

Document d'information DataHub

Date 08.03.2022

Version 1.1



Informations sur le document

Titre:	Document d'information DataHub
Date de publication:	8 mars 2022 (basé sur la version allemande)
Classification:	Non classifié
Projet:	Objectif de mise en œuvre 14 (élaborer et gérer l'architecture de la cyberadministration) de la stratégie suisse de cyberadministration 2020-2023
Nom du fichier:	Document d'information DataHub
Responsable du document:	Jürg Wüst, Transformation numérique et gouvernance de l'informatique
Vérifié par:	Comité d'architecture de la cyberadministration suisse Conseil de l'architecture de la Confédération Autres

Versions

Version	Date	Principale modification	Responsable
0.1	23.09.2021	Préparation du document d'information, élaboration initiale des contenus	AWK, collectif d'auteurs
0.2	01.10.2021	Restructuration des chapitres, élaboration de l'introduction	AWK, collectif d'auteurs
0.3	17.11.2021	Formulation du modèle cible, détail des piliers, mise en perspective des projets et initiatives	AWK, collectif d'auteurs
0.4	01.12.2021	Révision de l'ensemble du document	Jürg Wüst, Franziska Bühler, Andreas Gyger
0.5	03.12.2021	Mise au point du document	AWK, collectif d'auteurs
0.6	18.01.2022	Prise en considération des remarques du comité d'architecture de la cyberadministration suisse	Jürg Wüst, AWK, collectif d'auteurs
0.7	14.02.2022	Prise en considération des remarques du conseil de l'architecture de la Confédération	Jürg Wüst, AWK, collectif d'auteurs
1.0	24.02.2022	Mise au point de la version 1.0	Jürg Wüst, AWK, collectif d'auteurs
1.1	08.03.2022	Correction du chapitre 6.6: lien géodonnées cantonales; nouveau sous-chapitre 6.7 Infrastructure nationale de données géographiques (INDG), complément sur l'INDG apporté au chapitre 7.3	Jürg Wüst, AWK, collectif d'auteurs

Table des matières

1	Synthèse	5
2	Introduction.....	7
2.1	Données et échange de données	7
2.2	Mandat	7
2.3	Objectif du présent document	8
2.4	Méthodologie de la démarche.....	8
3	Modèle cible de l'écosystème DataHub	9
3.1	Exigences en matière d'échange de données.....	9
3.2	Vision et objectifs.....	9
3.3	Principaux cas d'application.....	9
4	Exigences relatives à l'écosystème DataHub.....	11
4.1	Participants à l'écosystème DataHub.....	11
4.2	Souveraineté en matière de données	11
4.3	Adhésion progressive et poursuite de l'utilisation des solutions existantes	11
4.4	Infrastructure favorisant une utilisation efficace des données	12
4.5	Protection des données (sécurisation dès la conception ou <i>security by design</i>), confidentialité, intégrité et disponibilité	12
4.6	Boîte à outils de nature technique, organisationnelle et méthodologique	12
5	Hub de données	14
5.1	Qu'est-ce que le hub de données?	14
5.2	Aperçu des éléments de l'écosystème DataHub.....	14
5.3	Gouvernance et architecture.....	15
5.4	Secteurs de données.....	15
5.5	Les quatre piliers DataHub	15
5.5.1	Répertoires: visibilité et accessibilité de l'information existante	16
5.5.2	Passerelles API: infrastructure pour les interfaces électroniques.....	16
5.5.3	Services connexes: services permettant de développer des prestations basées sur les données	17
5.5.4	Politiques et standards: conditions-cadres pour une utilisation optimale DataHub.....	17
5.6	Principes de base	18
6	Autres initiatives et projets pertinents	19
6.1	Gaia-X	19
6.2	sedex.....	21
6.3	NaDB: plateforme d'interopérabilité	22
6.4	Mise en place d'une gestion transversale des données de base des autorités (OMO9)	22
6.5	LINDAS	23

6.6	OGD Suisse.....	23
6.7	Infrastructure nationale de données géographiques	24
6.8	Infrastructure nationale de données mobilitaires	25
6.9	Hub de données du marché de l'énergie (OFEN)	26
6.10	Trusted Datahub (La Poste).....	28
6.11	Swissdec	28
6.12	Standards eCH	28
6.13	eDelivery.....	30
6.14	X-Road	30
7	Mise en perspective des projets et initiatives	32
7.1	Gaia-X	32
7.2	IOP NaDB, sedex, OGD Suisse, LINDAS, mise en place d'une gestion transversale des données de base des autorités, standards eCH.....	32
7.3	INDG, NaDIM, hub de données du marché de l'énergie, Trusted Datahub, Swissdec	33
7.4	eDelivery, X-Road.....	33
8	Champs d'action	34
8.1	Clarification de la gouvernance.....	34
8.2	Plan d'architecture	34
8.3	Prototypes	35
Annexes	36
A.1.	Personnes et organisations impliquées.....	36
A.2.	Cas d'application pertinents envisageables	37
A.3.	Exigences générales	38
A.4.	Abréviations et notions.....	39
A.5.	Documents référencés et autres sources.....	40

1 Synthèse

L'échange de données entre les différentes entités composant l'administration publique ainsi qu'entre l'administration publique et des tiers est un élément clé de la transformation numérique. La mise en œuvre du principe de la saisie unique des données (principe «*once only*») et l'utilisation multiple des données supposent que ces dernières puissent être échangées sans entrave.

Qu'elle relève de la Confédération, d'un canton ou d'une commune, une unité administrative se heurte aujourd'hui à d'importantes difficultés lorsqu'elle entend, dans le cadre de la transformation numérique, offrir à des tiers un large accès à ses données. Souvent, les unités administratives ne disposent pas des services de base nécessaires à l'échange de données, de sorte que ces services sont mis en place plusieurs fois, sans réelle coordination. En outre, les fournisseurs de services de base centraux sont confrontés à l'impossibilité d'organiser leur solution en fonction d'une architecture globale. Il en résulte qu'aujourd'hui les différents services de base ne sont pas harmonisés entre eux.

Le présent document d'information décrit un modèle cible d'écosystème DataHub, qui relève de manière globale les défis posés par l'échange de données. L'écosystème DataHub vise à créer les conditions nécessaires à un échange fluide de données et de prestations basées sur les données entre différents services administratifs ainsi qu'entre ceux-ci et des tiers, afin d'augmenter, au profit de la collectivité, l'efficacité des processus administratifs. Il s'agit pour l'essentiel d'encourager la mise en œuvre du principe de la saisie unique des données et l'utilisation multiple des données.

Le **DataHub est un écosystème** composé de toutes sortes d'organisations, de systèmes techniques, de méthodes et de standards qui permettent aux participants à l'écosystème d'échanger entre eux, de manière contrôlée et dans le respect des principes du fédéralisme, des données et des prestations basées sur les données. Le hub de données n'est pas un système technique centralisé servant à échanger des données administratives en Suisse. Il forme plutôt le cadre et la base architecturale permettant de rassembler un grand nombre de plateformes ou de systèmes existants ou à créer. L'écosystème DataHub est une construction complexe et ambitieuse, qui tient compte des particularités du fédéralisme suisse et de la gestion décentralisée des données qui en découle.

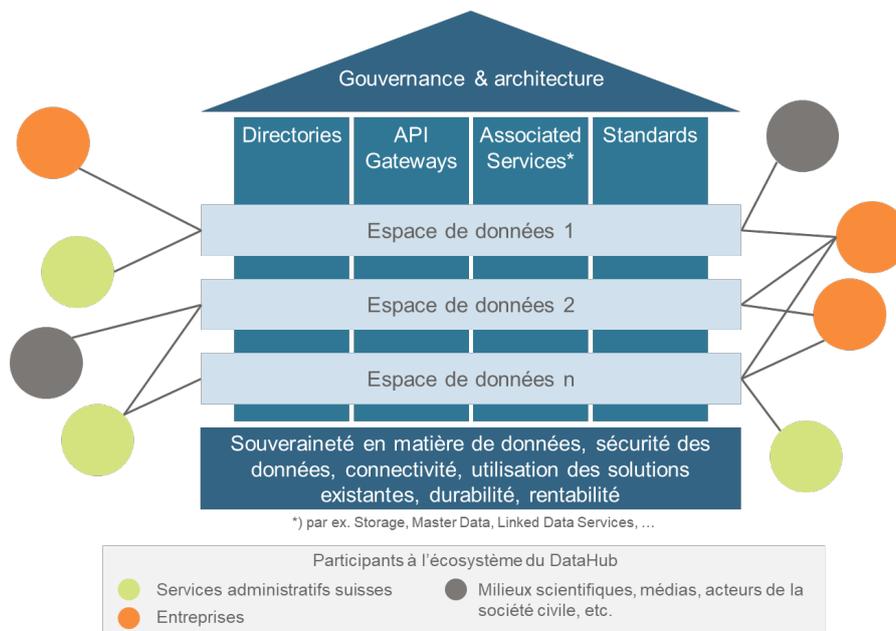


Figure 1: Schéma de l'écosystème DataHub et de ses composants

L'écosystème DataHub est ouvert à l'ensemble des services administratifs, entreprises et organisations tierces. Une gouvernance globale permet de contrôler et de développer la totalité de

l'écosystème. Une architecture globale décrit comment les différents éléments interagissent. Les participants à l'écosystème accèdent au DataHub par le biais de cas d'application concrets dans un secteur de données (espace de données). Les secteurs de données envisageables sont par exemple la «mobilité», la «géoinformation» ou l'«agriculture». Quatre piliers (Directories; API Gateways; Associated Services; Standards) représentent les éléments de systèmes permettant aux participants à l'écosystème de partager leurs données. L'écosystème repose sur des principes tels que ceux de la souveraineté en matière de données, de la sécurité des données, de la connectivité, de l'utilisation des solutions existantes, de la durabilité et de la rentabilité, qui sont essentiels pour assurer la participation à l'écosystème et l'échange de données.

Les réflexions sur le modèle cible d'écosystème DataHub et divers entretiens avec les parties prenantes ont permis de dégager trois champs d'action:

- **Clarification de la gouvernance**

Les questions de gouvernance doivent être clarifiées pour permettre la mise en place progressive, la gestion et le développement de l'écosystème. Il convient de définir quelle organisation assumera quelles responsabilités et comment le financement sera réglé.

- **Plan d'architecture**

Il importe que les composants de l'écosystème DataHub figurant dans le présent document soient décrits en détail et réunis dans une architecture commune.

- **Prototypes**

Pour que les concepts esquissés dans le présent document puissent être concrétisés, un ou plusieurs prototypes doivent être mis en œuvre. Parallèlement à la poursuite de la conception de l'écosystème, il s'agira ainsi d'acquérir très tôt une expérience pratique qui puisse être intégrée dans la suite des travaux.

2 Introduction

2.1 Données et échange de données

Les données, leur collecte, leur gestion, leur utilisation et leur échange sont au cœur de la transformation numérique. L'offre de prestations numériques des autorités et la mise en œuvre du principe de la saisie unique des données, l'accès ouvert aux données des autorités ainsi que l'utilisation multiple des données par l'administration fédérale présupposent des infrastructures et des processus, avec les standards correspondantes. Le Parlement de la Confédération l'a également reconnu et a exigé, par le biais de motions¹, une infrastructure de données durable (soit un hub de données) et un échange électronique sécurisé des données, au sein de l'administration fédérale, ainsi qu'avec les personnes physiques et morales.

L'échange numérique au sein de l'administration fédérale ainsi qu'avec l'économie et la société civile doit être amélioré au moyen d'interfaces interopérables, lisibles par machine et basées sur des standards ouverts. Un accès central, au sens d'un hub aux données, aux prestations basées sur les données de l'ensemble de l'administration doit être créé, tout en préservant l'autonomie des différentes unités administratives. Pour ce faire, il convient d'établir des règles et des standards qui soient obligatoires pour toutes les organisations, et dont le respect puisse être contrôlé.

Un hub de données doit permettre d'améliorer l'échange de données entre les services administratifs, les entreprises et la société civile ainsi que de créer une base importante pour les projets de cyberadministration. Il vise notamment à mettre en œuvre le principe de la saisie unique des données, de sorte que les habitants et les entreprises ne doivent transmettre qu'une seule fois les informations requises par les autorités.

L'échange de données entre les unités administratives et l'utilisation des données sont, de nos jours, confrontés à de nombreux défis. Il manque notamment souvent, des services de base d'une solution d'infrastructure aboutie, des directives et des procédures uniformes, ainsi qu'une gouvernance générale. Le présent document d'information aborde cette problématique et esquisse une solution possible, sous la forme d'un écosystème de hub de données.

2.2 Mandat

Dans le cadre de l'objectif de mise en œuvre 14 (OMO14) de la stratégie suisse de cyberadministration 2020-2023, une architecture cible globale est élaborée. Afin de coordonner, d'évaluer et de hiérarchiser les travaux, un comité d'architecture composé de représentants de l'économie, de la recherche scientifique, du public et de l'administration a été créé. Ce comité d'architecture a chargé l'organisation responsable de l'OMO14 d'étudier le thème d'un hub de données pour l'administration fédérale et de rédiger un document d'information à ce sujet.

Le mandat coïncide avec les objectifs de l'Administration numérique suisse (ANS), qui demande dans l'agenda «Infrastructures nationales et services de base de l'Administration numérique suisse» de diminuer le fardeau administratif de l'économie grâce à l'échange automatisé des données et aux interfaces avec l'administration ainsi que de créer et mettre en œuvre à tous les échelons étatiques les bases techniques, sémantiques et institutionnelles de la gestion des données. En outre, le mandat correspond aux objectifs de la déclaration signée par la Suisse sur la cyberadministration (*Tallinn Declaration*²).

¹ 20.4260, 18.4276 et 18.4238

² [Tallinn Declaration on eGovernment \(admin.ch\)](#)

2.3 Objectif du présent document

Le présent document d'information décrit un futur écosystème DataHub et les composantes organisationnelles et techniques de celui-ci. Pour rendre compte de la complexité de l'échange de données entre organisations, il ne suffit pas de mettre en lumière la seule infrastructure technique d'un hub de données. Seule la vision globale d'un écosystème DataHub, réunissant différents participants cherchant à échanger des données entre eux, permet de s'approcher d'un tel objectif.

L'écosystème DataHub a pour but de créer les conditions permettant aux différents services administratifs d'optimiser progressivement les échanges de données – entre eux ainsi qu'avec les autorités étrangères, les entreprises et les autres organisations. Le présent document s'adresse non seulement aux membres du comité d'architecture de la cyberadministration suisse, mais aussi aux décideurs de l'administration publique de tous les niveaux de l'État fédéral.

Le présent document place l'écosystème DataHub dans le cadre de l'architecture de la cyberadministration suisse. Il sert de base aux futures étapes de mise en œuvre de l'écosystème pour toutes les unités administratives fédérales, cantonales et communales, ainsi que pour les interlocuteurs de celles-ci. Le hub de données s'inscrit dans le sillage de projets déjà en cours. Ci-après sont décrits des champs d'action qui serviront de référence pour la suite des travaux.

Le cadre juridique et réglementaire n'est pas traité en détail ci-après. Pour la poursuite du projet, il faudra donc procéder à des clarifications juridiques supplémentaires afin, par exemple, de commencer à temps l'adaptation des bases juridiques. Les auteurs du présent document ne se prononcent pas non plus sur le futur financement d'infrastructures ou d'organisations techniques existantes ou à créer. Les deux aspects, à savoir les bases juridiques et la clarification du financement, sont en principe du ressort de l'ANS.

Le présent document ne prétend pas à l'exhaustivité, notamment en ce qui concerne les exigences relatives à un futur échange de données entre les autorités et les organisations concernées ou en ce qui concerne les composants d'infrastructure technique nécessaires. Il sert plutôt de base de discussion et constitue le point de départ des prochaines étapes devant mener à la création de l'écosystème DataHub.

2.4 Méthodologie de la démarche

Le présent document d'information a été rédigé par le domaine Transformation numérique et gouvernance de l'informatique (TNI) de la Chancellerie fédérale en collaboration avec l'Office fédéral de la statistique (OFS), dans le cadre des travaux liés à l'objectif de mise en œuvre 14 (élaborer et gérer l'architecture de la cyberadministration compte tenu du plan de mise en œuvre stratégique).

Il a été élaboré entre octobre 2020 et janvier 2022 sous la direction du domaine TNI de la Chancellerie fédérale, avec le concours de différents experts issus des milieux scientifiques et techniques. De nombreux représentants de l'administration, du pouvoir politique, d'organisations proches de l'administration et d'entreprises ont été associés, par le biais d'ateliers et d'entretiens, à l'élaboration de ce document. La liste de ces représentants se trouve à l'annexe A.1.

3 Modèle cible de l'écosystème DataHub

3.1 Exigences en matière d'échange de données

Les différents entretiens menés avec les parties prenantes (voir annexe A.1) ont montré qu'une prise en considération des seuls composants techniques de l'infrastructure d'un hub de données n'était pas suffisante pour garantir le succès d'une utilisation commune des données entre les organisations. Outre l'infrastructure technique, les parties échangeant des données ont notamment besoin de règles communes applicables à l'utilisation des données et d'une compréhension commune du contenu des données («sémantique»). La complexité est encore accrue par le fait qu'un service administratif ignore aujourd'hui avec quelles autres organisations il sera amené à échanger des données à l'avenir.

3.2 Vision et objectifs

La vision suivante d'un écosystème DataHub a été élaborée en collaboration avec les parties prenantes. Elle décrit l'objectif à atteindre et sa raison d'être.

Vision de l'écosystème DataHub

- L'écosystème DataHub crée les conditions d'un échange efficace de données et de prestations basées sur les données entre les différents services administratifs, ainsi qu'entre les services administratifs et des tiers, afin d'améliorer l'efficacité des processus administratifs au bénéfice de tous. Il permet une saisie unique des données dans le contexte d'une utilisation multiple de celles-ci.

Les objectifs suivants précisent la vision et les bénéfices attendus:

Objectifs de l'écosystème DataHub

- L'écosystème DataHub crée les conditions permettant à ses membres de mettre en place un accès standardisé aux données des autorités et aux prestations des autorités basées sur les données.
- Le hub de données met à disposition les éléments techniques et organisationnels du dispositif permettant l'échange de données tout en donnant aux différents services administratifs la possibilité d'optimiser progressivement leur propre traitement des données.
- Le hub de données fait appel à des solutions techniques et organisationnelles existantes.
- Le hub de données facilite la collaboration et la coopération entre les différents échelons de l'État.
- Le hub de données resserre progressivement les liens unissant les différentes administrations de l'État fédéral suisse, en se développant graduellement et en maintenant à un bas niveau les frais d'entrée dans l'écosystème.
- Le hub de données permet aux acteurs de l'écosystème de se connecter à des systèmes, des organismes et des comités suisses et internationaux.

3.3 Principaux cas d'application

Pour les acteurs de l'écosystème DataHub, les principaux avantages des cas d'application sont les suivants (voir également l'annexe A.2):

- Les participants à l'écosystème peuvent utiliser des éléments techniques et organisationnels du dispositif fourni de manière centralisée et développer ainsi progressivement leurs propres échanges de données avec des tiers, sans devoir mettre en place leur propre dispositif dédié.
- Les participants trouveront dans l'écosystème un cadre organisationnel ainsi que des méthodes et des directives leur permettant de promouvoir le plus efficacement possible le partage de données avec d'autres participants. Cela comprend notamment un accord sur des standards communs pour l'échange de données ainsi que la création d'une compréhension commune du contenu des données (sémantique).

Les participants à l'écosystème, et donc en particulier les services administratifs, ont la possibilité de mettre à disposition d'autres services administratifs, des données produites et gérées au sein de leur propre organisation, dans le respect de toutes les conditions-cadres. Le hub de données permet la publication et garantit la visibilité et la disponibilité des données. Les processus peuvent ainsi être organisés de manière plus efficace et la qualité des données peut être améliorée.

4 Exigences relatives à l'écosystème DataHub

L'écosystème DataHub doit répondre aux exigences décrites ci-après. Une liste plus détaillée des cas d'application et des exigences générales figure dans les annexes A.2 et A.3.

4.1 Participants à l'écosystème DataHub

Plus le nombre de participants à l'écosystème sera élevé, plus les bénéfices seront importants pour les parties prenantes.

L'écosystème DataHub doit donc être ouvert à tous les services administratifs, entreprises et autres organisations, en Suisse et à l'étranger, qui souhaitent interagir et échanger des données avec des services administratifs suisses. Les participants à l'écosystème doivent pouvoir échanger et partager des données, à condition de disposer des autorisations nécessaires et de respecter les bases juridiques correspondantes. On distingue les deux rôles³ suivants:

- Un **fournisseur de données** (*data provider*) est un participant à l'écosystème qui fournit, via le hub de données de l'écosystème, des données et des prestations liées aux données. Il définit et décrit sa propre offre de données et de prestations, y compris les conditions d'utilisation.
- Un **consommateur de données** (*data consumer*) est un participant à l'écosystème qui recherche et utilise, via le hub de données de l'écosystème, des données et des prestations liées aux données, par exemple pour permettre la création de nouvelles offres numériques pour les utilisateurs finaux.

Les participants à l'écosystème peuvent endosser les deux rôles suivant la situation.

L'écosystème doit en principe être ouvert à tous. Le fournisseur de données décide qui a accès à quelles données et à quel moment (voir également le chapitre 4.2).

4.2 Souveraineté en matière de données⁴

Les fournisseurs de données de l'écosystème doivent en principe pouvoir à tout moment décider eux-mêmes de leurs données. Il importe donc qu'ils puissent contrôler en tout temps leurs propres données et déterminer quels participants à l'écosystème peuvent y accéder et à quelles conditions. Mais pour permettre aux utilisateurs d'avoir un accès fiable et à long terme aux données, les fournisseurs passent préalablement un accord avec les utilisateurs. Ce faisant, les fournisseurs doivent définir, par exemple, une garantie en matière de disponibilité des données. Les accords tiennent compte du cadre juridique.

La souveraineté en matière de données prend en considération le fédéralisme de la Suisse et, donc, le fait que la responsabilité relative aux différentes entités de données relève de services administratifs décentralisés.

4.3 Adhésion progressive et poursuite de l'utilisation des solutions existantes

L'écosystème DataHub offre aux services administratifs et aux participants la possibilité d'une adhésion progressive, à un rythme choisi, ainsi que la possibilité d'une utilisation graduelle de ses éléments techniques et organisationnels (*opting-in*).

³ Ces deux rôles sont compatibles avec les rôles définis dans la vision architecturale. Les entreprises, les partenaires et les autorités peuvent tous agir comme fournisseurs et/ou utilisateurs de données.

⁴ La souveraineté en matière de données désigne le contrôle total des données stockées et traitées, ainsi que la décision autonome de savoir qui peut y accéder (Digital Summit, 2015).

Les solutions et applications existantes doivent, dans la mesure du possible, pouvoir continuer à être utilisées et intégrées dans l'écosystème, pour autant que les exigences en matière de sécurité, notamment, soient remplies. Cela permet par exemple de protéger les investissements déjà réalisés et de limiter les problèmes d'adhésion pour les nouveaux participants. L'écosystème DataHub met à disposition des outils permettant de se raccorder aux solutions existantes et de les intégrer.

4.4 Infrastructure favorisant une utilisation efficace des données

Le hub de données met à la disposition des participants à l'écosystème des composants d'infrastructure technique afin de permettre l'échange de données et de prestations basées sur les données. De nombreux composants sont déjà disponibles aujourd'hui de manière décentralisée dans certains services administratifs. Ils doivent pouvoir continuer à être utilisés dans la mesure du possible. Le hub de données rend ces composants visibles et accessibles à tous les participants à l'écosystème. Il propose des stratégies d'intégration afin que les composants d'infrastructure puissent être partagés à l'avenir, qu'ils soient utilisés de manière centralisée ou décentralisée.

Pour les services administratifs qui ne disposent pas encore de leur propre infrastructure, le hub de données fournit de manière centralisée des composants d'infrastructure. Cela réduit notablement les ennuis d'adhésion des nouveaux participants, en particulier les plus petits d'entre eux. Ces composants d'infrastructure, disponibles de manière centralisée, comprennent notamment des solutions d'interfaces sécurisées pour la communication avec les tiers, ainsi que de l'espace de stockage, qui est par exemple demandé par les petits services administratifs dans le domaine des données publiques en libre accès (*open government data*). Mais surtout, les participants à l'écosystème doivent avoir la possibilité de vérifier si des données pertinentes sont déjà disponibles chez un autre participant à l'écosystème et comment elles peuvent être obtenues (catalogue de données).

4.5 Protection des données (sécurisation dès la conception ou *security by design*), confidentialité, intégrité et disponibilité

Des données sensibles tombées en de mauvaises mains peuvent avoir de graves conséquences. Les composants de l'infrastructure DataHub répondent donc aux exigences les plus élevées en matière de protection et de sécurité de l'information. Les composants techniques centraux DataHub sont disponibles en tout temps et autant que possible à l'abri des pannes, afin que leur fonctionnement puisse être garanti. La cybersécurité est assurée dans l'écosystème DataHub.

Seuls les utilisateurs et systèmes authentifiés et autorisés peuvent accéder aux données de l'écosystème. Dans ce contexte, le service administratif qui fournit les données détermine les utilisateurs éligibles (voir chapitre 4.2). Les autorisations se basent sur les conditions-cadres légales et sur les accords conclus entre les participants à l'écosystème. Mais l'infrastructure DataHub peut également être utilisée pour mettre ses propres données à la disposition du public, par exemple dans le cadre de l'*open government data*.

L'intégrité des données, c'est-à-dire la traçabilité des modifications et la garantie que les données n'ont pas été modifiées par des personnes non autorisées sans qu'on le remarque, est essentielle pour différentes activités des autorités. Le principe de la confiance zéro doit valoir pour l'obtention de données non ouvertes. Là encore, il faut que l'écosystème DataHub offre des solutions adéquates.

4.6 Boîte à outils de nature technique, organisationnelle et méthodologique

Le hub de données fournit des outils techniques, organisationnels et méthodologiques pour aider les administrations à utiliser l'écosystème et, plus généralement, à continuer de mettre en œuvre le principe de la saisie unique des données. Le soutien est disponible aux quatre niveaux

d'interopérabilité du cadre d'interopérabilité européen (*European Interoperability Framework EIF*: droit, organisation, sémantique et technique).

Ce soutien comprend par exemple des instructions et des bonnes pratiques sur l'utilisation et la mise à disposition d'interfaces de programmation d'application (*application programming interfaces*, API), sur l'option d'ouverture par défaut (*open by default*) et de partage par défaut (*share by default*) dans le cadre des bases légales, sur la publication structurée et standardisée de métadonnées et de données, sur les prestations de conseil, ainsi que sur les modèles de référence architecturaux et les modèles de rôles organisationnels.

5 Hub de données

5.1 Qu'est-ce que le hub de données?

Le hub de données est un écosystème composé de toutes sortes d'organisations, de dispositifs techniques, de méthodes et d'autres standards qui permettent aux participants à l'écosystème d'échanger entre eux, de manière contrôlée et dans le respect des principes du fédéralisme, des données et des prestations basées sur les données.

Le hub de données n'est donc pas un système technique centralisé servant à échanger des données administratives en Suisse. Il forme plutôt le cadre permettant de rassembler un grand nombre de systèmes, d'applications, de plateformes et de données existants ou à créer, et d'en assurer l'utilisation commune au moyen de directives, de concepts ainsi que d'une gouvernance correspondante.

L'écosystème DataHub est une entreprise complexe et ambitieuse, qui tient compte des particularités du fédéralisme suisse et de la gestion décentralisée des données qui en découle. Le critère central pour assurer le succès de ce futur écosystème est le même que pour les autres écosystèmes: le nombre de participants. Le hub de données sera attrayant pour les participants s'il permet de traiter rapidement les premiers cas d'application (voir annexe A.2) et de réduire au maximum les entraves à l'entrée.

5.2 Aperçu des éléments de l'écosystème DataHub

Le modèle de l'écosystème DataHub, élaboré en tenant compte des exigences, peut être représenté de la façon suivante:

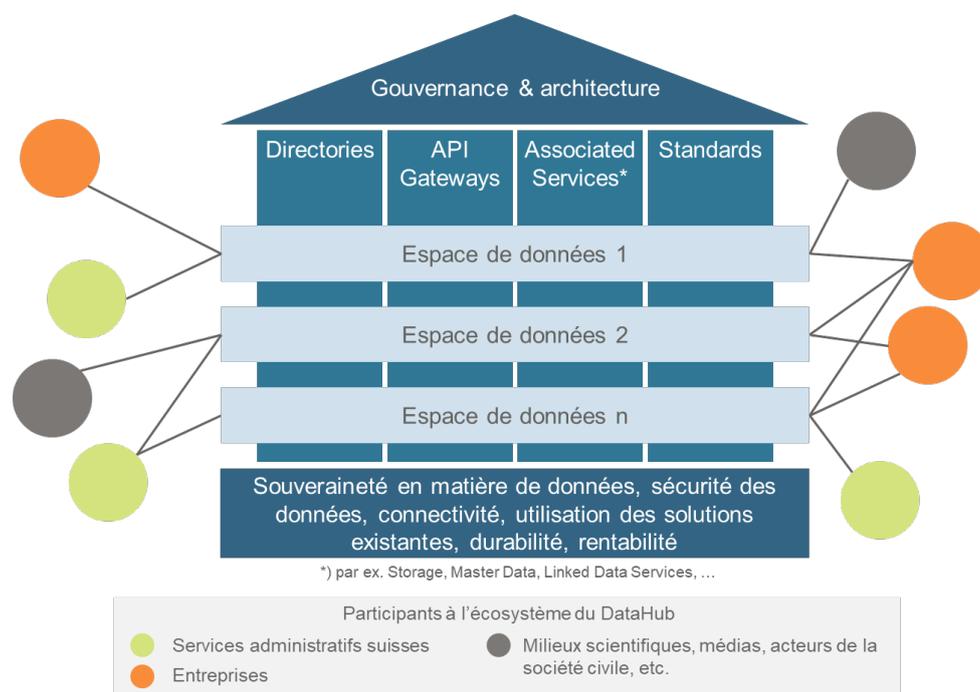


Figure 2: Schéma de l'écosystème de hub de données et de ses composants

Une gouvernance et une architecture globales forment le toit DataHub. Les quatre piliers Directories, API Gateways, Associated Services et Standards décrivent les éléments du dispositif

nécessaires pour que les services administratifs, les entreprises et les tiers puissent participer à l'écosystème et partager des données.

Les participants accèdent à l'écosystème en premier lieu par le biais de cas d'application concrets. Ils ont des défis concrets à relever dans un secteur spécifique de données et se retrouvent ainsi avec leurs pairs dans l'écosystème. Il peut s'agir de la mobilité, de la géo information, du réseau national de sécurité ou encore de l'agriculture.

L'écosystème lui-même repose sur des principes fondamentaux communs qui doivent être respectés, comme par exemple ceux qui concernent la souveraineté en matière de données et l'application des standards les plus élevées de sécurité des données. Les différents éléments DataHub sont décrits plus en détail ci-dessous.

5.3 Gouvernance et architecture

Une gouvernance commune permet de piloter l'ensemble de l'écosystème DataHub et de le faire évoluer en s'appuyant sur une architecture elle aussi commune. L'architecture est axée sur les concepts, les processus et la sécurité. Elle définit l'interconnexion, l'interopérabilité, l'intégration, la surveillance ainsi que l'exploitation des prestations dans l'écosystème et, en particulier, l'interaction organisationnelle entre les participants. Ceux-ci doivent se conformer aux règles du jeu découlant de ce contexte.

L'objectif de la gouvernance et de l'architecture consiste à définir le cadre pour les participants, de manière à laisser de la place à l'innovation, tout en privilégiant sans compromis la confiance et l'interopérabilité.

5.4 Secteurs de données

Un secteur de données décrit un domaine spécifique au sein duquel différents participants à l'écosystème souhaitent partager et échanger des données. Les secteurs de données envisageables sont par exemple des domaines comme «Espace et environnement» (voir NOGA⁵) ou des champs thématiques tels que «Mobilité», «Géo-information» ou «Agriculture».

Ces secteurs de données permettent aux participants à l'écosystème de se réunir autour d'un «thème de données». La communauté réunie autour d'un thème décide, dans le respect des règles de la gouvernance et de l'architecture globales, de la forme et de la conception concrètes des quatre piliers applicables à son thème de données et aussi, par exemple, de standards communs d'échange de données ou encore de questions touchant la sémantique (voir également le chapitre 5.5 ci-dessous). L'objectif est de permettre l'accès aux données à des participants extérieurs au secteur des données et de générer ainsi une plus-value globale.

De nouveaux participants rejoindront probablement l'écosystème DataHub par le biais d'un secteur de données spécifique, à l'exemple d'un service administratif ayant un besoin concret d'améliorer l'utilisation des données lors du remplacement prochain d'une application.

5.5 Les quatre piliers DataHub

Pour les quatre piliers, le «degré de maturité» ou le «degré de développement» varie d'un secteur de données à l'autre et, dans de nombreux cas, les éléments de base doivent encore être mis en place de manière centralisée ou par les participants à l'écosystème.

⁵ <https://www.bfs.admin.ch/bfs/fr/home/statistiken/industrie-dienstleistungen/nomenklaturen/noga.html>

5.5.1 Répertoires: visibilité et accessibilité de l'information existante

Les répertoires (*directories*) permettent à tous les participants de l'écosystème de visualiser et de trouver les données, les prestations, les informations et les éléments de solution existants.

L'écosystème DataHub est une entité décentralisée regroupant un grand nombre de participants, de données et de prestations basées sur les données. Le pilier Directories permet d'assurer la transparence dans ce réseau complexe et de trouver les informations souhaitées. Il offre aux participants la possibilité non seulement de publier et de faire connaître des données et des prestations au sein de l'écosystème, mais aussi de trouver d'emblée les données, prestations, sources d'approvisionnement et autres participants recherchés.

Pour ce faire, l'écosystème de hub de données met par exemple à disposition un regroupement de répertoires reliés entre eux et échangeant des informations. Les répertoires sont en principe spécifiques à un secteur de données. Le lien entre différents répertoires a pour but de rendre les métadonnées et leurs modifications interchangeables et de permettre les référencements croisés. L'accès à un répertoire peut être ouvert à tous les participants à l'écosystème, ou seulement à une partie d'entre eux.

Un participant à l'écosystème a la possibilité de publier des méta-informations sur ses données via un répertoire approprié. Dans le répertoire, il décrit les caractéristiques de ses données et de sa prestation sous forme de métadonnées (par ex.: contenu des données, volume des données, formatage des données, qualité des données, origine, etc.). Le même standard de description des données et des prestations s'applique à tous les répertoires. Cela permet l'interopérabilité des métadonnées au niveau sémantique. Le participant qui publie est responsable de la description exacte de la prestation qu'il propose et du respect des directives. En outre, il détermine lui-même qui peut consulter et obtenir ses données.

Le pilier des répertoires permet donc de retrouver les données et les prestations existantes en veillant à ce que celles-ci soient accessibles, interopérables et réutilisables⁶. Pour les participants à l'écosystème, des données et des prestations fournies de manière décentralisée peuvent ainsi être trouvées et interprétées.

5.5.2 Passerelles API: infrastructure pour les interfaces électroniques

Le pilier API Gateways (passerelles API) offre une infrastructure pour l'échange de données entre les participants à l'écosystème. Il couvre par exemple les interfaces électroniques (API), l'authentification des utilisateurs (*identity and access management, IAM*) et la notification automatique des modifications (*publish subscribe*).

Le pilier API Gateways comprend l'infrastructure nécessaire à l'échange de données entre les participants à l'écosystème.

Les passerelles API servent de plaques tournantes pour l'orchestration des API des différents participants qui sont présents dans l'écosystème. Ces passerelles peuvent être exploitées de manière centralisée ou décentralisée et servent de point unique d'information sur les interfaces existantes d'un secteur de données déterminé. Elles mettent à la disposition des participants des informations sur l'accès, l'utilisation et la gestion du cycle de vie des API correspondantes. Avec ces informations, les participants peuvent utiliser les API en tenant compte des conditions existantes.

Une interface est également une porte d'entrée pour les codes malveillants. La sécurisation dès la conception joue donc un rôle central. Une gestion réfléchie des identités et des accès, comprenant une authentification mutuelle et des autorisations sélectives, est nécessaire pour promouvoir un écosystème numérique sûr, sans devoir s'appuyer sur une autorité centrale. L'écosystème DataHub utilise les groupes d'identités existants (identités fédérées).

⁶ Cela correspond aux principes FAIR; voir par exemple https://en.wikipedia.org/wiki/FAIR_data.

La confiance dans l'identité d'un participant et, donc, dans l'exactitude des données, est obtenue par une authentification mutuelle entre le fournisseur et le consommateur de données, qui peuvent ainsi s'assurer de l'identité de leur interlocuteur. Étant donné qu'un participant doit pouvoir faire confiance aux autres participants, la transparence concernant les aspects de sécurité et une gestion rigoureuse du cycle de vie des identités sont nécessaires, cette dernière garantissant la validation, la gestion et l'invalidation correctes des identités. Le DataHub n'introduira pas de mécanisme central de contrôle des accès. Dans ce domaine, la responsabilité incombe toujours aux fournisseurs de données et de prestations. Les fournisseurs et les consommateurs ont le même accès non discriminatoire à l'écosystème DataHub.

Le pilier API Gateways aide à repérer et à utiliser les différentes interfaces des participants. Les passerelles API proposent une multitude d'adaptateurs pour permettre l'intégration de différents systèmes, indépendamment des protocoles techniques ou spécifiques à l'application. Lors de la transmission des données, elles assurent une éventuelle conversion dans un format interopérable. Elles permettent de retracer l'utilisation des API. Les fournisseurs d'API peuvent ainsi réaliser des analyses d'utilité.

Pour les participants à l'écosystème, la fonctionnalité «publier ou s'abonner» (*publish subscribe*) du pilier API Gateways apporte une plus-value. Les abonnés (*subscribers*) peuvent souscrire à des mises à jour auprès d'un fournisseur (*publisher*) et reçoivent un message avec les informations correspondantes en cas de modification des données. Ils peuvent ainsi réagir rapidement aux modifications des données.

5.5.3 Services connexes: services permettant de développer des prestations basées sur les données

Les services connexes (*associated services*) représentent les services nécessaires pour que les participants à l'écosystème puissent fournir leurs propres prestations basées sur des données. Cela comprend par exemple l'utilisation de l'espace de stockage, des données de base, des prestations basées sur des données liées ou le calcul d'une valeur de hachage pour certaines données.

Le pilier Associated Services offre aux participants à l'écosystème les services d'infrastructure et de soutien nécessaires à la production de données et de prestations étatiques basées sur les données. Initialement, ce pilier comprend les principaux services techniques, tels ceux qui gèrent l'espace de stockage, les données de base et les données liées. En fonction des besoins des utilisateurs de chaque secteur de données, ces services peuvent prendre différentes formes. Ils peuvent être fournis à des utilisateurs particuliers à partir d'une infrastructure locale ou être mis à disposition de l'ensemble des utilisateurs dans le cadre d'un service centralisé.

Les données de base et les registres sont localisés au niveau de ce pilier. En tant que services de base pour les participants à l'écosystème, les services connexes apportent une contribution essentielle à la mise en œuvre du principe de la saisie unique des données. D'autres services, comme la mise à disposition d'une puissance de calcul ou de services analytiques, seraient encore envisageables.

5.5.4 Politiques et standards: conditions-cadres pour une utilisation optimale DataHub

Les politiques et les standards forment le cadre de toutes les règles et directives communes que les participants à l'écosystème doivent respecter pour pouvoir partager leurs données de manière optimale.

Le respect de règles et standards définis en commun est décisif pour la mise en place d'un écosystème fonctionnel, dans lequel les données doivent pouvoir circuler librement entre les différents participants à l'écosystème. Le pilier Standards vise à définir des règles pratiques applicables aux données à échanger, à l'échange de données lui-même et à la sécurité des données. Les participants à l'écosystème de chaque secteur de données se mettent d'accord sur une compréhension commune des données (sémantique), ce qui constitue la condition préalable à

une utilisation commune des données. Dans la mesure du possible, il est fait usage des standards internationales reconnues et des meilleures pratiques, afin d'assurer notamment la connexion internationale des données.

Dans l'écosystème DataHub, le respect des règles est un élément central de la validation et de l'assurance qualité des prestations. Toutes les données et toutes les prestations basées sur les données proposées dans l'écosystème doivent être conformes aux directives. Le non-respect des directives peut entraîner l'exclusion du DataHub.

Ces règles sont, dans la mesure du possible, définies de manière transversale pour l'ensemble de l'écosystème. Elles sont notamment systématiquement reprises pour les piliers Directories et API Gateways, tandis que d'autres sont adaptées à un secteur de données spécifique. Le degré de maturité et l'étendue des politiques et des standards varieront en fonction des secteurs de données considérés.

5.6 Principes de base

L'écosystème de hub de données repose sur quelques principes qui revêtent une importance cruciale pour la participation à l'écosystème et pour le fonctionnement de l'échange de données. Ces principes répondent aux objectifs (voir chapitre 3) et aux exigences (voir chapitre 4): souveraineté en matière de données, sécurité des données, connectivité, utilisation des solutions existantes, durabilité et rentabilité.

6 Autres initiatives et projets pertinents

Les projets brièvement décrits dans le présent chapitre montrent l'importance du thème de l'échange de données. On a sélectionné de manière ciblée des initiatives dans le contexte de l'écosystème DataHub. De ce fait, le présent chapitre ne prétend pas à l'exhaustivité. La figure suivante montre où se situent approximativement les différents projets par rapport au concept de plateforme d'échange de données (cf. figure 2).



Figure 3: Situation des projets pertinents dans le contexte du DataHub. Les composants mis en exergue en bleu sont les éléments du DataHub qui importent pour le projet considéré ou qui sont déjà concernés (partiellement) par le projet.

6.1 Gaia-X⁷

Gaia-X est un projet de développement d'une infrastructure de données fédérée, sûre et centrée sur l'Europe, qui est promu par plus de 500 représentants de l'administration, de la science et de l'économie. Ce projet a été créé en 2020 par 22 organisations allemandes et françaises, dont le ministère allemand de l'économie et de l'énergie, le ministère français de l'économie et des finances, BMW, Bosch, Deutsche Telekom, Orange, SIEMENS, Safran et SAP. Depuis lors, les *hyperscalers* Microsoft, Amazon, Google et Alibaba les ont rejointes. L'écosystème Gaia-X est organisé en «hubs» nationaux (un hub par pays). L'Allemagne, la France, l'Italie et l'Autriche, notamment, disposent d'une telle plateforme. Mais la Suisse n'en a pas encore.

⁷ Source: voir <https://www.gaia-x.eu/>

Gaia-X a pour but de créer un écosystème dans lequel les données sont récoltées, mises à disposition et partagées dans un environnement fiable. Dans ce cadre, les participants conservent en tout temps la souveraineté sur leurs données. Il ne s'agit donc pas d'un cloud informatique, mais d'un système fédéré qui relie de nombreux prestataires et utilisateurs de services cloud.

Gaia-X consiste en quatre éléments clés: les *federation services* (services de fédération) définissent les exigences et services techniques minimaux nécessaires à l'exploitation de l'écosystème fédéré. Les *data spaces* (espaces de données) traduisent un concept d'intégration des données sans stockage centralisé: les données demeurent à leur source et ne sont utilisées en commun qu'au besoin. Un espace de données est constitué par l'ensemble de ses participants, qui sont tous soumis aux mêmes règles. À cet égard, citons en particulier l'International Data Spaces Association (IDSA)⁸, qui participe activement à la conception d'espaces de données internationaux en mettant l'accent sur la souveraineté en matière de données. L'IDSA vise à promouvoir des standards et des interfaces globales pour les espaces de données internationaux ainsi que pour les technologies et modèles d'affaires correspondants. Les standards décrivent une architecture de standards, directives et codes de conduite harmonisés concernant la fourniture et l'utilisation de données. Les services représentent des règles et schémas communs de coopération sur la base desquels des services peuvent être développés et proposés.

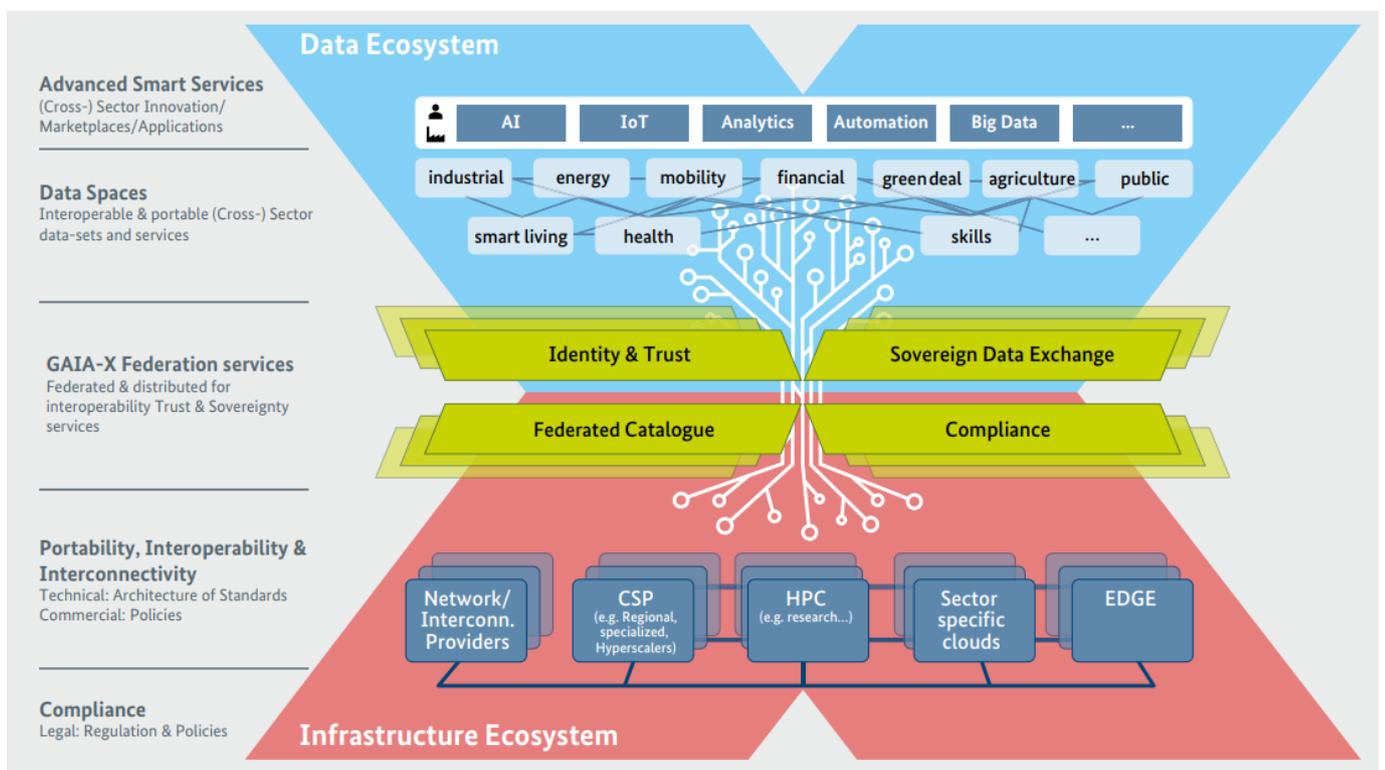


Figure 4: Vue d'ensemble de Gaia-X

On peut comparer la structure de Gaia-X avec les concepts du DataHub:

⁸ <https://internationaldataspaces.org>

Tableau 1: Comparaison des éléments de Gaia-X et DataHub

Élément de Gaia-X	Élément DataHub
Advanced Smart Services	Aucun élément correspondant
Data Spaces	Secteurs de données
Gaia-X Federation services: Identity & Trust, Sovereign Data Exchange, Federated Catalogue	Directories, API Gateways, gouvernance & architecture
Portability, Interoperability, Interconnectivity	API Gateways, Standards
Compliance	Standards, gouvernance & architecture

6.2 sedex

sedex («secure data exchange») est un service de l'OFS visant à garantir l'échange sécurisé de données. Cette plateforme intègre les applications métiers dans le domaine de la cyberadministration suisse

Actuellement, plus de 8000 unités organisationnelles (offices fédéraux, cantons, communes, associations, etc.) utilisent sedex dans plus de 90 domaines organisés thématiquement. En 2021, plus de 20 millions de messages ont été échangés de manière sûre par sedex.

La plateforme sedex est un intergiciel orienté message hautement disponible (24/7) qui est destiné à l'échange asynchrone sécurisé de données. Elle permet aussi les échanges de données synchrones (par ex. connexion du service UPI de la Centrale de compensation et autres services Internet RESTful et SOAP)

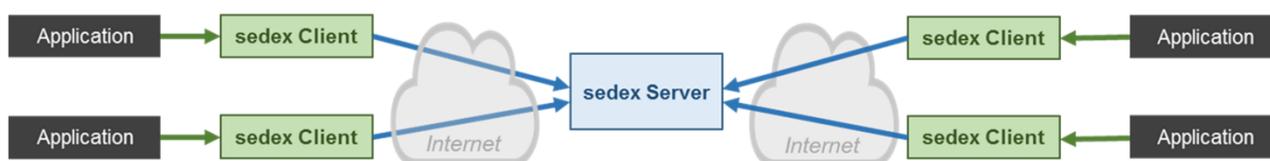


Figure 5: Vue d'ensemble de sedex

Comme des données sensibles sont échangées, la plateforme satisfait à de hautes exigences en termes de sécurité et de traçabilité:

- Le contenu des messages transmis par sedex n'est visible que par l'expéditeur et le destinataire. Ni le système sedex ni l'équipe d'assistance n'ont accès aux contenus, car l'identification des participants est assurée par une enveloppe normalisée par eCH et séparée du message. Pour le cryptage, sedex recourt aux procédures y relatives et aux certificats de sécurité de la Swiss Government PKI.
- Le système sedex génère pour chaque message une confirmation grâce à laquelle tant l'expéditeur que le destinataire peuvent suivre précisément l'envoi et la remise d'un message.

Le client sedex est exploité localement par les participants à sedex. En tant que plateforme d'échange de messages, il constitue l'interface entre une application métier et le serveur. L'Office fédéral de l'informatique et de la télécommunication développe continuellement les serveurs centraux, il les exploite de manière à en assurer la haute disponibilité et il les surveille en permanence. Application basée sur Java, le client peut être installé sur divers systèmes d'exploitation (y c. Windows, macOS, Linux, AIX, etc.) à l'aide d'un programme d'installation de logiciels. sedex met une image de conteneur officielle à disposition sur Docker Hub pour l'exploitation du client sedex sur une plateforme de conteneur telle que Docker, Kubernetes ou OpenShift.

Avec cette offre de services, sedex peut être considéré comme la colonne vertébrale actuelle de la transmission numérique de données de la cyberadministration suisse. En réponse aux exigences du groupe croissant de clients, cette offre est continuellement développée.

6.3 NaDB: plateforme d'interopérabilité

Le Conseil fédéral a mandaté l'OFS, dans le cadre du programme de gestion nationale des données (NaDB), pour qu'il simplifie et rende plus efficace la gestion des données des pouvoirs publics par l'utilisation multiple des données.

La mise en œuvre concrète du principe de la saisie unique des données est testée dans le cadre de quatre projets pilotes. Afin de promouvoir l'utilisation multiple des données sur le long terme, on utilise un catalogue de données qui donne des indications sur les données, leur qualité et leur lieu de stockage. Cette plateforme d'interopérabilité (I14Y-IOP) est en ligne depuis la mi-2021⁹, et elle est développée selon les directives du Conseil fédéral. En 2022, parallèlement à la poursuite de l'extension des jeux de données et nomenclatures documentés, un répertoire des interfaces électroniques (API) sera incorporé à l'I14Y-IOP.

Le service d'interopérabilité (IOS), rattaché à l'OFS, exploite la plateforme I14Y-IOP. Il accompagne et soutient l'harmonisation et la standardisation des données. Son assistance est à la disposition de toute l'administration, à commencer par l'OFS.

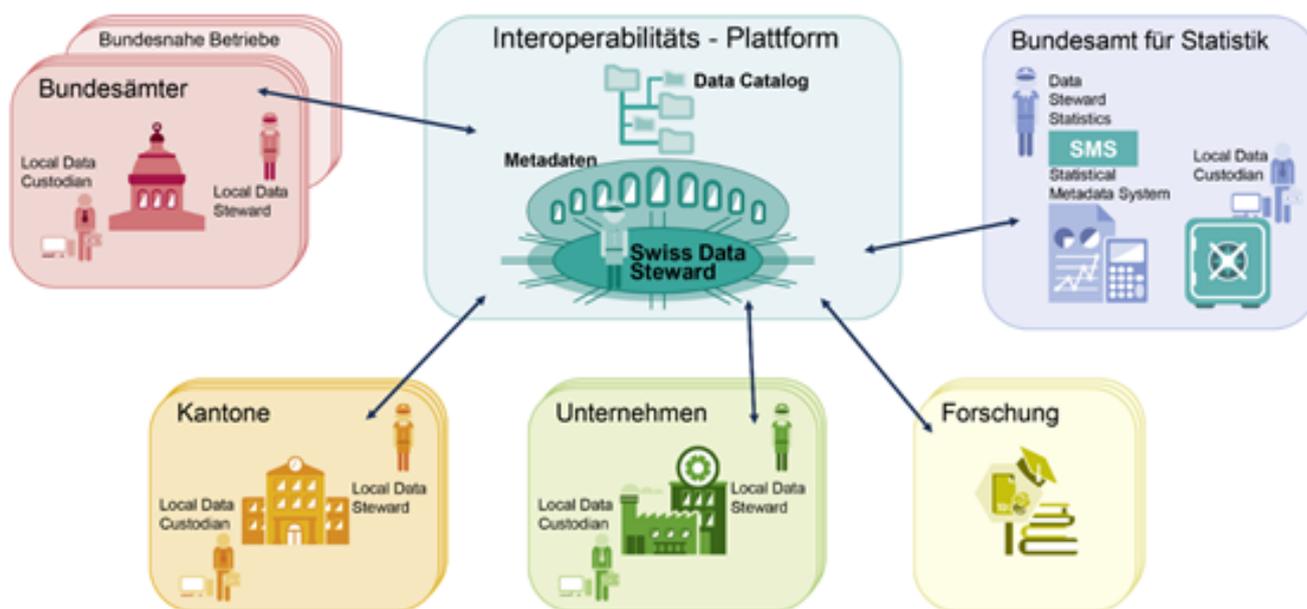


Figure 6: Vue d'ensemble de la plateforme d'interopérabilité et des organisations participantes

6.4 Mise en place d'une gestion transversale des données de base des autorités (OMO9)

Dans le cadre de l'objectif de mise en œuvre 9 (mettre en place une gestion transversale des données de base des autorités), il s'agit d'établir une vue d'ensemble des mesures juridiques, organisationnelles et techniques nécessaires pour permettre aux administrations cantonales et communales d'accéder aux données de base de la Confédération. En outre, il faut identifier et documenter dans quels registres de base ou services nationaux les données de base et les identifiants univoques correspondants doivent être gérés. Les registres de base et systèmes de

⁹ <http://www.i14y.admin.ch>

références suivants revêtent un intérêt particulier: le registre des entreprises et des établissements (REE); le registre d'identification des entreprises (IDE); le registre des bâtiments et des logements (RegBL); l'infrastructure fédérale des géodonnées (IFDG); et le service national des adresses.

6.5 LINDAS¹⁰

Les Archives fédérales suisses (AFS) mettent à disposition LINDAS (Linked Data Service). Ce service comprend un instrument de conversion et d'intégration des données ainsi qu'une infrastructure de stockage des données. Il permet aux autorités suisses et à d'autres organisations mandatées par l'État de mettre leurs données à disposition sous forme de données liées (*linked data*). Un catalogue de métadonnées fournit la vue d'ensemble nécessaire sur les données disponibles en recourant aux standards établis. On peut trouver les données par une requête de recherche et les appeler au moyen d'une API.

Actuellement, LINDAS contient notamment des données fournies par les CFF, l'Office de la statistique de la ville de Zurich, le service OGD, les AFS, l'Office fédéral de l'énergie (OFEN), la Commission fédérale de l'électricité, l'Office fédéral de la sécurité alimentaire et des affaires vétérinaires et le Secrétariat à l'économie.

6.6 OGD Suisse¹¹

L'OFS coordonne, en étroite collaboration avec le Département fédéral de l'intérieur, la mise en œuvre de la stratégie OGD Suisse 2019-2023. Cette stratégie vise à mettre à la disposition du public des données de l'administration sur le portail [opendata.swiss](https://www.opendata.swiss), en libre accès. Les données administratives en libre accès favorisent la transparence, la participation et l'innovation dans tous les domaines sociétaux. Pour soutenir et ancrer durablement ce processus, les données des services fédéraux sont publiées de plus en plus sur [opendata.swiss](https://www.opendata.swiss) dans des formats ouverts et lisibles par ordinateur (*open government data*). Les données existantes (par ex. les données des cantons, des communes, des villes ou des entreprises liées à la Confédération) sont elles aussi mises peu à peu à disposition. Font exception les données qui présentent un intérêt de protection légitime prépondérant (par ex. la protection des données).

On trouve des portails correspondants dans les administrations des trois niveaux de l'État fédéral, parmi lesquels les suivants:

- Données générales à l'échelle de la Suisse:
 - données publiques générales en libre accès: <https://www.opendata.swiss>
 - données de transport: <https://www.Opentransportdata.swiss>
- Géodonnées à l'échelle de la Suisse:
 - Confédération: <https://www.geo.admin.ch>
 - cantons: <https://www.geodienst.ch>
- Données générales cantonales:
 - <https://www.zh.ch/de/politik-staat/opendata.html?keyword=ogd#/home>
 - <https://ge.ch/sitg/donnees/demarche-open-data>
 - <https://daten.sg.ch/>
- Géodonnées cantonales:
 - <https://www.zh.ch/de/politik-staat/opendata/offene-geodaten.html#/home>

¹⁰ <https://lindas.admin.ch> (dernier accès le 26.01.2022)

¹¹ <https://www.bfs.admin.ch/bfs/fr/home/services/ogd.html>

- Données générales communales:
- <https://data.stadt-zuerich.ch>
- <https://daten.stadt.sg.ch/>
- Données internationales
 - UE: <http://data.europa.eu/euodp/fr/home>
 - Autriche: <https://www.data.gv.at/>
 - France: www.api.gouv.fr
 - Norvège: <https://www.ssb.no/en/api> et <https://data.norge.no/>

6.7 Infrastructure nationale de données géographiques

Afin de mettre en œuvre au sein de l'administration fédérale la stratégie suisse pour la géoinformation, approuvée par le Conseil fédéral en 2001, et de mettre à disposition les vastes ressources que représentent les géodonnées, il a été proposé en 2003 d'élaborer un système convivial, interconnecté et décentralisé constitué de composantes politiques, organisationnelles, financières, juridiques et techniques, soit une infrastructure nationale de données géographiques (INDG). Cette infrastructure doit garantir en tout temps à l'administration, à l'économie et à la société civile de toute la Suisse un accès simple, rapide et à faible coût à des géoinformations fiables. La loi sur la géoinformation entrée en vigueur en 2008 a agi comme un accélérateur et comme un guide pour l'élaboration d'une INDG. Elle a grandement contribué à établir la géoinformation au sein de l'administration publique.

La Confédération a développé une infrastructure fédérale de géodonnées qui enregistre chaque année plus de 28 millions d'utilisateurs. Cette infrastructure fournit trois pétaoctets de données couvrant plus de 800 ensembles de données issus de 35 services fédéraux, qu'elle met à la disposition des utilisateurs. Dans les cantons, des infrastructures cantonales ont été mises en service, et la Conférence des services cantonaux de la Géoinformation et du Cadastre a mis sur pied une infrastructure d'agrégation des cantons qui permet de regrouper des géodonnées de base et de les publier.

Il s'agit désormais d'améliorer l'interconnexion de ces composantes existantes avec l'INDG, de les développer, d'exploiter les synergies entre la Confédération et les cantons et de garantir durablement le financement.

Le développement de l'INDG est inscrit dans le plan d'action de la stratégie «Suisse numérique» de l'Office fédéral de la communication¹². Ce plan d'action prévoit la mise en place d'une organisation nationale pour le pilotage et la conception de l'INDG.

Le 11 décembre 2020, le Conseil fédéral a approuvé la stratégie suisse pour la géoinformation¹³, dont la vision et les champs d'action sont les suivants:

- **Vision**
Avec la vision «**Des géodonnées à la connaissance: des informations géolocalisées et interconnectées pour la Suisse**», la communauté de la géoinformation dispose d'une nouvelle vision de la direction dans laquelle elle doit se développer.
À cette fin, il est prévu de poursuivre la consolidation en cours de l'infrastructure nationale de données spatiales. Celle-ci devra devenir une plateforme de collaboration sur laquelle les données peuvent être conjointement produites et partagées. La plateforme permettra et accélérera la création de nouvelles connaissances sur notre espace de vie.
- **Sept champs d'action**
Les principales orientations des activités de la géoinformation pour les années à venir sont définies dans sept domaines d'action:

¹² <https://www.bakom.admin.ch/bakom/fr/page-daccueil/suisse-numerique-et-internet/strategie-suisse-numerique.html>

¹³ <https://www.geo.admin.ch/strategie>

- **promotion de l'écosystème de la géoinformation** résultant d'efforts conjoints coordonnés déployés par les autorités de tous les niveaux administratifs de même que les acteurs économiques, politiques, scientifiques et la société civile grâce à une coopération et une volonté active de dialogue;
- **mise à disposition de géodonnées interconnectables, appropriées et fiables** sous une forme utilisable et pérenne pour tous les utilisateurs;
- **facilitation des processus métiers** par l'intégration de la référence spatiale et une conception intégrée, une automatisation et une orientation vers les besoins de l'utilisateur;
- **développement et application de la science des géodonnées** en étendant l'expertise en matière d'analyse des géodonnées ainsi que le développement et l'utilisation d'algorithmes;
- **encouragement de l'innovation** en améliorant les conditions-cadres et les processus en vue de tester et de mettre en œuvre de nouvelles idées et approches ainsi que de nouveaux concepts;
- **acquisition et renforcement des compétences** des spécialistes en vue d'une meilleure maîtrise de la géoinformation, d'une gestion optimale du changement numérique ainsi que d'une diffusion des connaissances sur l'utilisation des géoinformations;
- **création de géoplateformes numériques performantes, ouvertes et partagées**, basées sur des géodonnées, des géoinformations et des géoservices normalisés et mis en réseau.

6.8 Infrastructure nationale de données mobilitaires¹⁴

L'infrastructure nationale de données mobilitaires (NaDIM) a pour but de simplifier l'échange de données mobilitaires entre les entreprises et les prestataires de mobilité et de services. L'accès aux transports publics doit gagner en attractivité pour les nouveaux clients en facilitant les offres supramodales.

À l'avenir, un seul clic sur Internet ou dans une application mobile devra permettre de composer et d'acheter directement des offres sur mesure incluant plusieurs moyens de transport. Ainsi, les offres de transport à disposition pourront être utilisées et exploitées de manière plus ciblée. Le système de transport global deviendra de ce fait plus efficace et plus durable. La NaDIM doit répondre aux principales conditions préalables d'un accès simple et sûr aux données.

Par sa conception, la NaDIM doit être indépendante, fiable, ouverte, non discriminatoire, transparente, sans but lucratif, de haute qualité et techniquement flexible. Son utilisation sera facultative. Les données resteront chez leurs propriétaires, qui en assureront la gestion. Le déploiement sera progressif. Grâce à la NaDIM, les développeurs d'applications et les exploitants de plateformes devront être en mesure de proposer des offres interconnectées à leurs clients tout en rendant celles-ci visibles aux prestataires de mobilité qui pourront les réserver. Actuellement, dans de nombreux domaines, les données de base sur les réseaux de transport, les moyens de transport et les systèmes de distribution ne sont pas encore accessibles, et la connexion entre les développeurs d'applications, les prestataires de mobilité et les autres propriétaires de données n'est pas généralisée, ce qui génère un coût important pour tous les acteurs concernés. La NaDIM doit remédier à ce problème.

¹⁴ <https://www.bav.admin.ch/bav/fr/home/publications/communiqués-de-presse.sg-id-79690.html>

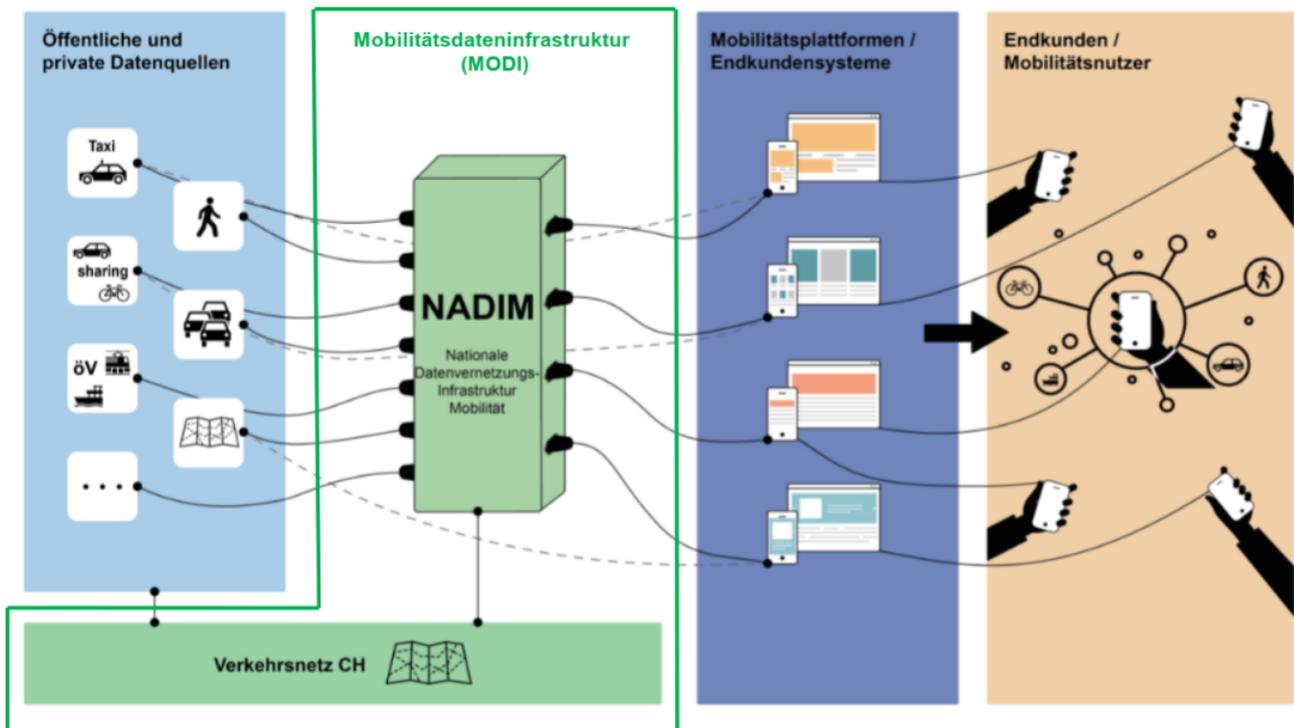


Figure 7: Infrastructure de données sur la mobilité (MODI)

6.9 Hub de données du marché de l'énergie (OFEN)

Dans le cadre de l'étude «Datahub Schweiz. Kern zukünftiger Dateninfrastruktur digitalisierter Strom- und Gasmärkte»¹⁵, l'OFEN a examiné des principes d'amélioration de l'échange de données sur le marché de l'énergie.

La numérisation, qui progresse très vite simultanément, permet de répondre à la complexité croissante du système énergétique, de réaliser des potentiels d'optimisation, d'accroître l'efficacité et de soutenir l'innovation. Pour exploiter ces possibilités qu'offre la numérisation, il est fondamental d'améliorer sensiblement l'accès et l'utilisation des données énergétiques. L'actuel système d'échange de données du marché de l'énergie se caractérisant par une forte décentralisation, il y a lieu de se demander dans quelle mesure il peut encore répondre à la complexité naissante et aux nouvelles exigences. Actuellement, l'échange de données, c'est-à-dire la communication sur le marché suisse, se caractérise par un modèle fortement bilatéral où chaque partie au marché développe et entretient elle-même ses interfaces avec les autres acteurs. Sans automatisation et sans numérisation des processus et des échanges de données, il faut s'attendre à une forte augmentation des charges liées au temps consacré par les acteurs au traitement manuel.

Il faut donc concevoir la future infrastructure de données de manière que l'automatisation des processus soit simple et que la connexion des nouveaux acteurs du marché soit nettement simplifiée. Il faut standardiser les interfaces et en réduire le nombre.

L'étude susmentionnée a débouché sur la recommandation de développer un hub de données national pour les marchés du gaz et de l'électricité. Celui-ci n'aura pas pour vocation de stocker des données de mesure. Il permettra le routage des données de comptage, la mise à disposition des données de base et des registres de données, ainsi que la fourniture de fonctions à valeur ajoutée comme l'agrégation des données de comptage. La mise en œuvre est peu risquée et

¹⁵ <https://www.bfe.admin.ch/bfe/fr/home/approvisionnement/numerisation.html>

pourra servir d'étape préliminaire à un hub de données complet qui stocke des données de mesure ou qui repose sur la technologie des registres distribués (TRD) / chaînes de blocs (*blockchain*).

Le hub de données assurera les fonctions suivantes:

- gestion centralisée des accès aux données disponibles de manière centralisée et décentralisée;
- routage des données disponibles de manière décentralisée;
- assurance de la qualité et validation des processus;
- création de valeur par agrégation.

Il s'agira de standardiser fortement les interfaces DataHub light, les processus et le format des données à échanger. En ce qui concerne les interfaces, il faudra surtout miser sur la technologie du futur que représentent les API.

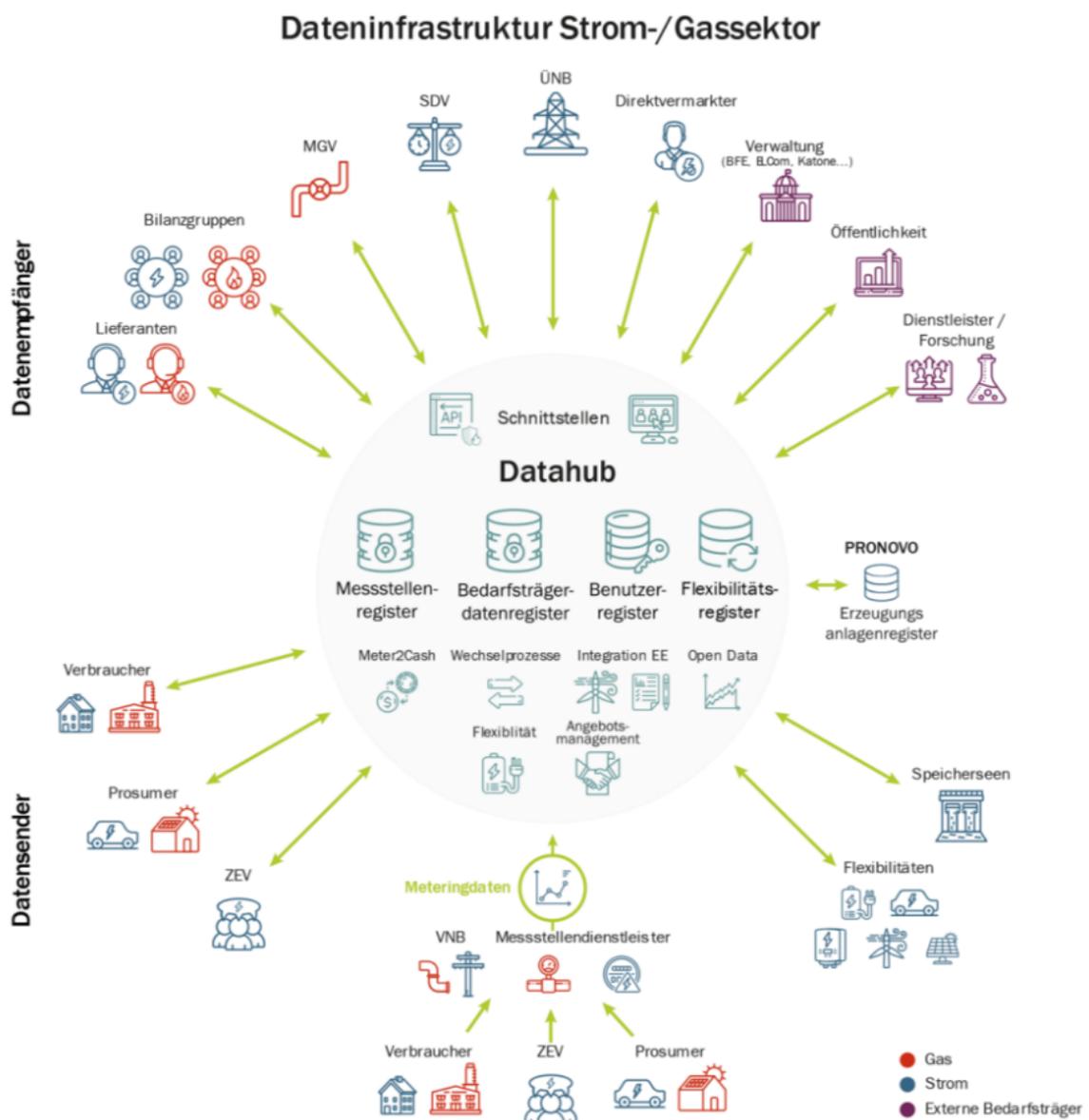


Figure 8: Conception d'une infrastructure de données pour les marchés du gaz et de l'électricité, centrée sur un hub de données

6.10 Trusted Datahub (La Poste)

La Poste développe une plateforme peer-to-peer, le Trusted DataHub (TDH), capable de confirmer la livraison de données électroniques entre ses clients de manière à satisfaire aux exigences des tribunaux. Dans ce cadre, elle ne connaît pas le contenu des données communiquées et garantit donc la souveraineté en matière de données. Le TDH est une plateforme qui, avec ses capacités clés, est un vecteur d'écosystèmes.

- Divers acteurs peuvent échanger et enregistrer des conventions et des informations dans un environnement sûr, cohérent et fiable (permettant la confiance multipartite).
- Les participants peuvent rechercher efficacement dans le réseau les informations et services déjà existants dont ils ont besoin pour leurs processus et services, ou les mettre à la disposition des autres participants (mise en relation des publics et efficacité de la réutilisation des informations).
- Le stockage des données appartient à chaque acteur, qui en assure lui-même le contrôle en toute autonomie. La Poste intervient comme exploitant unique, neutre et fiable du réseau.
- S'ils constatent des modifications pertinentes des informations, les participants au réseau peuvent en informer les autres participants autorisés et améliorer ainsi significativement la qualité de leurs processus (pertinence du signalement des changements).

6.11 Swissdec¹⁶

L'association Swissdec est un projet commun à but non lucratif regroupant plusieurs partenaires indépendants dont l'activité porte sur l'échange électronique de données entre les entreprises, les assureurs et les administrations publiques. Dans ce contexte, Swissdec assure pour l'essentiel les trois tâches suivantes: standardisation, certification des systèmes connectés et exploitation de l'infrastructure d'échange de données (distributeur Swissdec).

L'association Swissdec a été créée pour permettre la fourniture uniforme et sans rupture médiatique de données salariales. S'agissant de données salariales, Swissdec met à disposition la norme suisse en matière de salaire (ELM). Les entreprises qui recourent à un système ERP certifié par Swissdec (plus de 100 systèmes ERP sont certifiés) peuvent transmettre leurs communications de salaire sans rupture médiatique aux divers services demandeurs¹⁷. Cette transmission s'effectue dans le cadre d'une seule communication que le distributeur Swissdec répartit de telle manière que les destinataires des données salariales ne reçoivent que les données pertinentes pour eux. Les données salariales ne sont pas sauvegardées dans le distributeur Swissdec.

Outre la norme suisse en matière de salaire, Swissdec fournit aussi la norme suisse en matière de prestations (KLE), qui s'adresse à toutes les entreprises qui souhaitent traiter les cas d'absence suite à une maladie ou à un accident par voie électronique, directement à partir de leur programme de comptabilité salariale.

Les parties «standardisation» et «transmission» (distributeur) de l'écosystème de Swissdec se situent dans les piliers Standards et API Gateways.

6.12 Standards eCH

L'association eCH promeut, développe et adopte des standards dans le domaine de la cyberadministration à l'appui d'une collaboration électronique efficace entre les autorités, les entreprises et les particuliers. Les membres d'eCH sont la Confédération, tous les cantons,

¹⁶ <https://www.swissdec.ch/fr/>

¹⁷ Les destinataires des données sont les caisses de compensation AVS, les assureurs LPP, les administrations fiscales (certificat de salaire, impôt à la source), l'OFS (enquête suisse sur la structure des salaires), ainsi que les assureurs accidents et indemnités en cas de maladie.

diverses communes, quelque 120 entreprises, ainsi que diverses hautes écoles, associations et personnes privées. L'organisation se compose d'un comité directeur, d'un comité d'experts et d'une administration de deux personnes. Vingt groupes spécialisés viennent s'y ajouter. Les membres de l'association s'engagent à titre bénévole.

Les standards sont essentielles pour coordonner la coopération entre les différents acteurs en Suisse, comme eCH l'explique sur son site Internet:

«Une coopération est nécessaire pour qu'un processus fonctionne en continu par-dessus les limites administratives. Et une standardisation est nécessaire pour que les systèmes des diverses unités administratives puissent coopérer les uns avec les autres dans un processus constant. Dans un État fédéral comme la Suisse, ces deux composants sont indispensables à l'administration numérique.»

Les travaux d'eCH produisent notamment des standards, des meilleures pratiques et des instruments applicables au domaine de la cyberadministration, notamment au domaine de l'échange de données pour les télécommunications, les applications policières, etc. Ces produits sont accessibles au public.

Les standards approuvés d'eCH ont le statut de recommandations. La Confédération a repris dans son processus de standardisation les standards eCH concernant l'administration fédérale. Par la convention-cadre de droit public concernant la collaboration en matière de cyberadministration en Suisse 2016-2019, la Confédération, les cantons et les communes se sont en outre engagés à déclarer contraignantes en règle générale les standards de l'Association eCH, en particulier pour les acquisitions et le développement de solutions.

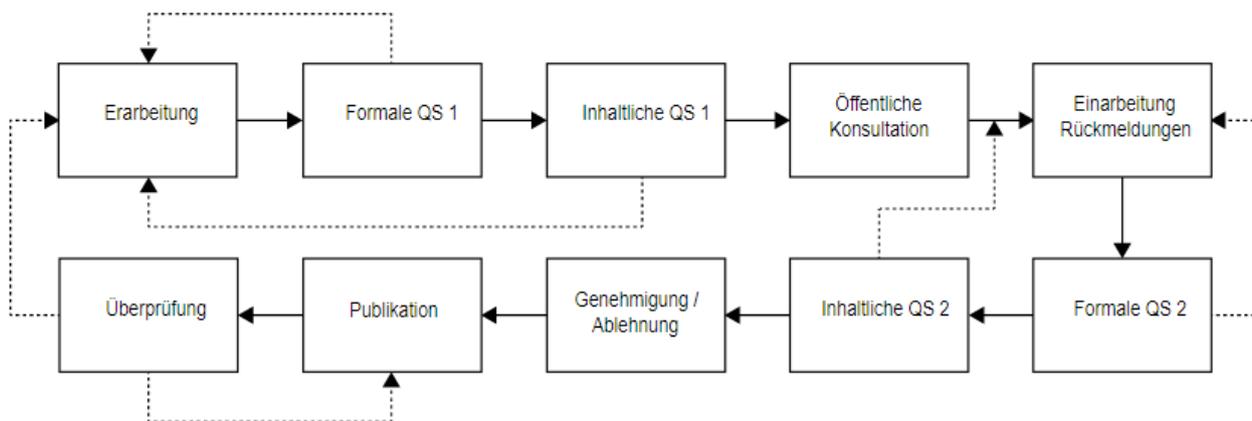


Figure 9: Processus de standardisation d'eCH selon eCH-0003 (Guide pour l'approbation des propositions)

En résumé, on peut dire que l'association eCH a fourni un précieux travail de pionnier en vue d'une concertation et d'une coopération plus étroites entre les trois niveaux de l'État. Grâce à l'implication de nombreux experts, les standards eCH sont de haute qualité et sont reconnus. On pourrait encore accroître leurs effets et leur utilité. Elles ne sont contraignantes que pour l'administration fédérale.

L'existence de standard ne signifie pas que les thèmes / sont complètement standardisés, mais seulement que certains de leurs aspects le sont. Dans certains cas, les domaines sont couverts par des standards d'application générale. S'agissant des domaines non soumis à la standardisation de l'eCH ou d'autres organismes, il en résulte que les données personnelles sauvegardées dans les systèmes des diverses administrations ne sont en général ni standardisées

ni harmonisées¹⁸. Selon la variante, ce volume potentiellement important de données non standardisées entrave l'éventuelle mise en œuvre d'une solution de traçabilité.

6.13 eDelivery¹⁹

Tout domaine politique de l'UE qui nécessite l'échange sûr, fiable, transfrontalier et intersectoriel (structuré ou non) de documents et de données peut recourir aux spécifications techniques prescrites par eDelivery. Développé par des organisations internationales comme l'Organization for the Advancement of Structured Information Standards et l'Institut européen de normalisation des télécommunications, le réseau eDelivery repose sur les principes juridiques et organisationnels eIDAS. La gestion technique d'eDelivery est assurée par la Direction générale de l'informatique de la Commission européenne. La Direction générale des réseaux de communication, du contenu et des technologies a compétence pour mettre en œuvre la politique de l'UE concernant eDelivery. eDelivery est utilisé dans le cadre de diverses grandes initiatives transfrontalières, par exemple PEPPOL, SPOCS, e-CODEX et e-SENS.

eDelivery est un module qui fournit des spécifications et standards techniques, des logiciels installables et des services supplémentaires en vue de permettre la réalisation de projets visant à créer un réseau de nœuds pour l'échange sûr de données numériques. Grâce au développement que permet eDelivery, des organisations publiques et privées actives dans différents domaines peuvent établir sans problème un canal sûr et interopérable pour se transmettre des documents et des données par un réseau public ou privé.

eDelivery, qui permet l'échange de documents et de données entre systèmes informatiques hétérogènes grâce à un protocole standardisé, crée ainsi la base d'une interopérabilité intersectorielle et interprojets. Cette solution garantit l'intégrité et la confidentialité des données dans le cadre de toute transmission de données au moyen de la signature numérique et du cryptage.

6.14 X-Road²⁰

X-Road est un système développé par le gouvernement estonien en coopération avec des chercheurs, programmeurs et entreprises de l'Estonie. Il comprend des concepts, des dispositions juridiques, des prescriptions de procédure, des standards et règles techniques, divers logiciels, de multiples banques de données décentralisées et des serveurs de sécurité dédiés. Il sert à échanger facilement, rapidement et sûrement par Internet des données et des services décentralisés en étant protégé contre les accès et les modifications de données non autorisés. Le système assure une forte protection non seulement aux échanges de données, mais aussi aux données elles-mêmes contre les manipulations et les suppressions non autorisées.

X-Road Data Exchange Layer est une couche d'échange de données standardisée, cohérente, collaborative, interopérable et sûre. Elle offre aux prestataires une nouvelle possibilité de proposer des services aux citoyens, aux entreprises et aux autorités. Elle permet de créer, simplement et efficacement en termes de coûts, des offres combinant différents services et sources de données.

X-Road Data Exchange Layer permet:

- d'améliorer la qualité des services et produits existants;

¹⁸ Les termes «standardisation» et «harmonisation» sont utilisés selon la définition que l'OFS en donne dans le cadre de la NaDB et de l'IOP:

Standardisation: définition d'une norme généralement valable.

Harmonisation: regroupement des données existantes sous une forme répondant à la norme, de manière que leur signification soit uniforme.

¹⁹ <https://ec.europa.eu/cefdigital/wiki/display/CEFDIGITAL/eDelivery>

²⁰ <https://github.com/nordic-institute/X-Road>

- d'innover en fournissant de nouveaux types de services;
- d'économiser sur les infrastructures, l'archivage et les autres coûts;
- de fournir une sécurité des données standardisée et de protéger la sphère privée;
- d'assurer une mise en œuvre simple et d'accéder aux données via des interfaces (une fois la connexion établie, tous les services compris sont disponibles).

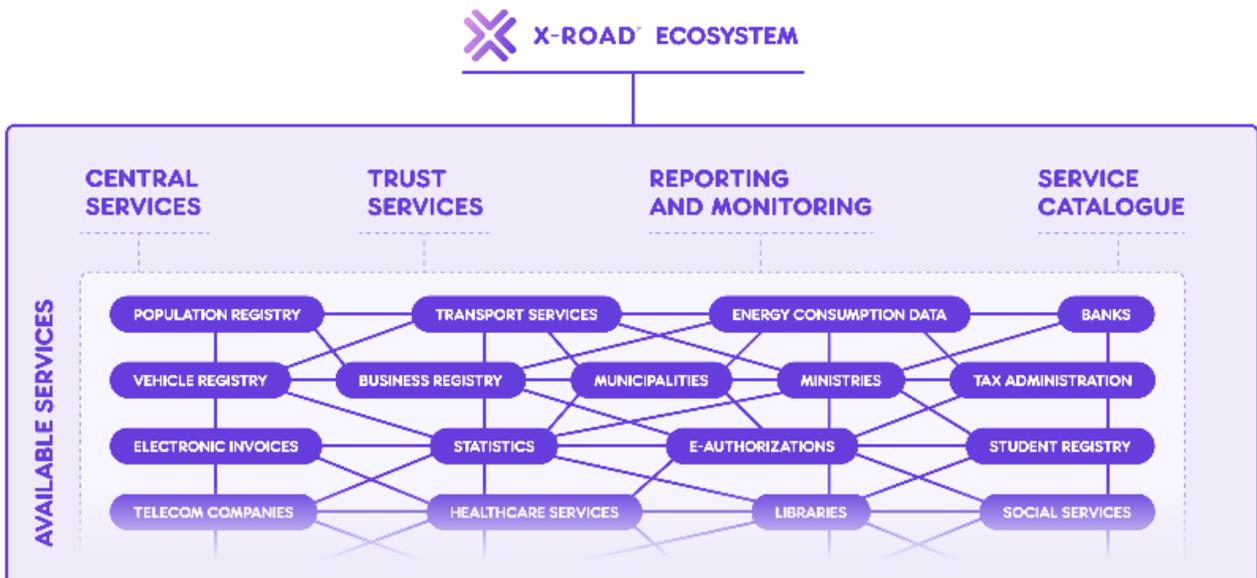


Figure 10: Vue d'ensemble de l'écosystème X-Road

7 Mise en perspective des projets et initiatives

Le présent chapitre examine les projets et initiatives évoqués dans le chapitre précédent au regard du contexte de l'écosystème de hub de données.

7.1 Gaia-X

Le projet international Gaia-X s'inscrit dans un contexte très semblable à celui du DataHub. Gaia-X est un écosystème réunissant des participants très divers qui se sont également engagés à échanger leurs données en minimisant les entraves. Gaia-X suppose notamment que le participant à un écosystème souhaite faire une utilisation commune des données avec d'autres participants sans toutefois renoncer à la souveraineté sur ses propres données.

Outre la similitude des contextes, certains objectifs, concepts et idées du DataHub concordent avec ceux de Gaia-X. Gaia-X a élaboré et publié diverses spécifications concernant des thèmes tels que l'architecture, les règles de politique, la gestion des standards, les fonctionnalités essentielles d'un catalogue, l'authentification, la gestion du cycle de vie par les participants ou la gestion des API²¹. Ces résultats conceptuels sont déjà très avancés, réfléchis, concrets et applicables. Les spécifications de Gaia-X accessibles au public (*open by default*) pourront servir de point de départ pour les nécessaires travaux de conception du DataHub. On pourra reprendre certains résultats (il ne faut pas réinventer la roue) et mettre à profit les initiatives dont la mise en œuvre a été un succès.

Certes, le projet est très jeune, mais on trouve parmi les participants à l'écosystème (outre des services étatiques) des poids lourds comme les hyperscalers Google, Amazon, Microsoft et Alibaba, de même que de grands groupes comme BMW et SAP. Cette situation porte à supposer, d'une part, que des acteurs importants font déjà partie de l'écosystème Gaia-X et, d'autre part, que la mise en œuvre des idées s'engage.

Catena-X, un écosystème concret de Gaia-X, poursuit la vision d'un échange de données de bout en bout de la chaîne de production de valeur automobile pour tous les participants. À cet effet, ce projet vise trois objectifs principaux: mettre à disposition des cas d'application initiaux pour les actuels thèmes prioritaires de l'industrie automobile; mettre à disposition la technologie et les réseaux; et assurer la mise à l'échelle en connectant les utilisateurs et les fournisseurs (accent mis sur les PME). Catena-X propose les premiers services opérationnels fondés sur Gaia-X.

Gaia-X est organisé en hubs de données nationaux. Contrairement aux pays voisins, la Suisse ne dispose pas encore d'une telle plateforme. Si les circonstances le permettent, l'écosystème de hub de données pourrait se fondre dans une plateforme Gaia-X suisse. Dans le cadre de la clarification des besoins d'un «Swiss cloud» (Département fédéral des finances), on a également identifié le besoin d'examiner la participation de la Suisse au projet Gaia-X. À l'heure où le présent document d'information est en voie d'élaboration, le Département des affaires étrangères examine diverses possibilités.

7.2 IOP NaDB, sedex, OGD Suisse, LINDAS, mise en place d'une gestion transversale des données de base des autorités, standards eCH

Ces projets sont propres à s'inscrire dans un futur écosystème de hub de données:

- Le catalogue de métadonnées de l'IOP (I14Y) pourrait servir de répertoire pour le hub de données.
- sedex permet d'ores et déjà l'échange sûr de données entre les services administratifs et constitue de ce fait un élément du pilier API Gateways. En outre, sedex offre une liste des participants à sedex et un transport asynchrone de données qui est sûr, traçable et fiable.

²¹ Les spécifications sont disponibles sous le lien suivant: <https://www.gxfs.de/federation-services/overview-specification-documents/>.

- OGD Suisse, avec opendata.swiss, sert de portail vers les données publiques en libre accès. OGD Suisse pourrait donc fournir des modules pour les piliers API Gateways et Associated Services.
- Le service LINDAS propose une solution qui couvre les quatre piliers. Il offre un catalogue de métadonnées (Directories) et une infrastructure de stockage de données (Associated Services) tout en indiquant les directives à suivre lors de la publication (Standards). LINDAS repose sur des standards et des meilleures pratiques établies aux niveaux national et international. La publication et l'obtention de données relèvent du pilier API Gateways, puisque des API sont déjà disponibles. Ainsi, LINDAS propose des éléments de base susceptibles d'être réutilisés dans le contexte DataHub.
- Les données de base et les registres jouent un rôle essentiel dans le contexte de la transformation numérique des pouvoirs publics. Les données de base définies au niveau fédéral doivent aussi être à la disposition des administrations publiques cantonales et communales. Les données de base communes seront pertinentes pour tous les participants à l'écosystème, tandis que les données de base spécifiques aux secteurs seront surtout utiles dans les secteurs de données et que les données de base locales ne seront appliquées que dans des cercles d'utilisateurs restreints. Les données de base et les registres influent sur l'écosystème dans son ensemble. Les données sont affichées dans des répertoires, elles peuvent être obtenues par des passerelles API et elles suivent des standards définies. La cyberadministration suisse, par son objectif de mise en œuvre 9, et l'écosystème de hub de données ont en commun de viser une large mise à disposition des données.
- Les standards eCH sont d'ores et déjà une composante indispensable de l'échange de données administratives. Dans le futur écosystème DataHub, elles feront partie intégrante du pilier Standards.

7.3 INDG, NaDIM, hub de données du marché de l'énergie, Trusted Datahub, Swissdec

L'INDG, la NaDIM, le hub de données du marché de l'énergie, le Trusted Datahub de la Poste, l'association Swissdec et les organisations qui leur sont rattachées peuvent être considérés comme des participants à un écosystème dans les secteurs des données relatives à la mobilité, à l'énergie et aux salaires. Le DataHub peut bénéficier de ces initiatives: dans le futur écosystème, elles pourront donner naissance à des secteurs de données propres et contribueront ainsi finalement de manière déterminante à la conception de l'écosystème.

7.4 eDelivery, X-Road

eDelivery est un module individuel d'échange sûr de données par voie numérique. Pour que l'écosystème DataHub soit mis à l'échelle dans l'espace européen, il importe de garantir la connectivité à cet égard.

X-Road est un bon exemple de comment mettre en œuvre la cyberadministration avec succès de manière centralisée. La structure fédéraliste de la Suisse et le degré de maturité de l'infrastructure s'opposent toutefois à l'utilisation d'X-Road. X-Road repose notamment sur diverses composantes de solution à code source ouvert. Ces composantes pourraient éventuellement être utilisées et développées dans le cadre de l'élaboration DataHub.

8 Champs d'action

Sur la base des entretiens menés avec les acteurs concernés et compte tenu des explications qui précèdent, on peut dégager les champs d'action suivants.

8.1 Clarification de la gouvernance

Il faut créer le cadre organisationnel en vue de la mise en place progressive et du développement de l'écosystème DataHub. La gouvernance et la gestion de l'écosystème doivent être clarifiées préalablement à la mise en place de l'écosystème.

Dans ce champ d'action, il faut procéder aux clarifications suivantes:

- clarifier la question de savoir quels services administratifs doivent se voir conférer quelles tâches, compétences et responsabilités dans cet écosystème, en particulier quant au pilotage, à la gestion et au contrôle du respect des règles;
- clarifier quelle organisation mettra sur pied, exploitera et développera les éléments centraux du DataHub (par ex. la passerelle API);
- clarifier le financement du projet et identifier les sponsors;
- clarifier les bases juridiques;
- esquisser la communauté DataHub et sa trajectoire de croissance;
- clarifier la question de savoir si, par exemple, une plateforme Gaia-X doit être visée pour la Suisse.

8.2 Plan d'architecture

Pour mettre en œuvre l'écosystème DataHub, il faudra élaborer les concepts architecturaux de l'écosystème global sous la forme d'un plan d'architecture. L'objectif de ce document est de décrire l'architecture de haut niveau en mettant l'accent sur les concepts, l'indépendance technologique, la souveraineté en matière de données des participants à l'écosystème et des fédérations décentralisées plutôt que sur des solutions centralisées. Du point de vue architectural, il faut garantir la connectivité aux écosystèmes de données nationaux et internationaux et la cohérence avec l'architecture des API de la Confédération et la vision architecturale. Le fédéralisme de la Suisse et les exigences qu'il pose à l'architecture font partie du plan d'architecture.

Le plan d'architecture s'adresse à toutes les personnes et à tous les groupes d'acteurs intéressés par le DataHub, par exemple aux responsables informatiques et aux architectes des services administratifs ou des organisations qui s'intéressent aux bases conceptuelles du DataHub. Il doit servir de point de départ pour qui veut étudier les concepts de base et leurs interactions, de même que l'élaboration et la spécification de l'architecture du DataHub. Il peut aussi servir de base à la participation concrète à l'écosystème. Il doit permettre aux décisionnaires de l'administration publique de viser dès un stade précoce la *datahub compliance*.

Le plan d'architecture couvre les thèmes suivants:

- description des cas d'application pertinents, par exemple participation à un secteur de données existant de l'écosystème, description des données propres, publication des données propres dans le répertoire, localisation des données et des services, conclusion d'un accord entre le fournisseur et l'utilisateur de données;
- présentation détaillée des exigences sous plusieurs angles: administration, entreprises, organisations, particuliers, fournisseur et utilisateur de données;
- principes architecturaux;

- sécurité des données;
- définitions des rôles et de leurs interdépendances, par exemple: fournisseur de données, utilisateur de données, propriétaire de données, orchestrateur de l'écosystème;
- description détaillée des concepts/piliers sur lesquels s'appuiera l'écosystème: catalogues permettant de localiser les données et les services, contrôle de l'identité des participants, confiance dans la transmission des données, compliance et auditabilité, passerelle API (portails d'échange et interfaces).

8.3 Prototypes

Des prototypes seront utiles pour concrétiser la suite du processus de conception. Ils permettront:

- de réunir des données d'expérience pratique parallèlement au développement conceptuel de l'écosystème DataHub;
- de tester les concepts des piliers quant à leur adéquation à la pratique;
- de contrôler et de concrétiser les concepts architecturaux et de gouvernance en vue de la coopération dans l'écosystème;
- d'accroître la notoriété de l'écosystème DataHub.

Un prototype doit être développé pour la passerelle API. Élément essentiel du DataHub, la passerelle API permettra d'utiliser en commun les données et les services de données. Elle fonctionnera comme une «porte d'entrée» et permettra d'utiliser plusieurs interfaces et services de données. Le prototype devra mettre en œuvre l'authentification des utilisateurs, les droits d'accès, une limite de débit, le monitoring et les statistiques. Il servira en particulier à clarifier la question de la fédération des API (comment les organisations tierces pourraient permettre l'utilisation en commun de leurs API, par ex. avec l'authentification des utilisateurs). Il faudra aussi déterminer comment lister automatiquement les API dans un catalogue (comme par ex. le catalogue national des données I14Y) pour que les API soient localisables et, donc, utilisables en tout temps.

S'agissant de mettre en œuvre le prototype, on se référera aux initiatives et projets mentionnés dans le chapitre [6](#). On reprendra les résultats déjà obtenus lorsque cela s'avérera possible et judicieux.

En outre, il faudra développer un prototype de cas d'application pour les données administratives en libre accès. L'objectif est d'identifier un service administratif (par ex. une commune) qui souhaite publier ses données en libre accès par une API plutôt que par un fichier Excel. Le prototype s'appuiera sur les piliers du DataHub. Il devra répondre aux questions suivantes: 1) Où peut-on héberger les données publiques à publier (Associated Services)? Comment peut-on utiliser le prototype de passerelle API pour permettre l'utilisation publique des données publiques en libre accès au moyen d'une API (API Gateways)? 3) Comment peut-on lister dans un catalogue de données les données publiques en libre accès et l'API correspondante (Directories)?

Annexes

A.1. Personnes et organisations impliquées

Le tableau suivant fournit la liste alphabétique des personnes et des organisations qui ont participé à l'établissement du présent document d'information.

Organisation	Personne	Rôle
AFC	Haggai Bass	Personne interrogée
ANS	Peppino Giarritta, Marcel Kessler	Participants à l'atelier
AVS / AI	Christoph Beer	Personne interrogée
AWK Group	Anthony Buchard, Thomas Kathriner, Andreas Meier, Christian Senning, Adrian Wägli	Collaborateurs de projet
Canton de Berne	Armin Herzog	Personne interrogée
Canton de Fribourg	Stéphane Schwab	Personne interrogée
Canton de Zurich	Lukas Weibel	Personne interrogée
Chancellerie fédérale	Andreas Spichiger Jürg Wüst	Mandant et collaborateur de projet Chef de projet
Comité d'architecture (OMO14)	Andreas Spichiger, ChF Jürg Wüst, ChF Peter Kolbe CFF Économie Thomas Auer, La Poste Nicolas Degen, Swissmem Stefan Agosti, E3 / eCH Thomas Bächler, Swissdec Matthias Günter, CH Open Alessia Neuronni / Matthias Stürmer, BFH Martin Huber, CSI Martin Godel, SECO / easyGov Daniel Krebs, DETEC Thomas Riedo, AFC Matthieu Overney, canton de Fribourg / iGovPortal.ch Lukas Weibel, canton de Zurich Philipp Achermann, canton de Berne	Mandants
Haute école spécialisée bernoise (BFH)	Alessia Neuronni	Personne interrogée, participante à l'atelier
La Poste suisse SA	Thomas Auer, Michael Rieder, Stefan Guggisberg	Personnes interrogées
OFS	Manuela Lenk, Franziska Bühler, Andreas Gyger, Matthias Steffen	Participants à l'atelier, collaborateurs de projet
Opendata.ch	André Golliez	Personne interrogée
SWICO	Judith Bellaiche	Personne interrogée

Tableau 2: Personnes et organisations impliquées

A.2. Cas d'application pertinents envisageables

Les cas d'application (CA) suivants revêtent de l'intérêt dans le contexte du DataHub (la liste ne prétend pas à l'exhaustivité).

ID	Catégorie	Cas d'application
CA01	Fonctionnalités	Comme représentant d'un service administratif , je souhaite par exemple, lors du remplacement d'une application, m'assurer que la nouvelle application utilise également les données existantes provenant de systèmes tiers (au lieu de gérer ces données isolément), de manière à réduire le nombre d'entrées de données, à accroître la transparence et à améliorer la qualité.
CA02	Fonctionnalités	Comme représentant d'un service administratif , je souhaite mettre à la disposition des autres services administratifs les données produites/gérées dans ma propre organisation aux fins de leur utilisation en commun, en respectant toutes les conditions-cadres, le but étant d'organiser plus efficacement les processus, d'accroître la transparence et d'améliorer la qualité des données (<i>share by default</i>).
CA03	Fonctionnalités	Comme représentant d'un service administratif , je souhaite mettre à la disposition du public les données produites/gérées dans ma propre organisation, en respectant toutes les conditions-cadres, afin d'accroître la transparence de nos propres activités administratives conformément au principe d'ouverture par défaut et de permettre la nouveauté et l'innovation dans l'utilisation des données et l'élaboration de modèles d'affaires par les particuliers et les entreprises.
CA04	Fonctionnalités	Comme représentant d'un service administratif , je souhaite associer les particuliers et les entreprises à la gestion de leurs propres données et dossiers, en respectant toutes les conditions-cadres, afin de permettre un self-service, de simplifier les processus, de réduire au minimum les entrées de données et d'améliorer la qualité des données.
CA05	Fonctionnalités	Comme représentant d'un service administratif , je souhaite exploiter avec d'autres services administratifs un hub / guichet central pour l'échange de données, afin d'éviter de devoir gérer individuellement d'innombrables interfaces avec de nombreux autres services administratifs.
CA06	Fonctionnalités	Comme représentant d'une entreprise ou comme particulier , je souhaite pouvoir utiliser toutes les données de l'administration via un seul hub central, afin d'éviter de devoir gérer individuellement d'innombrables interfaces avec de nombreux services administratifs. .
CA07	Fonctionnalités	Comme représentant d'un service administratif , je souhaite pouvoir continuer d'utiliser ma passerelle API actuelle et la relier par exemple à une passerelle centrale, afin de garantir mes investissements et de gérer ma propre passerelle en fonction de mes exigences.
CA08	Fonctionnalités	Comme représentant d'un service administratif , je souhaite utiliser l'espace de stockage sûr, évolutif et bon marché du DataHub pour éviter d'avoir à développer et à gérer un espace de stockage propre et, par exemple, pour utiliser cet espace de stockage pour des données publiques en libre accès.
CA09	Fonctionnalités	Comme représentant d'un service administratif , je souhaite pouvoir recourir à des bibliothèques et à des instruments dans le contexte de l'échange de données, afin que nos ressources informatiques puissent évoluer dans la bonne direction.
CA10	Gouvernance	Comme représentant d'un service administratif , je souhaite disposer d'une feuille de route pour l'utilisation du DataHub ou la participation à celui-ci. Je souhaite pouvoir commencer à petites doses et maîtriser mon degré d'engagement en fonction de mon appréciation, de manière à garantir que mes besoins soient toujours pris en compte.
CA11	Gouvernance	Comme représentant d'un service administratif , je souhaite pouvoir utiliser des directives, meilleures pratiques, modèles architecturaux de référence, modèles de rôle organisationnels, etc. dans le contexte de l'échange de données, afin que nos ressources informatiques puissent évoluer dans la bonne direction.

Tableau 3: Sélection de cas d'application dans le contexte DataHub

A.3. Exigences générales

Les exigences générales (EG) suivantes ont été formulées sur la base des motions, de la situation initiale identifiée et des entretiens menés avec divers acteurs (la liste ne prétend pas à l'exhaustivité).

ID	Mot-clé	Exigence
EG01	Saisie unique des données	Le DataHub fournit des composantes de solution de nature technique et organisationnelle / méthodologique afin que les services administratifs puissent poursuivre la mise en œuvre du principe de la saisie unique des données.
EG02	Consolidation des données	Le DataHub met à disposition les moyens techniques permettant aux divers participants à l'écosystème de développer progressivement une logique spécialisée consolidée. Le développement proprement dit de celle-ci génère un coût important, il requiert une coordination et se concrétise dans des structures de projet dédiées supplémentaires.
EG03	Confidentialité	Seuls les utilisateurs et systèmes autorisés, dotés des droits correspondants, peuvent accéder aux données et aux fonctions. Les données publiques en libre accès peuvent constituer une exception.
EG04	Protection des données et sécurité de l'information	Le stockage et l'échange de données répondent en tout temps aux exigences de la loi sur la protection des données et de l'ordonnance concernant la protection des informations, aux besoins de protection définis par les propriétaires de données et aux éventuelles directives juridiques supplémentaires.
EG05	Intégrité	Les échanges de données et les modifications apportées à celles-ci sont traçables en tout temps.
EG06	Disponibilité	Les composantes techniques essentielles du DataHub satisfont aux plus hautes exigences en termes de disponibilité. Dans la mesure du possible, leur stockage est géoredondant. Les participants à l'écosystème peuvent ainsi utiliser ces composantes comme des modules pour leurs propres services critiques.
EG07	Interopérabilité	Le DataHub constitue la base juridique, organisationnelle, sémantique et technique de l'interopérabilité.
EG08	IAM fédérée	Une gestion des identités et des accès (<i>identity and access management</i> , IAM) centralisée règle l'accès aux données. Les autres systèmes IAM des participants à l'écosystème peuvent y être incorporés.
EG09	Organisations	En principe, toutes les organisations (administrations, entreprises privées, universités, organisations non gouvernementales, organisations à but non lucratif, etc.) et les particuliers peuvent participer aux échanges de données s'ils remplissent les conditions d'autorisation.
EG10	Coûts	Le DataHub permet l'échange de données à peu de frais par les participants. Les coûts sont répercutés sur les utilisateurs. Les données publiques en libre accès peuvent constituer une exception.
EG11	Vitesse	Les vitesses de téléchargement vers l'amont et l'aval satisfont aux plus hautes exigences. Elles devront être définies ultérieurement de manière plus précise.
EG12	Latence	La latence de communication satisfait aux plus hautes exigences. Il faudra la définir ultérieurement de manière plus précise.
EG13	Connectivité	En tant qu'écosystème, le DataHub assure la connectivité de nouvelles solutions en provenance de la Suisse et de l'étranger.
EG14	Technologie agnostique	Le DataHub doit correspondre à une technologie agnostique qui garantit un accès sans discrimination à l'écosystème.
EG15	Cycle de vie des données	Le DataHub met à disposition une fonctionnalité de gestion des données publiées au cours du cycle de vie.

Tableau 4: Exigences générales dans le contexte DataHub

A.4. Abréviations et notions

Abréviation	Description
API	Une interface de programmation d'application (<i>application programming interface</i>) est une partie d'application qu'un système logiciel met à la disposition des autres programmes pour permettre la connexion à ladite application.
Catalogue de métadonnées	Un catalogue de métadonnées est une liste de métadonnées qui contient les définitions et les règles de représentation de toutes les données d'application et qui indique la relation entre les divers objets de données.
Consommateur de données	Le consommateur de données (<i>data consumer</i>) est un participant à l'écosystème qui recherche et utilise des données et des prestations liées aux données dans l'écosystème DataHub, par exemple afin de composer de nouvelles offres numériques destinées aux utilisateurs finaux.
Données publiques en libre accès (OGD)	Les données publiques en libre accès (<i>open government data</i> , OGD) correspondent aux ensembles de données du secteur public auxquels il est possible d'accéder librement sans aucune restriction, dans l'esprit d'une administration ouverte répondant à l'intérêt de la communauté.
Écosystème	Un écosystème est une association d'organisations présentant les formes juridiques les plus diverses (les participants à l'écosystème) qui, par leur coopération, génèrent une valeur ajoutée grâce à leur orchestration de rang supérieur.
Fournisseur de données	Par fournisseur de données (<i>data provider</i>), on entend un participant à l'écosystème qui met à disposition des données et des prestations liées aux données dans l'écosystème DataHub. Le fournisseur de données définit et décrit sa propre offre de données et de services, y compris les conditions d'utilisation.
Métadonnées	Les métadonnées sont des données qui fournissent un contexte spécialisé et technique pour d'autres données. Elles décrivent les caractéristiques / modalités des données correspondantes.
Principe de la saisie unique des données	Le principe de la saisie unique des données (principe « <i>once only</i> ») vise à ce que les entreprises et les particuliers ne doivent indiquer qu'une seule fois certaines informations à l'administration.
Sémantique	La sémantique étudie la signification exacte des termes dans les langues. S'agissant d'échange de données, il s'agit pour les parties à l'échange de données de s'entendre – et de consigner leur consensus – sur la signification commune des données échangées, c'est-à-dire sur la signification du contenu des données.

Tableau 5: Abréviations et notions

A.5. Documents référencés et autres sources

Titre	Auteur / éditeur	Date	Lien / fichier
[1] Architekturvision	Chancellerie fédérale	État en août 2021	-
[2] eCH-0138 Concept organisationnel pour la description et la documentation de tâches, prestations, processus et structures d'accès de l'administration publique en Suisse	eCH	12.04.2012	http://www.ech.ch/fr/standards/60366
[3] Gaia-X	Gaia-X European Association for Data and Cloud AISBL	27.10.2021	gaia-x.eu
[4] eCH-0126 Concept cadre «Administration interconnectée en Suisse»	eCH	n/a	https://www.ech.ch/fr/standards/60339
[5] Rapport sur l'évaluation des besoins d'un «Swiss Cloud»	Chancellerie fédérale	Déc. 2020	https://www.bk.admin.ch/bk/fr/home/digitale-transformation-ikt-lenkung/bundesarchitektur/cloud.html
[6] Trusted Data Hub Introductory White Paper	La Poste	Avril 2020	-

Tableau 6: Documents référencés et autres sources