Décision de l'Assemblée plénière | 22 juin 2023

Fédération des données dans la formation professionnelle; clôture de la phase de conception et non-réalisation du projet: décision

Considérations du Secrétariat général

1 Le 24 juin 2021, l'Assemblée générale de la CDIP a chargé l'agence spécialisée Educa d'élaborer, conjointement avec les organes et les organisations spécialisées de la formation professionnelle, un concept relatif à une fédération des données dans la formation professionnelle, visant à simplifier l'échange des données dans la formation professionnelle à l'échelle suisse et à permettre une meilleure utilisation de ces dernières.

2 L'Assemblée plénière a débloqué un montant de 375 000 francs destiné à l'élaboration de ce concept, le Secrétariat d'État à la formation, la recherche et à l'innovation (SEFRI) y contribuant pour un même montant.

3 Educa a terminé la phase de conception le 1er mars 2023 et remis au SEFRI et au Secrétariat général de la CDIP un concept détaillé en trois parties. Le concept montre de quelle manière il sera possible d'organiser à l'avenir la circulation des flux de données dans la formation professionnelle. Il présente une architecture système applicable à toute la Suisse et discute de l'incidence des nouveaux principes régissant la protection des données, tels qu'ils figurent notamment dans la nouvelle loi fédérale sur la protection des données. Le concept souligne globalement la nécessité de tirer au clair diverses questions qui présentent un enjeu pour l'avenir, comme la vérification numérique des titres de formation.

4 En raison de la complexité initiale, de l'ampleur du projet et d'autres priorités dans la formation professionnelle, le SG CDIP propose toutefois, en concertation avec le SEFRI, de ne pas réaliser la fédération des données. La discussion menée dans les organes de la formation professionnelle à propos du concept de fédération des données a montré qu'Educa n'était pas parvenu à convaincre du bien-fondé du projet les représentantes et représentants cantonaux du domaine de la formation professionnelle. En revanche, les projets en cours ayant pour thème la gestion des données dans la formation professionnelle seront priorisés.

5 Les cantons ont déjà reçu des factures d'un montant de 83 333 francs au titre des contributions cantonales pour la phase de réalisation de la fédération des données, initialement prévue à partir d'octobre 2022. Il est prévu de provisionner ces contributions pour les affecter à de futurs projets de numérisation. Les montants inscrits au budget 2023 (500 000 francs) ne seront pas facturés aux cantons. Le budget 2024 et le plan financier seront modifiés en conséquence.

6 Des discussions menées à partir de juillet 2023 entre l'agence spécialisée Educa et les acteurs de la formation professionnelle permettront de déterminer s'il est possible de réaliser ultérieurement des parties du concept présenté en tant qu'éléments du projet («incréments»). Ceux-ci seront également soumis au SEFRI et à la CDIP en vue d'une décision.

Décision de l'Assemblée plénière

1 Le Comité prend connaissance du concept détaillé qui marque la clôture de la phase de conception du projet relatif à la fédération des données dans la formation professionnelle.

2 La fédération des données dans la formation professionnelle n'entre pas dans sa phase de réalisation. Le montant de 83 333 francs déjà facturé aux cantons sera provisionné dans la perspective d'autres projets de numérisation.

Berne, le 22 juin 2023

Conférence des directrices et directeurs cantonaux de l'instruction publique
Au nom de l'Assemblée plénière:

sig.
Susanne Hardmeier | Secrétaire générale

Annexe:
– Concept de la fédération des données dans la formation professionnelle

Notification:
- Membres de la Conférence
- CSFP
- COP
- Educa
- SEFRI

La présente décision sera publiée sur le site web de la CDIP.

232.32-14.3 bee/bop

educa

Erlachstrasse 21
3012 Bern

+41 31 300 55 00
info@educa.ch

Concept de la fédération des données dans la formation professionnelle
1ère partie – fondement et architecture

15.11.2022

Digitaler Bildungsraum Schweiz
Espace numérique suisse de formation
Spazio formativo digitale svizzero
Spazi da furmazium digital svizzer
Swiss digital education space

educa

Inhalt

Management Summary **3**

1. Introduction **4**
1.1 Mandat de projet sur la conception d'une fédération des données pour la formation professionnelle et la préparation des bases légales 4
1.2 Implication des partenaires 5
1.3 Procédure 6

2. Fondement **7**
2.1 Analyse du contexte de la formation professionnelle initiale 7
2.2 La situation initiale sur plan juridique 10
2.3 Développements technologiques et sociétaux 11

3. L'architecture d'une fédération des données **12**
3.1 L'architecture en tant que cas d'application: justificatifs numériques dans la formation professionnelle initiale 13
3.2 Le point de vue technique sur l'architecture 17
3.3 Prise en compte des principes d'architecture 19

4. Conclusion intermédiaire **21**

Annexe **23**
Définitions 23

Management Summary

Depuis juin 2021, sur mandat de la Conférence suisse des directrices et directeurs cantonaux de l'instruction publique (CDIP) et du Secrétariat d'État à la formation, à la recherche et à l'innovation (SEFRI), l'agence spécialisée Educa se charge de la conception d'une fédération des données pour la formation professionnelle. Cette fédération des données doit simplifier l'accès aux données et garantir leur échange de manière sécurisée. Les données sont considérées comme une ressource stratégique qui présente un grand potentiel d'innovation pour la formation professionnelle. L'objectif final est de pouvoir échanger efficacement des données dans toute la Suisse, de les rendre accessibles grâce à un espace de confiance et utilisables par l'ensemble du système, respectivement par l'ensemble des partenaires. Cette fédération des données doit être comprise comme une architecture globale qui définira la manière dont les données seront à l'avenir échangées au sein de la formation professionnelle initiale. Elle sert donc également d'architecture cible pour les services en place.

L'accès aux données revêt une importance capitale pour tous les partenaires de la formation professionnelle initiale s'agissant des moyens décisionnels, de développement et d'innovation. L'accès aux données est un moteur d'innovation. La vitesse avec laquelle certains événements sont disponibles dans toute la Suisse, tels que les diplômes de formation des apprenantes et apprenants, les décisions d'affectation aux écoles professionnelles, les données de cours pour les formatrices et formateurs, les obligations financières, les informations portant sur de nouveaux contenus de formation..., limite considérablement le potentiel d'innovation du monde du travail dans le cadre de la formation professionnelle.

Le concept de fédération des données hisse, en faveur de la formation professionnelle initiale, à un nouvel échelon les moyens de mise à disposition des informations. Ainsi, la fédération des données se met entièrement au service de la capacité d'innovation du système suisse de formation **professionnelle en connectant directement tous les partenaires, de manière autonome et sans détour, par des événements survenant dans un contexte d'utilisation.** Les apprenantes et apprenants ainsi que les sites de formation sont placés au centre du concept.

Les objectifs du projet, conformément au mandat reçu, ont pu être repris dans le concept d'architecture. Le cadre futur du potentiel d'innovation pour la formation professionnelle initiale a été renforcé. En outre, les exigences des apprenantes et apprenants ainsi que celles du monde du travail ont été plus fortement considérées dans le concept de fédération des données. Enfin, le concept d'architecture tient compte des besoins liés à la nouvelle Loi sur la protection des données.

Ce document est la première des trois parties du concept. Il porte sur l'architecture technique d'une future fédération des données. Les deux autres parties traitent du modèle de gouvernance nécessaire à cette architecture et des bases légales éventuellement nécessaires à sa mise en oeuvre. Conjointement avec la planification de réalisation (notamment le calendrier de mise en œuvre et la budgétisation), les trois parties de ce concept constituent ensemble le dossier de décision destiné au SEFRI et à la CDIP.

1. Introduction

L'espace suisse de formation est diversifié et perméable. Les individus évoluent à la fois dans des univers analogiques et numériques, ont de nombreux contacts et produisent des données dans des lieux les plus divers. Depuis juin 2021, sur mandat de la Conférence suisse des directrices et directeurs cantonaux de l'instruction publique (CDIP) et du Secrétariat d'État à la formation, à la recherche et à l'innovation (SEFRI), l'agence spécialisée Educa se charge d'élaborer un concept de fédération des données pour la formation professionnelle. Dans ce contexte, les données sont considérées comme une ressource d'importance stratégique qui présente un grand potentiel d'innovation pour la formation professionnelle. L'objectif final poursuivi est de pouvoir échanger efficacement des données dans toute la Suisse, de les rendre accessibles grâce à un espace de confiance et utilisables par l'ensemble du système, respectivement par l'ensemble des partenaires.

Par le biais d'une structure décentralisée, une fédération des données offre, dans le cadre de l'échange de données, des voies de communication plus directes, accroît ainsi la capacité de résilience du système de la formation professionnelle et réduit les risques y afférents (par ex. le risque de panne d'un service central). Par le terme de «fédération des données», on entend un système d'applications et de banques de données participatives et faiblement reliées, soutenu par une infrastructure décentralisée. Cette infrastructure décentralisée met à disposition les services de confiance et les mécanismes de recherches de consensus nécessaires au bon fonctionnement de la fédération des données.

1.1 Mandat de projet sur la conception d'une fédération des données pour la formation professionnelle et la préparation des bases légales

En tant qu'agence spécialisée de la CDIP et du SEFRI, Educa a obtenu en juin 2021 de l'Assemblée plénière de la CDIP et de la Direction du SEFRI le mandat de projet[1] qui a pour objet la *conception d'une fédération des données pour la formation professionnelle ainsi que dans la préparation des bases légales correspondantes*. Ce projet a été simultanément intégré dans le plan d'action de la stratégie «Suisse numérique»[2]. Le mandat de projet formule les objectifs de projet suivants pour la fédération des données :

1. L'échange de données entre les acteurs de la formation professionnelle doit avoir lieu sans obstacles et sans retard, et il doit être de bonne qualité.
2. Les potentiels d'innovation basés sur les données doivent être rendus accessibles au système de la formation professionnelle à l'échelle de toute la Suisse.
3. Les données sont utilisées en tant que ressource d'importance stratégique pour le système de la formation professionnelle à l'échelle de toute la Suisse.
4. Le projet majeur fournit une contribution à la capacité d'innovation du système de la formation professionnelle.

[1] https://edudoc.ch/record/220333/files/PB240621-federation-f.pdf

[2] https://www.digitaldialog.swiss/fr/politique-dutilisation-des-donnees-pour-lespace-suisse-de-formation

Le projet est piloté par le Comité de coordination Numérisation de l'éducation (CC N). La direction du projet incombe à Educa. Le concept élaboré relatif à la fédération des données pour la formation professionnelle devra à nouveau être discuté par l'Assemblée plénière de la CDIP ainsi que par la Direction du SEFRI et, le cas échéant, fera l'objet d'une décision.

1.2 Implication des partenaires

Pour l'élaboration du concept de fédération des données en faveur de la formation professionnelle initiale, les points de vue de tous les partenaires sont pris en compte. La Confédération est compétente pour le pilotage stratégique et le développement ; les cantons sont compétents pour la mise en œuvre et la surveillance. Les organisations du monde du travail (OrTrA) s'occupent des contenus de la formation et des places de formation. Le constat selon lequel l'échange de données au sein de la formation professionnelle initiale présente un potentiel d'amélioration considérable était – et est toujours – incontesté par tous les partenaires. Au cours de la phase d'élaboration du concept, Educa s'est basée sur les prémisses suivantes, après implication des partenaires :

1. Les projets d'échange de données actuellement en cours s'agissant des processus d'échange de données (notamment HAKA, norme eCH concernant les données Formation professionnelle, divers registres) sont des bases importantes.
2. Une série de plateformes exploitées de manière centralisée (à l'échelon cantonal/national) relèvent d'une thématique «end-of-life».
3. Il n'existe aucune proposition d'architecture à l'échelle de toute la Suisse qui permet de relier entre eux des services existants et de futurs services numériques.
4. L'architecture de la fédération des données doit tenir compte des spécificités fédérales de la Suisse et des compétences respectives des partenaires.
5. Du point de vue de l'accès aux données, les exigences des cantons sont fortement représentées ; les exigences du monde du travail et, au sein de ces dernières, également les exigences des apprenantes et apprenants sont en revanche, jusqu'à présent, faiblement représentées.

Pour pouvoir formuler ces prémisses, les exigences d'un grand nombre de parties prenantes issues des rangs des partenaires ont été identifiées. Un premier aperçu à ce sujet avait déjà été établi pendant la phase d'initialisation entre la mi-2020 et le printemps 2021. Avec le début de la phase de conception à la mi-2021, les échanges se sont intensifiés et sont devenus plus systématiques.

Les échanges avec les partenaires peuvent être résumés comme suit :

1. Échanges intensifs avec les membres de la Conférence suisse des offices de la formation professionnelle (CSFP), notamment lors de deux rondes d'échanges qui ont eu lieu début mars et fin mai 2021
2. Participation aux séances de commissions de la CSFP (comité et plénière) ainsi qu'à la Commission Organisation et Processus (COP)

3. Entretiens bilatéraux avec tous les offices de la formation professionnelle de Suisse romande et du Tessin
4. Entretiens avec des offices cantonaux de la formation professionnelle de Suisse alémanique
5. Échanges continuels avec le Centre suisse de services Formation professionnelle (CSFO)
6. Échange continuel avec une multitude d'OrTra, de prestataires de services et d'entreprises formatrices
7. Petit nombre d'entretiens avec des écoles professionnelles
8. Échanges récurrents avec l'Union patronale suisse (UPS), l'Union suisse des arts et métiers (USAM), ainsi qu'avec l'Union syndicale suisse (USS) et Travail.Suisse.

Au niveau stratégique, Educa a participé à deux séances de la Conférence tripartite de la formation professionnelle (CTFP) et a mené des entretiens avec le SEFRI. En outre, elle a systématiquement évalué des documents de stratégie au niveau de la Confédération (par ex. plan d'action «Suisse numérique», divers rapports du Conseil fédéral), des documents fondamentaux cantonaux et intercantonaux (par ex. Stratégie suisse de cyberadministration 2020–2023) et les a pris en compte pour ses travaux de conception.

1.3 Procédure

La formation professionnelle initiale repose sur une série de prescriptions légales qui sont également mises en œuvre au moyen de processus d'échanges de données. En bref, on peut affirmer que ces données reflètent la formation professionnelle initiale. Cette idée est placée au cœur de la conception de la fédération des données ; dès lors, il s'agit quasiment de la vision architecturale de la fédération des données. Pour l'élaboration du concept, Educa applique la méthode TOGAF (The Open Group Architecture Framework). Il s'agit d'un modèle de marche à suivre déjà bien établi qui permet d'élaborer des architectures d'entreprise. L'objectif de l'architecture est d'établir les structures de base de la fédération des données et de définir des règles qui coordonnent l'interaction dynamique de tous les composants. Le système de formation professionnelle doit être représenté le plus précisément possible. Par le biais de la discussion sur les principes de l'architecture, et au travers de leur formulation, on crée ainsi les fondements des nouvelles structures. Ensuite, divers exemples de cas d'application peuvent être établis. Ces scénarios d'avenir permettent de prendre en compte la diversité propre au système de la formation professionnelle. Ils permettent aussi de comprendre les modalités de l'échange de données et de les rendre visibles. On peut en déduire le type de solution technique et le type de gouvernance appropriés requis pour pouvoir simplifier l'accès aux données et améliorer l'échange de données dans le système de la formation professionnelle.

Lors de la conception d'une fédération des données, l'implication de l'ensemble des parties prenantes revêt une importance particulière (voir chapitre 1.2.). En outre, Educa a analysé de manière intensive les processus existants de l'harmonisation de l'échange de données entre les offices cantonaux de la formation professionnelle (HAKA) et, lors d'une deuxième ronde,

elle a étudié les projets de processus-CIBLES HAKA que le CSFO a élaborés (voir chapitre 2.1.). Sur la base de ces processus HAKA, on peut déduire l'ensemble du paysage des parties prenantes qui participent aux processus d'échange de données. Des solutions techniques existantes ont également été analysées et discutées avec les parties prenantes. De plus, des bases légales ont été évaluées et analysées (voir chapitre 2.2.).

Afin de réduire la complexité des flux de données dans le système de la formation professionnelle, des cas d'application concrets et tournés vers l'avenir feront l'objet d'une description. Le premier exemple de cas d'application est centré sur un apprenant que nous appellerons Luca. Dans l'explication de ce cas d'application, nous adoptons des hypothèses simplifiées, ce qui permet d'analyser l'échange de données dans une future fédération des données. Cela détermine la solution technique, et cela rend visibles les questions de gouvernance qui se posent dans ce cas de figure.

La première partie du présent concept met l'accent sur l'architecture technique d'une future fédération des données. Cette architecture technique est d'abord présentée sur la base d'un premier cas d'application (voir chapitre 3.1.). Vient ensuite une description technique de l'architecture (voir chapitre 3.2.). Enfin, les principes d'architecture sont énoncés en troisième lieu (voir chapitre 3.3.).

Dans les deux autres parties prévues du concept (pour le 10 janvier 2023 et pour le 1er mars 2023), nous présenterons d'abord le modèle de gouvernance, à savoir les prescriptions et mécanismes qui piloteront à l'avenir la fédération des données. Des cas d'application supplémentaires seront également élaborés. Sur cette base, les bases légales qui existent déjà seront mises en évidence, ainsi que, le cas échéant, de futures bases légales qu'il faudrait créer pour que la fédération des données puisse être introduite dans la formation professionnelle initiale. La planification de réalisation sera également incluse dans la troisième partie du concept.

2. Fondement

Le prochain chapitre analyse l'échange de données existant dans la formation professionnelle initiale, présente les bases légales actuelles, et met en lumière des développements technologiques et sociétaux.

2.1 Analyse du contexte de la formation professionnelle initiale

Les acteurs de la formation professionnelle se préoccupent déjà depuis un certain temps de la question de savoir comment il serait possible de simplifier et d'améliorer l'échange de données dans la formation professionnelle à un niveau qui dépasserait les sites de formation individuels, les responsabilités respectives et les frontières cantonales. Ces dernières années, à la différence des projets et programmes nationaux, de nombreux cantons ont pu réaliser en partie des progrès déterminants dans l'échange de données pour la formation professionnelle.

L'application technique utilisée dans un canton donné pour la gestion des données définit et limite simultanément les possibilités de l'échange de données automatisé. Pour rendre

possible ce dernier, quelques cantons ont décidé d'utiliser les mêmes applications techniques, alors que d'autres cantons ont installé des interfaces. Néanmoins, à l'heure actuelle, il n'existe aucun échange de données automatisé entre l'ensemble des applications techniques, et il n'y a pas non plus d'échange de données automatisé contraignant au sein de la même application technique.

Le graphique 1 présente les applications techniques utilisées dans les offices cantonaux de la formation professionnelle et les groupes d'échange qui en résultent. La Suisse alémanique est répartie en deux groupes de cantons : ceux qui utilisent ESCADA2, et ceux qui utilisent KOMPASS 3 (organisés au sein de l'association «Communauté d'intérêts Informatique dans le système éducatif (IGIB)»). En Suisse romande et dans le canton du Tessin, la situation est plus hétérogène. Les cantons de Neuchâtel et du Jura utilisent la même solution (CLOEE2), Fribourg et le Tessin utilisent IS-Academia, le canton du Valais utilise ESCADA2, et les cantons de Genève et Vaud utilisent leurs propres applications. Quelques-unes de ces applications ne sont utilisées que par les offices cantonaux de la formation professionnelle, alors que d'autres sont également utilisées pour la gestion des établissements scolaires (en partie également en dehors de la formation professionnelle initiale).

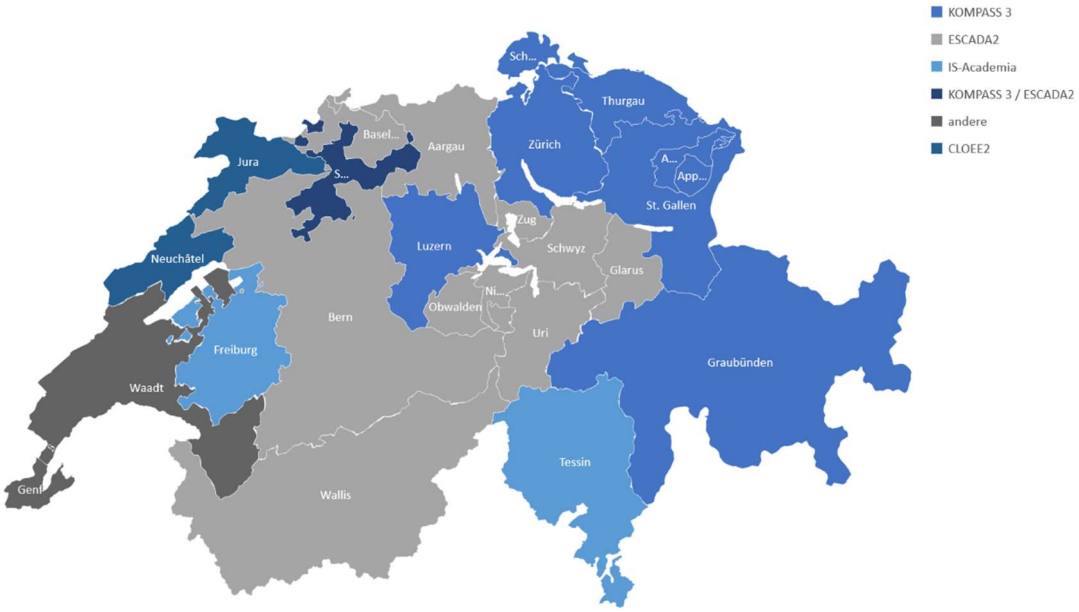

Figure 1 : Applications techniques utilisées en Suisse

Après que les cantons se sont prononcés le 27 mai 2020, lors de l'Assemblée des membres de la CSFP, pour un rattachement des écoles professionnelles aux systèmes cantonaux en vue de l'échange de données, un serveur d'échange de données appelé «plaque tournante de données» a été développé par les cantons membres de la communauté d'intérêts IGIB, et il a été mis en service à l'automne 2022. Cela rend possible pour la première fois l'échange automatisé de données entre les deux applications techniques ESCADA et KOMPASS.

En règle générale, les offices cantonaux de la formation professionnelle sont la plaque tournante de l'échange de données dans la formation professionnelle initiale. En dernière analyse, c'est aux offices cantonaux qu'incombe l'exécution de la loi sur la formation professionnelle, et c'est par le biais des offices cantonaux que fonctionne une grande partie de l'échange de données, même s'ils n'en sont pas les premiers bénéficiaires. Jusqu'à présent, l'automatisation de l'échange de données n'est souvent mise en œuvre qu'en partie. Il arrive fréquemment que des données qui sont envoyées par courriel ou sous forme papier soient échangées via des fichiers Excel ou PDF entre des offices cantonaux de la formation professionnelle et d'autres acteurs de la formation professionnelle initiale. Et cela nécessite une charge de travail administratif considérable qui accroît également les risques d'erreur. Cette situation est due au manque de sécurité juridique, au manque d'harmonisation entre les lignes directrices sur la protection des données, ainsi qu'à l'hétérogénéité des applications techniques utilisées. En outre, la même application technique est parfois utilisée différemment dans différents cantons, et même au sein d'un même canton, à plusieurs niveaux de la formation professionnelle. Cela s'explique par l'existence de modes de travail et de processus individuels différents. Des données sont souvent saisies plusieurs fois et sont sauvegardées dans des systèmes différents, d'où une baisse de la qualité des données. Cette situation a été décrite dans l'«analyse préliminaire Optima», établie par la Haute école spécialisée de Suisse occidentale (HES-SO) en 2016. Des enquêtes réalisées par la COP auprès des offices cantonaux de la formation professionnelle en octobre 2019 et en mars 2020 ont fait état d'un tableau similaire.

Il existe un large consensus sur le fait que la situation actuelle n'est pas satisfaisante et que les mesures prises ne sont pas encore suffisantes pour organiser avec efficacité l'échange de données à l'échelle de toute la Suisse dans la formation professionnelle initiale. Pour cette raison, la CSFP a confié au CSFO le mandat consistant à mettre en œuvre le projet intitulé «Harmonisation de l'échange de données intercantonal HAKA». Le CSFO fournit déjà des prestations centralisées en lien avec l'échange de données, il dispose d'un savoir-faire sur les processus opérationnels, et il constitue à ce jour l'unique ressource opérationnelle commune à l'échelle de toute la Suisse pour les offices cantonaux de la formation professionnelle dans ce contexte. Dans le cadre de phase de conceptualisation du projet HAKA, une analyse de la situation actuelle des processus d'affaires et processus d'échange de données existants de la formation professionnelle initiale a été réalisée dans les cantons. Sur cette base, des processus-CIBLES harmonisés d'échange de données ont été ensuite élaborés. Ils constituent la base des futurs projets de mise en œuvre. Toutefois, comme le CSFO l'a lui-même déjà constaté dans le mandat relatif à l'initialisation du projet, une utilité de HAKA pour la formation professionnelle ne pourra ensuite être générée qu'à l'avenir, «[…] à condition que les résultats élaborés dans le présent projet soient effectivement utilisés et mis en œuvre. Cela devra [faire] partie de projets subséquents».

À ce jour, aucune réglementation juridiquement contraignante à l'échelle de toute la Suisse ne s'applique à la transmission technique et à la sécurité des données dans la formation professionnelle initiale. En revanche, avec les «lignes directrices sur l'échange de données pour la formation professionnelle initiale», il existe depuis des années des lignes directrices nationales avec des instructions sur le contenu et la forme des données à transmettre dans la formation professionnelle initiale. Ces lignes directrices, actuellement disponibles dans la

version 3.07a, sont basées sur un consensus des parties impliquées et sont mises en œuvre dans tous les cantons, étant précisé qu'à l'heure actuelle, dans un petit nombre de canton, l'échange de données a encore lieu selon des versions plus anciennes. Pour succéder aux actuelles lignes directrices sur l'échange de données, une norme eCH pour la formation professionnelle a fait l'objet d'un projet développé par la CSFO. Initialement, il était toutefois prévu de reprendre de manière inchangée, pour l'essentiel, les lignes directrices sur l'échange de données, et de ne les définir que sur le plan technique en tant que norme eCH-XML. Mais comme les processus actuels d'échange de données présentent d'importants déficits, un remaniement fondamental basé sur les résultats du projet HAKA s'est avéré nécessaire. Au moment où nous rédigeons le présent concept, le standard eCH-0260 se trouve en consultation publique, et l'on vise un engagement formel («commitment») des directrices et directeurs des offices cantonaux pour la session plénière de la CSFP en février 2023.

Les projets décrits constituent certes un net progrès par comparaison à la situation décrite dans le projet «Optima» de l'année 2016. Mais les solutions réalisées jusqu'à présent demeurent fragmentaires. Lors de plusieurs échanges qui ont eu lieu avec Educa, tous les acteurs impliqués ont exprimé le besoin latent d'une amélioration et d'une automatisation de l'échange de données dans la formation professionnelle à l'échelle de toute la Suisse. Cela concerne l'échange d'informations entre des administrations cantonales, notamment avec la Suisse latine, mais aussi à l'échelon national avec les entreprises formatrices et entre ces dernières, ainsi qu'avec les OrTra. Par le passé, ces dernières ont déjà fait part de manière répétée, et en plusieurs endroits, de la nécessité de pouvoir disposer d'un échange de données harmonisé via les trois sites de formation de la formation professionnelle initiale (entreprises, écoles professionnelles, cours interentreprises). Les réformes imminentes de la SEC Suisse et du commerce de détail renforceront à titre supplémentaire le besoin de recevoir des données valides des écoles professionnelles sur les apprenantes et les apprenants.

2.2 La situation initiale sur plan juridique

Les prescriptions juridiques, comme par exemple le cadre juridique existant applicable à la formation professionnelle (à savoir la loi fédérale sur la formation professionnelle, l'ordonnance sur la formation professionnelle, etc.) ainsi que la protection des données, surtout au niveau cantonal, revêtent une importance prioritaire pour la fédération des données. En ce qui concerne les technologies blockchain, à l'heure actuelle, il n'existe guère de prescriptions juridiques concrètes au niveau national. Différents projets de loi sont toutefois en cours d'élaboration (par ex. loi fédérale sur l'utilisation des moyens électroniques pour l'exécution des tâches des autorités, loi sur l'e-ID). Les conséquences sur le droit de la protection des données d'un tel échange de données décentralisé sont discutées de manière intensive depuis assez longtemps. Ainsi, par exemple, dans une évaluation de la technologie blockchain, le Parlement européen parvient à la conclusion selon laquelle cette technologie est plus appropriée que des technologies centralisées pour garantir et exécuter des normes de protection des données. En sus du stockage réparti des données personnelles, cette évaluation a mis en évidence des avantages supplémentaires de cette technologie comme la minimisation des données (par ex. par le biais de ce qu'on appelle les «zero-knowledge-proof») et une automatisation de l'exécution des lois sur la protection des données (par ex. au moyen des «contrats intelligents»). Conjointement avec d'autres études sur le sujet, un consensus

semble se dessiner sur la manière dont des technologies de ce genre peuvent être utilisées de manière conforme au droit. Dans la perspective de la révision de la loi sur la protection des données, des projets pilotes sont en cours dans les cantons et étudient également la question de savoir comment de tels avantages peuvent être réalisés dans le contexte politique et juridique fédéral de la Suisse.

Au vu des caractéristiques principales de l'architecture requise qui deviennent toujours plus claires au fil du temps, nous verrons dans quels domaines il existe un besoin de création de bases légales. De surcroît, il est possible de recourir aux instruments usuels du système de formation professionnelle (par ex. règlement d'organisation, contrats destinés aux membres, etc.). De plus, sur la base de l'article 48 Cst., la possibilité existe de conclure des contrats intercantonaux. Nous traiterons cette situation de fait dans la prochaine partie du présent concept.

2.3 Développements technologiques et sociétaux

Introduire un échange de données plus efficace et, sur cette base, simplifier l'utilisation des bases de données existantes constitue un objectif central de la transformation numérique des administrations au sein de la Confédération et des cantons. Les processus automatisés d'échange de données ont en effet le potentiel qui permet de décharger de certaines tâches les «sujets de données» («data subjects»), et peuvent ainsi améliorer la qualité des bases de données. Cela permet aux autorités et aux institutions publiques de fournir plus rapidement leurs prestations, avec davantage d'efficacité, et en étant centrés sur les besoins des clients. À ce jour, des efforts importants ont été investis dans la scalabilité de tels systèmes, mais ils se focalisaient en priorité sur les besoins de ces autorités.

Dans le même temps, on prend de plus en plus conscience du fait qu'à long terme, l'acceptation sociale et la légitimité démocratique des systèmes d'échange automatisé de données ne pourront être garanties qu'à condition que, lors de leur conception et de leur mise en œuvre, on tienne plus fortement compte des questions qui ont trait à la sécurité des données et à l'autodétermination informationnelle. Dès lors, tant en Suisse que dans le contexte international, des pistes de solution qui donnent à l'individu (au «sujet de données») davantage de contrôle sur ses données gagnent en importance. Ainsi, par exemple, sur la base du rapport relatif à la consultation publique sur le «Document de travail concernant le projet d'identité électronique (e-ID)» de l'Office fédéral de la justice, le Conseil fédéral parvient à la conclusion selon laquelle les principes de la «souveraineté des données chez l'utilisateur (self sovereign identity)», de la «sobriété des données», «du privacy by design» (donc de l'implémentation de la protection des données par l'architecture des données) et du stockage décentralisé des données devraient constituer les piliers clés d'une future solution e-ID de la Suisse. L'«image cible e-ID» a déclenché plusieurs projets en Suisse qui étudient la faisabilité des justificatifs numériques (par ex. permis de conduire numérique, carte d'identité fédérale numérique, etc.). Également dans le domaine de la formation, différents projets sont en cours afin d'analyser le potentiel des technologies blockchain, par ex. pour la vérification de la validité des diplômes des hautes écoles chez SWITCH.

Au plan international, on poursuit des pistes de réflexion similaires, comme par ex. au sein de l'Union européenne avec le projet «European Self Sovereign Identity Framework» en tant que

partie intégrante de l'European Blockchain Services Infrastructure (EBSI). À cet égard, des infrastructures de données qui reposent sur des principes décentralisés ne conviennent pas seulement à la gestion des identités (identity management). Par exemple, dans un white paper récemment publié, la Commission économique des Nations Unies pour l'Europe (United Nations Economic Commission for Europe, UNECE) propose d'utiliser une telle architecture afin d'améliorer la traçabilité et la transparence dans le trafic transfrontalier des marchandises. En outre, une série de fondations et de consortiums qui opèrent à l'échelon national et international, dont le World Wide Web Consortium, digitalswitzerland, la fondation Hyperledger, IDunion ou la fondation Trust Over IP, ont identifié un large éventail de normes et de possibilités d'application pour des technologies qui reposent sur les principes de base des approches décentralisées. Parmi les applications possibles figureraient l'échange international de données sur la santé, la lutte contre la fraude lors du versement de subventions, voire l'établissement de certificats de hautes écoles à validité vérifiable. Ce qui est commun à toutes ces possibilités d'application, c'est qu'elles rendent possibles un échange d'informations sécurisé, transparent, contrôlable et traçable, et qu'elles génèrent ainsi de la confiance dans l'espace numérique. En raison de ces caractéristiques, des architectures qui implémentent ces principes sont également attractives pour l'échange standardisé de données.

3. L'architecture d'une fédération des données

L'architecture d'une fédération des données repose sur des principes d'architecture communs de la formation professionnelle (voir ci-dessous). Afin de pouvoir analyser le flux de données, un cas d'application concret sera développé. Celui-ci sera ensuite représenté dans la solution technique afin de mieux comprendre les relations entre les composants. Alors que l'architecture technique permet de décrire les processus et d'intégrer les composants, les responsabilités respectives doivent également être définies et les mesures y relatives doivent être localisées. Des détails supplémentaires sur la gouvernance de l'architecture globale figurent dans la deuxième partie du présent concept.

L'objectif de l'architecture est d'établir les structures de base de la fédération de données et de définir des règles qui coordonnent l'interaction dynamique de tous les composants. Le système de formation professionnelle doit être représenté le plus précisément possible. La discussion et la formulation des principes d'architecture permettent de créer les fondations des nouvelles structures.

Étant donné que dans une fédération de données, bien qu'il n'y ait pas d'organisation ou de service général qui contrôle ou coordonne toutes les activités mais qu'il existe toutefois une multitude d'institutions qui participent, cet arrangement peut tout à fait être qualifié d'écosystème. La nature décentralisée des composants clés de l'architecture garantit en outre un niveau élevé de résilience et de disponibilité, précisément là où elle est le plus nécessaire, afin d'ancrer la confiance dans cet écosystème numérique.

L'architecture proposée place les apprenantes et apprenants au centre de toutes les relations de confiance et leur permet de faire valoir leur autodétermination informationnelle. Il s'agit d'une contribution importante à la protection des données puisqu'il est possible, dans une mesure raisonnable, de procéder à une divulgation sélective. Il est même possible, sous

certaines conditions, de formuler des énoncés sans devoir divulguer la moindre donnée. Des descriptions plus détaillées à ce propos figurent dans la prochaine partie du présent rapport.

3.1 L'architecture en tant que cas d'application: justificatifs numériques dans la formation professionnelle initiale

Le justificatif numérique est un concept central de cette architecture. Il s'agit d'une ou plusieures déclarations, exprimées sous une forme numérique et vérifiable, d'une organisation au sujet d'une personne. En règle générale, ces justificatifs numériques sont considérés comme le pendant numérique des certificats et des diplômes physiques. Dans le cadre de la fédération des données, les transferts de données entre les parties ont lieu, le plus souvent, par la délivrance de justificatifs numériques aux détentrices et aux détenteurs. L'autorisation d'agir en tant qu'organisme émetteur est enregistrée sur le registre décentralisé des données. La détentrice ou le détenteur du justificatif numérique peut le transmettre à son tour à des tiers aux fins de vérification (sans impliquer l'organisme émetteur). Cela se fait grâce au registre décentralisé des données. De cette façon, il est possible de contrôler si le justificatif numérique provient d'un organisme émetteur autorisé. En règle générale, l'organisme émetteur délivre le justificatif numérique à la personne pour lequel le justificatif a été établi, de sorte que le sujet et la détentrice ou le détenteur sont généralement la même personne. Cela n'est toutefois pas toujours le cas, notamment lorsqu'un organisme émetteur délivre un justificatif numérique aux tuteurs légaux d'un enfant ; dans ce cas, les détentrices ou les détenteurs sont les personnes exerçant l'autorité parentale et le sujet est l'enfant.

Avant qu'un tel échange de données puisse avoir lieu, une relation de confiance doit être établie entre les parties qui communiquent entre elles. Cela se fait au moyen d'un registre décentralisé des données. Sur la base d'un registre décentralisé des données, des institutions de formation peuvent établir des justificatifs numériques et les apprenantes ou les apprenants peuvent les présenter. Grâce aux informations publiques que le registre contient, des tiers pourront procéder à une vérification des justificatifs. L'authenticité des justificatifs numériques est ainsi garantie par une procédure de codage cryptographique, ce qui instaure la confiance. Différentes technologies peuvent être envisagées pour la mise en oeuvre fiable d'un tel registre des données. Dans le présent exemple, il est prévu de faire usage d'une blockchain publique et permissionnée qui est un type de stockage décentralisé des données. Les organismes émetteurs auraient la permission d'écrire sur cette blockchain (d'où l'expression «blockchain permissionnée») et les vérificatrices et vérificateurs pourraient lire les données sur cette blockchain afin de vérifier les justificatifs (d'où l'expression «blockchain publique»). L'approche décentralisée du stockage des données – qui constitue l'un des concepts clés d'une blockchain – reflète et intègre le caractère fédéraliste du système de la formation professionnelle dans cette architecture.

Toutes les transactions exécutées via des nœuds correspond à des événements réels (voir aussi chapitre 3.2). Elles sont rendues abstraites par des identifiants décentralisés de telle sorte qu'elles ne peuvent être prouvées que par les personnes concernées. Pour cette raison, la charge de la preuve incombe à l'apprenante ou à l'apprenant (donc à la détentrice ou au détenteur) lorsqu'un justificatif numérique est présenté. L'instance de contrôle valide uniquement le justificatif numérique ainsi que les transactions qui y sont liées dans le registre

décentralisé des données. Lors de la vérification, aucune information sur les apprenantes ou les apprenants ne peut être consultée sur le registre décentralisé des données (voir graphique 2).

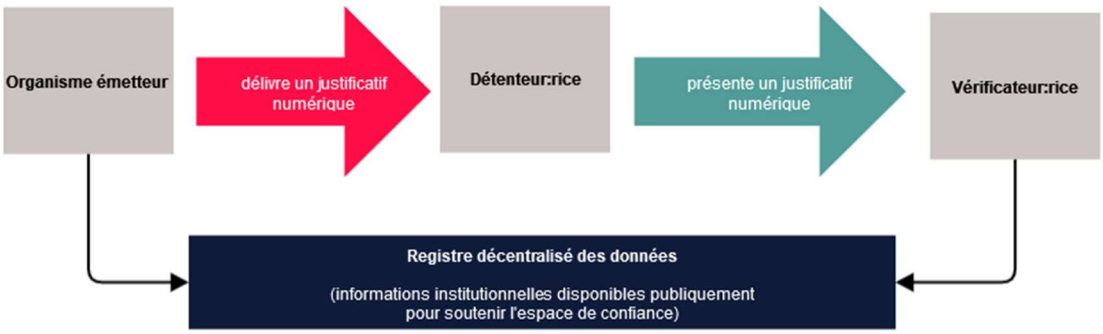

Figure 2 : Justificatifs numériques

Les justificatifs numériques de Luca

L'exemple de cas d'application suivant commence avec le contrat d'apprentissage d'un apprenant. Appelons-le Luca et supposons qu'il aimerait devenir menuisier avec certificat fédéral de capacité (CFC).

Luca reçoit l'approbation de son contrat d'apprentissage par l'office cantonal de formation professionnelle par voie ordinaire. Suite à l'approbation de son contrat d'apprentissage, l'office de formation professionnelle a l'intention de lui délivrer sa pièce d'identité pour apprenantes et apprenants, mais cette fois, sous forme numérique (voir graphique 3). Luca reçoit, par exemple par courrier postal ou par E-mail, un code QR généré pour lui qu'il scanne grâce à un portefeuille électronique. Il s'agit d'un logiciel d'application sur son téléphone mobile. Afin d'éviter que ce code QR ne puisse être scanné par une autre personne qui a pris possession de ce code QR, l'office de la formation professionnelle demande à Luca qu'il prouve son identité. Cela serait par exemple possible pour Luca de présenter un justificatif signé déjà disponible dans son portefeuille électronique et délivré par une source digne de confiance (par ex. une e-ID). Lorsque l'apprenant scanne le code QR et qu'il confirme son identité au moyen de l'e-ID, une paire de clés cryptographique est générée en arrière-plan. Cela signifie qu'une clé privée et une clé publique sont créées pour chaque partie. Ensuite, les clés publiques sont échangées. Il en résulte une connexion directe et cryptée entre l'office de formation professionnelle et Luca (relation de confiance dite peer-to-peer). Luca n'a donc plus besoin de créer un compte auprès de l'office de formation professionnelle. Par conséquent, les risques liés à la gestion d'identités d'un compte sont également supprimés (par ex. piratage pour voler des mots de passe). Une telle approche décentralisée qui permet d'établir une connexion

sécurisée entre les parties est ainsi nettement plus sûre que les approches classiques basées sur un service central (par ex. un prestataire d'identité (identity provider)).

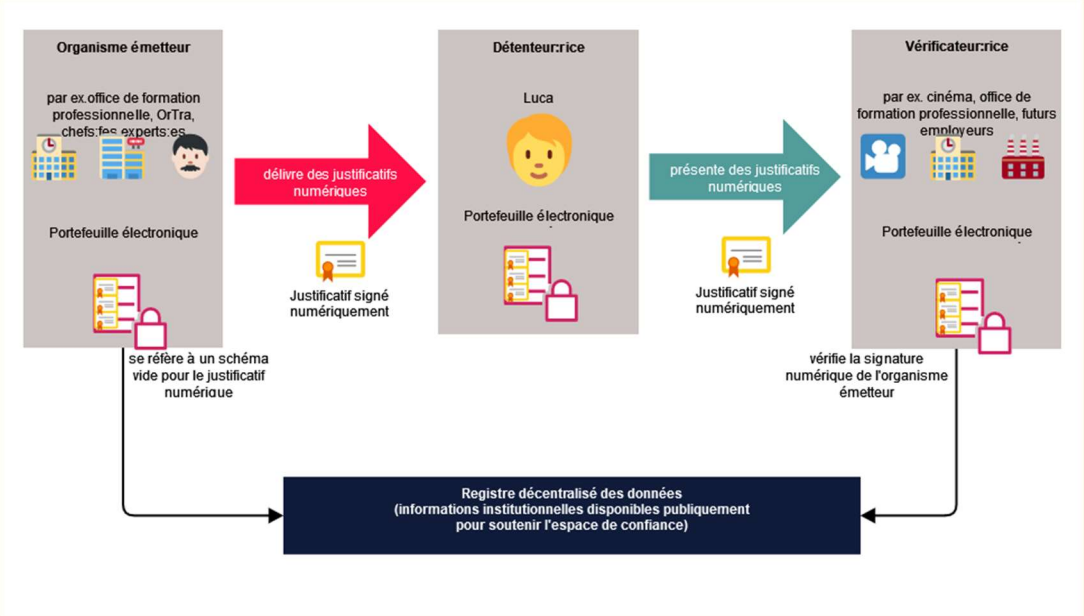

Figure 3 : Exemples de cas d'application de justificatifs numériques

Si, comme décrit ci-dessus, une connexion sécurisée a été établie entre deux acteurs, un justificatif numérique peut alors être transmis. Dans ce cas, l'office de formation professionnelle (organisme émetteur) pourrait transmettre à Luca (détenteur) la pièce d'identité numérique pour apprenantes et apprenants. Luca conserve cette pièce d'identité dans son portefeuille électronique (en sus d'autres justificatifs numériques). Luca peut par exemple présenter cette pièce d'identité lors d'une visite au cinéma afin d'obtenir un rabais. Ce faisant, Luca ne fait que confirmer qu'il est actuellement en formation. D'autres détails, comme par exemple la profession qu'il a choisie ou son âge, ne sont pas nécessaires, raison pour laquelle ils ne doivent pas être divulgués.

La validité de la pièce d'identité de Luca résulte des faits suivants :

- Elle a été établie par une source digne de confiance, à savoir par un organisme dûment autorisé à cet effet (par ex. par l'office de formation professionnelle),
- Elle n'a pas été révoquée entre-temps (étant donné qu'on ne trouve aucune transaction de révocation dans un registre décentralisé des données consultable par toutes les parties) ainsi que
- d'autres conditions pertinentes (elle découle, par ex., durée de validité de la preuve).

Pendant sa formation, Luca reçoit toutes les informations nécessaires, dont celles sur l'école professionnelle, sur les cours interentreprises et sur l'entreprise formatrice, sous forme de justificatif numérique dans son portefeuille électronique. Si Luca doit, par exemple suivre un cours interentreprise, il s'inscrit auprès de l'OrTra compétente au moyen des informations qu'il a déjà dans son portefeuille électronique sous forme de justificatifs numériques (par ex. pièce d'identité pour apprenantes et apprenants ou autres combinaisons de justificatifs numériques que Luca peut réunir dans un dossier). Il n'est donc pas nécessaire de remplir un formulaire. En revanche, comme ci-dessus, une relation sécurisée doit premièrement être établie avec l'office de la formation

professionnelle afin que l'identité de Luca soit confirmée. Par la suite, Luca reçoit un justificatif numérique qui confirme sa participation au cours. Ce justificatif est conservé tant dans le portefeuille électronique institutionnel de l'établissement de formation compétent que dans le portefeuille électronique personnel de Luca.

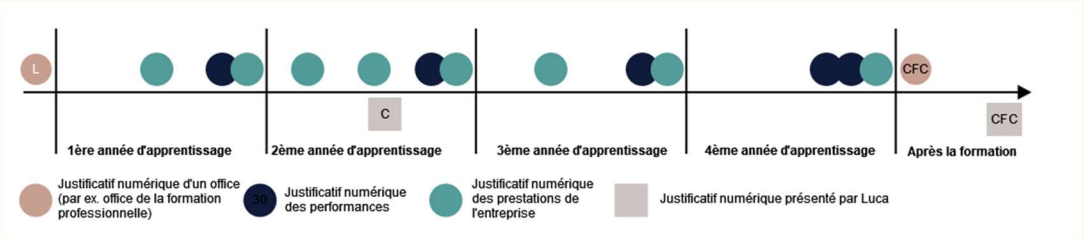

Figure 4 : Les justificatifs numériques dans un contexte scolaire, resp. dans un contexte d'entreprise

D'autres exemples suivent le même schéma : on s'assure premièrement qu'une relation de confiance directe, sécurisée et cryptée est établie. Ce n'est qu'ensuite que des déclarations sont communiquées sous forme de justificatifs numériques sur le parcours scolaire de Luca (voir les points bleus dans le graphique 4) ou au sein de l'entreprise (voir les points verts dans le graphique ci-dessus). Cela s'applique, par exemple, à l'établissement du carnet de notes, y compris les notes d'examens ou aux attestations de travaux pratiques individuels et ce, jusqu'aux certificats semestriels ou à l'obtention d'un certificat fédéral de capacité de « menuisier/menuisière avec CFC ».

La possession de ces justificatifs numériques rend possible une prise de décision entre plusieurs parties sans que l'intervention d'une autorité centrale soit nécessaire. Cela serait le cas dans la formation professionnelle, où des attestations de l'école professionnelle, des OrTra, des entreprises formatrices et des expertes et experts aux examens sont requises afin de pouvoir délivrer le CFC à Luca. En outre, une telle prise de décision peut être automatisée (par ex. en tenant compte des transactions enregistrées dans le registre décentralisé des données), ce qui permet de réduire considérablement les erreurs administratives et les retards.

La délivrance d'un CFC peut à son tour déclencher d'autres événements de transaction (par ex. la révocation d'une pièce d'identité pour apprenantes et apprenants). Une fois que les faits sont établis, ils peuvent être vérifiés à tout moment dans le futur (par ex. si, après sa formation, Luca postule à un poste auprès d'un employeur (voir le carré «CFC» dans le graphique ci-dessus). Si, en revanche, il devait s'avérer que le justificatif numérique qui a été établi n'était pas correct, cela peut être corrigé par la révocation de ce justificatif (à savoir par une transaction de révocation enregistrée dans le registre décentralisé des données).

Au final, le résultat est que l'apprenante ou l'apprenant est placé au centre d'un réseau de relations de confiance avec tous les sites de formation et autres institutions concernés. Elle ou lui reçoit des justificatifs numériques des événements importants de la procédure de qualification et peut, le cas échéant, présenter ces justificatifs afin de justifier leur validité. Toutes ces données sont conservées de manière sécurisée dans le portefeuille électronique personnel des acteurs concernés. Si une obligation étatique l'exige, ces justificatifs peuvent être également conservés par les autorités. Dès lors que les apprenantes ou apprenants exercent un meilleur contrôle sur leurs données, il est possible de mieux garantir le respect des dispositions sur la protection des données.

3.2 Le point de vue technique sur l'architecture

Une architecture comprend l'ensemble des processus, des applications et des banques de données ainsi que les infrastructures requises pour pouvoir soutenir un système d'exploitation donné. Ce chapitre décrit les éléments clés de l'architecture d'une fédération des données. D'un point de vue technique, une fédération des données est une association de différents systèmes qui conservent autant que possible leur autonomie et qui sont en mesure d'organiser leurs activités sans recours à un organe central de coordination. Sur le plan fonctionnel, l'architecture est décrite de façon à pouvoir comprendre les cas d'application susmentionnés et le traitement des justificatifs numériques. Une description plus détaillée du paysage des processus, des graphiques illustrant le fonctionnement des systèmes d'information et des infrastructures, figure dans la 3ème partie de la conception d'une fédération des données pour la formation professionnelle.

Le diagramme suivant décrit une architecture à quatre niveaux afin de présenter la mise en œuvre de la fédération des données proposée :

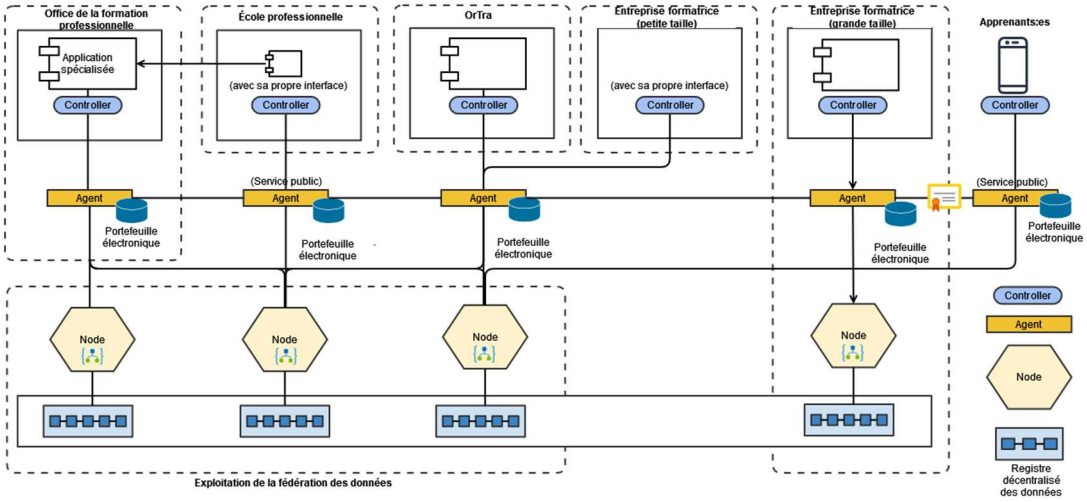

Figure 5 : Architecture à quatre niveaux de la fédération des données

D'un point de vue technique, les composants suivants revêtent une importance particulière :

- Les *controllers* permettent aux personnes et aux applications existantes d'interagir avec la fédération des données. Cela comprend la validation des entrées, la coordination des agents et le traitement des erreurs. Les controllers peuvent offrir une interface utilisateur (de manière similaire aux navigateurs qui permettent aux utilisatrices et aux utilisateurs d'interagir avec Internet) ou être intégrés ans des applications spécialisées déjà existantes.
- Les *agents* sont contrôlés par les controllers et traduisent les instructions qu'ils ont reçues en transactions ordonnées chronologiquement pour permettre l'exécution des transactions sur les nœuds. Les agents agissent en qualité d'intermédiaire entre différents programmes et ne nécessitent pas d'interactions humaines directes. Les

agents utilisent leurs propres banques de données (appelées «portefeuilles électroniques») afin de conserver des données personnelles, y compris les justificatifs. De tels portefeuilles électroniques existent sous différentes formes, que cela soit pour l'établissement et la gestion en masse de justificifatifs numériques (par exemple par des institutions) ou pour la gestion individuelle de justificatifs numériques (par exemple par des personnes privées).

- Les *nœuds* sont responsables de l'exécution des transactions dans un registre décentralisé de données. Chaque nœud gère sa propre copie du registre des données et l'actualise sur la base des activités de tous les nœuds. Ainsi, les nœuds forment un réseau qui a pour but de stocker de manière décentralisée des données transactionnelles. Il est également possible que les nœuds hébergent des contrats intelligents qui exécutent des transactions en suivant une logique prédéfinie. Un code de ce genre peut aussi envoyer des notifications à d'autres parties sous la forme d'événements.
- La *fédération des données* utilise un registre décentralisé des données. Il s'agit d'un registre accessible au public, non modifiable, qui enregistre et vérifie les justificatifs numériques. Un registre des données assume le rôle de Grand Livre et ne peut pas être comparé à une banque de données traditionnelle. Son unique objectif est d'enregistrer des transactions dans un ordre chronologique approprié et de protéger ces enregistrements contre toutes manipulations ultérieures. Par exemple, les clés publiques d'organismes émetteurs, les schémas des justificatifs numériques ou une liste de révocation de justificatifs numériques sont lisibles par tous. Il s'agit cependant toujours de données factuelles dépourvues de caractère personnel – les données personnelles ne sont pas mémorisées dans le registre décentralisé des données.

Il existe de nombreuses configurations imaginables pour une fédération des données. Le diagramme ci-dessus n'en présente que quelques-unes d'entre elles. Il se peut par exemple qu'une institution de formation gère ses propres agents ou qu'elle utilise un agent public (en tant que mandataire). De plus, les nœuds de la fédération des données peuvent être exploités par un prestataire de services techniques. Il se peut aussi que des organisations plus importantes souhaitent apporter leurs propres nœuds et traiteraient ainsi également le registre décentralisé de données.

Les systèmes existants devraient, dans la mesure du possible, être intégrés dans l'architecture susmentionnée (via des controllers/agents). C'est notamment le cas d'applications qui sont des sources déterminantes pour des données critiques (par exemple DBLAP2). Le transfert des données ne doit toutefois pas suivre les processus usuels qui sont en partie manuel. Il serait possible de prévoir des transactions automatisées déclenchées par des événements ou par une logique d'application interne déterminée (par exemples par contrats intelligents). Cela permettrait notamment de créer automatiquement un justificatif numérique. Ainsi, une architecture basée sur le déclenchement d'événements et l'établissement de connexions directes semble être une bonne approche.

3.3 Prise en compte des principes d'architecture

L'architecture proposée tient compte de plusieurs principes d'architecture (PA) qui doivent guider l'utilisation des données dans la formation professionnelle et qui a déjà été thématisée par la politique de formation. L'un des principes les plus importants est la promotion d'une gestion fédéraliste des données (principe d'architecture Gouvernance PA.G.1) qui considère les données comme un actif pour la prise de décision (PA.G.3). Pour ce faire, les événements pertinents d'un parcours de formation sont mis à disposition des parties concernées sous la forme de justificatifs numériques au moyen d'un registre décentralisé des données. Grâce à l'accessibilité simplifiée et sans délai de ces informations (PA.G.2) grâce à l'architecture, les données peuvent être échangées et utilisées efficacement. Le recours à des contrats intelligents en tant qu'instruments de prise de décision permet également d'améliorer la gouvernance en termes de transparence, de cohérence et de responsabilités (PA.G.4). C'est notamment le cas pour des transactions impliquant plusieurs parties (par exemple lorsqu'une école professionnelle et une entreprise formatrice prennent une décision en commun).

La confiance dans la formation professionnelle naît de toutes les parties (PA. G.5) et pas seulement des autorités. Celle-ci est instaurée par le contrôle d'identité (PA.S.2, PA.S.3) qui permet une communication directe entre les parties (PA.N.1), sans qu'il soit nécessaire que les données soient centralisées avant le traitement. Les mesures spécifiques de sécurité permettent de plus que seuls les acteurs légitimes participent à la fédération des données (PA.G.5, PA.S.1, PA.S.2 et PA.S.3). Ceux-ci assument la responsabilité des transactions exécutées en leur nom (AP.G.4, AP.D.1). Les mesures techniques peuvent être soutenues par d'autres instruments organisationnels (par ex. par des règlements d'organisation, des contrats destinés aux membres) et servir au respect de la conformité juridique en matière de protection des données (PA.S.5). Alors que par exemple la délivrance de certificats relève de la compétence des institutions de formation actuelles (par ex. l'office de la formation professionnelle, l'école professionnelle), leur vérification n'est pas seulement une question de remise de justificatifs numériques mais suppose également que la personne concernée soit habilitée à décider elle-même du type de données à transmettre. Il existe donc dans l'architecture des mécanismes qui prévoient une divulgation sélective des données. D'autres mécanismes aident à contrôler la conformité de la conservation des données personnelles conformément au droit de la protection des données.

En 2022, les principes d'architecture suivants ont été élaborés :

Réference	Principe architectural (PA)
Gouvernance dans le sens de structures de contrôle et de gestion : ce principe veille à ce que tous les transferts de données au sein de l'architecture dans son ensemble soient licites. Ce principe prévoit des mécanismes de contrôle et de gestion afin de garantir la légalité des transferts de données. La gouvernance a un caractère hiérarchique.	
PA.G.1	**Conservation fédéraliste des données :** La fédération des données n'a pas besoin d'un stockage centralisé des données et est organisée de manière fédéraliste. Il s'agit d'une association de systèmes individuels qui conservent leur autonomie.

PA.G.2	**Modèle de gouvernance consensuel :** cela sert à l'acceptation et à la légitimation des personnes autorisées à prendre des décisions de manière autonomne. Ce principe doit favoriser l'agilité et le fonctionnement du système, en aucun cas le freiner.
PA.G.3	**Disponibilité de données lors de prises de décision :** cela part du principe que les décisions sont prises correctement et légalement lorsque toutes les informations pertinentes sont disponibles.
PA.G.4	**Transparence, intégrité et responsabilité concernant les événements pertinents liés à la formation** : ce principe garantit la révocabilité des données à tout moment. La révocabilité des données résulte d'un besoin potentiel légitime de la personne concernée qui doit obéir à des règles claires.
PA.G.5	**Accès légitimé à la fédération des données :** seuls ceux qui ont un intérêt légitime pour la fédération peuvent y accéder. Ce qui constitue un intérêt légitime doit être clairement défini. Les rôles des différentes participantes et des différents participants doivent être définis, leurs droits dans le système doivent être décrits en conséquence.
Sécurité dans sens d'état général, qui est jugé aussi exempt que possible de risques (inacceptables). La sécurité poursuivie ici est la sécurité objective (par opposition à la sécurité subjective), qui identifie les risques, évalue correctement leur potentiel de danger et les contrecarres efficacement.	
PA.S.1	**Contrôle d'accès basé sur les rôles (RBAC) :** cela permet de garantir que seules les personnes autorisées (légitimées - voir AP.G.5) ont accès ou peuvent effectuer des modifications dans le cadre des rôles qui leur sont attribués.
PA.S.2	**Garantie de l'identité des composantes :** Sécurité de l'identité des composants : L'accès à la fédération de données doit être limité aux participants dignes de confiance, qui doivent être en mesure de prouver leur identité. Ce n'est qu'ainsi que les principes nécessaires à la protection des données peuvent être respectés. Ce principe d'architecture est une condition pour accéder à la Fédération.
PA.S.3	**Garantie de l'identité des parties concernées :** au-delà de la garantie générale de l'identité de l'AP.S.2, ce principe architectural assure la légitimation confirmée de la partie concernée. Ce seuil supplémentaire est important car la partie concernée est en quelque sorte propriétaire des données traitées. Elle est la seule à pouvoir requérir des modifications (traitements).
PA.S.4	**Monitoring continu des échanges de données :** des instruments/mécanismes sont intégrés dans le système afin de contrôler et de garantir en permanence la mise en œuvre du PA.S.1-3. Ce PA tente de détecter et d'empêcher les abus en amont.
PA.S.5	**Échange de données plus sécurisé et davantage conforme à la législation :** conformément à ce principe architectural, un «state of the art» technologique est fixé comme référence. Cela permet de garantir des normes élevées.
Données dans le sens de signes contenant une information. Les informations importantes dans ce contexte sont celles qui concernent les personnes dans le domaine de l'éducation.	
AP.D.1	**Les données sont un bien commun au service de la formation professionnelle :** les apprenantes et les apprenants sont les seuls propriétaires des données dans l'éducation. Ce principe exprime la nécessité que les données dans l'éducation ne soient pas accumulées par des services administratifs isolés. Toutes les institutions de formation (entreprises formatrices, écoles professionnelles, etc.) sont au service d'un traitement efficace et ciblé des données. En tant qu'émetteurs de documents d'identité, ces composantes ne doivent pas faire obstacle au système. .

AP.D.2	**Les composantes ont pour tâche de garantir la qualité des données :** la responsabilité individuelle et la responsabilité en matière de protection des données se distinguent. En identifiant de manière sûre les composants (PA S.2-3) et en leur attribuant un rôle clair (PA S.1), il est possible d'identifier rapidement les responsabilités en cas d'erreur. Ainsi, chaque composant doit agir sous sa propre responsabilité au sein de la fédération.
PA.D.3	**Interprétabilité sémantique des données :** ce principe garantit l'interprétation correcte des informations (données); il suppose un langage commun et la possibilité de traduire à tout moment des termes différents dans le langage commun. L'interprétabilité sémantique des données assure la continuité, c'est-à-dire que les mêmes significations sont toujours données aux données saisies de la même manière.
Réseau dans sens d'un système avec une structure sous-jacente qui peut être modélisée et qui contient des mécanismes d'auto-organisation.	
PA.N.1	**Échange direct de données :** il est important que les différents composants du réseau puissent communiquer directement entre eux (sans intermédiaire). Ce n'est qu'ainsi qu'ils peuvent réagir en temps utile et de manière efficace. L'échange direct de données soutient également l'intégrité des données, dans la mesure où les données sont directement traitées par l'initiative de l'utilisatrice ou de l'utilisateur (sans intermédiaires). Ce principe suppose que les PA S.1-5 soient respectés à tout moment.
PA.N.2	**Atomicité d'événements de formation :** il est important d'imaginer et de mettre en œuvre des règles qui définissent quand, dans quelle mesure et dans quel but contraignant les événements éducatifs peuvent être communiqués ouvertement. Cela permet de garantir la tenue à jour des données à tout moment.
PA.N.3	**Asynchronité de la communication :** le moment où les données sont transmises n'a pas d'importance pour le réseau; les données doivent pouvoir être consultées à tout moment par tout le monde. Dès qu'elles sont saisies, elles s'appliquent.

4. Conclusion intermédiaire

Cette première partie explique l'importance de l'implication des partenaires, la procédure ainsi que le point de départ pour la conception d'une fédération des données pour la formation professionnelle. En outre, l'architecture conçue à cet effet est présentée sous différents angles. D'une part, les cas d'application aident à comprendre cette thématique et, d'autre part, à fournir des informations techniques plus approfondies. Le concept repose sur des principes d'architecture qui ont été discutés avec les partenaires. Il a pour but de simplifier l'accès aux données et de garantir la sécurité des échanges. Dans ce contexte, les données sont considérées comme une ressource stratégique d'importance qui présente un grand potentiel d'innovation pour la formation professionnelle. L'objectif final poursuivi est de pouvoir échanger efficacement des données dans toute la Suisse, de les rendre accessibles grâce à un espace de confiance et utilisables par l'ensemble du système, respectivement par l'ensemble des partenaires.

Les objectifs du projet, conformément au mandat reçu, ont pu être repris dans le concept d'architecture. Le cadre futur du potentiel d'innovation pour la formation professionnelle initiale a été renforcé. En outre, les exigences des apprenantes et apprenants ainsi que celles du monde du travail ont été plus fortement considérées dans le concept de fédération des

données. Enfin, le concept d'architecture tient compte des besoins liés à la nouvelle Loi sur la protection des données.

Ce document est la première des trois parties du concept. Il porte sur l'architecture technique d'une future fédération des données. Les deux autres parties traitent du modèle de gouvernance nécessaire à cette architecture et des bases légales éventuellement nécessaires à sa mise en œuvre. Conjointement avec la planification de réalisation (notamment le calendrier de mise en œuvre et la budgétisation), les trois parties de ce concept constituent ensemble le dossier de décision destiné au SEFRI et à la CDIP.

Annexe

Définitions

Concept	Définition
Agent	Un «agent», au sens technique du terme, est un composant d'un logiciel qui traite les instructions d'un ou de plusieurs «controllers», qui les rediffuse en conséquence au sein de la fédération des données, et qui communique avec le node ou qui mémorise les justificatifs numériques (et, de ce fait, les données personnelles).
Organisme émetteur	Par le terme d'organisme émetteur (ou «issuer»), on entend un organisme qui, par ex., est autorisé par une institution étatique à établir des justificatifs numériques.
Exploitante / exploitant	Il peut s'agir d'institutions, de personnes privées ou de personnes morales qui exploitent les nodes.
Blockchain	Un système de stockage de ledger dans lequel les données de transaction sont attachées et ne peuvent pas être modifiées ultérieurement. Les blockchains sont réalisées via un réseau de node, raison pour laquelle elles sont, en règle générale, décentralisées.
Controller	Par le terme de «controller», on entend un composant d'un logiciel qui permet à des personnes et à des institutions d'accéder à l'«agent», via une interface-utilisateur, ce qui donne ainsi accès à des justificatifs numériques et permet d'utiliser le registre décentralisé des données, resp. de communiquer avec d'autres membres ou participantes et participants de la fédération des données.
Registre décentralisé des données	Un registre décentralisé des données est une mémoire publiquement accessible et lisible par toutes les actrices et tous les acteurs. Il contient toutes les informations nécessaires pour pouvoir exploiter une fédération des données en général, et un écosystème destiné à l'établissement de justificatifs numériques en particulier.
Identificateurs décentralisés	Il s'agit d'une forme d'identification qui est utilisée dans des systèmes décentralisés. Elle rend possible la mise en liens d'événements avec des données de telle sorte que seules les parties concernées puissent établir ces mises en liens.
Portefeuille électronique	Par le terme de portefeuille électronique, on entend un logiciel d'application qui sert à conserver de manière sécurisée des justificatifs numériques.
Détentrice / détenteur	Par le terme de détentrice ou détenteur, on entend une personne ou une institution qui possède électroniquement un justificatif numérique.
Contrats intelligents	Les contrats intelligents sont des programmes informatiques qui traitent des données de manière automatisée et qui, sur cette base, ont pour effet de déclencher une prise de décision automatisée selon des critères prédéfinis.
Node	Un node (nœud) est une instance de traitement du registre décentralisé des données. Les nodes réalisent, contrôlent et mémorisent des transactions sur le registre décentralisé des données et peuvent en outre également héberger des contrats intelligents.

Membre	Il s'agit d'une partie institutionnelle de la fédération des données qui est autorisée à collaborer de manière active (par ex. création de transactions). De telles parties sont tenues de signer un contrat avec l'exploitante/avec l'exploitant de la fédération des données.
Participants	Il s'agit d'une partie de la fédération des données qui est autorisée à collaborer de manière passive (par ex. consultation d'informations publiques dans le registre décentralisé des données ou vérification des justificatifs). En règle générale, de telles parties ne doivent pas avoir signé de contrat avec l'exploitante/avec l'exploitant de la fédération des données.
Transaction	Une transaction est une opération qui permet de reprendre des données dans le registre décentralisé des données. Les transactions ne contiennent aucune donnée personnelle.
Vérificatrice / vérificateur	Par le terme de vérificatrice ou vérificateur, on entend des personnes ou des institutions qui vérifient la validité des justificatifs numériques qui leur sont présentées au moyen d'informations disponibles sur le registre décentralisé des données.
Justificatif numérique (vérifiable)	Dans le présent contexte, un justificatif numérique contient un ou plusieurs énoncés d'une institution ou d'une autorité (par ex. d'une école ou d'une instance étatique) à propos d'un sujet donné (par ex. à propos d'une apprenante ou d'un apprenant). Il est possible de vérifier si ces déclarations proviennent d'une source fiable à l'aide des données figurant sur le registre décentralisé des données – d'où l'expression «justificatif numérique vérifiable».

educa

Erlachstrasse 21
3012 Bern

+41 31 300 55 00
info@educa.ch

Concept de la fédération des données dans la formation professionnelle
2ème partie – Nouveau paradigme des données et modèle de gouvernance

10.1.2023

Digitaler Bildungsraum Schweiz
Espace numérique suisse de formation
Spazio formativo digitale svizzero
Spazi da furmazium digital svizzer
Swiss digital education space

educa

Contenu

Management Summary **3**

1. Introduction **5**
1.1 Récapitulation de l'architecture technique 5
1.2 Collaboration avec les parties prenantes 6

2. Nouveaux paradigmes des données **6**
2.1 Regard sur la nouvelle Loi sur la protection des données 7
2.2 Données personnelles: où sont-elles dans l'architecture de la fédération des données ? 8
2.3 Stratégies de gestion pour un traitement juridique conforme et encadrant les toxic data 10
2.4 Cas d'application: les offices et les sites d'apprentissage face aux données après une interruption de l'apprentissage 16

3. Modèle de gouvernance pour la fédération des données **17**
3.1 Modèle organisationnel: rôles et instruments dans la fédération des données 19
3.2 Importance d'une infrastructure des données pour la qualité et l'efficacité du système éducatif 21
3.3 Gestion des données de base 22
3.4 Qualité des données 23
3.5 Sécurité informationnelle et protection des données 25
3.6 Intégration des données et interopérabilité 26

4. Conclusion intermédiaire **27**

Annexe **28**
Définitions 28

Management Summary

Dans la première partie du concept de la fédération des données dans la formation professionnelle, nous avons mis l'accent sur la future infrastructure technique. Ce faisant, nous avons souligné les aspects de l'amélioration de l'efficacité, de la disponibilité des données et de la connectivité. La deuxième partie du concept présente une autre perspective. Outre la nécessité d'améliorer l'efficacité à cause de l'actuel changement de paradigme dans la gestion des données personnelles, il existe un autre facteur qui joue désormais un rôle moteur essentiel et qui explique pourquoi l'échange des données dans la formation professionnelle initiale devrait être repensé.

Suite aux progrès de la transformation numérique, les « données » sont en effet devenues un thème sociopolitique, en particulier lorsqu'il s'agit des données personnelles. La nouvelle Loi fédérale sur la protection des données annonce à cet égard un renforcement des obligations imparties aux responsables des données. Dans ce contexte, ils seront tenus d'informer les personnes concernées de manière appropriée sur la collecte et le traitement de leurs données personnelles. À l'avenir, la personne concernée verra son pouvoir décisionnel nettement renforcé. Parallèlement, la gestion des données personnelles doit être évaluée encore plus fortement à partir d'un point de vue individuel. Une future fédération des données, vraisemblablement organisée sous la forme d'une institution de droit public, devra tenir compte de ces changements au vu de leur potentiel d'innovation. Le cas d'application intitulé « Les offices et les sites d'apprentissage face aux données après une interruption de l'apprentissage » montre comment une telle procédure peut être concrétisée au sein de la fédération des données. On étude ici notamment la question des toxic data. Par toxic data, on entend des données personnelles qui sont traitées, au sens le plus large du terme – volontairement ou non – de manière non conforme à la loi. De nombreux acteurs chargés du traitement des données dans la formation professionnelle initiale rendent cet aspect particulièrement pertinent. L'architecture technique proposée aidera à élaborer des stratégies qui permettront de maîtriser les aspects juridiques requis pour une gestion des données conformes à la future loi sur la protection des données.

Par conséquent, l'infrastructure technique d'une fédération des données dans la formation professionnelle requiert un mode de gouvernance adéquat. Pour ce faire, il y a lieu de mettre en œuvre un modèle de gouvernance des données. Ce modèle est un cadre général qui décrit comment les organisations membres gèrent, échangent et utilisent leurs données. Ce cadre comprend des règles, lignes directrices, processus et systèmes destinés à la saisie, l'enregistrement, l'évaluation, l'archivage, la communication et l'effacement de données, ainsi que des lignes directrices permettant de garantir la qualité de ces données. Le modèle de gouvernance des données a donc pour but de permettre à tous les acteurs impliqués de la fédération d'utiliser les données avec cohérence et efficacité, et de les échanger de façon sécurisée.

Le concept de la fédération des données dans la formation professionnelle s'inscrit dans un paysage d'espaces de données fiables, organisé selon les principes du fédéralisme, qui est actuellement en cours de développement. L'interopérabilité entre ces espaces de données est structurée à l'échelon national ainsi que, de plus en plus, également à l'échelon international.

Ce document est la deuxième des trois parties du concept. Conjointement avec une planification de réalisation (notamment le calendrier de mise en œuvre et la budgétisation), les trois parties du concept constituent ensemble le dossier destiné au Secrétariat d'Etat à la formation, à la recherche et à l'innovation (SEFRI) et à la Conférence des directrices et directeurs cantonaux de l'instruction publique (CDIP).

1. Introduction

La première partie du concept expliquait l'importance de l'implication de l'ensemble des partenaires, le mode de procédure ainsi que la situation initiale pour la conception d'une fédération des données dans la formation professionnelle. En outre, la conception de l'architecture des données a été présentée sous différents angles. Cette deuxième partie du concept propose un modèle de gouvernance pour la fédération des données qui est conforme aux exigences de la future législation sur la protection des données. Sur la base d'un cas d'application en lien avec des toxic data, le potentiel de l'architecture proposée est étudié en mettant en évidence des stratégies de résolution de problèmes correspondantes.

1.1 Récapitulation de l'architecture technique

L'objectif de cette architecture consiste à fixer les structures de base de la fédération des données et à définir les règles qui coordonnent les interactions dynamiques entre tous les composants. Ce faisant, il y a lieu de répliquer le système de la formation professionnelle avec le maximum de précision possible. Une architecture comprend la totalité des processus, applications et bases de données, ainsi que les infrastructures qui sont requises afin de soutenir un modèle d'exploitation déterminé. L'architecture technique de la fédération des données esquissée dans la première partie du concept est organisée de manière décentralisée (voir chapitre 3 de la première partie).

Un aspect essentiel de l'approche architecturale du concept réside dans la notion de justificatif numérique vérifiable. Il s'agit d'un ou plusieurs énoncés, exprimés au sujet d'une personne, ces énoncés étant consignés sous une forme numérique et vérifiable. En outre, les justificatifs numériques se basent sur une relation de confiance qui est établie entre les parties qui communiquent. Les participantes et participants de la fédération des données reçoivent les justificatifs numériques et peuvent les transmettre.

Du point de vue technique, une fédération de données est un regroupement de différents systèmes qui, dans ce contexte, gardent en majeure partie leur autonomie et qui, en l'absence d'une instance centrale de coordination, peuvent organiser leurs activités. À cet égard, ce sont notamment les composants suivants qui ont une importance décisive: le controller, l'agent, le node et le registre décentralisé des données.

Par la discussion sur les principes d'architecture et au travers de leur formulation, on crée ainsi le fondement des nouvelles structures. L'architecture proposée tient compte de plusieurs principes architecturaux qui doivent piloter l'utilisation des données dans la formation professionnelle. L'un de ces principes architecturaux les plus importants consiste à favoriser une conservation fédéraliste des données (principe de gouvernance de l'architecture) qui considère les données comme un actif pour la prise de décision. De plus, ces principes d'architecture ont été élaborés en mettant l'accent sur la sécurité, le type et la qualité des données et en adoptant un système en réseau.

1.2 Collaboration avec les parties prenantes

Alors que, dans la première partie du concept, nous avons intégré dans une forte mesure les connaissances de l'ensemble des partenaires au sein de la formation professionnelle, la thématique de la deuxième partie ne traite qu'en partie de manière spécifique de la formation professionnelle initiale. Grâce à la collaboration avec le Centre suisse de services Formation professionnelle (CSFO), nous avons pu rédiger en particulier de manière détaillée le cas d'application présenté ici intitulé « Les offices et les lieux d'apprentissage face aux données après une interruption de l'apprentissage » (voir chapitre 4). En revanche, le thème d'importance centrale, à savoir celui de la protection des données notamment son implémentation dans des architectures telles que celle de la fédération des données, est un axe prioritaire d'analyse de notre agence spécialisée. Educa possède en effet une expertise significative dans ce domaine, tout comme sur la question de savoir quelle doit être la configuration idéale d'un modèle de gouvernance des données destiné à une infrastructure de données décentralisée dans un contexte de formation fédéraliste. En outre, en étroite collaboration avec l'office du Préposé à la protection des données du canton de Zurich, nous avons pu échanger sur le potentiel de la fédération des données sur la protection des données personnelles dans la formation professionnelle. En particulier, l'approche que nous avons poursuivie a été d'examiner des solutions aux traitements de données non conformes et qui pourraient engager la responsabilité des acteurs concernés. Pour ce faire, il faut élaborer des stratégies de résolution de problèmes qui peuvent être mises en œuvre au moyen de la fédération des données (voir chapitre 2.3).

L'architecture de la fédération des données est conforme aux normes internationales qui ont standardisé l'approche pour les justificatifs numériques, ainsi que pour les identificateurs décentralisés qui en font partie. Le respect de ce type de normes garantit un degré élevé de cohérence technique et d'interopérabilité avec d'autres écosystèmes qui sont en train de se développer en Suisse (voir chapitre 3.6).

Au vu de l'évolution des espaces de données de confiance et de la garantie de connectivité que doit apporter une future fédération de données dans la formation professionnelle, une collaboration a été initiée avec le secteur « Transformation numérique et gouvernance de l'informatique » (TNI) de la Confédération. Du côté de la Confédération, l'intérêt dont elle fait preuve pour la fédération des données doit être mis en lien avec le projet majeur « Écosystème DataHub du TNI ». Il existe une forte congruence en matière d'objectifs et de méthodologie entre ces deux projets majeurs, avec un potentiel de synergies correspondant. À l'heure actuelle (état au 10.01.2023), des entretiens ont lieu sur les modalités de cette collaboration.

2. Nouveaux paradigmes des données

Le rôle que jouent les données au sein de la société a changé, tout comme la manière dont les données sont gérées. L'époque où les données étaient quelque chose qu'il fallait rassembler et gérer de manière centralisée est une époque révolue. En lieu et place de ce qui précède, on constate de plus en plus que les données sont mises à disposition de manière décentralisée, et qu'elles sont agrégées tant dans des formats structurés que non structurés. Afin de pouvoir les exploiter – par exemple en tant que base de décision – elles doivent être

préparées en conséquence. Cela permet en outre à plusieurs parties de prendre des décisions au niveau local sans dépendre de l'intermédiation d'une unité de coordination centrale. Les données sont toujours plus souvent exploitées pour en tirer des déductions ou des pronostics. Cela signifie qu'on n'utilise pas seulement les données pour établir des rapports, mais qu'elles sont aussi utilisées en temps réel. De nouveaux paradigmes de gestion des données s'appliquent dans de tels cas. Ces nouvelles approches nécessitent de pouvoir disposer d'une architecture qui offre un potentiel d'innovation et qui permet l'expérimentation (par ex. la création d'un jeu de données afin de tester des approches analytiques des données).

Compte tenu du nombre croissant de productrices et producteurs ainsi que de consommatrices et consommateurs de données et aussi au vu de la diversité croissante de leurs besoins de traitement de données, une gestion centralisée des données axée sur les besoins d'une minorité d'acteurs devient toujours moins justifiable. Dans un monde numérique de plus en plus décentralisé, la responsabilité partagée et la gouvernance polycentrique revêtent un rôle plus important. Ces développements remettent en question l'idée selon laquelle la gouvernance constitue soit un processus défini de manière uniforme, soit relève du domaine de compétence d'une autorité centrale unique (gouvernance monocentrique). Bien au contraire, des décisions sont prises en commun par différents acteurs qui ont besoin d'un accès à des sources de données différentes. S'y ajoute le fait que les dispositions de la nouvelle Loi sur la protection des données (nLPD) prévoient des contrôles nettement plus stricts pour l'enregistrement, le traitement, la conservation et l'effacement des données. La nLPD mettra en outre les institutions sous pression et les contraindra à gérer avec grand soin leurs données. Il faudra respecter non seulement le principe de minimisation des données, mais encore celui du renforcement de l'autodétermination informationnelle individuelle. Avec les solutions actuelles, il ne sera que difficilement possible de mettre en œuvre ces deux principes. De nouvelles obligations et de nouvelles responsabilités prendront ainsi naissance qui rendront nécessaires des changements du côté de la formation professionnelle.

2.1 Regard sur la nouvelle Loi sur la protection des données

Le texte de la nouvelle Loi sur la protection des données (nLPD) qui entrera en vigueur le 1er septembre 2023 implique de nouveaux enjeux pour les responsables du traitement des données. Tant le secteur public que le secteur privé doivent s'y préparer. La révision a notamment pour but de renforcer l'autodétermination informationnelle et la sphère privée des personnes concernées. Elle ne se rapporte qu'aux personnes physiques dont les données sont traitées. La révision consolide donc les droits des personnes concernées et, parallèlement, étend les obligations des responsables du traitement, en distinguant toujours le statut d'organe fédéral de la personne privée. Les responsables des données doivent anticiper la mise en place de nouvelles obligations, notamment par de nouvelles mesures organisationnelles et techniques. Il appartient en effet aux responsables des données, particulièrement aux organes publics, de sauvegarder les droits de la personne concernée. Ils doivent identifier les risques sur la base d'une analyse d'impact, tout en assurant la sécurité des données afin de garantir la confidentialité, la disponibilité, l'intégrité et la traçabilité des données. Les projets numériques devront être conçus conformément aux principes « Privacy by Design » et

« Privacy by Default ». Autrement dit, le responsable du traitement des données doit standardiser la protection des données dès la conception.

Afin de garantir notamment la transparence du traitement, l'article sur le droit d'accès a été mis à jour et complète davantage le devoir d'information. Ce nouvel article liste un certain nombre d'informations supplémentaires qui doivent être communiquées à la personne concernée lorsqu'elle demande l'accès à ses données, notamment la durée de conservation des données personnelles. La nLPD intègre également de nouvelles notions comme celles de profilage et de décisions individuelles automatisées, auxquelles on relie des obligations telles que le devoir d'information. Le profilage, par exemple par un organe fédéral, implique le consentement exprès de la personne concernée. Les responsables des données devront également tenir un registre dans lequel seront répertoriées toutes les activités de traitement et le déclarer au Préposé fédéral à la protection des données et à la transparence (PFPDT), dans tous les cas lorsqu'il s'agit d'organes fédéraux. De même, les responsables des données auront l'obligation d'annoncer les cas de violation de la sécurité des données présentant un risque élevé pour la personnalité ou les droits fondamentaux de la personne concernée. À choix pour les personnes privées, et de manière obligatoire pour les organes fédéraux, les responsables des données devront désigner un conseiller à la protection des données. Finalement, la nLPD complète les dispositions pénales, notamment la violation du devoir de discrétion, en chiffrant les amendes à CHF 250'000 au plus.

2.2 Données personnelles: où sont-elles dans l'architecture de la fédération des données ?

La conception de la fédération de données doit tenir compte des droits et des obligations imposés par la nLPD afin d'assurer la conformité des communications de données. Le stockage des données personnelles présente un intérêt particulier. L'architecture présentée dans le premier rapport tient compte des principes « Privacy by Design » et « Privacy by Default ». Ces principes représentent des obligations de diligence inscrites dans la nLPD et imposent au responsable du traitement des données de mettre également en place les mesures techniques nécessaires pour assurer la protection des données et pour les sécuriser.

Avant d'apporter des précisions sur le cycle de vie des données dans le concept de fédération des données, il est important de retenir les points suivants:

- Le registre décentralisé des données ne contient pas de données personnelles.
- Seuls des acteurs autorisés peuvent ajouter des transactions au registre décentralisé des données (blockchain publique et permissionnée).
- La communication entre les parties est cryptée.
- Chaque acteur (organisme émetteur, détenteur, vérificateur) est responsable du respect de la protection des données qu'il traite, qu'il communique et qu'il sauvegarde directement ou par l'intermédiaire de son controller, de son agent et de son portefeuille électronique.

Dans le concept de fédération des données, les données sont traitées par:

- l'utilisateur (organisme émetteur, détenteur, vérificateur): lorsqu'il communique des données au controller ou en prend connaissance par son biais.

- le controller: lorsqu'il communique les données de l'utilisateur à l'agent, et de l'agent à l'utilisateur.
- l'agent: par exemple lorsqu'il les génère, les collecte ou les communique.
- le portefeuille électronique: lorsqu'il conserve les données.

Ces trois derniers acteurs peuvent être hébergés localement ou sur le cloud. Le choix, par exemple, d'un cloud agent implique d'examiner les questions juridiques habituelles, notamment celles qui se posent concernant la conformité du traitement des données.

La communication de données repose sur l'établissement d'une relation de confiance, soit d'une connexion directe et cryptée. Celle-ci est possible grâce à une clé privée et une clé publique générées par chaque agent. Seule la clé publique est échangée entre les agents. Sans cette connexion, les acteurs ne peuvent pas communiquer entre eux. Ces données sont uniques à la connexion établie, renouvelables, uniquement connues et accessibles par les parties à la relation de confiance. Cette relation de confiance Peer-to-Peer permet d'assurer la confidentialité, la disponibilité, l'intégrité, mais également la traçabilité des données communiquées.

Les données communiquées se trouvent dans les justificatifs numériques. Le contenu d'un justificatif numérique dépend du type de données requises par le schéma. Les schémas ne contiennent aucune donnée personnelle, mais indiquent le type de données personnelles qui doivent être saisies, à l'image d'un formulaire vierge. Les schémas sont établis par des personnes autorisées (membres de la fédération des données) et sont enregistrés sur le registre décentralisé des données. Les schémas permettent d'assurer le respect du principe de proportionnalité et d'économie de données en standardisant les données et en limitant les données au minimum nécessaire. Les données personnelles sont ensuite communiquées par l'organisme émetteur à son controller, puis transmises par ce dernier à l'agent afin qu'un justificatif soit généré. Le degré de sensibilité des données contenues dans un justificatif numérique dépend du type de données qui doivent être saisies selon le schéma y afférent. Lors de la communication d'un justificatif numérique, les données personnelles sont décryptées par l'agent destinataire, conservées dans le portefeuille électronique et communiquées au détenteur par le controller. Aucune donnée personnelle n'est enregistrée dans le registre décentralisé des données. À ce stade, seul le détenteur et l'organisme émetteur ont connaissance du contenu du justificatif numérique.

La détentrice ou le détenteur d'un justificatif numérique peut à son tour le transmettre, par le biais d'une présentation, à une autre personne après qu'une relation de confiance cryptée soit établie. Le vérificateur a la possibilité de vérifier l'authenticité du justificatif numérique en comparant la clé publique de l'organisme émetteur indiqué sur le justificatif numérique avec celui enregistré sur le registre centralisé des données. Cet identificateur décentralisé public identifie un organisme émetteur, et n'est donc pas une donnée personnelle au sens de la nLPD.

Grâce à cette architecture et notamment aux différentes stratégies présentées au chapitre suivant, le concept de la fédération des données permet de renforcer les droits de la personne concernée en lui offrant davantage de contrôle sur ses données. Il permet également aux responsables du traitement des données de gagner en efficacité et de se conformer à la nLPD, notamment en évitant la toxicité des données.

2.3 Stratégies de gestion pour un traitement juridique conforme et encadrant les toxic data

Sur la base de l'architecture de la fédération des données dans la formation professionnelle, les données peuvent être échangées avec une efficacité accrue et de manière plus sécurisée. Le chapitre suivant présente des stratégies appropriées, explique comment les données peuvent être échangées de manière conforme à la loi au sein de la fédération des données, et selon quels principes ce traitement des données est réalisé.

Par toxic data, on entend des données personnelles qui ont été traitées, au sens le plus large du terme – volontairement ou non – de manière non conforme à la loi.

Les traitements de données ci-après constituent des exemples de création de toxic data. Il s'agit de traitements de données qui:

- sont mis en œuvre sans le consentement des intéressées et intéressés ou sans base légale suffisante (par ex. transmission non autorisée de données) ;
- outrepassent l'objectif poursuivi requis (par ex. rétention de données) ;
- dépassent un délai imparti par la loi (par ex. délai d'archivage de données manqué).

Comme l'illustre le dernier exemple, le traitement des données non conforme à la loi peut aussi avoir pour effet que des données auparavant traitées de manière conforme à la loi deviennent toxiques si, au fil du temps, elles sont conservées en dépassant les délais impartis.

Si la problématique des toxic data n'est certes pas limitée au domaine de la formation professionnelle, elle y joue néanmoins un rôle important. Premièrement parce que les données personnelles des apprenantes et des apprenants sont réparties entre différents cantons (lieu de l'établissement scolaire, lieu de l'apprentissage, cantons d'examens), entre les écoles professionnelles, les organisations du monde du travail (OrTrA), les entreprises formatrices ou les cheffes-expertes et les chefs-experts. Comme les structures de ces acteurs sont très hétérogènes, les mesures de protection des données qu'ils ont mis en œuvre et les types de traitements de données qu'ils appliquent sont, eux aussi, très différents. Deuxièmement, lors du traitement des données dans la formation professionnelle, il faut respecter une multitude de délais impartis par la loi pour éviter que ces données ne deviennent toxiques.

Les exigences croissantes assignées au traitement des données posent posent un défi également aux entreprises formatrices, aux écoles professionnelles et aux OrTra de petite taille. De surcroît, avec la nouvelle loi sur la protection des données, le thème des toxic data gagnera encore en importance (voir chapitre 2.1). Dès lors, la fédération des données décrite dans la première partie du concept (novembre 2022) se focalise aussi sur la question de savoir comment on peut éviter l'apparition de toxic data. Cela se fait, d'une part, déjà au niveau de l'architecture via l'introduction du principe « Privacy by Design » (voir chapitre 2.2). D'autre part, la fédération des données offre aux acteurs concernés des stratégies de résolution de problèmes qui empêchent ou réduisent fortement la probabilité de survenance de toxic data.

Les stratégies proactives de résolution de problèmes les plus prometteuses qui permettent d'éviter les toxic data sont discutées de manière détaillée dans la figure 1 ci-après. Il s'agit des stratégies auxquelles on recourt d'habitude au sein d'une fédération des données dans la formation professionnelle. Dans ce contexte, les avantages pour tous les acteurs du

système de la formation professionnelle sont également mis en évidence, et il est procédé à une appréciation juridique de la situation. S'agissant des stratégies de résolution des problèmes précédemment décrites, il est important de retenir les deux points suivants:

Alors que l'architecture des données combat la survenance de toxic data à l'échelle de l'ensemble du système, les stratégies de résolution des problèmes décrites dans la figure 1 doivent, en revanche, être implémentées par les acteurs individuels afin de lutter contre la problématique des toxic data.

Un modèle d'importance décisive de stratégies de résolution des problèmes réside dans l'« automatisation des processus » (voir encadré 1). En bref: pour les membres d'une fédération des données, lorsqu'on empêche l'apparition de toxic data, cela entraîne aussi un gain d'efficacité dans l'échange des données.

Figure 1 : Sélection de stratégies visant à empêcher la survenance de toxic data, avantages possibles et appréciations juridiques

N°	Stratégie de résolution des problèmes	Exemples	Appréciation juridique	Avantages pour les membres	Avantages pour les participant.e.s
1	**Schémas économes en données:** afin d'éviter de manière préventive l'apparition de toxic data, les schémas sont structurés de telle sorte que l'on enregistre (et que l'on échange) le plus petit volume de données possible dans un justificatif numérique.	On procède à l'inscription d'une apprenante ou d'un apprenant à un cours interentreprises au moyen d'un justificatif numérique. Le schéma spécifique au métier est structuré de telle façon que seules les données strictement nécessaires soient communiquées aux OrTra concernées. Les données qui ne sont pas impérativement nécessaires à l'inscription aux cours interentreprises sont exclues de la communication au moyen du schéma correspondant.	Les schémas sont enregistrés sur le registre décentralisé. Ceux-ci sont toutefois dépourvus de données personnelles. Les schémas permettent de renforcer le principe de proportionnalité en standardisant et en réduisant les données requises au minimum.	Les *auteurs.trices de schémas* peuvent faire figurer dans des schémas les types de données que d'autres acteurs sont autorisés à partager. Ainsi, par ex., un office de formation professionnelle peut créer un schéma qui sera utilisé en priorité pour l'échange de données entre les entreprises formatrices et les OrTra. *Organisme émetteur*: parce qu'il se conforme au principe de minimisation des données, il réduit ainsi les conséquences possibles des toxic data. Le/la *vérificateur.trice* réduit à un strict minimum le volume des données qui peuvent devenir potentiellement toxiques (tout simplement parce qu'elle ne sont pas du tout communiquées).	Le/la *détenteur.trice* profite de la garantie selon laquelle seul sera enregistré le volume de données impérativement nécessaire pour un but déterminé.

2	Les **conditions d'utilisation** sont définies par l'organisme émetteur, resp. par la détentrice ou le détenteur, et peuvent contenir des règles spécifiant i) ce que doit faire avec les données un autre membre au sein de la fédération des données ; ii) ce qu'il est autorisé à faire avec les données, resp. iii) ce qu'il n'est pas autorisé à faire avec celles-ci. Les conditions d'utilisation sont utilisées, d'une part, afin d'imposer l'exécution de dispositions légales et de les lire de manière automatisée et, d'autre part, on peut définir, là où la loi l'autorise, des règles individuelles pour la gestion de ses propres données.	Les certificats semestriels d'une école professionnelle (organisme émetteur) sont attribués sous forme de justificatifs numériques. Les conditions d'utilisation de l'organisme émetteur contiennent les délais légaux de conservation des données applicables à l'organisme émetteur et à la vérificatrice ou au vérificateur (par ex. un office de formation professionnelle). Pour la transmission du certificat semestriel à des tiers (par ex. dans le cadre d'une postulation), une apprenante définit ses propres conditions d'utilisation (dans les limites du cadre légal en vigueur).	Par la communication de données, le responsable du traitement peut ainsi directement informer la personne concernée sur le traitement de ses données, par exemple sur la durée de conservation. En transférant des données, la personne concernée consent au traitement de ses données. En accompagnant le transfert de ses données de conditions d'utilisation, celle-ci peut préciser la portée de son consentement au traitement de ses données.	L'*organisme émetteur* intègre directement les délais légaux d'archivage dans le justificatif numérique au moyen des conditions d'utilisation et crée ainsi la transparence. L'*organisme émetteur*, resp. le/la *vérificateur.trice* lit de manière automatisée les conditions d'autorisation en automatisant le modèle de cycle de vie des données (par ex. afin de respecter les délais légaux). Pour pouvoir profiter des avantages apportés par les conditions d'utilisation, une condition préalable doit être remplie: les données doivent être gérées uniquement par l'intermédiaire de l'agent ou la lecture automatique des conditions d'utilisation doit être paramétrée dans l'application technique.	*Détenteur.trice*: si des données comportant des lieux d'apprentissage privés sont partagées, elles peuvent être dotées de conditions d'utilisation qui auront pour effet que ces données seront automatiquement effacées. Notamment dans des solutions de raccordement, on évite ainsi de manière préventive l'apparition de toxic data (par ex. utilisation de certificats numériques dans des dossiers de candidature).

3	**Zero Knowledge Proof** (ZKP): là où c'est possible, des informations sont réduites à un énoncé « vrai » ou « faux » au moyen d'un ZKP. On peut ainsi formuler des énoncés sur la question de savoir si une information figurant dans un certificat numérique remplit ou non un critère déterminé, mais sans devoir divulguer l'information effective elle-même.	De nombreux apprenantes et apprenants atteignent l'âge de la majorité pendant leur apprentissage professionnel. Dès lors, la situation juridique change aussi pour tous les acteurs impliqués. On utilise un ZKP pour contrôler chaque fois si la personne concernée a atteint l'âge de la majorité (mais sans divulguer la date de naissance précise), que ce soit en lien avec la signature relative aux absences au sein de l'école professionnelle, avec les renseignements communiqués aux parents par les formatrices:teurs professionnels, ou en cas de modification du contrat d'apprentissage.	Le ZKP permet à la personne concernée d'apporter la preuve d'un fait, sans en révéler le contenu, à un destinataire qui pourra vérifier l'exactitude de cette preuve. La personne concernée garde donc un fort contrôle sur ses données, et le destinataire obtient l'information souhaitée alors que le strict minimum lui a été communiqué. Grâce à la méthode du ZKP, le principe de minimisation des données est renforcé.	*Organisme émetteur*: grâce au ZKP qui permet de minimiser les données, l'identité de l'organisme émetteur lui-même sera gardée secrète lors de la vérification d'un justificatif numérique effectuée par la vérificatrice ou le vérificateur. Cela permet aux organismes émetteurs de préserver leur secret de fonction ou leur secret professionnel là où c'est nécessaire.	*Détenteur.trice*: là où c'est possible, des informations sont réduites à un énoncé « vrai » ou « faux » au moyen d'un ZKP, ce qui permet à la détentrice ou au détenteur de transmettre des informations à son sujet en respectant le principe de minimisation des données. Protection contre la corrélation en cas de révocation: le/la *vérificateur.trice* ne peut procéder à aucun profilage.

4	**Dossiers numériques économes en données**: en rassemblant ou en consultant un dossier numérique à partir d'un ou plusieurs justificatifs numériques, seules seront sélectionnées les informations qui sont impérativement nécessaires pour une prise de décision. La manière dont il faut structurer un dossier et le degré de minimisation des données dépendent des bases légales en vigueur.	Lors d'un processus de postulation une fois l'apprentissage professionnel terminé, seule la note finale est communiquée sans divulguer le reste des informations du certificat. Il est facile de ne pas divulguer également d'autres informations, par ex. si, dans un premier temps, l'employeur potentiel ne veut pas encore savoir le nom et le sexe de la candidate ou du candidat.	La communication sélective renforce le principe de l'autodétermination informationnelle. La personne concernée a, par ce biais, la possibilité de communiquer uniquement le minimum d'informations nécessaires pour atteindre le but du traitement des données.	*Vérificateur.trice*: gain d'efficacité grâce à des consultations plus ciblées des données. Pour la prise de décision, seules sont consultées les informations nécessaires sur une apprenante ou un apprenant. Cela permet d'alléger les processus. *Vérificateur.trice*: le risque d'apparition de toxic data est fortement réduit puisque le volume de données qui parviendront à la vérificatrice ou au vérificateur sera d'emblée nettement réduit grâce à un dossier économe en données.	Le/la *détenteur.trice* décide lui-même/elle-même, dans des situations où la loi ne prévoit aucune autre disposition, quels types de données seront communiquées.

2.4 Cas d'application: les offices et les sites d'apprentissage face aux données après une interruption de l'apprentissage

Le cas d'application ci-dessous montre clairement que dans la gestion des données toxiques, la fédération des données rend possibles non seulement des stratégies proactives comme celles présentées dans la figure 1, mais encore des stratégies réactives, comme par ex. la révocation de justificatifs numériques. Cette stratégie de révocation respecte, elle aussi, les principes de proportionnalité, d'adéquation à l'objectif et d'autodétermination informationnelle, et elle est conforme à la loi. Les membres ainsi que les participantes et participants de la fédération des données sont ainsi informés de la survenance d'un événement rapidement et sans complications. Le critère d'importance cruciale à cet égard est le fait que les processus peuvent être déclenchés automatiquement dans le système sur la base d'un événement, et que les processus subséquents peuvent être déclenchés chez les acteurs concernés.

Les offices et les sites d'apprentissage face aux données après une interruption d'apprentissage

Prenons un exemple: la fédération des données vient d'informer l'office de formation professionnelle du canton Quantus – un canton fictif du Plateau suisse – du fait que l'apprenant Alex a interrompu son apprentissage d'employé de commerce parce que le contenu de cet apprentissage ne lui plaisait pas. Comme l'illustre la figure 1, pendant la première année d'apprentissage d'Alex, des données au sujet d'Alex ont été enregistrées dans différents sites du système de la formation professionnelle.

À cause de la résiliation de son contrat d'apprentissage, ces données doivent maintenant être traitées en respectant les prescriptions légales afin d'éviter qu'elles ne deviennent toxiques. Étant donné qu'Alex n'a aucune vue d'ensemble de tous les sites où ses données sont enregistrées, ni des modalités de leur sauvegarde, il se sent dépossédé de son droit de contrôle sur ses données. Dès lors, sans hésiter, il décide de révoquer tous les justificatifs numériques déjà établis à son sujet. Cette transaction déclenche l'envoi d'annonces (ou notifications) qui informent tous les acteurs pertinents dans la formation professionnelle sur la nécessité respective d'intervenir s'agissant des données à propos d'Alex. Dans son école professionnelle et à l'office de formation professionnelle, les transactions déclenchent le processus d'archivage prescrit qui fait en sorte que ses données ne seront plus gérées en tant que dossier courant. L'ancienne entreprise formatrice d'Alex prend aussi connaissance de cette annonce et procède chez elle aux démarches nécessaires, à savoir qu'elle efface toutes les données dont la conservation, en vertu de la loi, n'est plus autorisée. L'organisation du monde du travail (OrTra) compétente, qui possède les premières données sur les notes de cours interentreprises d'Alex ainsi que d'autres données de base pertinentes sur Alex, lance les processus d'archivage requis. En outre, la commission d'examen est aussi informée du fait qu'en raison de la résiliation du contrat d'apprentissage, les données d'Alex doivent faire l'objet d'un traitement. Même des fournisseurs de services privés (comme par ex. des prestataires de tests d'évaluation ou des organisations privées de cours interentreprises) pourraient être informées du fait que les données d'Alex, après la résiliation de son contrat d'apprentissage, devront être traitées conformément aux prescriptions légales. Enfin, l'annonce sur la résiliation du contrat d'apprentissage d'Alex sera intégrée – sous une forme pseudonymisée – dans la statistique cantonale auprès de l'office de formation professionnelle aux fins de monitoring de la formation.

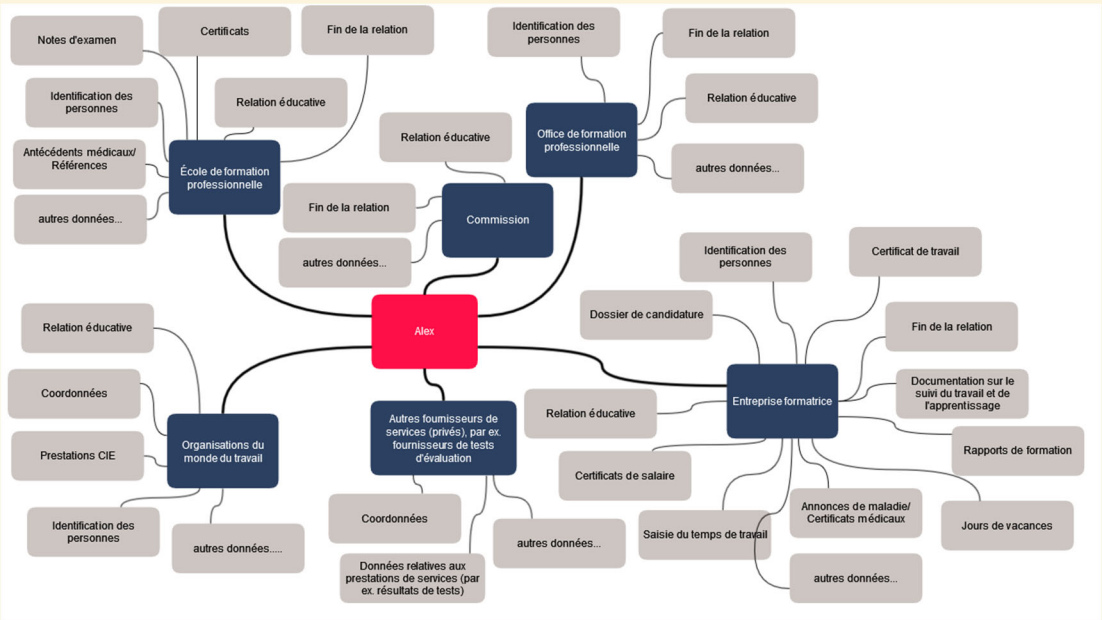

Figure 1 : Les données sur Alex dans le système de la formation professionnelle au moment d'une résiliation du contrat d'apprentissage

Indications: les données enregistrées chez les acteurs concernés ne constituent que des exemples ; il ne s'agit donc pas d'une liste exhaustive. En outre, il est souvent difficile de vérifier quel est le nombre d'attributs qui se trouvent effectivement derrière un point de données. Le standard d'échange de données dans la formation professionnelle proposé, à savoir la « Norme eCH-0260 concernant les données Formation professionnelle », donne ici un bon aperçu de la quantité de données qui est liée aux différentes annonces.

3. Modèle de gouvernance pour la fédération des données

Une fédération des données est un projet de grande ampleur de différentes organisations (par ex. offices de la formation professionnelle, écoles professionnelles, entreprises formatrices, prestataires de services, etc.) ayant chacune leurs propres priorités et leurs propres intérêts. La formation professionnelle initiale dispose déjà d'une gouvernance bien établie (par ex. la loi sur la formation professionnelle, les commissions et organes spécialisés, etc.). Mais pour la fédération des données, il est nécessaire de préciser les modalités de la gouvernance existante. Cela se fait au moyen d'un modèle de gouvernance des données. Ce modèle définit comment il faut organiser l'échange de données, et comment il faut gérer au mieux des problèmes. Il décrit comment des organisations membres gèrent, échangent et utilisent leurs données. Dès lors, il comprend des règles, des lignes directrices, des processus et des systèmes destinés à la saisie, la sauvegarde, l'évaluation, l'archivage, la transmission et l'effacement de données. Il inclut également des normes qui garantissent la sécurité et la qualité de ces données. L'objectif du modèle de gouvernance des données consiste à permettre à tous les acteurs impliqués de la fédération des données d'utiliser les données avec cohérence et efficacité et de les échanger de manière sécurisée. De cette façon, il est possible de

garantir que les données sont utilisées de manière uniforme et responsable, et que les droits et les intérêts de toutes les instances impliquées sont sauvegardés. Des mesures techniques et organisationnelles de l'infrastructure des données permettent de faire respecter et d'imposer les règles de cette gouvernance. Par conséquent, le modèle de gouvernance des données est l'instrument avec lequel on peut mettre en œuvre, dans des actions concrètes, les principes d'architecture présentés dans la première partie du concept.

À cet effet, nous distinguons entre les domaines de données suivants qui sont pilotés par le modèle de gouvernance selon le graphique ci-après:

Figure 2: Domaines de données du modèle de gouvernance des données

Chaque domaine décrit des normes et des pratiques qui, en combinaison avec d'autres domaines, peuvent être utilisées afin d'élaborer des stratégies de résolution de problèmes. Ainsi, on peut s'attaquer par exemple à la duplication des données dans la formation professionnelle en garantissant le fait que les données ne proviennent que de sources faisant autorité (qualité des données) qui sont rendues directement accessibles au moyen de réseaux Peer-to-Peer (infrastructure des données) et en faisant appel à des standards prédéfinis (intégration des données et interopérabilité). Ces différentes notions sont brièvement expliquées ci-dessous et sont présentées plus en détail dans les chapitres suivants:

- **Modèle organisationnel**: il s'agit de la définition des rôles et des compétences, du pouvoir de décision et de la délégation de l'autorité, y compris de toutes les formes d'instruments permettant d'imposer ces accords (par ex. des contrats) (voir chapitre 3.1).
- **Infrastructure des données**: cette notion désigne la totalité des processus, applications et bases de données ainsi que les infrastructures sous-jacentes qui sont requises pour pouvoir garantir l'atteinte des objectifs de la fédération des données (voir chapitre 3.2).
- **Gestion des données de base**: la définition des plus importantes sources faisant autorité au sein de la formation professionnelle et la détermination des méthodes d'accès aux données (y compris Discovery-Service). Les données de base sont donc considérées comme correctes de manière standard. Elles sont reprises et utilisées sans autre examen (voir chapitre 3.3).
- **Qualité des données**: ce domaine contient la question de savoir comment on définit et comment on impose la qualité des données, y compris les mesures appropriées qui permettent de rendre des comptes sur la qualité des données (voir chapitre 3.4).
- **Sécurité informationnelle et protection des données**: il s'agit des mesures techniques et organisationnelles qui permettent de garantir la sécurité et l'intégrité des données ainsi

que la protection des droits et libertés des personnes via des actions appropriées visant à protéger vos données personnelles (voir chapitre 3.5).

- **Intégration des données et interopérabilité**: toutes les normes (par ex. W3C, eCH, etc.), y compris les dispositions qui en font partie, qui garantissent l'échange des données. Dans ce contexte, il faut garantir une intégration appropriée des données dans des systèmes tiers ainsi que leur interopérabilité entre productrices et producteurs et consommatrices et consommateurs de données (voir chapitre 3.6).

Le modèle de gouvernance de données garantit que la fédération des données agit dans l'intérêt de ses membres au travers de la réglementation sur la responsabilité partagée. Cela garantit à son tour l'obligation de rendre des comptes et la traçabilité des actions dans un système distribué et décentralisé. La transparence de l'infrastructure des données (par ex. registre décentralisé des données et de ses transactions) accroît la confiance et réduit les risques systémiques (par ex. protection des données). En outre, le mode de transmission des données permet de réduire les temps de latence, ce qui entraîne une plus grande rapidité dans la prise de décision, et ce qui améliore simultanément la participation. Dans le paragraphe suivant, nous décrivons le modèle de gouvernance des données de manière plus détaillée.

3.1 Modèle organisationnel: rôles et instruments dans la fédération des données

Pour le modèle de gouvernance des données dans la fédération des données, on distingue les rôles respectifs suivants:

- **Les membres** sont des organisations impliquées qui sont liées contractuellement au sein de la fédération des données. Ils mettent à disposition certains services et interagissent avec l'infrastructure sous-jacente (c'est-à-dire qu'ils ont le contrôle sur leurs agents et leurs controllers et qu'ils assument la responsabilité respective y afférente). Exemples de membres: offices de formation professionnelle, écoles professionnelles, entreprises formatrices et autres entreprises qui proposent des prestations de services existantes ou de nouvelles prestations de services.
- **Les participantes et participants** sont des consommatrices ou des consommateurs des services fournis par les membres. Ils n'ont (en règle générale) aucune interaction directe avec l'infrastructure sous-jacente, bien que sous certaines conditions, on leur accorde un droit d'accès en lecture seule (par ex. aux fins de validation de justificatifs numériques). Exemples de participantes et participants: les apprenantes et les apprenants ainsi que les personnes détentrices de l'autorité parentale qui veulent accéder aux données dans la formation professionnelle ou qui veulent mettre à disposition des données (via les services fournis par les membres).
- L'**exploitation technique** de la fédération des données est compétente pour l'élaboration des prescriptions sur l'architecture technique ainsi que pour leur mise en œuvre, leur actualisation et leur maintenance. Cela se fait sous la surveillance du bureau, et en collaboration avec les membres de la fédération des données (par ex. parce qu'ils hébergent un node, etc.).

- Le **bureau** de la fédération des données est la plaque tournante administrative qui est compétente pour l'administration de la fédération des données (par ex.: communication, processus d'exploitation, gestion des contrats, etc.).
- Le **pilotage de la politique de l'éducation** est compétent pour la gouvernance, la stratégie, ainsi que la fixation des objectifs et des priorités de la fédération des données. Sont représentés au sein du pilotage de la politique de l'éducation tous les groupes de membres essentiels de la fédération des données (à savoir: les offices de formation professionnelle, les écoles professionnelles, les OrTrA, etc.).

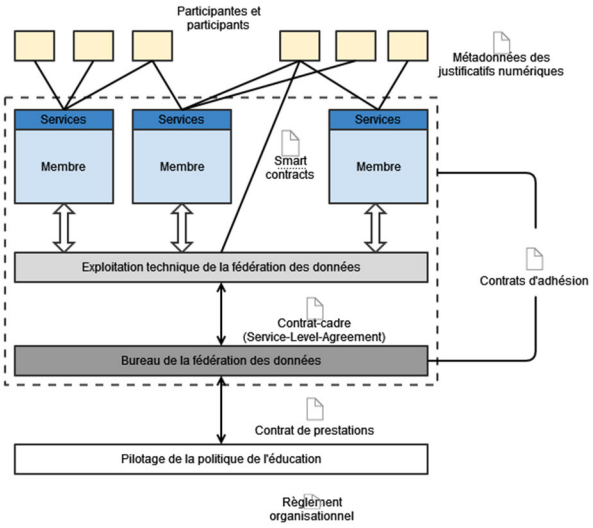

Figure 3: Les rôles par rapport aux instruments de gouvernance

Les instruments de gouvernance suivants seront utilisés:

- **Règlement organisationnel**: une déclaration de principe transsectorielle sur la fonction de la fédération des données, et sur les rôles et compétences respectifs qu'elle contient.
- **Contrat de prestations**: une concrétisation des tâches fonctionnelles du bureau dans le cadre des objectifs politiquement visés de la fédération des données. Cet accord existe entre le pilotage de la politique de formation et le bureau dans le cadre de leurs fonctions respectives existantes.
- **Contrat-cadre** (Service-Level-Agreement): un instrument qui définit la prestation attendue ainsi que d'autres caractéristiques non fonctionnelles (par ex. sécurité, scalabilité, disponibilité, etc.). Il s'agit d'un contrat passé entre le bureau et l'exploitation technique.
- **Contrats d'adhésion**: la qualité de membre de la fédération des données est régie par des contrats qui sont mis à disposition par le bureau. Les participantes et participants n'ont en revanche aucune relation contractuelle (écrite) avec la fédération des données.
- **Métadonnées des justificatifs numériques**: toutes les métadonnées qui servent à réglementer l'utilisation des justificatifs numériques d'autorisation (par ex. des conditions d'utilisation qui réglementent les autorisations, les interdictions et les obligations).

- **Smart contracts:** des décisions algorithmiques spécifiques peuvent être déléguées au moyen de contrats logiciels qui exécutent les ordres sur la base d'événements de transactions dans le registre décentralisé des données. Ce logiciel peut émettre des notifications auxquelles peuvent s'abonner des membres (ou d'autres parties). Ces notifications peuvent être utilisées par les membres pour des actions supplémentaires.

3.2 Importance d'une infrastructure des données pour la qualité et l'efficacité du système éducatif

La création d'une infrastructure de données qui comprend la structure, les standards et l'administration des données du système éducatif a pour but de garantir leur qualité, leur exactitude et leur fiabilité. Elle aide, par exemple, à identifier des possibilités d'amélioration dans les processus d'administration des données, et à devenir ainsi plus efficace et meilleur marché, mais aussi à améliorer la qualité de la prise de décisions. Mais cela ne peut être atteint que par le biais d'une implication plus forte des acteurs qui fournissent les données pertinentes pour la prise de décisions, tout en renforçant également le pouvoir des acteurs qui participent à cette prise de décisions. Ce renforcement du pouvoir des acteurs résulte de la situation structurelle créée par l'infrastructure des données qui rend possible des relations directes entre les acteurs (par ex. synchronicité de la communication, transparence des transactions, etc.).

Les membres de la fédération des données sont visibles en tant que nodes, en tant qu'agents et en tant que controllers de ces agents (voir figure 4). Ils sont contractuellement liés aux règles communes destinées au traitement des données (y compris la création, la mise à disposition, l'utilisation, la sauvegarde et l'effacement des données). Ces obligations comprennent l'utilisation des composants techniques (par ex. les agents), y compris les mesures des composants techniques ayant pour but de protéger la fédération des données ainsi que de garantir le respect des prescriptions sur l'architecture. Ces dernières sont prescrites par l'exploitation technique et ont été approuvées par le bureau. Des contrats de ce genre transfèrent ainsi des droits et des obligations. En fonction du type de qualité de membre concerné, ils accordent certaines autorisations techniques (par ex. des droits d'accès en lecture seule, resp. l'accès à certaines données déterminées, etc.). De cette façon, on crée en priorité une responsabilité commune qui est fondée sur la confiance. Ainsi, par ex., au travers du processus d'adhésion à la fédération des données, on dépose dans le registre décentralisé des données le matériel de confiance nécessaire, comme par exemple des clés publiques, des informations institutionnelles et autres artefacts.

L'organisation qui est responsable de l'exploitation, de la maintenance et de l'actualisation de l'infrastructure commune de la fédération des données fonctionne en tant qu'exploitation technique. Certaines parties de cette responsabilité peuvent être déléguées à des organisations membres qui sont tenues de respecter les directives de l'exploitation technique. Des composants communs, comme par ex. les nodes et le registre décentralisé des données, sont gérés selon des critères uniformes. Cela ne concerne pas seulement les spécifications techniques (par ex. l'utilisation de containers virtualisés pour l'hébergement du logiciel), mais cela concerne également la gouvernance des composants de logiciels (à savoir: des lignes directrices sur l'architecture des logiciels destinés à la création de smart contracts ainsi que

sur la manière de gérer leurs mises à jour, etc.). Pour cette raison, le modèle de gouvernance est basé sur des technologies dites « permissionnées » pour lesquelles les organisations impliquées disposent de droits de participation qui se fondent sur la confiance qui existe au sein du système éducatif. Ce mode de faire s'oppose à d'autres méthodes, comme par ex. à des approches problématiques du point de vue écologique, comme par exemple à de coûteux algorithmes de consensus Proof-of-Stake, qui sont très gourmands en énergie (par ex. le bitcoin).

3.3 Gestion des données de base

Les données de base sont les données qui sont les plus importantes pour la formation professionnelle, raison pour laquelle elles sont considérées comme critiques. Les données de base ne comprennent pas seulement les données de référence de la formation professionnelle (comme par exemple des informations institutionnelles), mais elles incluent également des données importantes pour l'exploitation (par ex. des notes de cours et autres justificatifs de prestations). À l'heure actuelle, au sein de la formation professionnelle, il n'existe guère de consensus sur ce que sont exactement les données de base, et sur la question de savoir quelles règles d'entreprise doivent être appliquées afin de préserver leur intégrité. Néanmoins, de nombreux services existants pourraient constituer des interfaces avec des données importantes, dont le Registre des places d'apprentissage (RPA) et les « Profession Services » (y compris les registres professionnels) qui sont liés au RPA. Avec le concours de la fédération des données, ces données peuvent être extraites des sources officielles correspondantes. Mais dans la pratique, il est actuellement difficile d'identifier automatiquement ces sources et d'entrer en relation avec elles, de sorte qu'on est souvent obligé de saisir ces données à plusieurs reprises (par ex. les indications sur le lieu de domicile). Cela entraîne des doublons et des difficultés lors de l'actualisation des données (par ex. lors de changements d'adresse). C'est pourquoi la fédération des données offre un service d'information (appelé « Discovery-Service »). Dans ce contexte, les sources précitées ne doivent pas être explicitées au sein du système de la formation professionnelle (par ex. des attributs de personnes provenant d'un système d'identité numérique (e-ID), des données sur le lieu de domicile auprès d'un service de contrôle des habitants (CdH), etc.). La fédération des données propose d'introduire une « topologie d'architecture de registry », c'est-à-dire que les données de base sont réparties en différents lieux, qui sont toutefois indiqués dans un registre décentralisé.

Par exemple, un office de formation professionnelle (en tant que membre de la fédération des données) aimerait savoir quelle est l'adresse de domicile correcte d'une apprenante ou d'un apprenant. Le controller de l'office de formation professionnelle concerné apprend automatiquement, par le biais de son agent, via le Discovery-Service, à partir de quelle source ces données doivent être obtenues pour cette apprenante ou cet apprenant (par ex. à partir du service de contrôle des habitants (ou CdH) du canton), et il prend ainsi directement contact avec le CdH. Les données de référence techniques requises à cet effet sont archivées dans le registre décentralisé des données (par ex. le type de contact avec le « Service Point ») et doivent être gérées par la source concernée (par ex. par le CdH). L'échange de données se fait au moyen de justificatifs numériques qui sont signés numériquement par le CdH et qui sont directement envoyés à l'office de formation professionnelle. Si un changement d'adresse survient ultérieurement, le service du CdH en informe automatiquement

l'office de formation professionnelle via l'infrastructure de l'agent. Il est important de souligner qu'aucune donnée interrogée au sein du registre décentralisé des données n'est traitée, ni mémorisée.

Une instruction peut ensuite être donnée à l'agent afin qu'il interroge directement les informations dont on a besoin, et qu'il les envoie en retour au controller. On y parvient par le biais de protocoles basés sur la norme eCH, ainsi que via des interfaces standardisées qui mettent à disposition les applications techniques (ou FA en allemand). La fédération des données permet de contacter les services correspondants via le registre décentralisé des données, de crypter les données et de les échanger de manière sécurisée. Dans ce cas, les controllers servent d'interface avec les systèmes existants. Dans de nombreux cas (par ex. si l'e-ID utilise la Self Sovereign Identity (SSI)), aucune interface n'est nécessaire à cet effet. L'e-ID existerait dans le même écosystème de sorte que les données peuvent être interrogées directement chez les participantes et participants.

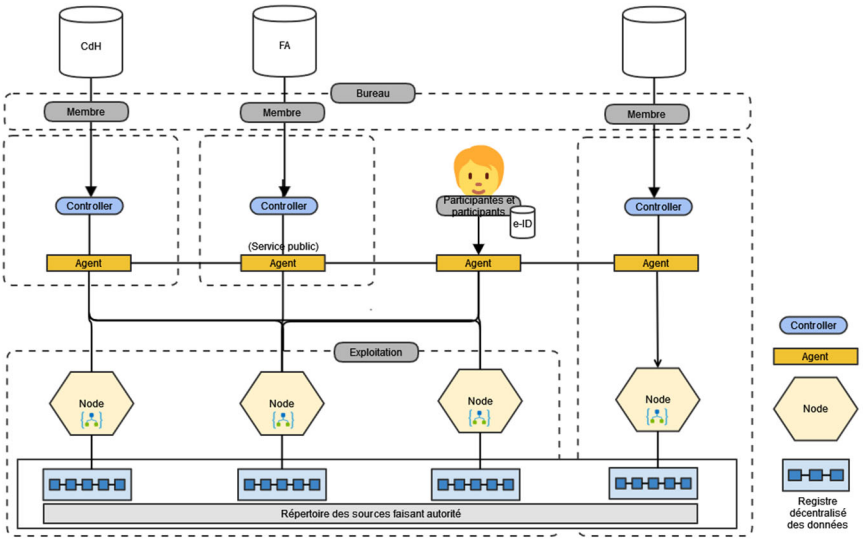

Figure 4 : Gestion des données de base (sources faisant autorité)

3.4 Qualité des données

La qualité des données est liée à leur affectation à un objectif déterminé. Dès lors, la définition de la qualité des données doit être fixée en fonction de la situation. Les sources de données dans la formation professionnelle ont pour mission de fixer une définition appropriée de la qualité des données en tenant compte des besoins de tous les groupes de parties prenantes, et notamment en tenant compte des normes eCH correspondantes. D'habitude, une telle définition se réfère aux éléments suivants:

- Actualité: les données sont conformes à l'état actuel concerné de la réalité enregistrée.
- Absence d'ambiguïté: ce qui a été enregistré avec les données est interprétable sans aucune ambiguïté.

- Uniformité: les faits identiques sont enregistrés de manière identique.
- Exactitude: les faits sont enregistrés avec suffisamment d'exactitude.
- Cohérence: les données enregistrées sont en elles-mêmes non contradictoires.
- Caractère correct: les faits sont enregistrés de manière conforme à la réalité.
- Absence de redondance: les données ne présentent aucun doublon.
- Pertinence: il faut enregistrer les données avec les informations importantes et nécessaires pour un processus.
- Compréhensibilité: dans leur terminologie et leur structure, les données sont en accord avec les connaissances des utilisatrices et utilisateurs.
- Exhaustivité: les données sont, pour l'essentiel, dépourvue de valeurs manquantes.
- Fiabilité: la manière dont les données ont pris naissance ainsi que leur origine sont vérifiables.

Pour atteindre les objectifs de la fédération des données, on considère que les données sont d'une qualité (acceptable) si elles sont appropriées à l'utilisation visée tant au sein de l'exploitation que pour la prise de décisions. Dans un environnement réparti comme celui de la formation professionnelle initiale, cela constitue une exigence considérable posée aux productrices et producteurs de données qui doivent anticiper les besoins en matière d'utilisation des données de la part de (plusieurs) consommatrices et consommateurs différents. Pour cette raison, la qualité des données est considérée comme un critère de qualité du service. Ce critère sera publié via le registre décentralisé des données et permettra aux consommatrices et consommateurs des données de sélectionner, au moyen d'un service d'information, les sources de données qui conviennent le mieux à leurs besoins. Ainsi, lorsque des agents interagissent avec le registre décentralisé des données, la responsabilité quant au caractère correct des données et des opérations exécutées par ces agents est assumée par les membres concernés qui ont actionné ces agents.

Afin de réduire les problèmes liés au caractère actuel de la qualité des données (à savoir des données qui étaient correctes au moment de leur enregistrement, mais qui ne sont plus correctes au moment de leur utilisation), il est nécessaire de créer des chemins les plus directs possibles entre les acteurs de la formation professionnelle (voir chapitre 3.2). Cela signifie que la centralisation des données qui intervient avant leur traitement constitue un risque pour la qualité des données, mais que ce risque peut être éliminé pour une grande part via l'architecture proposée par la fédération des données. Par ailleurs, le caractère actuel et la disponibilité en temps réel des données ne jouent pas toujours un rôle pour la qualité des données, comme par exemple lors de la recherche portant sur des données historiques.

La fédération des données peut améliorer la qualité des données. Et ce, notamment en raison d'une gouvernance des données qui tient compte de la totalité de leur cycle de vie (et, partant, de l'obligation de rendre des comptes à ce sujet), tout comme de l'actualité des données obtenue par l'échange direct entre productrices et consommatrices de données. En outre, la fédération des données se fonde sur des sources fiables qui se sont engagées à respecter une définition commune de la qualité des données et qui garantissent l'intégrité des données. Cela se fait par le biais du contrôle de la provenance des données, à savoir en garantissant la traçabilité de données déterminées qui sont reliées à leurs sources originelles, ainsi que par le biais de règles d'entreprise automatisées et de l'introduction de conditions d'intégrité imposées aux bases de données. De surcroît, des schémas destinés aux

justificatifs numériques peuvent être mémorisés dans le registre décentralisé des données. Ces schémas peuvent être utilisés afin de contrôler la conformité aux standards de qualité des données. On s'assure ainsi du fait que les données sont dignes de confiance, et on accroît ainsi la probabilité d'utiliser les bonnes données pour la prise de décisions.

3.5 Sécurité informationnelle et protection des données

La fédération des données offre de nouvelles manières d'échanger et d'utiliser les données. En évitant l'existence de silos centralisés de traitement des données, on garantit ainsi qu'il existe moins de risques systémiques (cluster risks) au sein du système éducatif. De cette façon, il est possible de réduire considérablement les risques liés à la cybercriminalité. En effet, les données sont enregistrées, traitées et réparties par les instances qui sont effectivement impliquées dans la prise de décision opérationnelle. Un élément d'importance centrale de l'infrastructure des données réside dans la mise à disposition d'un matériel de confiance (par ex. des clés publiques) qui se réfère à des institutions de formation bien établies et qui peut être utilisé à des fins de validation. Cela ressemble certes aux approches hiérarchiques classiques utilisées d'habitude sur Internet (par ex. Public-Key-Infrastructures), mais il s'agit en réalité d'une approche décentralisée, et donc plus facile à gérer. Le degré élevé de connectivité qui en résulte au sein du réseau est sécurisé par des relations de confiance protégées par paires cryptographiques. Dès lors, il n'est pas nécessaire de centraliser la confiance sous la forme d'une solution classique de gestion des identités, ce qui réduit nettement le risque de vol d'identité. Ces mesures réduisent les risques de cyberattaques qui essaient d'intercepter, de lire ou de manipuler de manière inaperçue l'échange de données entre deux acteurs. La décentralisation garantit une résilience plus élevée et une sécurité accrue contre les pannes, et contre des stratégies d'attaque usuelles et parfois déjà survenues (par ex. attaque par déni de service (Denial-of-Service), rançongiciel (ransomware), etc.).

En outre, la fédération des données peut prendre aussi des mesures spécifiques afin de garantir une meilleure protection des données. Le caractère décentralisé du réseau améliore tant l'affectation des données à un objectif déterminé que la minimisation des données. Cela tient au fait que les données peuvent être échangées pour l'essentiel entre les acteurs participants, voire peuvent être uniquement échangées entre ces derniers, et ce, sans qu'une centralisation dans des silos « gourmands en données » ne soit nécessaire. Cela permet d'éviter les risques secondaires liés à l'existence de données agrégées. Ces risques prennent naissance lorsque certains éléments sont certes enregistrés de manière légitime, mais qu'ils constituent néanmoins un risque collectif lorsqu'ils sont mémorisés avec d'autres données. Au sein de la fédération des données, on ne trouve des données communes que dans le registre décentralisé des données. Mais ces données sont référencées par des identificateurs décentralisés qui ne sont connus que par les parties concernées. Cela rend possible la communication d'événements pertinents pour le système éducatif, mais sans que les parties ne soient contraintes de divulguer des informations inconnues à ce jour. Très souvent, il suffit aussi de formuler des énoncés sur des données sans devoir divulguer les données elles-mêmes. C'est notamment le cas lorsqu'il faut formuler des énoncés binaires (par ex. pour répondre à la question de savoir si une apprenante ou un apprenant a atteint ou non l'âge de la majorité). Dans de tels cas, l'approche dite « Zero Knowledge Proof » permet de formuler des énoncés sous la forme de « preuves » cryptographiques (par ex. certifier qu'un apprenant est

majeur sans devoir divulguer sa date de naissance). Il est prévu que des services de ce genre fassent partie intégrante de l'interface avec la fédération des données.

Lorsque des données sont communiquées, cela peut se faire sous la forme d'un justificatif numérique dans lequel des conditions d'utilisation peuvent être stipulées. Ainsi, par exemple, il est possible de définir et de fixer des conditions d'utilisation qui spécifient ce qui est autorisé ou interdit, resp. qui précisent quelles autres obligations s'appliquent aux acteurs concernés. Ces données servent de justificatif avec lequel la personne chargée du traitement exerce ses obligations légales, et celle-ci peut ainsi attester que les dispositions de la loi sur la protection des données sont respectées. En outre, la fédération des données offrira un service de notifications qui permettra d'informer les parties concernées d'une éventuelle révocation d'un justificatif numérique déjà établi. Dès lors, cela nous permettra de savoir clairement quand et où il y a nécessité d'intervenir, mais sans devoir dépendre constamment d'une logique d'application dont l'implémentation et le rodage peuvent prendre énormément de temps.

3.6 Intégration des données et interopérabilité

L'objectif de la fédération des données est de faciliter et simplifier le processus de combinaison de données issues de sources différentes (par ex. à partir de trois sites d'apprentissage différents). Mais l'intégrabilité de jeux de données différents au sein de la fédération ne repose pas seulement sur l'infrastructure susmentionnée, mais dépend encore du respect de certaines normes et de l'intégration de solutions existantes de gestion des données chez les acteurs de la formation professionnelle initiale. Entre-temps, ces normes sont désormais standardisées et constituent la base qui rend possible leur interopérabilité. Elles favorisent la croissance et l'intégration des écosystèmes qui, à l'heure actuelle, sont en train de naître tant en Suisse qu'à l'international (par ex. l'European Blockchain Services Infrastructure (EBSI)). Ainsi, il sera possible de rendre compatibles des solutions dans le domaine de la formation professionnelle avec des solutions d'autres branches, comme par ex. celles de l'e-Gov, de l'administration numérique et des services sociaux.

Dans de tels écosystèmes, le support de données prioritaire – comme dans le présent concept de la fédération des données – est le justificatif numérique (comme décrit dans la première partie du concept). Ce faisant, il est ainsi possible (et judicieux) que plusieurs institutions établissent leurs certificats numériques (par ex. pour des certificats, pour des annonces de places d'apprentissage ou pour des autorisations de formation) sur la base de schémas définis en commun. Puisque tous les canevas doivent être déposés dans le registre commun décentralisé des données, l'architecture proposée facilite considérablement l'utilisation de tels canevas dans l'ensemble des systèmes. Pour la définition de certains formats dans les canevas, on fait appel à des standards de données existants, comme par ex. à la norme de données eCH-0260 « Formation professionnelle ». Des schémas utilisés en commun sur la base de standards existants garantissent ainsi un degré élevé d'interopérabilité et, de ce fait, permettent d'échanger des données dans l'ensemble des cantons et des institutions.

De surcroît, le recours à des standards de données existants et leur mise en œuvre systématique constituent le cœur de la capacité d'intégration des données dans les systèmes

existants de gestion des données au sein de la fédération. Cette intégration est garantie par la spécification des interfaces entre les controllers et les solutions existantes de bases de données des membres de la fédération. Il est important de souligner que, dans l'architecture proposée, il reste possible d'échanger des informations entre les différents membres sous une forme non standardisée, ou sans faire usage de justificatifs numériques (par ex. sous la forme de notifications ou d'autres interactions). Par exemple, il reste possible que, dans le cas d'application décrit ci-dessus, l'office de formation professionnelle informe de l'interruption d'apprentissage d'un apprenant l'entreprise formatrice, l'école professionnelle et les prestataires de cours interentreprises. Cela peut se faire sous la forme de l'envoi d'un simple message. En outre, grâce à l'utilisation de canaux de communication Peer-to-Peer sécurisés de manière cryptographique, cette forme d'échange est nettement plus sûre, par exemple, qu'un échange qui serait effectué par courriel.

4. Conclusion intermédiaire

Cette deuxième partie du concept de la fédération des données dans la formation professionnelle offre un aperçu du potentiel de l'architecture technique décrite dans la première partie et du mode de gestion des futurs défis liés à la loi sur la protection des données. Avec l'introduction de la nouvelle législation sur la protection des données, on assiste à un véritable changement de paradigme: à l'avenir, des droits de décision nettement plus étendus seront dévolus à l'individu. Cette architecture technique permet de soutenir les stratégies de résolution des problèmes dans la gestion des données personnelles.

Le modèle de gouvernance des données proposé pour la future fédération des données est conçu de manière polycentrique. Compte tenu du nombre croissant de productrices et producteurs ainsi que de consommatrices et consommateurs de données, et aussi au vu de la diversité accrue de leurs besoins en matière de traitement des données, une gestion centralisée des données axée sur les besoins d'une minorité d'acteurs devient toujours moins justifiable. Dans un monde numérique de plus en plus décentralisé, la responsabilité et la gouvernance partagées jouent un rôle toujours plus important. Le modèle de gouvernance des données est interprété en tant que complément à l'architecture technique décentralisée.

Le concept de la fédération des données dans la formation professionnelle s'inscrit dans un paysage d'espaces de données fiables, organisé selon les principes du fédéralisme, qui est actuellement en cours de développement. L'interopérabilité entre ces espaces de données est structurée à l'échelon national ainsi que, de plus en plus, également à l'échelon international.

Annexe

Définitions

Concept	Définition
Agent	Un « agent », au sens technique du terme, est un composant d'un logiciel qui traite les instructions d'un ou de plusieurs « controllers », qui les rediffuse en conséquence au sein de la fédération des données, et qui communique avec le node ou qui mémorise les justificatifs numériques (et, de ce fait, les données personnelles).
Organisme émetteur	Par le terme d'organisme émetteur (ou « issuer »), on entend un organisme qui, par ex., est autorisé par une institution étatique à établir des justificatifs numériques.
Source faisant autorité	Source de données qui possède un degré élevé de crédibilité et de confiance.
Exploitante / exploitant	Il peut s'agit d'institutions, de personnes privées ou de personnes morales qui exploitent les nodes.
Blockchain	Un système de stockage de ledger dans lequel les données de transaction sont attachées et ne peuvent pas être modifiées ultérieurement. Les blockchains sont réalisées via un réseau de nodes, raison pour laquelle elles sont, en règle générale, décentralisées.
Controller	Par le terme de « controller », on entend un composant d'un logiciel qui permet à des personnes et à des institutions d'accéder à l'« agent », via une interface-utilisateur, ce qui donne ainsi accès à des justificatifs numériques et permet d'utiliser le registre décentralisé des données, resp. de communiquer avec d'autres membres u participantes et participants de la fédération des données.
Provenance des données	Elle permet de garantir la traçabilité de certaines données en les rapportant à leurs sources initiales. On peut ainsi constater, par exemple, si certaines données proviennent effectivement ou non de sources faisant autorité.
Registre décentralisé des données	Un registre décentralisé des données est une mémoire publiquement accessible et lisible par toutes les actrices et tous les acteurs. Ils contient toutes les informations nécessaires pour pouvoir exploiter une fédération des données en général, et un écosystème destiné à l'établissement de justificatifs en particulier.
Identificateurs décentralisés des données (DID)	Il s'agit d'une forme d'identification qui est utilisée dans des systèmes décentralisés. Elle rend possible la mise en lien d'événements avec des données de telle sorte que seules les parties concernées puissent établir ces mises en liens.
Portefeuille électronique	Par le terme de portefeuille électronique, on entend un logiciel d'application qui sert à conserver de manière sécurisée des justificatifs électronique.

Service d'informations (appelé Discovery-Service)	Il s'agit d'un service technique qui, au sein du registre décentralisé des données, rend possible l'identification (automatique) de prestations de services qui sont proposées par des membres.
Détentrice / détenteur	Par le terme de détentrice ou de détenteur, on entend une personne ou une institution qui possède électroniquement un justificatif numérique.
Smart Contracts	Les smart contracts (contrats intelligents) sont des programmes informatiques qui traitent des données de manière automatisée et qui, sur cette base, ont pour effet de déclencher une prise de décision automatisée selon des critères prédéfinis.
Node	Un node (nœud) est une instance de traitement du registre décentralisé des données. Les nodes réalisent, contrôlent et mémorisent des transactions sur le registre décentralisé des données et peuvent en outre également héberger des contrats intelligents.
Membre	Il s'agit d'une partie institutionnelle de la fédération des données qui est autorisée à collaborer de manière active (par ex. création de transactions). De telles parties sont tenues de signer un contrat avec l'exploitante/avec l'exploitant de la fédération des données.
Conditions d'utilisation	Les conditions d'utilisation sont des métadonnées qui sont écrites sur des justificatifs numériques aux fins de réglementation des autorisations, des interdictions et des obligations imparties aux acteurs concernés.
Permissioned (Blockchain)	Une infrastructure de blockchain dont les droits de lecture et d'écriture sont restreints.
Gouvernance polycentrique	Une approche de gouvernance où plusieurs parties négocient des règles et des stratégies au sein d'un écosystème afin de résoudre des problèmes communs.
Proof-of-Stake	Une procédure avec laquelle un réseau de blockchain peut atteindre un consensus ; il s'agit souvent d'une procédure au sein de laquelle des calculs à haute intensité énergétique doivent être résolus.
Privacy by Default	La conception d'un système dans lequel des réglages prédéfinis sont mémorisés de telle sorte que le traitement des données personnelles est limité au strict minimum nécessaire pour le but d'utilisation déterminé de ces données dans la mesure où la personne concernée n'en décide pas autrement.
Privacy by Design	Une approche avec laquelle on garantit de manière préventive la protection des données par le biais de la conception technique et organisationnelle d'un système.
Public-Key-Infrastructure	Une infrastructure destinée à la délivrance, à la distribution et à la validation d'un matériel d'information d'importance décisive dont l'appartenance à certaines organisations déterminées a été certifiée.
Schéma	Un canevas destiné à un justificatif numérique qui spécifie son contenu (à savoir des métadonnées, y compris des conditions d'utilisation, des déclarations et une méthode de preuve).

Self Sovereign Identity (SSI)	Une approche destinée à la gestion des identités où l'on utilise des justificatifs numériques afin de transmettre des informations d'attributs au sujet d'une personne (par ex. nom, prénom, date de naissance).
Directives sur l'architecture logicielle	Lignes directrices fixées par l'exploitation technique et approuvées par l'organisme émetteur. Elles déterminent comment le logiciel (par ex. des smart contracts) doit être développé pour la fédération des données (par ex. des canevas destinés à des procédures standards, ou la mise en œuvre de mesures de sécurité, etc.).
Participants	Il s'agit d'une partie de la fédération des données qui est autorisée à collaborer de manière passive (par ex, consultation d'informations publiques dans le registre décentralisé des données ou vérification des justificatifs). En règle générale, de telles parties ne doivent pas à avoir signé de contrat avec l'exploitante/avec l'exploitant de la fédération des données.
Transaction	Une transaction est une opération qui permet de reprendre des données dans le registre décentralisé des données, Les transactions ne contiennent aucune donnée personnelle
Vérificatrice / vérificateur	Par le terme de vérificatrice ou vérificateur, on entend des personnes ou des institutions qui vérifient la validité des justificatifs numériques qui leur sont présentés au moyen d'informations disponibles sur le registre décentralisé des données.
Justificatif numérique (vérifiable)	Dans le présent contexte, un justificatif numérique contient un ou plusieurs énoncés d'une institution ou d'une autorité (par. ex. d'une école ou d'une instance étatique) à propos d'un sujet donné (par ex. d'une apprenante ou d'un apprenant). Il est possible de vérifier si ces déclarations proviennent d'une source fiable à l'aide des données figurant dans le registre décentralisé des données – d'où l'expression « justificatif numérique vérifiable ».
Zero Knowledge Proof	Un ZKP ou « Zero Knowledge Proof » est un protocole cryptographique qui permet de formuler un énoncé au sujet de certaines données, mais sans divulguer les données proprement dites (par ex. le fait qu'on est âgé de plus de 18 ans, mais sans devoir divulguer la date de naissance).

educa

Erlachstrasse 21
3012 Bern

+41 31 300 55 00
info@educa.ch

Concept de la fédération des données dans la formation professionnelle
3ème partie – Plan de réalisation et coûts

1.3.2023

Digitaler Bildungsraum Schweiz
Espace numérique suisse de formation
Spazio formativo digitale svizzero
Spazi da furmazium digital svizzer
Swiss digital education space

educa

Table des matières

Management Summary **3**

1. Introduction **4**

2. L'ensemble des processus de la fédération des données **5**

3. Cas d'application: l'échange direct de données entre l'office de la formation professionnelle, les prestataires de cours interentreprises et les entreprises formatrices **7**

4. Besoin de régulation pour la fédération des données: conventions intercantonales **16**
4.1 Création de la fédération des données 16
4.2 Échange de données intercantonal 17

5. Plan de réalisation de la fédération des données **17**
5.1 Présentation détaillée du plan de réalisation 18
5.2 Mesures d'accompagnement pour la réalisation du concept 23

6. Coûts et financement du plan de réalisation **24**
6.1 Modèle de comptabilité analytique de la fédération des données 24
6.2 Coûts de réalisation et coûts d'exploitation 25
6.3 Financement par les unités finales d'imputation 29

7. Gestion des risques de la fédération des données **30**
7.1 Liste des risques 31

8. Conclusion intermédiaire **37**

Annexe **38**
Définitions 38
Principes d'architecture 41
Sources principes du chapitre 4 43

Management Summary

Dans la première partie du concept de la fédération des données pour la formation professionnelle, nous avons mis l'accent sur la future infrastructure technique. Ce faisant, nous avons souligné les effets sur l'efficience, la disponibilité des données et la connectabilité. La deuxième partie du concept se concentre sur la protection des données et sur les règles de gouvernance adéquates. Le modèle de gouvernance des données développé dans la deuxième partie a pour objectif de permettre à tous les acteurs impliqués dans la fédération des données d'utiliser les données de manière cohérente, efficace et conforme à la loi.

Cette troisième partie présente le calendrier, les coûts ainsi que la mise en œuvre des différentes étapes du concept de la fédération des données. Le potentiel de la fédération des données est en outre présenté de manière détaillée sur la base d'un deuxième cas d'application: l'échange direct de données et automatisé entre l'office de formation professionnelle, les prestataires des cours interentreprises et les entreprises formatrices. La troisième partie se termine par une analyse détaillée des risques afférents à la phase de réalisation.

Pour que la future fédération des données puisse être couronnée de succès, il convient de créer les bases réglementaires nécessaires. D'une part, la création de la fédération des données requiert de la part des cantons et, le cas échéant, par la Confédération, qu'ils élaborent conjointement la base légale adéquate et conforme au droit public. D'autre part, la fédération des données ne pourra déployer pleinement son potentiel qu'à condition que – compte tenu des dispositions cantonales relatives à la protection des données – la communication des données entre les cantons ainsi qu'au sein de la fédération des données puisse s'appuyer sur une convention intercantonale stable.

Le plan de réalisation se base sur une mise en œuvre par étapes de la fédération des données. La réalisation doit nécessairement commencer par l'élaboration des bases légales. Lors de cette première étape, ce sont en particulier les éléments de la gouvernance et de l'architecture qui revêtent une grande importance. Il s'ensuit la phase d'exploitation proprement dite, à savoir une phase où sera ajoutée, incrémentalement, soit petit à petit, la valeur ajoutée de la fédération des données pour ses utilisatrices et ses utilisateurs. Les quatre premiers incréments prévus comprennent *l'établissement de CFC sous forme de justificatif numérique, l'inscription aux cours interentreprises, les certificats des écoles professionnelles sous forme de justificatifs numériques et l'octroi de l'autorisation de formation.* La réalisation de ces bases légales et de ces quatre premiers incréments est prévue sur une durée de trois ans maximum. Ensuite, des incréments supplémentaires pourront suivre.

Le présent document est la dernière partie, correspondant à la troisième partie du document final de présentation du concept d'une fédération des données pour la formation professionnelle. Les trois parties de ce concept constituent conjointement le dossier de décision destiné au Secrétariat d'Etat à la formation, à la recherche et à l'innovation (SEFRI) et à la Conférence des directrices et directeurs des départements cantonaux de l'instruction publique (CDIP).

1. Introduction

Le plan de réalisation (voir chapitre 5) est le cœur de cette troisième partie. Il illustre la mise œuvre du concept par étapes et sous la forme d'incréments. L'idée de base de ce plan est que, dans un contexte déterminé, les acteurs concernés soient libres de déterminer l'ordre chronologique des incréments relatifs à la mise en place de la fédération des données. Ces acteurs sont toutefois impérativement tenus de commencer par l'incrément initial correspondant aux bases du plan de réalisation.

Les différents incréments décrits dans le plan de réalisation sont systématiquement subdivisés en quatre sous-catégories: la gouvernance, l'architecture, les activités de management et les mesures d'accompagnement. Ces sous-catégories sont systématiquement classées par activités principales en tenant compte de l'estimation des coûts correspondants (voir chapitre 5). Le besoin d'un cadre légal pour la création de la fédération des données est analysé au chapitre 4.

La fédération des données s'intègre parfaitement dans les stratégies poursuivies par la Confédération et par les cantons en matière de cyberadministration. Plusieurs exemples cantonaux démontrent en effet que la technologie du registre décentralisé des données ou de la blockchain est une solution choisie voire déjà appliquée en matière de cyberadministration.

Pouvoirs publics	Projet
Canton de Genève	par son université, délivre des justificatifs numériques permettant d'authentifier des diplômes par la blockchain; par son registre du commerce, délivre des justificatifs numériques en utilisant la blockchain
Canton de Zurich	par son université, teste des portefeuilles numériques et la délivrance de justificatifs numériques basés sur la technologie blockchain
Canton du Jura	par son registre des poursuites, délivre des justificatifs numériques attestant l'authenticité des documents grâce à la technologie blockchain
Canton d'Argovie	dans le cadre d'un projet, a recouru à la blockchain afin de délivrer des attestations de domicile sous forme de justificatifs numériques
Canton de Schaffhouse	a mis en place pour son registre des poursuites un projet pilote basé sur la blockchain et son projet eID
Confédération	par l'Office fédéral du personnel, examine la faisabilité d'une nouvelle carte de légitimation électronique pour les employés de l'administration fédérale en rapport avec l'eID

2. L'ensemble des processus de la fédération des données

La gouvernance de la fédération des données (voir chapitre 3 de la 2ème partie) relève le besoin de structures et de pratiques définies pour pouvoir garantir l'exécution des tâches par la fédération des données. Ce modèle organisationnel détermine pour l'essentiel, les structures, la nature des responsabilités et le devoir réciproque de rendre des comptes (voir 2ème partie du chapitre 3.1). A son tour, cette structure nécessite de reposer sur un ensemble de processus – autrement dit, de dresser un portrait de la totalité des processus et tâches existants qui devront être mis en œuvre par le secrétariat. Le graphique 1 est un résumé de cet ensemble de processus. La structure représentée dans le graphique 1 s'inspire en partie du modèle-cadre COBIT (en anglais: «control objectives for information and related technology»). Il a toutefois été adapté afin de tenir compte des spécificités de la formation professionnelle en Suisse. Le tableau suivant fournit un aperçu des principaux groupes de processus de cette structure:

Tableau 1: groupes de processus de la fédération des données

Groupe de processus	Compétence	Description
Gouvernance et pilotage	Commissions chargées de la politique de formation	L'entière responsabilité politique et institutionnelle, la coordination entre la Confédération, les cantons et les acteurs de la formation professionnelle ainsi que la surveillance du financement de la fédération des données.
Planification, alignement et organisation	Secrétariat	Toutes les tâches de gestion qui sont requises pour garantir le succès de la fédération des données, y compris la planification des ressources, la gestion de la qualité et l'obligation de rendre des comptes.
Acquisition et développement	Secrétariat	Toutes les activités en rapport avec le développement continu de la fédération des données: de la garantie d'une gouvernance des données appropriée à sa régulation jusqu'à la conception et au développement de nouvelles fonctionnalités.
Service de prestations et de support	Secrétariat (lifecycle management) Exploitation technique (exploitation informatique, continuité de l'exploitation)	Toutes les activités liées au cycle de vie d'un membre de la fédération (c.-à-d. adhésion, soutien, sortie) ainsi que l'exploitation technique, la maintenance de l'infrastructure et la continuité des services y afférents.
Mesure et garantie des performances	Secrétariat en collaboration avec des services d'audit externes	Toutes les tâches liées à l'évaluation de la fédération des données et au respect des exigences légales afin de fonctionner conformément aux normes et pratiques en vigueur.

Afin que cet ensemble de processus du secrétariat puisse se développer au même rythme que la mise en œuvre progressive de la fédération des données (voir chapitre 5.1), il devra être mis à jour et être adapté à la situation après le développement de la base et avant l'implémentation de chaque incrément. En outre, au sein de cet ensemble de processus, des articulations existent entre ces différents processus. Ces boucles de feedback sont représentées sous forme de flèche dans le graphique 1 ci-dessous.

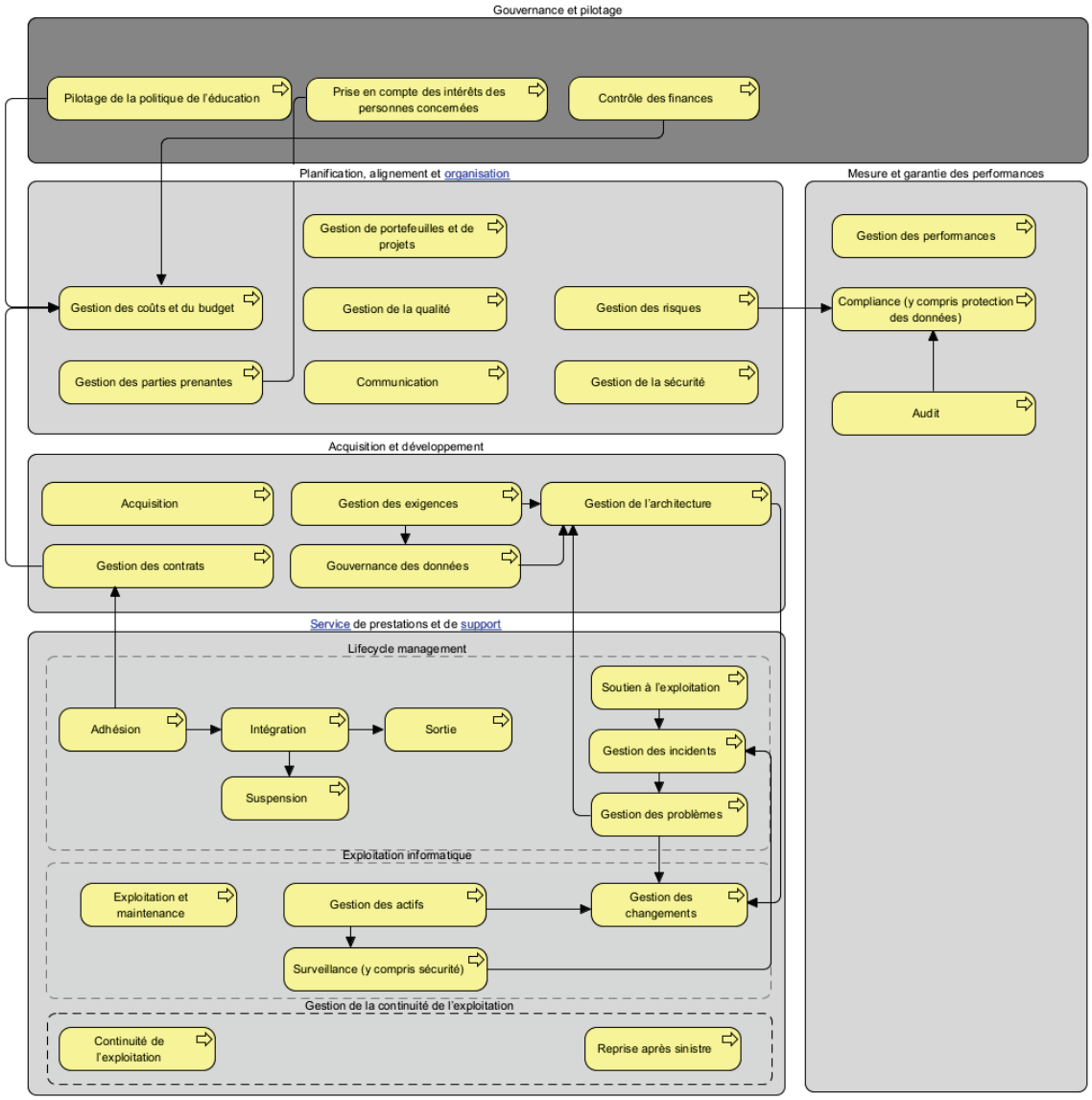

Figure 1: ensemble des processus pour le secrétariat de la fédération des données

3. Cas d'application: l'échange direct de données entre l'office de la formation professionnelle, les prestataires de cours interentreprises et les entreprises formatrices

Au sein de la fédération des données qui est proposée, différentes institutions et différents individus interagissent mutuellement. Cette interaction mutuelle est notamment mise en œuvre par l'échange de justificatifs numériques – à savoir des justificatifs qui contiennent la plupart du temps un ou plusieurs énoncés d'une institution ou d'une autorité (par ex. une école professionnelle) sur un sujet (par ex. sur une apprenante ou un apprenant). Le chapitre suivant, notamment par le deuxième cas d'application qui s'y trouve, illustre d'une part l'utilisation de portefeuilles électroniques par des entreprises et des institutions au sein de la fédération des données (à la différence du premier cas d'application du chapitre 3.1 de la 1ère partie qui est centré sur les apprenantes et les apprenants). D'autre part, ce deuxième cas d'application concrétise les principes d'architecture décrits au chapitre 3.3 de la 1ère partie dans la mise en œuvre pratique de la fédération des données. Enfin, ce chapitre et le cas d'application qui en fait partie expliquent la manière dont la fédération des données accélère et automatise les processus actuels d'échange de données.

Les exigences qui s'appliquent aux portefeuilles électroniques internes aux organisations, présentées dans cette partie, diffèrent nettement des exigences afférentes à un portefeuille électronique privé (voir le chapitre 3.1 de la 1ère partie). Concrètement: le portefeuille électronique interne à une organisation appartient à l'entreprise ou à l'institution concernée. Certes, les collaboratrices et collaborateurs de cette institution sont tenus d'interagir avec ce portefeuille électronique pour établir, maintenir et vérifier des justificatifs numériques. Cette interaction a lieu via une application spécifique dont l'apparence est comparable à une application commune, ou via une solution traditionnelle de l'administration scolaire. La nouveauté se trouve dans la manière dont les processus d'échange de données sont gérés dans l'arrière-plan du système.

Pour que les collaboratrices et les collaborateurs puissent travailler avec le portefeuille électronique interne à l'organisation, des réglementations internes à cette organisation sont nécessaires. En font partie, par exemple, le droits d'accès des collaboratrices et collaborateurs à des justificatifs existants spécifiques à l'entreprise (comme par ex. l'autorisation de formation), les mécanismes de répartition des justificatifs réceptionnés entre les collaboratrices et collaborateurs compétents, le droit des collaboratrices et collaborateurs de représenter l'entreprise dans une relation directe avec un autre membre ou avec un autre participant de la fédération des données (découle des principes d'architecture PA.G.5: «Accès légitimé à la fédération des données»; PA.S.1: «Contrôle d'accès basé sur des rôles» et PA.S.2: «Garantie d'identité des composants»). Il est envisageable que des collaboratrices et des collaborateurs s'annoncent auprès du portefeuille électronique interne à l'organisation en présentant le justificatif numérique correspondant (pour des raisons de simplification, ce processus d'annonce ne figure pas dans le cas d'application figurant ci-dessous). La validité des justificatifs numériques sera ensuite contrôlée via un portefeuille électronique interne à l'organisation grâce à la consultation d'une base de données interne à l'entreprise. De cette façon, le portefeuille électronique interne à l'organisation connaît également les droits d'accès et les

droits de traitement de cette personne. Ainsi, il sera possible d'automatiser des processus tels que, par ex., l'envoi de demandes ou de questions à des personnes déterminées.

Toutes les institutions et entreprises doivent définir de manière autonome, dans un modèle de gouvernance, les droits d'accès, les mécanismes de répartition et les autres réglementations sur le portefeuille électronique interne à l'organisation. Elles doivent les mettre en œuvre au sein de l'architecture technique du portefeuille électronique interne à leur organisation. L'unique exigence concernant ces réglementations internes à l'organisation posée par la fédération des données est la connectabilité de ce portefeuille électronique interne à l'organisation, à la fédération des données.

En garantissant la connectabilité, les portefeuilles électroniques internes aux organisations offrent aux institutions et aux entreprises concernées le potentiel de simplifier et d'automatiser l'échange des données dans le domaine de la formation professionnelle. Ce qui précède est concrétisé par les deux cas d'application suivants: 1) L'approbation par l'office de la formation professionnelle du contrat d'apprentissage d'Alina, une peintre en formation visant l'obtention du CFC et 2) L'inscription d'Alina et de ses collègues d'apprentissage professionnel aux cours interentreprises (ou cours IE).

Échange direct de données au moyen de portefeuilles électroniques internes à l'organisation

Le tableau 2 ainsi que le graphique 2 ci-dessous décrivent comment la formatrice professionnelle d'Alina et un collaborateur de l'office cantonal de la formation professionnelle gèrent l'approbation du contrat d'apprentissage d'Alina par le biais de leurs portefeuilles électroniques respectifs et internes à leur organisation respective. Le formulaire de contrat d'apprentissage d'Alina dûment rempli est déposé par l'entreprise formatrice – comme prévu dans le processus HAKA correspondant – dans notre cas, directement par la future formatrice professionnelle d'Alina. Ce formulaire de contrat d'apprentissage est ensuite complété et déposé selon les règles usuelles qui prévalent jusqu'à présent. L'échange de données ultérieur entre l'office de la formation professionnelle, l'entreprise formatrice, le prestataire de cours interentreprises (ou cours IE) et Alina, a lieu par le biais de la fédération des données. Dans ce contexte, la dernière colonne du tableau 2 présente les principes d'architecture propres à la fédération des données qui sont appliqués lors de chaque étape du processus d'échange des données.

Tableau 2: Échange de données et principes d'architecture aux fins d'approbation d'un contrat d'apprentissage

Étapes dans le processus d'échange de données	Principes d'architecture qui sont appliqués (voir chapitre 3.3 de la 1ère partie du concept)
1. Avant que le contrat d'apprentissage ne puisse être approuvé, l'office cantonal de la formation professionnelle est tenu de contrôler si l'entreprise formatrice dispose de l'autorisation de formation requise. C'est pourquoi, lors de la réception du formulaire de contrat d'apprentissage, le collaborateur de l'office de la formation professionnelle adresse une demande, par le biais du portefeuille électronique de l'office cantonal interne à l'organisation, à l'entreprise formatrice pour savoir si elle dispose d'une telle autorisation et ce, grâce à une liaison déjà existante, sécurisée et directe (car Alina n'est pas la première apprenante de cette entreprise). Dans sa demande, le collaborateur requiert de l'entreprise formatrice qu'elle lui présente, sous forme de dossier numérique, l'autorisation de formation qui avait été déjà établie à son attention (ce mode de faire est différent du processus d'approbation actuel).	• PA.G.1: Conservation fédéraliste des données • PA.S.5: Échange de données plus sécurisé et davantage conforme à la législation • PA.D.1: Les données sont un bien commun au service de la formation professionnelle • PA.N.1: Échange direct de données
2. Le portefeuille électronique de l'entreprise formatrice interne à l'organisation est configuré de telle sorte que l'entreprise formatrice réagit automatiquement à cette demande, répondant en temps réel, et transmet l'autorisation de formation à l'office de la formation professionnelle par dossier numérique. Dès lors, aucune intervention manuelle de la part de la formatrice professionnelle d'Alina est nécessaire.	• PA.G.1: Conservation fédéraliste des données • PA.S.5: Échange de données plus sécurisé et davantage conforme à la législation • PA.D.1: Les données sont un bien commun au service de la formation professionnelle • PA.N.1: Échange direct de données

		• PA.N.2: Atomicité des événements de formation • PA.N.3: Asynchronité de la communication
3.	Le contrôle de validité de l'autorisation de formation s'effectue lui aussi automatiquement. Après réception dans le portefeuille électronique de l'office cantonal interne à l'organisation du dossier numérique concernant l'autorisation de formation, ce portefeuille électronique procède à une consultation automatique du registre décentralisé des données. La seule tâche qui reste à effectuer par le collaborateur de l'office cantonal est qu'il doit encore contrôler si l'office dispose de toutes les informations requises pour pouvoir contrôler la validité du contrat d'apprentissage – toutes les autres tâches se déroulent en arrière-plan.	• PA.G.3: Disponibilité de données lors de prises de décisions • PA.G.4: Transparence, intégrité et responsabilité relatives à des événements pertinents pour la formation • PA.N.1: Échange direct de données • PA.N.3: Asynchronité de la communication
4.	Une fois que ce contrôle a été effectué, le collaborateur de l'office de la formation professionnelle établit un justificatif numérique du contrat d'apprentissage par l'application qui permet d'accéder au portefeuille électronique interne à l'organisation. À cet effet, le collaborateur utilise un schéma archivé dans le registre de données accessible publiquement.	• PA.D.3: Interprétabilité sémantique des données • PA.N.1: Échange direct de données • PA.N.3: Asynchronité de la communication
5.	Ensuite, le collaborateur de l'office de la formation professionnelle envoie le justificatif numérique du contrat d'apprentissage, via la fédération des données, à l'entreprise formatrice d'Alina avec l'indication selon laquelle le contrat d'apprentissage est approuvé. A partir de ce moment-là, l'entreprise formatrice peut faire état officiellement du fait qu'elle est «l'entreprise formatrice d'Alina». La formatrice professionnelle d'Alina est informée via le portefeuille électronique interne à l'organisation de la réception du justificatif électronique.	• PA.G.1: Conservation fédéraliste des données • PA.G.3: Disponibilité de données lors de prises de décisions • PA.S.5: Échange de données plus sécurisé et davantage conforme à la législation • PA.N.1: Échange direct de données • PA.N.2: Atomicité des événements de formation
6.	L'approbation du contrat d'apprentissage génère automatiquement une invitation de mise en relation pour Alina dans le portefeuille électronique de l'office cantonal interne à l'organisation et elle est envoyée à Alina, par exemple sous la forme d'un code QR (pour des détails sur la mise en place de cette mise en relation, voir le cas d'application «Justificatifs numériques de Luca» dans la 1ère partie du concept de la fédération des données).	• PA.S.3: Garantie de l'identité des parties concernées • PA.N.1: Échange direct de données • PA.N.3: Asynchronité de la communication

7.	Alina scanne la demande et l'accepte, via son portefeuille électronique privé, sur son terminal mobile (pour les détails sur un contrôle éventuel de l'identité d'Alina, voir le cas d'application «Justificatifs numériques de Luca» dans la 1ère partie du concept de la fédération des données).	• PA.S.3: Garantie de l'identité des parties concernées • PA.N.1: Échange direct de données
8.	Par l'acceptation de l'invitation à la mise en relation, le contrat d'apprentissage approuvé est envoyé à Alina – à nouveau de manière automatisée – via le portefeuille électronique interne à l'organisation sous forme de justificatif numérique par une liaison sécurisée.	• PA.G.1: Conservation fédéraliste des données • PA.G.3: Disponibilité de données lors de prises de décisions • PA.S.5: Échange de données plus sécurisé et davantage conforme à la législation • PA.N.1: Échange direct de données • PA.N.2: Atomicité des événements de formation • PA.N.3: Asynchronité de la communication
9.	Avec l'établissement du justificatif électronique relatif au contrat d'apprentissage, le prestataire de cours IE reçoit lui aussi simultanément et donc en temps réel, via une liaison directe et sécurisée déjà existante, une notification (voir terminologie en annexe) concernant l'approbation du contrat d'apprentissage d'Alina. Cette notification servira de pré-inscription au cours IE et ne contiendra que le minimum de données requises telles que les indications sur la profession, l'orientation spécialisée et l'entreprise formatrice. Dans la mesure où le futur responsable de cours IE d'Alina a reçu l'autorisation par le portefeuille électronique interne à l'organisation, cette notification lui sera aussi directement transmise. Dès lors, tant le prestataire de cours IE que le futur responsable de cours IE d'Alina sauront suffisamment tôt le nombre d'heures de cours qu'il faudra dispenser à Alina. Cela les aidera à planifier et à mettre en œuvre leur offre de cours IE.	• PA.G.3: Disponibilité de données lors de prises de décisions • PA.S.5: Échange de données plus sécurisé et davantage conforme à la législation • PA.N.1: Échange direct de données • PA.N.2: Atomicité des événements de formation • PA.N.3: Asynchronité de la communication

Le tableau 3 décrit l'échange de données et les principes d'architecture concernés du deuxième exemple – à savoir l'inscription d'Alina aux cours IE via le prestataire des cours IE. Pour qu'Alina soit définitivement inscrite aux cours IE, le prestataire des cours IE envoie une demande aux apprenantes et aux apprenants qui se sont inscrits, et donc aussi à Alina. Pour ce faire, il est nécessaire de parcourir les étapes suivantes qui figurent également dans le graphique 2:

Tableau 1: Échange de données et principes d'architecture correspondants pour l'inscription définitive auprès des cours IE.

Étapes dans le processus d'échange de données	Principes d'architecture qui sont appliqués (voir chapitre 3.3 de la 1ère partie du concept)
1. Le futur responsable de cours IE d'Alina établit une invitation pour une liaison sécurisée et directe avec Alina au moyen du portefeuille électronique interne à l'organisation du prestataire de cours IE. Il envoie cette invitation à Alina, par exemple en tant que code QR dans un courriel (pour les détails sur un éventuel contrôle de l'identité d'Alina, voir le cas d'application «Justificatifs numériques de Luca» dans la 1ère partie du concept de la fédération des données).	• PA.S.3: Garantie de l'identité des parties concernées • PA.N.1: Échange direct de données • PA.N.2: Atomicité des événements de formation
2. Alina scanne ce code QR avec son portefeuille électronique personnel et accepte la mise en relation. Ce canal de communication sera maintenu pendant toute la durée de la formation puisqu'on pourra également transmettre ultérieurement, via cette mise en relation, des attestations de compétences sous forme de justificatifs numériques.	• PA.S.3: Garantie de l'identité des parties concernées • PA.S.5: Échange de données plus sécurisé et davantage conforme à la législation • PA.N.1: Échange direct de données • PA.N.2: Atomicité des événements de formation • PA.N.3: Asynchronité de la communication
3. Puis le portefeuille électronique interne à l'organisation du prestataire de cours IE envoie une demande automatique à Alina la priant de lui transmettre toutes les informations détaillées encore nécessaires pour les cours IE. Cette demande contient des indications claires sur les données précises requises de la part d'Alina.	• PA.S.5: Échange de données plus sécurisé et davantage conforme à la législation • PA.D.1: Les données sont un bien commun au service de la formation professionnelle • PA.D.3: Interprétabilité sémantique des données • PA.N.1: Échange direct de données • PA.N.2: Atomicité des événements de formation • PA.N.3: Asynchronité de la communication

4.	Alina reçoit ensuite une demande dans son portefeuille électronique personnel. Elle autorise la transmission des informations souhaitées. Ainsi, le dossier numérique est renvoyé au portefeuille électronique interne à l'organisation du prestataire de cours IE via la liaison directe cryptée. Le dossier comprend notamment des données sur la personne d'Alina, mais aussi sur sa situation en matière de formation (il contient aussi, par ex. des données de contact de sa formatrice professionnelle).	• PA.G.1: Conservation fédéraliste des données • PA.S.5: Échange de données plus sécurisé et davantage conforme à la législation • PA.D.1: Les données sont un bien commun au service de la formation professionnelle • PA.N.1: Échange direct de données • PA.N.2: Atomicité des événements de formation • PA.N.3: Asynchronité de la communication
5.	Une fois reçu par le portefeuille électronique interne à l'organisation du prestataire de cours IE, la validité du dossier est automatiquement contrôlée au moyen d'une consultation du registre décentralisé des données. Le dossier est basé sur un justificatif numérique valable. Le futur responsable de cours IE d'Alina est informé de la réception du dossier, et Alina est ainsi inscrite auprès des cours IE.	• PA.G.3: Disponibilité de données lors de prises de décisions • PA.G.4: Transparence, intégrité et responsabilité relatives à des événements pertinents pour la formation • PA.N.1: Échange direct de données • PA.N.3: Asynchronité de la communication
6.	Par cette inscription, une demande est automatiquement et immédiatement envoyée au portefeuille électronique interne à l'organisation afin de mettre en place une liaison directe et cryptée à l'attention de la formatrice professionnelle d'Alina, par exemple en tant que lien web dans un courriel.	• PA.S.3: Garantie de l'identité des parties concernées • PA.S.5: Échange de données plus sécurisé et davantage conforme à la législation • PA.N.1: Échange direct de données • PA.N.2: Atomicité des événements de formation • PA.N.3: Asynchronité de la communication
7.	La formatrice professionnelle d'Alina ouvre ce lien via son application qui permet d'accéder au portefeuille électronique interne à l'organisation, et elle accepte la mise en relation.	• PA.S.3: Garantie de l'identité des parties concernées • PA.S.5: Échange de données plus sécurisé et davantage conforme à la législation • PA.N.1: Échange direct de données • PA.N.3: Asynchronité de la communication

8.	La notification selon laquelle Alina est inscrite aux cours IE est immédiatement reçue par le portefeuille électronique interne à l'organisation de l'entreprise formatrice d'Alina. La formatrice professionnelle d'Alina est automatiquement informée de la réception de cette notification via le portefeuille électronique interne à l'organisation puisqu'elle a donné son consentement à la transmission de notifications de ce genre.	• PA.G.3: Disponibilité de données lors de prises de décisions • PA.S.5: Échange de données plus sécurisé et davantage conforme à la législation • PA.N.1: Échange direct de données • PA.N.2: Atomicité des événements de formation. • PA N.3: Asynchronité de la communication
9.	Aussitôt que les premières dates définitives des cours IE sont connues, le responsable des cours IE (ou l'administration du prestataire de cours IE) peut convoquer Alina à ces cours via la liaison directe.	• PA.G.3: Disponibilité de données lors de prises de décisions • PA.S.5: Échange de données plus sécurisé et davantage conforme à la législation • PA.N.1: Échange direct de données • PA.N.2: Atomicité des événements de formation • PA.N.3: Asynchronité de la communication
10.	L'entreprise formatrice d'Alina reçoit également une notification sur la convocation d'Alina aux cours IE et, par le biais du portefeuille électronique interne à son organisation, la formatrice professionnelle d'Alina reçoit elle aussi une notification sur cette convocation.	• PA.G.3: Disponibilité de données lors de prises de décisions • PA.S.5: Échange de données plus sécurisé et davantage conforme à la législation • PA.N.1: Échange direct de données • PA.N.2: Atomicité des événements de formation • PA.N.3: Asynchronité de la communication

Grâce à la fédération des données, le responsable de cours IE d'Alina garde toujours une vue d'ensemble actualisée sur le nombre de participantes et participants à son cours IE. De plus, si Alina vient à ne pas pouvoir participer un jour à un cours, par exemple parce qu'elle est tombée malade, elle peut annoncer son absence au prestataire de cours IE via la liaison directe existante (et, si nécessaire, elle pourra compléter cette annonce avec un certificat médical sous la forme d'un dossier numérique). Le responsable de cours IE d'Alina en sera ainsi immédiatement informé, et il pourra saisir cette absence. De cette façon, une annonce dans ce sens à l'entreprise formatrice d'Alina sera à son tour déclenchée (et, de ce fait, sera aussi envoyée à sa formatrice professionnelle) et les informera qu'Alina a manqué ses cours IE pour cause de maladie. Ainsi, la fédération des données contribue à ce que toutes les informations nécessaires soient connues des acteurs concernés en temps réel, et ce, via des liaisons directes cryptées.

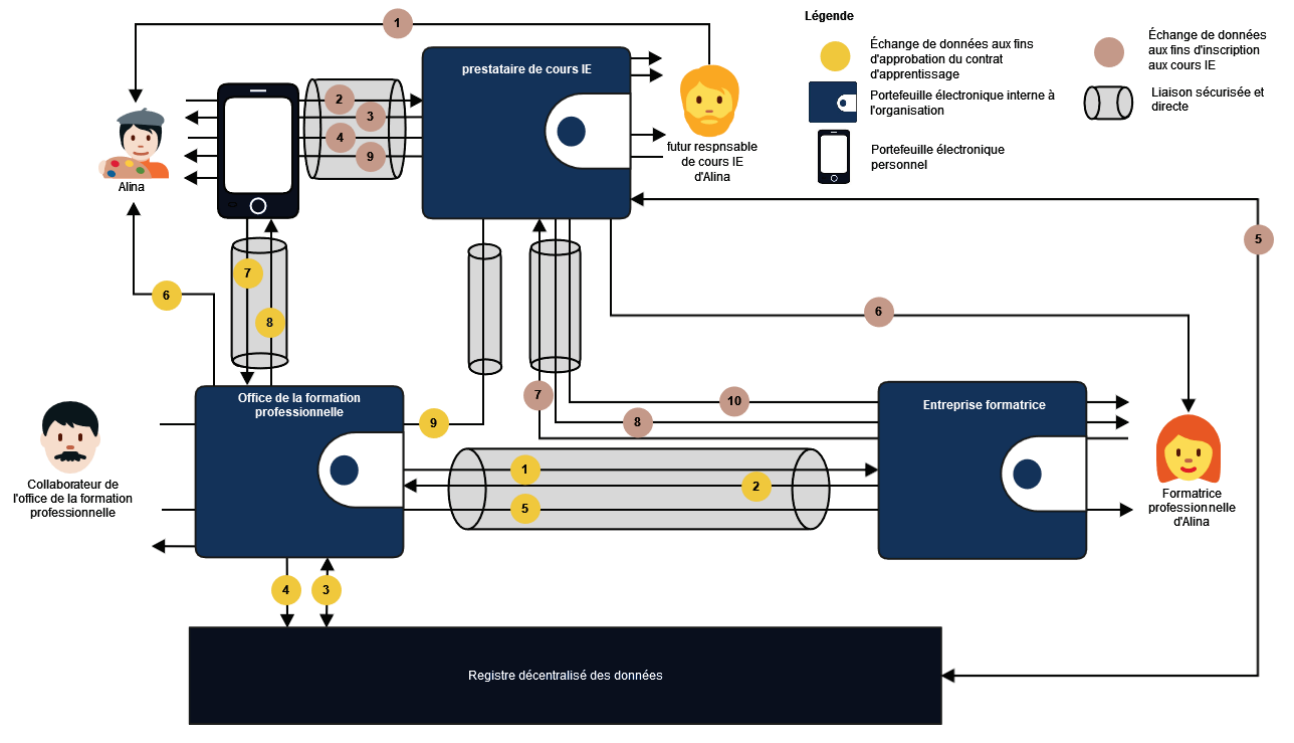

Figure 2: l'approbation du contrat d'apprentissage d'Alina et son inscription aux cours interentreprises (ou cours IE)

4. Besoin de régulation pour la fédération des données: conventions intercantonales

Selon le présent concept, une future fédération des données sera créée et exploitée dans un contexte de droit public. La création de l'organisation «fédération des données» nécessite la conclusion d'une convention entre tous les acteurs responsables. Une fois mise en exploitation, la fédération des données aura pour effet de simplifier la transmission des données par les cantons dans un cadre valable à l'échelon de toute la Suisse. Cette transmission de données et le traitement des données personnelles liées à cette dernière devront se faire en tenant compte des législations cantonales sur la protection des données. Une analyse montre en effet que le potentiel de la fédération des données pourrait mieux déployer ses effets sur la base d'une convention intercantonale commune sur la communication des données.

4.1 Création de la fédération des données

Conformément à sa compétence en matière de formation professionnelle, la Confédération a rédigé la Loi sur la Formation professionnelle (LFPr), loi complétée par des ordonnances. L'exécution est toutefois du ressort des cantons. Ainsi, il appartient à ces derniers de s'organiser de manière à permettre une application conforme de la LFPr. Les cantons sont en effet autonomes dans leur organisation, conformément à l'article 47 de la Constitution fédérale (Cst.). Il convient également de relever que, contrairement au domaine de l'instruction publique, la Confédération ne peut forcer les cantons à adhérer à une convention intercantonale relative à la formation professionnelle. En matière de cyberadministration, la Confédération est également dépourvue du pouvoir d'imposer des mesures organisationnelles aux cantons. C'est en effet ce qu'a rappelé la Chancellerie fédérale en prenant position en 2011 sur la question des bases légales pour la collaboration TIC entre la Confédération et les cantons. Cette prise de position a conduit la Confédération et les cantons à signer une convention-cadre relative à la cyberadministration afin de favoriser la collaboration en la matière entre les cantons et la Confédération.

La fédération des données poursuit différents buts, notamment celui de favoriser la collaboration entre les cantons en matière de formation professionnelle. Sa mise en œuvre implique la création et l'usage de ressources technologiques. La fédération des données requiert dès lors des changements communs d'organisation de la part des cantons afin que leurs institutions, membres de la fédération des données, puissent en faire usage de manière collaborative, uniforme et interopérable. La décision de créer la solution commune qu'est la fédération des données ainsi que celle de la rejoindre est une compétence qui, de notre avis, ressort du domaine de la cyberadministration et appartient aux cantons. Pour les raisons qui précèdent, il nous apparaît qu'une convention intercantonale (le terme «convention» est à comprendre dans ce texte dans le sens générique, sans distinction entre la notion de concordat, accord, règlement ou convention) est nécessaire afin de fonder la fédération des données et ainsi permettre aux cantons de se soutenir et d'unir leurs efforts organisationnels en matière de formation professionnelle. D'autres paramètres tels que la forme de la fédération des données, le droit applicable ainsi que le financement conviennent d'être réglés. Considérant que

la formation professionnelle est une compétence fédérale, la Confédération est un partenaire clé pour les cantons.

L'élaboration de cet élément est incluse dans le plan de réalisation de l'élaboration des bases de la fédération des données (voir chapitre 5).

4.2 Échange de données intercantonal

Afin que des données puissent être communiquées collaborativement entre les membres de la fédération des données de différents cantons à travers la fédération de données, l'harmonisation de dispositions cantonales en matière de protection des données est également recommandée. En effet, à mesure que les institutions cantonales sont soumises à leur propre loi-cadre en matière de protection des données, que la LFPr, en tant que loi sectorielle fédérale, est directement exécutée par ces mêmes institutions cantonales, que LFPr n'apporte cependant pas de précision concernant la communication intercantonale de données, un accord intercantonal sur ce point est attendu afin de permettre aux membres de la fédération des données de tous les cantons de profiter pleinement du potentiel de la fédération des données. Toutefois, l'absence d'une telle convention ne fait pas obstacle à la mise en œuvre de la fédération des données dès lors que les données peuvent être communiquées entre les institutions d'un même canton conformément aux législations applicables. Les cantons autorisant déjà la communication de données à d'autres cantons conformément à leur législation cantonale ou sur la base de conventions intercantonales pourront également bénéficier dans cette mesure des avantages de la collaboration intercantonale par le biais de la fédération des données.

L'élaboration de cet élément ne fait pas partie du plan de réalisation.

5. Plan de réalisation de la fédération des données

Le plan de réalisation prévoit une approche modulaire où de nouvelles fonctions sont résumées sous la forme d'incréments qui reposent sur des bases communes. Grâce à l'approche modulaire, les prestations à fournir peuvent être fournies de manière relativement indépendante les unes des autres, ce qui ouvre des possibilités de structuration flexible de la réalisation. Cette structuration fonctionnelle de la fédération des données est en accord avec le modèle d'exploitation retenu (voir par ex. le paysage des processus) et rend possible une extension sans accrocs des prestations sans donner lieu à des pannes ou défaillances dans les affaires courantes.

La mise en œuvre de la fédération des données a lieu par incrémentations. Chaque incrément est mis en œuvre en étroite collaboration avec les acteurs impliqués. Cela permet un apprentissage en continu et cela génère une plus-value directe pour la formation professionnelle. Les éléments suivants sont prévus dans le plan de réalisation:

- Bases: la création de bases de gouvernance et de bases architecturales sont requises pour la mise en place d'une première infrastructure pour les incréments présentés ci-après (y compris pour la mise au concours de l'exploitation technique).
- Incrément #1: établissement des CFC en tant que justificatifs numériques

- Incrément #2: inscription aux cours IE
- Incrément #3: attestations de l'école professionnelle en tant que justificatifs numériques
- Incrément #4: octroi de l'autorisation de formation
- Incréments supplémentaires en fonction des besoins (ces incréments ne sont pas inclus dans le plan de réalisation)

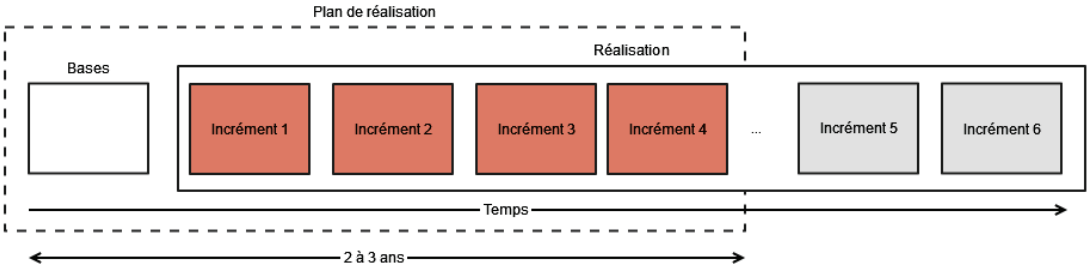

Figure 3: déroulement schématique de la réalisation

Chaque incrément contient les catégories suivantes:

- Gouvernance: toutes les affaires qui ont trait à la surveillance et à la régulation de la fédération des données, y compris celles qui concernent le pilotage de la politique de formation ainsi que des systèmes formels pour l'audit et la compliance (par ex. ISMS) et autres conventions (par ex. standards, règlement d'organisation, contrats).
- Architecture: toutes les mesures techniques et organisationnelles qui garantissent l'exploitation dans les règles de l'art de la fédération des données conformément aux principes d'architecture. Cela n'inclut pas seulement l'infrastructure technique et les composants (y compris d'autres interfaces au sein de la formation professionnelle), mais encore les processus du secrétariat.
- Exploitation: toutes les activités gravitant autour de l'exploitation quotidienne de la fédération des données (par ex. maintenance de l'infrastructure, activités du secrétariat, etc.).
- Gestion: activités de soutien en lien avec l'organisation, planification des ressources et contrôle du travail (par ex. gestion de projets et gestion des risques, contrôle des coûts et gestion du personnel).
- Mesures d'accompagnement: activités de soutien qui garantissent que les résultats peuvent être utilisés avec succès (par ex. documentation, formation et communication).

5.1 Présentation détaillée du plan de réalisation

Les tâches de projet prioritaires pour la réalisation de la fédération des données sont présentées dans les lignes qui suivent. Sont exclues de la présente liste toutes les tâches opérationnelles qui concernent le secrétariat et l'exploitation de l'infrastructure. L'estimation des coûts est calculée sous forme de «jours-personne» auxquels sont attribués ultérieurement différents pourcentages dans le modèle de coûts. De surcroît, des coûts supplémentaires sont occasionnés à des tiers dans quelques cas (par ex. prestations de services, licences, etc.) qui sont également pris en compte dans le modèle de coûts comme suit:

Plan de realization Bases	Coûts (Personnel) Jours-per-sonne/ int	Jours-per-sonne/ ext	Frais de tiers dû
Governance			
Pilotage de la politique de formation	24	0	
Convention de prestations, règlement d'organisation et modèle d'organisation	20	10	
Implication des parties prenantes	20	0	
Paysage des processus du secrétariat et concept d'exploitation	54	0	
Bases de la gestion des contrats (à savoir SLA, contrats de la fédération)	20	5	
Data Governance Framework & Standard	20	0	
Système de gestion de la sécurité de l'information (ISMS, ISO:IEC 27001:2022) et estimation des conséquences de la protection des données (DSFA)	10	10	
Conseil et instruction juridiques	0	20	
Coûts supplémentaires en lien avec des tâches déléguées (par ex. coordination)	5	25	
Architecture			
Infrastructure de base (nodes, registre décentralisé des données et réseau)	35	30	
Infrastructure de gestion des identités et de la sécurité (y compris Edulog, e-ID), y compris protection des données	10	20	
Gestion de la qualité, testing et validation	25	15	
Infrastructure technique de soutien (par ex. container repo, versioning, etc.)	5	1	X
Gestion			
Mise au concours de l'exploitation technique	10	5	
Gestion de projet (y compris gestion des risques)	24	0	
Administration et soutien (par ex. RH, comptabilité, etc.)	10	0	
Exploitation			
Exploitation de l'infrastructure informatique de base (infrastructure de base et prestations de réseau)	0	60	X
Managed servers & containers	0	0	X
Storage & backups	0	0	X
Exploitation du secrétariat	24	0	
Mesures d'accompagnement			
Documentation des directives techniques	24	0	
Présence web, y compris protection de la marque de la fédération des données	20	20	X
Communication avec les parties prenantes (y compris présence web, séances d'information, etc.)	2	0	X

Plan de realization Incrément #1: établissement des CFC en tant que justificatifs numériques	Coûts (Personnel) Jours-per-sonne/ int	Jours-per-sonne/ ext	Frais de tiers dû
Gouvernance			
Architecture des données du registre décentralisé de données aux fins d'établissement des CFC	5	5	
Concept de sécurité de l'information et de protection des données (ISDS) pour l'incrément 1	5	5	
Conseil pour les processus de la formation professionnelle	20	0	
Architecture			
Développement de base de l'infrastructure pour l'établissement des CFC	15	30	
Configuration de l'infrastructure pour le raccordement d'applications de standard wallet	5	10	
Procédures pour l'établissement et le contrôle des CFC numériques	10	20	
Onboarding des applications existantes	0	0	
Développement, configuration, y compris les droits d'accès de l'infrastructure des agents	10	5	
Intégration dans les processus administratifs du secrétariat	10	0	
Mesures techniques et organisationnelles dans le domaine de la sécurité de l'information et de la protection des données	20	5	X
Gestion de la qualité, testing et validation	10	15	
Gestion			
Gestion de projet (y compris gestion des risques)	24	0	
Administration et soutien (par ex. RH, comptabilité, etc.)	10	0	
Exploitation			
Coûts d'exploitation technique courants (au pro rata)	0	60	X
Exploitation du secrétariat	34	0	
Mesures d'accompagnement			
Spécification technique des agents	10	5	
Soutien aux offices de la formation professionnelle lors de l'adaptation des processus d'établissement	24	0	
Mesures de communication à l'attention des parties prenantes dans le contexte «CFC numériques»	20	0	
Mesures de sensibilisation pour les élèves/pour les jeunes	20	0	

Plan de realization Incrément #2: inscription aux cours IE	Coûts (Personnel) Jours-per-sonne/ int	Jours-per-sonne/ ext	Frais de tiers dû
Gouvernance			
Architecture des données du registre décentralisé des données aux fins d'inscription aux cours IE	5	2	
Conception de sécurité de l'information et de protection des données (ISDS) pour l'incrément 2	5	2	
Conseil pour les processus de la formation professionnelle	20	0	
Architecture			
Infrastructure technique et procédures aux fins d'inscription des apprenant.e.s aux cours IE	15	20	
Spécification et configuration de l'infrastructure des agents de cloud	10	0	
Développement d'une application standard de controllers pour les entreprises formatrices et pour les organisations du monde du travail (OrTrA)	10	50	
Développement, configuration, y compris les droits d'accès de l'infrastructure des agents	15	5	
Intégration dans les processus administratifs du secrétariat	10	0	
Mesures techniques et organisationnelles dans le domaine de la sécurité de l'information et de la protection des données	15	5	
Gestion de la qualité, testing et validation	10	15	
Gestion			
Gestion de projet (y compris gestion des risques)	15	0	
Administration et soutien (par ex. RH, comptabilité, etc.)	10	0	
Exploitation			
Coûts d'exploitation techniques courants (au pro rata)	0	60	X
Exploitation du secrétariat	34	0	
Mesures d'accompagnement			
Spécification technique pour les controllers et leurs agents	10	5	
Soutien aux OrTra et aux entreprises formatrices lors de l'adaptation des processus d'inscription	30	0	
Mesures de communication à l'attention des parties prenantes dans le contexte «inscription aux cours IE»	20	0	
Mesures de sensibilisation pour formatrices et formateurs professionnels et pour prestataires de cours IE	20	0	

Plan de realization Incrément #3: attestations d'écoles professionnelles en tant que justificatifs numériques	Coûts (Personnel) Jours-personne/ int	Jours-personne/ ext	Frais de tiers dû
Gouvernance			
Architecture de données du registre décentralisé des données aux fins d'établissement d'attestations d'écoles professionnelles	5	2	
Concept de sécurité de l'information et de protection des données (ISDS) pour l'incrément 2	5	2	
Conseil pour les processus de la formation professionnelle	20	0	
Architecture			
Infrastructure technique et procédures aux fins d'établissement de certificats numériques	15	40	
Développement, configuration, y compris les droits d'accès de l'infrastructure des agents	10	20	
Onboarding des applications existantes	0	0	
Intégration dans les processus administratifs du secrétariat	10	0	
Mesures techniques et organisationnelles dans le domaine de la sécurité de l'information et de la protection des données	15	5	
Gestion de la qualité, testing et validation	10	15	
Gestion			
Gestion de projet (y compris gestion des risques)	15	0	
Administration et soutien (par ex. RH, comptabilité, etc.)	10	0	
Exploitation			
Coûts d'exploitation technique courants (au pro rata)	0	60	X
Exploitation du secrétariat	51	0	
Mesures d'accompagnement			
Spécification technique pour les controllers et leurs agents	10	5	
Soutien aux écoles professionnelles lors de l'adaptation des processus d'établissement	20	0	
Mesures de communication à l'attention des parties prenantes dans le contexte «Attestations d'écoles professionnelles»	20	0	
Mesures de sensibilisation destinées aux responsables des écoles professionnelles	20	0	

Plan de realization Incrément #4: octroi de l'autorisation de formation	Coûts (Personnel) Jours-personne/ int	Jours-personne/ ext	Frais de tiers dû
Gouvernance			
Architecture de données du registre décentralisé des données pour l'octroi d'autorisation de formation	5	2	
Concept de sécurité de l'information et de protection des données (ISDS) pour l'incrément 2	5	2	
Conseil pour les processus de la formation professionnelle	15	0	
Architecture			
Infrastructure technique et procédures aux fins d'octroi d'autorisations de formation numériques	10	25	
Intégration dans les processus administratifs du secrétariat	10	0	
Mesures techniques et organisationnelles dans le domaine de la sécurité de l'information et de la protection des données	10	5	
Gestion de la qualité, testing et validation	7.5	10	
Gestion			
Gestion de projet (y compris gestion des risques)	12	0	
Administration et soutien (par ex. RH, comptabilité, etc.)	8	0	
Exploitation			
Coûts d'exploitation technique courants (au pro rata)	0	40	X
Exploitation du secrétariat	51	0	
Mesures d'accompagnement			
Spécification technique pour les controllers et leurs agents	5	5	
Soutien aux offices de la formation professionnelle lors de l'adaptation des processus d'établissement	20	0	
Mesures de communication à l'attention des parties prenantes dans le contexte «Autorisation de formation»	20	0	

5.2 Mesures d'accompagnement pour la réalisation du concept

Une caractéristique essentielle de la planification du projet réside dans la prise en compte de mesures d'accompagnement afin de garantir la durabilité de l'impact des incréments. Cela présuppose que l'on calcule l'utilité visée et que les facteurs d'impulsion nécessaires à cet effet soient identifiés. Les mesures d'accompagnement typiques sont la documentation relative aux interfaces, la formation des personnes qui utiliseront la fédération des données ainsi que des mesures de communication. Il faudra désigner les personnes responsables de l'exécution de ces mesures, et il faudra aussi garantir que ces personnes rendent des comptes sur l'efficacité de ces mesures. Il s'agit d'une tâche qui relève de la gouvernance.

Ces mesures d'accompagnement sont souvent soumises à une responsabilité partagée qui exige une volonté politique et l'engagement des personnes dirigeantes.

6. Coûts et financement du plan de réalisation

L'estimation des coûts est soumise à un certain degré d'incertitude. Il est possible de déterminer les principaux facteurs de coûts ainsi que la dynamique des coûts d'un tel projet majeur en se basant sur des projets d'infrastructure comparables réalisés dans le secteur de la formation. De plus, des hypothèses sont adoptées sur la stratégie en matière de coûts (par exemple via une stratégie «cloud first» où ce sont surtout les coûts d'exploitation, en lieu et place des coûts d'investissement, qui jouent un rôle). L'estimation des coûts relatifs à la mise en œuvre se réfère à un domaine de développement qui porte sur la mise à disposition de nouvelles fonctionnalités (considérés conjointement avec les mesures d'accompagnement correspondantes) ainsi que sur un domaine d'exploitation qui inclut le secrétariat et l'exploitation technique. En règle générale, les coûts de secrétariat sont des coûts uniques et variables et reflètent les exigences posées aux fonctionnalités à mettre en œuvre. Les coûts d'exploitation technique, pour leur part, demeurent relativement stables et s'accroissent légèrement avec l'augmentation de l'utilisation de la fédération des données. Les coûts annexes, comme par exemple les coûts qui sont engendrés par des mesures d'intégration, ne sont pas pris en compte dans le présent modèle de coûts puisqu'ils doivent être supportés par les membres (par ex. les entreprises formatrices, les OrTra, etc.).

6.1 Modèle de comptabilité analytique de la fédération des données

Le graphique suivant illustre les principes de base du modèle de comptabilité analytique de la fédération des données. La planification des coûts du secrétariat est basée sur le plan comptable classique d'une entreprise en Suisse.

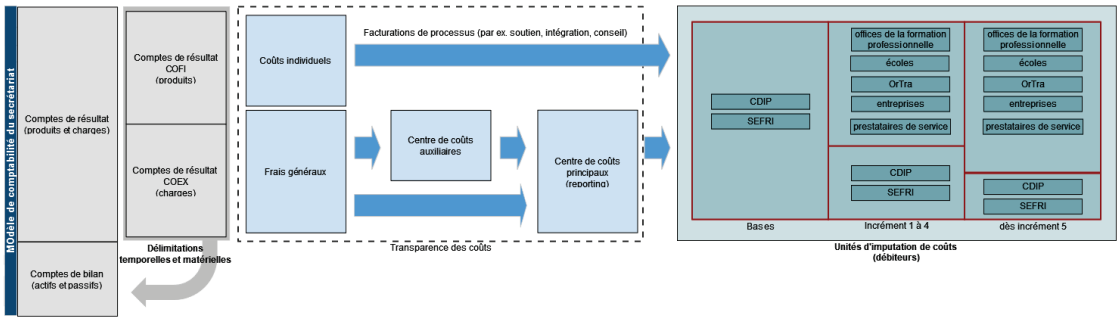

Figure 4: modèle de comptabilité analytique de la fédération des données

Les remarques suivantes sont pertinentes pour la gestion des coûts de la fédération des données:

- Les unités finales d'imputation reflètent le financement de la fédération des données, et pas la comptabilité analytique des prestations de services de la fédération. Le reporting destiné au pilotage s'effectue au moyen des centres de coûts principaux, à savoir: la gouvernance, le secrétariat et l'exploitation. Pour les raisons précitées, dans le présent modèle, les unités finales d'imputation sont assimilées aux débiteurs (du côté facturation). Le financement est décrit de manière plus détaillée ci-après (voir chapitre 6.3).
- Les coûts individuels sont en majorité directement attribuables et variables. Pour cette raison, en règle générale, ils sont gérés via des processus payants (par ex. prestations de soutien, intégration, etc.) ou autres prestations (par ex. conseil). Des coûts de ce genre sont directement facturés aux unités finales d'imputation (c.-à-d. aux débiteurs).
- Les frais généraux sont des coûts relativement fixes qui doivent être supportés par tous les acteurs. Ils sont subdivisés en deux sous-catégories comme suit:
    - Les centres de frais auxiliaires sont des frais indirects qui, comme leur nom l'indique, ne sont pas directement associés à la fourniture de prestations (par ex. coûts de management, de communication, de conseil juridique, etc.). Des coûts de ce genre figurent sous le centre de coûts principal et sont ventilés sur une seule et même base convenue pour tous les acteurs. Cette répartition des (coûts indirects non productifs) est décrite de manière plus détaillée ci-après.
    - En revanche, les coûts principaux sont des coûts indirects, mais qui ne sont pas causés via une participation immédiate aux coûts de la fédération des données (par ex. coûts d'administration, coûts d'utilisation de l'infrastructure, etc.). Ces coûts sont ventilés sur toutes les unités finales d'imputation sur une base de répartition équitable (à savoir via une clé de répartition), comme décrit de manière plus détaillée ci-après. Il faut toutefois prendre garde au fait que les clés de répartition peuvent être chaque fois différentes selon le centre de coûts principal concerné.
- Les frais généraux se gèrent le mieux via les dépenses d'exploitation en lieu et place des dépenses d'investissement. Ainsi, il serait par exemple judicieux d'envisager un achat de cloud pour l'hébergement de nodes et d'agents.
- Les délimitations matérielles provenant de la comptabilité d'exploitation (COEX) contiennent, par exemple, des impôts, des amortissements, des intérêts calculatoires, etc. et sont d'habitude délimitées dans le sens de la comptabilité financière (COFI).

6.2 Coûts de réalisation et coûts d'exploitation

Sur une période de trois ans, les coûts globaux s'élèvent à quelque 4,3 millions de francs. Cela correspond à un capital d'exploitation de 120'000 francs par mois en moyenne. Toutefois, dans la pratique, les coûts de développement et d'exploitation se situent en moyenne à moins d'un million de francs par année. Pendant la première année, on crée les bases pour tous les autres incréments de sorte que pour cette première année seulement, les coûts d'exploitation sont supérieurs à un million de francs. Cela s'explique notamment par le fait que c'est au début du projet que les coûts de gouvernance sont générés, mais que ces coûts diminueront nettement lors des incréments suivants. L'analyse des coûts – présentée ici sous une forme condensée – part du principe que la réalisation se fera par étapes au moyen de

ces incréments. Nous visons des effets de synergies des coûts dus à l'impact de l'apprentissage sur les acteurs concernés.

Conformément au plan de réalisation, les coûts sont subdivisés en différentes catégories (par ex. gouvernance, architecture, exploitation, etc.). Chaque incrément donne lieu également à un autre thème prioritaire architectural et, dès lors, engendre des mesures d'accompagnement liées à ce thème. Ces mesures d'accompagnement ont pour but de garantir que la nouvelle fonctionnalité soit correctement comprise, et qu'elle soit utilisée de manière appropriée. Enfin, en raison de l'existence de cette procédure incrémentale, les coûts d'exploitation évoluent de manière stable.

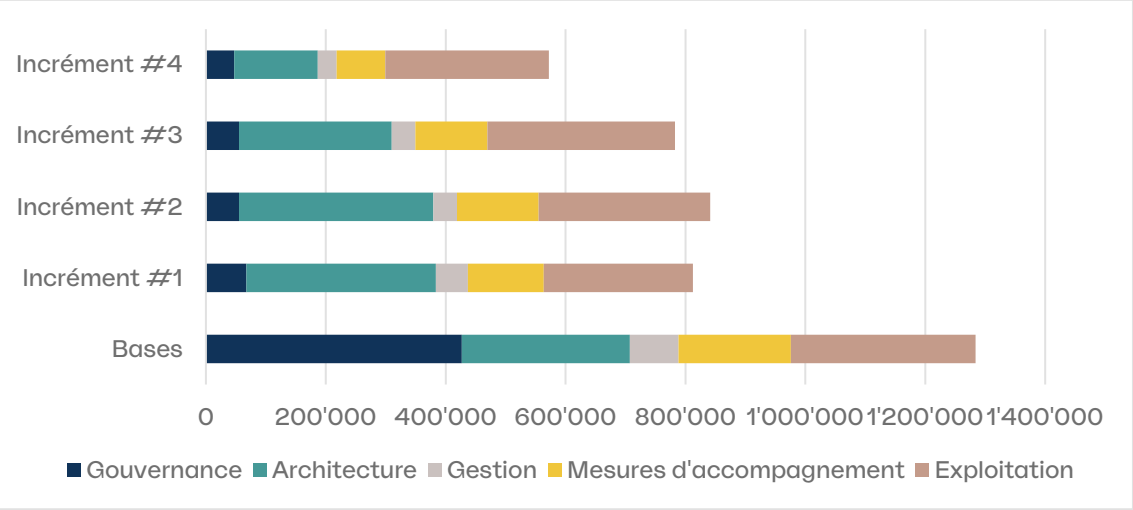

Figure 5: Les coûts selon la phase de projet

Le tableau de flux de trésorerie, qui est composé des coûts de développement et des coûts d'exploitation récurrents, est saisi de manière séparée pour la période s'étendant jusque et y compris à l'incrément 4, et présente un flux de trésorerie relativement stable. Ce dernier peut donc facilement s'intégrer dans les cycles budgétaires des organismes responsables de la fédération des données.

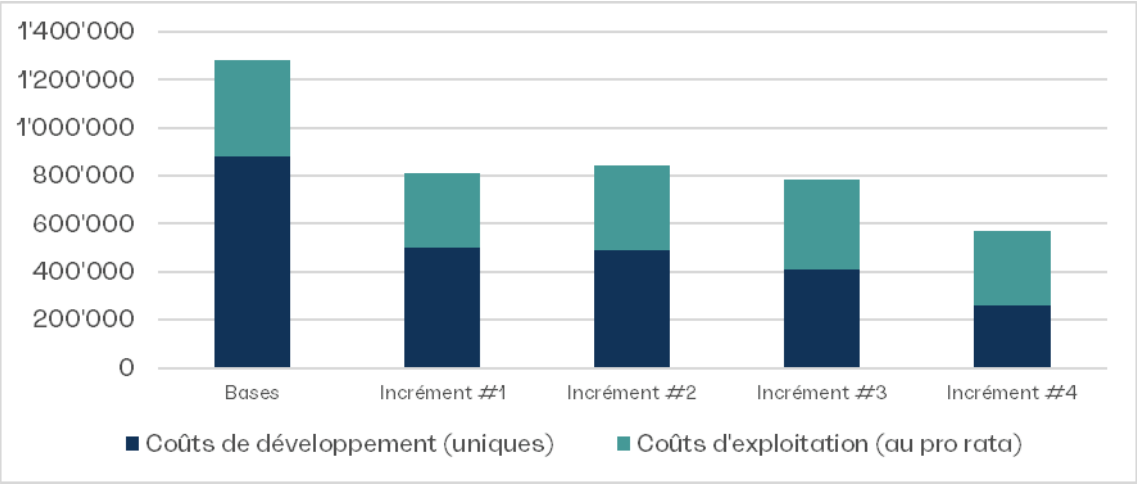

Figure 6: tableau de flux de trésorerie selon l'incrément concerné

Tableau 2a: coûts globaux de la réalisation de la fédération des données **(bases, incréments 1 à 4)**

	CHF (à l'excl. de la TVA)
Gouvernance	652'600
Architecture	1'314'500
Exploitation	1'428'800
Gestion	245'700
Mesures d'accompagnement	651'900
Coûts globaux	*4'293'500*

Tableau 4b: répartition des coûts globaux selon les phases de projet (bases et incréments 1 à 4)

	Durée estimée	**CHF (à l'excl. de la TVA)**
Bases	**6 mois**	
Gouvernance		426'900
Architecture		280'300
Exploitation		307'900
Gestion		81'400
Mesures d'accompagnement		187'600

Sous-total		*1'284'100*
Incrément #1: établissement de CFC en tant que justificatifs numériques	**6 mois**	
Gouvernance		67'400
Architecture		316'200
Exploitation		248'700
Gestion		53'700
Mesures d'accompagnement		126'700
Sous-total		*812'700*
Incrément #2: inscription aux cours IE	**6 mois**	
Gouvernance		55'400
Architecture		324'100
Exploitation		286'200
Gestion		39'500
Mesures d'accompagnement		136'200
Sous-total		*841'400*
Incrément #3: attestations d'écoles professionnelles en tant que justificatifs numériques	**6 mois**	
Gouvernance		55'400
Architecture		254'700
Exploitation		313'000
Gestion		39'500
Mesures d'accompagnement		120'400
Sous-total		*783'000*
Incrément #4: octroi de l'autorisation de formation	**4 mois**	
Gouvernance		47'500
Architecture		139'200
Exploitation		273'000
Gestion		31'600
Mesures d'accompagnement		81'000
Sous-total		*572'300*

Tableau 3: tableau de flux de trésorerie

	Coûts de développement (uniques) CHF	Coûts d'exploitation (au pro rata) CHF
1ère année		
Bases	882'400	401'700
Incrément #1: établissement de CFC en tant que justificatifs numériques	500'650	312'050
2ème année		
Incrément #2: inscription aux cours IE	487'100	354'300
Incrément #3: attestations d'écoles professionnelles en tant que justificatifs numériques	409'800	373'200
3ème année		
Incrément #4: octroi de l'autorisation de formation	258'800	313'500

6.3 Financement par les unités finales d'imputation

Le présent concept propose un modèle de financement à trois niveaux (voir graphique 4 du chapitre 6.1):

1. Bases: les coûts de réalisation sont supportés à raison de 50% par la CDIP et de 50% par le SEFRI. Il est présumé que les coûts supportés par la CDIP sont répartis entre les cantons conformément à la clé de répartition de la CDIP.
2. Incréments 1 à 4: pendant cette première phase d'exploitation, les coûts individuels seront facturés aux parties. Les coûts de soutien doivent être facturés lors de la conception, du développement et de l'intégration de leurs solutions respectives. Les autres frais généraux (qui correspondent à la plus grande partie des coûts) seront supportés à parts égales par la CDIP et par le SEFRI, par analogie à la base de financement 1.
3. À partir de l'incrément 5: lors de la phase d'exploitation qui suivra, les coûts individuels seront facturés comme jusqu'à présent. En outre, la part des frais généraux générés par l'exploitation et la maintenance de l'infrastructure technique et des autres tâches quotidiennes du secrétariat sera répartie entre les membres de la fédération. Le soin sera laissé aux cantons de décider s'il y a lieu de facturer leurs installations (par ex. office de la formation professionnelle, écoles professionnelles) ou si la totalité des coûts agrégés sera prise en charge par les cantons. En tous les cas, les organisations qui exercent leurs activités à l'échelon intercantonal facturent directement leur participation à la fédération des données, étant précisé que ces organisations sont libres de conclure des conventions avec des cantons de manière individuelle. Une participation financière nettement réduite de la part de la CDIP et du SEFRI sera maintenue. Celle-ci se limite à des prestations annexes et nécessaires afin de pouvoir garantir une gouvernance appropriée (par ex. pilotage de la politique de formation, gestion des risques, audit, etc.).

7. Gestion des risques de la fédération des données

L'analyse figurant ci-après tient compte des risques qui pourraient se manifester soit dans le cadre du projet, soit ultérieurement pendant la phase d'exploitation. Par conséquent, l'approche de l'analyse des risques retenue ne consiste pas seulement à utiliser de manière proactive les techniques de gestion des risques de projet, mais également de considérer des approches opérationnelles et nécessaires de gestion des risques (par ex. le recours à des contrôles informatiques afin de maîtriser les risques de sécurité). En fonction de leur degré de pertinence, de potentiels risques ont été identifiés et analysés principalement pour les domaines de la technologie, du droit et de l'innovation. Les potentielles conséquences de ces risques seront classées et priorisées sur la base de standards (à savoir: risque faible, modéré ou élevé, etc.) pour chacun de ces domaines. Cela permet d'assurer l'évaluation des différents risques de la manière la plus cohérente possible. Dans ce contexte, des mesures ont été définies afin de traiter les risques identifiés. La présente analyse de risques est le fruit d'un travail interne réalisé pendant la phase conceptuelle, qui résulte également d'entretiens avec les représentantes et les représentants des groupes d'acteurs de la formation professionnelle initiale. La gestion des risques est un processus continu qui requiert l'identification de nouveaux risques ainsi que la réévaluation des risques existants. La présente analyse de risques, qui résume la situation sous la forme d'un registre des risques (voir tableau 7), ne doit donc pas être considérée comme exhaustive et définitive.

Tous les risques sont saisis dans un registre des risques présenté visuellement et simplement par un profil de risque. La matrice des risques est subdivisée en différentes zones de couleur, chaque couleur reflétant une stratégie différente de gestion des risques. Les risques affichés dans la zone rouge sont considérés comme inacceptable et des mesures doivent être impérativement prises. Les risques qui se trouvent dans la zone jaune sont soumis à une surveillance active et permanente. Très souvent, dans ce cas, certaines mesures doivent être mises en place pour réduire les effets ou la probabilité de survenance de ce type de risques. Partant, il faut uniquement s'assurer que ces mesures déploient véritablement leurs effets. Les risques qui figurent dans la zone verte sont considérés comme généralement acceptables ou sont gérés de manière appropriée de sorte qu'ils ne devront être réévalués que si les circonstances laissent penser que leur niveau de risque a changé.

Parmi les risques les plus importants, il y a les risques qui sont liés aux réticences face aux innovations. Celles-ci s'expliquent par l'incertitude qu'apportent ces innovations pour le futur ou par l'absence de l'expertise nécessaire pour pouvoir les comprendre (par ex. les risques #2, #3 et #6 du registre des risques, voir plus loin). Ces réticences peuvent à leur tour ralentir la recherche d'un consensus sur la meilleure manière de procéder (par ex. les risques #5 et ceux listés dans le registre des risques). Des réticences de ce genre sont inévitables, car il arrive souvent que les conséquences de l'introduction d'innovations ne deviennent visibles qu'après une assez longue période. Cela ne peut se faire que par un dialogue ouvert et avec la volonté d'accepter les changements.

Les failles de sécurité inconnues ou celles qui résultent de la manière dont les composants de la fédération des données (par ex. les controllers et les agents) seront développés représentent une autre source importante de risques (par ex. risques #7, #9, #10, #11 et #19 du registre des risques). Il s'agit d'une catégorie de risques fréquente propre aux projets

informatiques. Pour des risques de ce genre, il existe différentes solutions bien établies, dont l'une des plus courantes est l'utilisation d'un système de gestion de la sécurité de l'information (ISMS). D'autre part, le risque lié à la protection des données peut être en grande partie limité étant donné qu'aucune donnée personnelle n'est stockée dans la blockchain, même si d'autres formes secondaires de risque peuvent toujours provenir de l'extérieur de la fédération des données (par ex. risque #15 dans le registre des risques).

D'autres risques, plutôt isolés, figurent dans le registre des risques. Ils s'appuient à leur tour sur une analyse et une enquête approfondies réalisées sur la base des informations disponibles. Il convient toutefois de relever qu'une évaluation plus détaillée des risques ne sera possible que lorsqu'une préférence politique sur certains choix architecturaux aura été exprimée.

7.1 Liste des risques

La matrice des risques ci-dessous permet de visualiser la répartition des risques résiduels selon les responsables des risques. En d'autres termes, ce graphique illustre l'impact estimé des dommages et la probabilité de réalisation des risques identifiés, en supposant que toutes les mesures proposées dans le tableau 7 aient été mises en œuvre. Pour plus de clarté, les responsables des risques sont divisés en deux groupes, à savoir:

1. Une partie des risques concerne des risques portant sur la mise en place et l'exploitation de l'infrastructure technique de la fédération des données. Les exploitants de l'infrastructure ainsi que le secrétariat assument la responsabilité de la surveillance et de la minimisation de ces risques.
2. Outre les risques relatifs aux infrastructures techniques des membres et des participants à la fédération des données, d'autres risques non techniques, à considérer surtout comme étant des défis à relever, menacent la mise en œuvre de la fédération des données. Le suivi et la lutte contre ces risques incombent aux acteurs concernés ainsi qu'au pilotage de la politique de l'éducation.

La numérotation des risques illustrés sur le graphique ci-dessous correspond à la numérotation des risques du tableau 7. Ce graphique montre que des risques résiduels subsistent, même si des mesures importantes ont été prises et mises en œuvre pour minimiser les risques. Pour ces raisons, une surveillance continue des risques connus et un screening des risques non encore identifiés restent nécessaires.

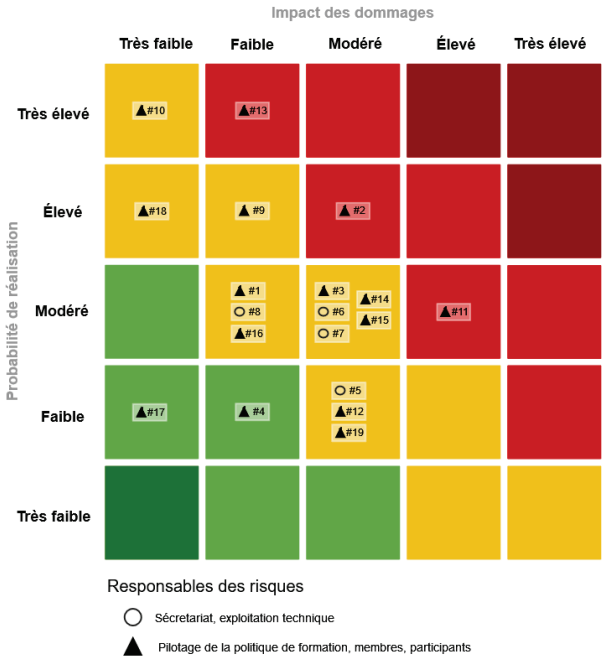

Figure 7: matrice des risques selon les responsables des risques

Le tableau 6 présente les risques et leur intitulé, le responsable des risques, l'impact, le degré de probabilité, les mesures, l'impact résiduel et la probabilité résiduelle. Parmi les responsables des risques, on trouve les acteurs auprès desquels le risque peut être localisé (voir chapitre 3.1 de la 2e partie du concept). La fédération des données n'a qu'une influence limitée, voire aucune influence sur certains de ces risques. Il est important que les acteurs concernés formulent, complètent et mettent systématiquement en œuvre les mesures prévues afin de minimiser le risque.

Tableau 4: analyse des risques

No	Description	Responsables des risques	Impact	Probabilité	Mesures	Impact résiduel	Probabilité résiduelle
Innovation							
1	Les membres potentiels refusent de se raccorder à la fédération des données de sorte que la fédération des données ne peut pas déployer ses effets.	membres exploitation technique	modéré	très élevé	Mesures de confiance par le biais de la communication, de projets pilotes et d'autres mesures.	faible	modéré
2	Les cantons hésitent à mettre en œuvre la fédération des données puisque ses effets sur le système actuel sont peu clairs.	divers	élevé	élevé	Établir une vue d'ensemble de la planification concernant les applications techniques existantes, modes d'emploi relatifs à l'intégration et autres mesures.	modéré	élevé
3	On en arrive à une résistance de la part des cantons à cause des changements en matière de rôles, de tâches et de responsabilités.	pilotage de la politique de formation	très élevé	élevé	Information, guides, manuels	modéré	modéré
4	Le manque ou l'absence de connaissances spécialisées et le manque ou l'absence de ressources chez tous les membres ont pour effet de retarder la mise en œuvre.	membres	modéré	modéré	Des mesures de soutien différentes pour l'intégration, par exemple des modes d'emploi et des formations ainsi que d'autres mesures.	faible	faible
5	Parce que l'état actuel de la technologie n'en est encore qu'à un stade précoce, il n'est pas encore clair de savoir quels standards techniques vont s'imposer.	secrétariat	élevé	modéré	Lors de la recherche d'une solution technique, une séance d'information sur les systèmes déjà existants a eu lieu. Échanges réguliers d'expériences et d'informations avec des instances publiques et privées qui poursuivent des projets similaires.	modéré	faible

Technologie							
6	Les entreprises informatiques ne disposent que de connaissances techniques limitées sur la mise en œuvre d'une fédération des données.	Exploitation technique	modéré	Modéré	Surveillance de l'écosystème du SSI ainsi qu'échanges d'informations réguliers avec la Confédération afin de garantir qu'à l'avenir, la formation professionnelle et la Confédération n'exploiteront pas chacune leur propre réseau de blockchain.	modéré	modéré
7	L'infrastructure de la blockchain s'avère peu sûre.	Exploitation technique	modéré	élevé	Introduction d'un ISMS et exécution de mises à jour (updates et upgrades) régulières afin de garantir la sécurité. Contrôle périodique des points faibles, y compris des mesures correspondantes. Surveillance et contrôle au moyen d'une vue d'ensemble des cybermenaces et des tendances actuelles.	modéré	modéré
8	L'aptitude à la scalabilité et la surcharge de la fédération des données sont insuffisamment testées.	Exploitation technique	modéré	modéré	On fait appel à une infrastructure de scalabilité avec l'administration correspondante. L'infrastructure mise en place permet d'éviter des goulots d'étranglement en matière de performance. En outre, on garantit l'utilisation optimale des ressources en recourant à un monitorage d'exploitation.	faible	modéré
9	Violation de la sécurité des portefeuilles électroniques basés sur le cloud.	participant.e.s membres	modéré	Élevé	Mesures techniques et organisationnelles. Une clause contractuelle pour les membres qui autorise des contrôles de sécurité des membres. Mettre en évidence des solutions alternatives aux portefeuilles électroniques basés sur le cloud (portefeuille électronique local en tant qu'alternative pour les participant.e.s).	faible	élevé

10	Violation de la sécurité des portefeuilles électroniques privés locaux.	participant.e.s	très faible	très élevé	Sensibilisation globale au moyen de mesures de communication, et rendre possibles des backups automatiques.	très faible	très élevé
11	Violation de la sécurité des portefeuilles électroniques internes à l'organisation.	membres	très élevé	élevé	Recours à des mesures de protection de l'infrastructure informatique ainsi qu'à des backups automatiques.	élevé	modéré
12	Absence ou manque d'unité s'agissant des standards et des processus au sein de la fédération des données.	membres	élevé	modéré	Tomber d'accord ensemble sur des standards revêtant force obligatoire en tenant compte de l'avis des OrTra et des entreprises formatrices.	modéré	faible
Droit							
13	L'approche de l'autodétermination informationnelle ne parvient pas à se réaliser.	participant.e.s	faible	faible	Recours à des mesures techniques et organisationnelles afin d'imposer l'application du principe de l'autodétermination informationnelle ainsi que sensibilisation de tous les acteurs impliqués à l'application du principe de l'autodétermination informationnelle.	faible	très élevé
14	Les participant-e-s ne disposent pas des compétences numériques requises afin de faire valoir et d'imposer leurs droits.	participant.e.s	élevé	élevé	Recours à des mesures techniques et organisationnelles afin de garantir la sécurité et l'intégrité des données ainsi que la protection des droits et des libertés des personnes (mineures). Sensibilisation au moyen de mesures de communication sur l'application du principe de l'autodétermination informationnelle portant sur ses propres données.	modéré	modéré
15	En raison de l'existence des infrastructures de gestion des identités, la possibilité d'une suridentification existe.	Confédération secrétariat	élevé	élevé	Mise à disposition d'un service d'identification par proxy qui accepte une e-ID, mais qui traduit un identificateur transitoire qui n'identifie pas la personne.	modéré	modéré

					Interdiction du couplage en cas de participation à la fédération des données.		
16	L'interaction avec la fédération des données est entravée par des obstacles.	membres	faible	faible	Prévoir des centres d'assistance pour les personnes vivant avec des handicaps par le biais des organismes émetteurs.	faible	modéré
17	Dans le registre décentralisé des données, il y a des données qui n'appartiennent pas au registre décentralisé des données.	membres	modéré	modéré	Recours à des mesures techniques afin d'empêcher qu'on n'écrive de fausses informations dans la blockchain ainsi que recours à ce qu'on appelle des «tombstones» (litt.: pierres tombales) qui rendent des contenus illisibles. Recours à des mesures de sécurité afin d'empêcher que des données personnelles n'arrivent dans la blockchain ou chez la mauvaise personne. Contrôle au niveau du node. Principe du «contrôle à 4 yeux»	très faible	faible
18	En raison du manque de volonté politique, les bases légales nécessaires à l'échange de données ne peuvent pas être créées.	pilotage de la politique de formation	modéré	modéré	Convention intercantonale sur l'harmonisation des données dans la formation professionnelle.	très faible	élevé
19	Un membre abuse de manière illicite de données	secrétariat membres	modéré	modéré	Recours à des mesures techniques et organisationnelles aux fins de contrôle. En outre, on fixe des mesures contractuelles, et l'on prévoit des audits de sécurité et sensibilisation au moyen de mesures de communication.	modéré	faible

8. Conclusion intermédiaire

Dans la première partie du concept de la fédération des données pour la formation professionnelle, nous avons mis l'accent sur la future infrastructure technique. Nous avons souligné les effets sur l'efficience, la disponibilité des données et la connectabilité. La deuxième partie du concept se concentre sur la protection des données et sur les règles de gouvernance adéquates. Le modèle de gouvernance des données développé dans la deuxième partie a pour objectif de permettre à tous les acteurs impliqués dans la fédération des données d'utiliser les données de manière cohérente, efficace et conforme à la loi.

Cette troisième partie présente le calendrier, les coûts ainsi que la mise en œuvre des différentes étapes du concept de la fédération des données. Le potentiel de la fédération des données est en outre présenté de manière détaillée sur la base d'un deuxième cas d'application: l'échange direct de données et automatisé entre l'office de formation professionnelle, les prestataires des cours interentreprises et les entreprises formatrices. La troisième partie se termine par une analyse détaillée des risques afférents à la phase de réalisation.

Comme le présente cette troisième partie, une durée de deux à trois ans est prévue pour la réalisation de la fédération des données pour la formation professionnelle. Dès la première année de réalisation, il sera possible de passer à un mode d'exploitation. La mise en place de la fédération des données se fera ensuite de manière incrémentielle. Les valeurs ajoutées pour les acteurs de la formation professionnelle seront mises à disposition par étapes. Cette approche a des effets positifs sur la panification temporelle des besoins de financement. Les coûts de développement seront plus importants au début de la phase de réalisation et diminueront avec le temps. Parallèlement, la part dédiée aux coûts d'exploitation évoluera en fonction de l'augmentation du nombre des utilisatrices et des utilisateurs de la fédération des données. Le plan de réalisation sera complété par une analyse sur les besoins de régulation pour la fédération des données ainsi que par une analyse des risques.

Annexe

Définitions

Concept	Définition
Agent	Un «agent», au sens technique du terme, est un composant d'un logiciel qui traite les instructions d'un ou de plusieurs «controllers», qui les rediffuse en conséquence au sein de la fédération des données, et qui communique avec le node ou qui mémorise les justificatifs numériques (et, de ce fait, les données personnelles).
Organisme émetteur	Par le terme d'organisme émetteur (ou «issuer»), on entend un organisme qui, par ex., est autorisé par une institution étatique à établir des justificatifs numériques.
Source faisant autorité	Source de données qui possède un degré élevé de crédibilité et de confiance.
Exploitante / exploitant	Il peut s'agir d'institutions, de personnes privées ou de personnes morales qui exploitent les nodes.
Blockchain	Un système de stockage de ledger dans lequel les données de transaction sont attachées et ne peuvent pas être modifiées ultérieurement. Les blockchains sont réalisées via un réseau de nodes, raison pour laquelle elles sont, en règle générale, décentralisées.
Controller	Par le terme de «controller», on entend un composant d'un logiciel qui permet à des personnes et à des institutions d'accéder à l'«agent», via une interface-utilisateur, ce qui donne ainsi accès à des justificatifs numériques et permet d'utiliser le registre décentralisé des données, resp. de communiquer avec d'autres membres, participantes et participants de la fédération des données.
Provenance des données	Elle permet de garantir la traçabilité de certaines données en les rapportant à leurs sources initiales. On peut ainsi constater, par exemple, si certaines données proviennent effectivement ou non de sources faisant autorité.
Registre décentralisé des données	Un registre décentralisé des données est une mémoire publiquement accessible et lisible par toutes les actrices et tous les acteurs. Il contient toutes les informations nécessaires pour pouvoir exploiter une fédération des données en général, et un écosystème destiné à l'établissement de justificatifs en particulier.
Identificateurs décentralisés des données (DID)	Il s'agit d'une forme d'identification qui est utilisée dans des systèmes décentralisés. Elle rend possible la mise en lien d'événements avec des données de telle sorte que seules les parties concernées puissent établir ces mises en liens.
Portefeuille électronique	Par le terme de portefeuille électronique, on entend un logiciel d'application qui sert à conserver de manière sécurisée des justificatifs électroniques.
Service d'informations (appelé Discovery-Service)	Il s'agit d'un service technique qui, au sein du registre décentralisé des données, rend possible l'identification (automatique) de prestations de services qui sont proposées par des membres.

Détentrice / détenteur	Par le terme de détentrice ou de détenteur, on entend une personne ou une institution qui possède électroniquement un justificatif numérique.
Smart Contracts	Les smart contracts (contrats intelligents) sont des programmes informatiques qui traitent des données de manière automatisée et qui, sur cette base, ont pour effet de déclencher une prise de décision automatisée selon des critères prédéfinis.
Node	Un node (nœud) est une instance de traitement du registre décentralisé des données. Les nodes réalisent, contrôlent et mémorisent des transactions sur le registre décentralisé des données et peuvent en outre également héberger des contrats intelligents.
Membre	Il s'agit d'une partie institutionnelle de la fédération des données qui est autorisée à collaborer de manière active (par ex. création de transactions). De telles parties sont tenues de signer un contrat avec l'exploitante/avec l'exploitant de la fédération des données.
Conditions d'utilisation	Les conditions d'utilisation sont des métadonnées qui sont écrites sur des justificatifs numériques aux fins de réglementation des autorisations, des interdictions et des obligations imparties aux acteurs concernés.
Permissioned (Blockchain)	Une infrastructure de blockchain dont les droits de lecture et d'écriture sont restreints.
Gouvernance polycentrique	Une approche de gouvernance où plusieurs parties négocient des règles et des stratégies au sein d'un écosystème afin de résoudre des problèmes communs.
Proof-of-Stake	Une procédure avec laquelle un réseau de blockchain peut atteindre un consensus ; il s'agit souvent d'une procédure au sein de laquelle des calculs à haute intensité énergétique doivent être résolus.
Privacy by Default	La conception d'un système dans lequel des réglages prédéfinis sont mémorisés de telle sorte que le traitement des données personnelles est limité au strict minimum nécessaire pour le but d'utilisation déterminé de ces données dans la mesure où la personne concernée n'en décide pas autrement.
Privacy by Design	Une approche avec laquelle on garantit de manière préventive la protection des données par le biais de la conception technique et organisationnelle d'un système.
Public-Key-Infrastructure	Une infrastructure destinée à la délivrance, à la distribution et à la validation d'un matériel d'information d'importance décisive dont l'appartenance à certaines organisations déterminées a été certifiée.
Schéma	Un canevas destiné à un justificatif numérique qui spécifie son contenu (à savoir des métadonnées, y compris des conditions d'utilisation, des déclarations et une méthode de preuve).
Self Sovereign Identity (SSI)	Une approche destinée à la gestion des identités où l'on utilise des justificatifs numériques afin de transmettre des informations d'attributs au sujet d'une personne (par ex. nom, prénom, date de naissance).

Directives sur l'architecture logicielle	Lignes directrices fixées par l'exploitation technique et approuvées par l'organisme émetteur. Elles déterminent comment le logiciel (par ex. des smart contracts) doit être développé pour la fédération des données (par ex. des canevas destinés à des procédures standards, ou la mise en œuvre de mesures de sécurité, etc.).
Participants	Il s'agit d'une partie de la fédération des données qui est autorisée à collaborer de manière passive (par ex., consultation d'informations publiques dans le registre décentralisé des données ou vérification des justificatifs). En règle générale, de telles parties ne doivent pas avoir signé de contrat avec l'exploitante/avec l'exploitant de la fédération des données.
Tombstone	Une transaction erronée qui se trouve dans le registre décentralisé peut être déclarée «tombstone» (litt. : pierre tombale). Dès lors, les contenus de la transaction seront recouverts de noir. Ainsi rendus illisibles, ils ne seront plus consultables.
Transaction	Une transaction est une opération qui permet de reprendre des données dans le registre décentralisé des données. Les transactions ne contiennent aucune donnée personnelle
Vérificatrice / vérificateur	Par le terme de vérificatrice ou vérificateur, on entend des personnes ou des institutions qui vérifient la validité des justificatifs numériques qui leur sont présentés au moyen d'informations disponibles sur le registre décentralisé des données.
Justificatif numérique (vérifiable)	Dans le présent contexte, un justificatif numérique contient un ou plusieurs énoncés d'une institution ou d'une autorité (par. ex. d'une école ou d'une instance étatique) à propos d'un sujet donné (par ex. d'une apprenante ou d'un apprenant). Il est possible de vérifier si ces déclarations proviennent d'une source fiable à l'aide des données figurant dans le registre décentralisé des données – d'où l'expression «justificatif numérique vérifiable».
Zero Knowledge Proof	Un ZKP ou «Zero Knowledge Proof» est un protocole cryptographique qui permet de formuler un énoncé au sujet de certaines données, mais sans divulguer les données proprement dites (par ex. le fait qu'on est âgé de plus de 18 ans, mais sans devoir divulguer la date de naissance).

Principes d'architecture

Référence Principe architectural (PA)

Gouvernance dans le sens de structures de contrôle et de gestion: ce principe veille à ce que tous les transferts de données au sein de l'architecture dans son ensemble soient licites. Ce principe prévoit des mécanismes de contrôle et de gestion afin de garantir la légalité des transferts de données. La gouvernance a un caractère hiérarchique.

PA.G.1	**Conservation fédéraliste des données:** La fédération des données n'a pas besoin d'un stockage centralisé des données et est organisée de manière fédéraliste. Il s'agit d'une association de systèmes individuels qui conservent leur autonomie.
PA.G.2	**Modèle de gouvernance consensuel:** cela sert à l'acceptation et à la légitimation des personnes autorisées à prendre des décisions de manière autonome. Ce principe doit favoriser l'agilité et le fonctionnement du système, en aucun cas le freiner.
PA.G.3	**Disponibilité de données lors de prises de décisio :** cela part du principe que les décisions sont prises correctement et légalement lorsque toutes les informations pertinentes sont disponibles.
PA.G.4	**Transparence, intégrité et responsabilité concernant les événements pertinents liés à la formation**: ce principe garantit la révocabilité des données à tout moment. La révocabilité des données résulte d'un besoin potentiel légitime de la personne concernée qui doit obéir à des règles claires.
PA.G.5	**Accès légitimé à la fédération des données:** seuls ceux qui ont un intérêt légitime pour la fédération peuvent y accéder. Ce qui constitue un intérêt légitime doit être clairement défini. Les rôles des différentes participantes et des différents participants doivent être définis, leurs droits dans le système doivent être décrits en conséquence.

Sécurité dans sens d'état général, qui est jugé aussi exempt que possible de risques (inacceptables). La sécurité poursuivie ici est la sécurité objective (par opposition à la sécurité subjective), qui identifie les risques, évalue correctement leur potentiel de danger et les contrecarre efficacement.

PA.S.1	**Contrôle d'accès basé sur les rôles (RBAC):** cela permet de garantir que seules les personnes autorisées (légitimées - voir AP.G.5) ont accès ou peuvent effectuer des modifications dans le cadre des rôles qui leur sont attribués.
PA.S.2	**Garantie de l'identité des composantes.** Sécurité de l'identité des composants: l'accès à la fédération de données doit être limité aux participants dignes de confiance, qui doivent être en mesure de prouver leur identité. Ce n'est qu'ainsi que les principes nécessaires à la protection des données peuvent être respectés. Ce principe d'architecture est une condition pour accéder à la Fédération.
PA.S.3	**Garantie de l'identité des parties concernées:** au-delà de la garantie générale de l'identité de l'AP.S.2, ce principe architectural assure la légitimation confirmée de la partie concernée. Ce seuil supplémentaire est important car la partie concernée est en quelque sorte propriétaire des données traitées. Elle est la seule à pouvoir requérir des modifications (traitements).
PA.S.4	**Monitoring continu des échanges de données:** des instruments/mécanismes sont intégrés dans le système afin de contrôler et de garantir en permanence la mise en œuvre du PA.S.1-3. Ce PA tente de détecter et d'empêcher les abus en amont.

PA.S.5	**Échange de données plus sécurisé et davantage conforme à la législation:** conformément à ce principe architectural, un «state of the art» technologique est fixé comme référence. Cela permet de garantir des normes élevées.

Données dans le sens de signes contenant une information. Les informations importantes dans ce contexte sont celles qui concernent les personnes dans le domaine de l'éducation.

PA.D.1	**Les données sont un bien commun au service de la formation professionnelle:** les apprenantes et les apprenants sont les seuls propriétaires des données dans l'éducation. Ce principe exprime la nécessité que les données dans l'éducation ne soient pas accumulées par des services administratifs isolés. Toutes les institutions de formation (entreprises formatrices, écoles professionnelles, etc.) sont au service d'un traitement efficace et ciblé des données. En tant qu'émetteurs de documents d'identité, ces composantes ne doivent pas faire obstacle au système. .
PA.D.2	**Les composantes ont pour tâche de garantir la qualité des données:** la responsabilité individuelle et la responsabilité en matière de protection des données se distinguent. En identifiant de manière sûre les composants (PA S.2-3) et en leur attribuant un rôle clair (PA S.1), il est possible d'identifier rapidement les responsabilités en cas d'erreur. Ainsi, chaque composant doit agir sous sa propre responsabilité au sein de la fédération.
PA.D.3	**Interprétabilité sémantique des données:** ce principe garantit l'interprétation correcte des informations (données); il suppose un langage commun et la possibilité de traduire à tout moment des termes différents dans le langage commun. L'interprétabilité sémantique des données assure la continuité, c'est-à-dire que les mêmes significations sont toujours données aux données saisies de la même manière.

Réseau dans sens d'un système avec une structure sous-jacente qui peut être modélisée et qui contient des mécanismes d'auto-organisation.

PA.N.1	**Échange direct de données:** il est important que les différents composants du réseau puissent communiquer directement entre eux (sans intermédiaire). Ce n'est qu'ainsi qu'ils peuvent réagir en temps utile et de manière efficace. L'échange direct de données soutient également l'intégrité des données, dans la mesure où les données sont directement traitées par l'initiative de l'utilisatrice ou de l'utilisateur (sans intermédiaires). Ce principe suppose que les PA S.1-5 soient respectés à tout moment.
PA.N.2	**Atomicité d'événements de formation:** il est important d'imaginer et de mettre en œuvre des règles qui définissent quand, dans quelle mesure et dans quel but contraignant les événements éducatifs peuvent être communiqués ouvertement. Cela permet de garantir la tenue à jour des données à tout moment.
PA.N.3	**Asynchronité de la communication:** le moment où les données sont transmises n'a pas d'importance pour le réseau; les données doivent pouvoir être consultées à tout moment par tout le monde. Dès qu'elles sont saisies, elles s'appliquent.

Sources principes du chapitre 4

1. Constitution fédérale (Cst): principalement les articles 47, 48, 48a, 61a, 62, 63;
2. Loi fédérale sur la formation professionnelle (LFPr): principalement l'article 66;
3. Loi fédérale sur la coopération entre la Confédération et les cantons dans l'espace suisse de formation (LCESF): principalement article 1;
4. Convention entre la Confédération et les cantons sur la coopération dans l'espace suisse de formation (CCoop-ESF): principalement article 6a;
5. Chancellerie fédérale, Prise de position du 22 décembre 2011 concernant les bases légales pour la collaboration TIC entre la Confédération et les cantons
6. Convention-cadre de droit public concernant la collaboration en matière de cyberadministration en Suisse 2020
7. Convention-cadre de droit public concernant l'Administration numérique suisse
8. Département fédéral des finances, Rapport explicatif du 11 décembre 2020 concernant la Loi fédérale sur l'utilisation des moyens électroniques pour l'exécution des tâches des autorités (LMETA)
9. Malinverni Giorgio/ Hottelier Michel/ Hertig Randall Maya/ Flückiger Alexandre, les conventions intercantonales in: Droit constitutionnel suisse - Volume I: l'Etat, 2021, pp. 623 - 640.
10. Egli Patricia, art. 47, in: Die schweizerische Bundesverfassung St. Galler Kommentar, 3ème édition, 2014, pp. 991-998.
11. Montavon Michael, Cyberadministration et protection des données - Etude théorique et pratique de la transition numérique en Suisse du point de vue de l'Etat, des citoyen-ne-s et des autorités de contrôle, 2021.
12. Epiney Astrid, Zur Abgrenzung des Anwendungsbereichs des Datenschutzgesetzes des Bundes und der kantonalen Datenschutzgesetzte - Aufgezeigt am Beispiel der für Mitarbeiterdaten privater Spitäler massgeblichen gesetzlichen Vorgaben, in: Jusletter 2 mars 2015, www.weblaw.ch