quelle fréquence les utilisateurs accèdent à ces données). Pour évaluer cette relation, nous pouvons calculer le coefficient de corrélation de Pearson, qui quantifie la relation entre les deux variables. Un coefficient



proche de +1 indique une forte corrélation positive, tandis qu'un coefficient proche de -1 indique une forte corrélation négative. Un coefficient proche de 0 suggère qu'il n'y a pas de relation linéaire entre les variables.

Si le coefficient entre la fréquence de mise à jour et la fréquence d'accès est élevé et positif, cela indique que des mises à jour plus fréquentes des données sont associées à une utilisation plus fréquente de ces données. Cela peut suggérer que les utilisateurs trouvent les données plus utiles et fiables lorsqu'elles sont régulièrement actualisées. À l'inverse, un coefficient négatif indiquerait que des mises à jour fréquentes pourraient, paradoxalement, réduire l'utilisation des données, peut-être en raison de perturbations causées par les mises à jour constantes.

6.3 Comparaison avec les objectifs initiaux

La comparaison des résultats obtenus avec les objectifs initiaux permet d'évaluer le degré de réussite du projet de transformation des données liées. Cette comparaison permet :

- d'identifier les écarts entre les attentes et la réalité
- d'analyser les causes de ces écarts
- de proposer des mesures correctives pour améliorer les performances futures

Afin d'être le plus objectif possible, il faudrait d'une part identifier les domaines où les résultats dépassent les attentes et ceux où ils sont en deçà et d'autre part comprendre pourquoi certains objectifs n'ont pas été atteints et proposer des mesures correctives. L'exemple ci-dessous propose un exemple fictif de résultat et d'analyse.

Indicateur	Objectif initial	Résultat obtenu	Ecart	Analyse
Performance des requêtes SPARQL	Améliorer le temps de réponse des requêtes SPARQL pour qu'il soit inférieur à 2 secondes.	Le temps de réponse moyen des requêtes SPARQL est de 1,8 secondes, ce qui respecte l'objectif fixé.	Aucun écart, l'objectif est atteint.	L'objectif de performance des requêtes SPARQL a été atteint grâce à des investissements significatifs dans l'infrastructure technique et à l'optimisation des requêtes. Les outils de monitoring ont permis de détecter et de corriger rapidement les goulots d'étranglement.
Fréquence d'utilisation des données liées	Augmenter l'utilisation des données liées de 25% dans les six premiers mois.	L'utilisation des données liées a augmenté de 15% seulement.	L'objectif n'est pas atteint, avec un écart de -10%.	La sous-utilisation des données liées peut être attribuée à une insuffisance de communication et de formation. Les utilisateurs potentiels ne sont pas suffisamment sensibilisés aux avantages des données liées et peuvent rencontrer des difficultés techniques pour y accéder.
Qualité des données	Atteindre une satisfaction des utilisateurs de 90% en termes de qualité des données.	La satisfaction des utilisateurs est de 85%.	L'objectif n'est pas atteint, avec un écart de -5%.	La légère insatisfaction en termes de qualité des données est due à des problèmes d'intégration de différentes sources de données. Des incohérences et des erreurs mineures dans les données importées ont été signalées, nécessitant une amélioration des processus de nettoyage et de validation des données.
Nombre de spécialistes formés	Former 50 spécialistes à l'utilisation	45 spécialistes ont été formés.	L'objectif n'est pas atteint, avec	Le nombre de spécialistes formés est légèrement en deçà de l'objectif en raison de contraintes



des technologies Linked Data.

un écart de - 5 spécialistes.

de calendrier et de disponibilité des formateurs. De plus, certains programmes de formation n'ont pas pu être complétés dans les délais prévus en raison de la complexité des sujets abordés.

6.4 Recommandations et ajustements

L'évaluation des résultats obtenus par rapport aux objectifs initiaux peuvent révéler des domaines où des améliorations peuvent être apportées. Afin d'optimiser les processus et de maximiser l'impact des projets de Linked Data sur l'administration fédérale, des recommandations et des ajustements stratégiques sont nécessaires. Ces recommandations visent non seulement à corriger les écarts identifiés, mais également à promouvoir une culture d'amélioration continue et d'innovation.

Le tableau ci-dessous propose des exemples de recommandation basés sur notre exemple du chapitre 6.3.

Recommandation	Action	Exemple
Augmenter la sensibilisation et la formation	Lancer des campagnes de sensibilisation pour promouvoir les avantages des données liées.	Organiser des journées portes ouvertes et des démonstrations interactives des projets pour montrer concrètement les bénéfices apportés.
	Organiser des sessions de formation supplémentaires	Développer un programme de formation avec des modules spécifiques sur l'utilisation de SPARQL et la modélisation RDF.
Améliorer les processus de nettoyage et de validation des données	Mettre en place des processus plus rigoureux de nettoyage et de validation des données avant leur intégration.	Implémenter un système de contrôle de la qualité des données utilisant des outils d'automatisation pour corriger les erreurs avant l'importation dans le triplestore.
	Développer des protocoles standards pour le nettoyage et la validation des données.	Etablir des check-lists et des guides pour assurer une préparation homogène des données à travers l'administration.
Optimiser la planification des programmes de formation	Réviser les calendriers de formation pour s'adapter à la disponibilité des formateurs et des participants.	Adapter les horaires des formations pour qu'ils coïncident avec les périodes de moindre activité des participants.
	Développer des modules de formation en ligne pour permettre une formation flexible.	Créer une plateforme d'e-learning avec des cours interactifs, des vidéos tutoriels, et des exercices pratiques.
Renforcer le support technique	Augmenter le support technique disponible pour les utilisateurs afin de résoudre les problèmes techniques.	Mettre en place un service de support technique dédié aux données liées avec une équipe de spécialistes disponible via une hotline et un chat en direct.
	Former une équipe de support technique spécialisée dans les données liées.	Organiser des sessions de formation avancée pour les membres de l'équipe de support.



7 Phase 5 : Référencement

La documentation et la communication aide à assurer la transparence et l'accessibilité des initiatives visant à développer les données liées au sein de l'administration. Pour maximiser l'impact et favoriser l'adoption de ces technologies, nous proposons de mettre en place une stratégie de publication et de partage des documents. Ce chapitre présente une proposition de mise en œuvre qui sera discutée plus en détail lors des prochaines phases des activités liées aux données liées.

Phase	Objectif	Action	Détail
Publication sur l'internet	Centraliser toutes les informations, guides, manuels, et résultats relatifs aux initiatives de Linked Data sur une plateforme accessible au public.	Créer une page web dédiée aux données liées	https://www.bk.admin.ch/bk/fr/home/digitale-transformation-ikt-lenkung/bundesarchitektur/daten-und-informationsmanagement/linked-data.html
Utilisation de GitHub pour la collaboration et la transparence	Héberger les scripts, outils, et configurations utilisés dans les initiatives de Linked Data pour encourager la collaboration et la réutilisation.	Créer un dépôt GitHub pour chaque projet majeur	Contenu du dépôt : <ul style="list-style-type: none">• Scripts de transformation des données.• Modèles RDF et configurations SPARQL.• Documentation technique et guides d'utilisation.
Intégration avec les portails de données ouvertes	Assurer une large diffusion et accessibilité des données publiées.	Publier les jeux de données et leurs métadonnées	<ul style="list-style-type: none">• Décrire chaque jeu de données avec des métadonnées complètes.• Fournir des liens vers la documentation et les outils associés. Exemple de mise en œuvre : <ul style="list-style-type: none">• Nettoyer et préparer les données pour la publication.• Utiliser les API d'opendata.swiss ou LINDAS pour automatiser la publication des données.• Annoncer les nouvelles publications via le site web.