



Programme de gestion nationale des données

Projet Assurance qualité des données des entreprises

Rapport de juin 2022

Neuchâtel, 2022

Éditeur:	Office fédéral de la statistique (OFS)	Concept de mise en page:	Section DIAM
Renseignements:	Service central d'informations OFS, Tél. +41 58 463 60 11	Images:	Division Interopérabilité et registres (IOR)
Rédaction:	Division Interopérabilité et registres (IOR)	Graphiques:	Division Interopérabilité et registres (IOR)
Domaine:	00 Bases statistiques des généralités	Cartes:	Division Interopérabilité et registres (IOR)
Langue du texte original:	Allemand	Téléchargement:	www.statistique.ch
Traduction:	Services linguistiques de l'OFS	Copyright:	OFS, Neuchâtel 2022 La reproduction est autorisée, sauf à des fins commerciales, si la source est mentionnée

Table des matières

Aperçu.....	3
Nouveaux indicateurs de qualité pour les données administratives publiées sur la plateforme d'interopérabilité I14Y	4
Nouveaux indicateurs de qualité pour le REE.....	6
Qualité de l'authentification des entreprises.....	8
Annexe technique A: Questionnaire qualité pour la plateforme I14Y – projet	9
Annexe technique B: Tableau de bord des indicateurs de qualité du REE	10
Références	14

Aperçu

Conformément au mandat du Conseil fédéral (1), les travaux du projet relatif à l'assurance qualité des données des entreprises vont bon train à l'Office fédéral de la statistique. Après une première phase, consacrée à l'analyse et décrite dans le rapport (2), le projet se trouve à présent en **phase de réalisation**.

Depuis la remise du dernier rapport, les efforts se sont concentrés sur trois domaines: l'élaboration et l'application de nouveaux indicateurs de qualité sur la plateforme d'interopérabilité I14Y, le registre des entreprises et des établissements (REE) et les questions liées à l'authentification électronique des entreprises. Dans ce dernier cas, le travail consiste surtout à déterminer quels éléments de la **gestion des identités et des accès** de la Confédération (IAM, *identity and access management*), initialement conçus pour les personnes physiques, peuvent être repris pour les entreprises.¹

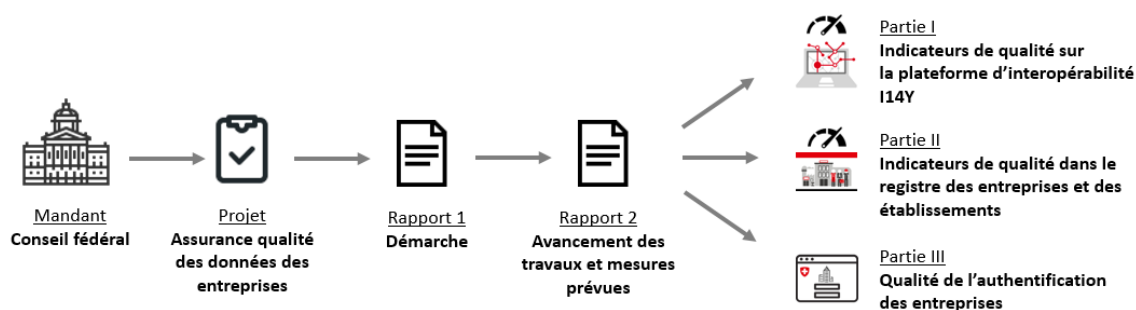


Figure 1: Aperçu

¹ Le présent document constitue une suite au rapport de novembre 2020 et doit être considéré et lu comme tel.

Nouveaux indicateurs de qualité pour les données administratives publiées sur la plateforme d'interopérabilité I14Y

La plateforme d'interopérabilité I14Y est l'élément central du programme de gestion nationale des données (NaDB). C'est aussi un système ouvert, accessible à tous (3) et optimisé en continu, afin de fournir une **cartographie** aussi exhaustive que possible du **paysage des données** de l'administration suisse. Cette cartographie repose essentiellement sur des métadonnées, de sorte que la plateforme constitue aussi un «catalogue des métadonnées suisses».

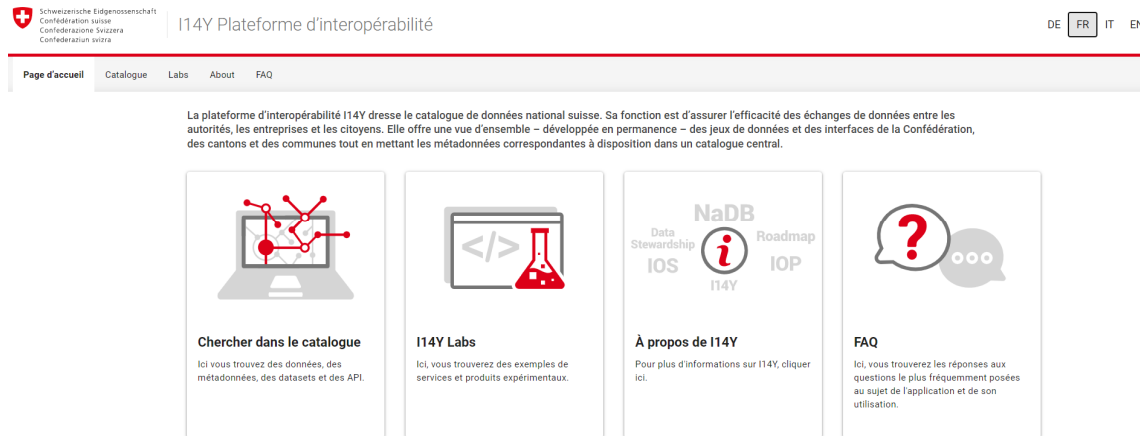


Figure 2: L'interface utilisateur de la plateforme d'interopérabilité I14Y: élément clé de la gestion nationale des données.

Les **métadonnées** décrivent les données elles-mêmes. Ces dernières ne se trouvent donc pas sur la plateforme I14Y : les données restent stockées chez leur propriétaire, leur transmission entre les différents organismes intervenant de manière bilatérale à l'aide des interfaces électroniques sur la plateforme (interfaces de programmation permettant d'appeler des fonctions et d'obtenir des données²). Tout service administratif qui requiert des données ne peut donc évaluer leur qualité qu'après les avoir obtenues. Dès le lancement du programme NaDB, ses responsables se sont dès lors demandé s'il ne faudrait pas étendre le modèle des métadonnées de la plateforme I14Y de telle sorte que ces dernières incluent également des informations sur la qualité du contenu des données proprement dites.

Le projet relatif à l'assurance qualité des données des entreprises a fourni l'occasion d'inclure sur la plateforme I14Y les premières métadonnées qui concernent le contenu des données. Un questionnaire sur la qualité a été élaboré à cet effet, que les propriétaires des données remplissent lorsqu'ils décrivent un nouveau jeu de données³. Les informations ainsi réunies sur la qualité des données sont ensuite publiées sous forme de métadonnées, au même titre que les caractéristiques déjà existantes⁴. À l'avenir, les nouveaux **indicateurs de qualité** se référeront aussi bien aux jeux de données qu'à leurs caractéristiques.

² Application Programming Interface API.

³ Le questionnaire comprend par exemple des questions sur les identificateurs utilisés ou sur l'historisation (voir annexe technique A).

⁴ Ces indications respectent la norme DCAT (voir <https://www.w3.org/TR/vocab-dcat-3/>).

The screenshot shows the I14Y Platform for Interoperability website. The header includes the Swiss Confederation logo and the text 'Schweizerische Eidgenossenschaft', 'Confédération suisse', 'Confederazione Svizzera', and 'Confederaziun svizra'. The main navigation bar has links for 'Page d'accueil', 'Catalogue', 'Labs', 'About', and 'FAQ'. The 'Catalogue' tab is active, showing a list of datasets on the left: 'CHISCO-19', 'NOGA2008', and 'Données de référence des entreprises' (highlighted). The main content area is titled 'Données de référence des entreprises' and has three tabs: 'Description' (active), 'Structure', and 'Distributions'. The 'Description' tab contains a paragraph explaining that reference data is defined as base information needed for administrative tasks, sourced from the Register of Enterprises and Establishments (REE) and the Register of Identification Numbers (RIDE). It also mentions that the REE manages data for legal entities, while the RIDE manages data for local entities. Below the text is a table of metadata:

Identifiant	ENTERPRISE_MASTER_DATA
Date de publication	14.12.2021
Date de modification	-
Editeur	Office fédéral de la statistique
Points de contact	Office fédéral de la statistique

Figure 3: Métadonnées d'un jeu de données. À l'avenir, les métadonnées incluront des indicateurs de qualité.

Pour ne pas imposer aux services administratifs une surcharge de travail démesurée et vérifier également la faisabilité du système, la saisie des informations sur la qualité se fondera dans un premier temps sur une série limitée d'indicateurs. Ces indicateurs seront tout d'abord soumis à un **test en conditions réelles**, puis le cas échéant adaptés (voire, au besoin, élargis). Ils subiront ensuite une évaluation finale et feront l'objet d'une large consultation avant d'être intégrés dans le modèle d'information de la plateforme I14Y. Cette manière de procéder garantira, malgré la nécessité d'accroître la qualité des données des entreprises, que le travail supplémentaire imposé aux propriétaires de données se limitera à l'indispensable lorsque la saisie des informations sur la qualité des données sera élargie à tous les indicateurs.

Dans le cadre de ce projet, on a également examiné si l'**approche mise en œuvre en Autriche** pouvait être utilisée (4) et dans quelle mesure la Suisse pourrait s'en inspirer. Aucune des méthodes étudiées ne peut cependant être reprise telle quelle.

Au terme du test en conditions réelles mentionné plus haut, l'analyse commencée sera poursuivie et assortie de réflexions d'ordre organisationnel. Afin de mettre en place rapidement une solution susceptible de faire l'unanimité, les étapes suivantes seront menées en étroite concertation avec les principaux services concernés. Grâce aux nouveaux indicateurs, ces services pourront à l'avenir évaluer d'un coup d'œil la qualité d'un jeu de données. Une fois en place, ce système favorisera l'**utilisation multiple des données** et contribuera ainsi grandement à l'application du principe *once-only* en Suisse.

Nouveaux indicateurs de qualité pour le REE

En ce qui concerne la gestion nationale des données, l'introduction d'indicateurs de qualité sur la plateforme I14Y relève de l'objectif général qui consiste à garantir une qualité aussi élevée que possible pour les données des entreprises. Du point de vue du REE, ces nouveaux indicateurs visent concrètement à garantir une **transparence maximale** quant au contenu des données et aux éventuels problèmes inhérents aux informations enregistrées : les utilisateurs et les utilisatrices doivent pouvoir se faire aisément et rapidement une idée des points forts et des faiblesses des données figurant dans le registre (5).

La nécessité théorique d'introduire de tels indicateurs de qualité dans le REE a été identifiée dès le lancement du projet et décrite dans le 1^{er} rapport de novembre 2020. Se fondant sur les normes de qualité d'Eurostat (6) et après avoir examiné nombre de solutions décrites dans la littérature spécialisée (7, 8), l'équipe du projet a retenu **cinq dimensions de la qualité – complétude, précision, cohérence, fiabilité et rapidité d'exécution** – pour lesquelles de nouveaux **indicateurs** seront définis dans un premier temps. Les quatre premières de ces dimensions sont celles qui reviennent le plus souvent dans la littérature spécialisée et fournissent en général une image assez précise de la qualité d'un jeu de données.

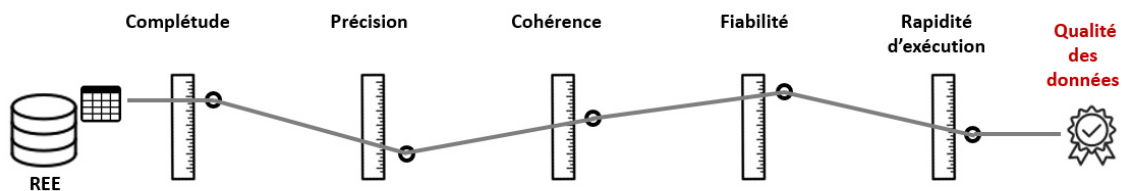


Figure 4: Les cinq dimensions qui serviront à mesurer la qualité des données du REE.

Bien que la dimension **rapidité d'exécution** constitue un aspect assez inhabituel, elle présente un intérêt particulier dans le cas du REE. Comme l'échange de données avec les partenaires administratifs est au cœur du fonctionnement du registre, les annonces de modifications reçues dans le cadre du REE doivent être impérativement traitées dans un certain délai. La rapidité du traitement de ces annonces constitue dès lors un indicateur important de la qualité de ce registre. Le REE vise non seulement à donner une image précise du passé, mais également à être aussi proche que possible de la réalité économique actuelle.

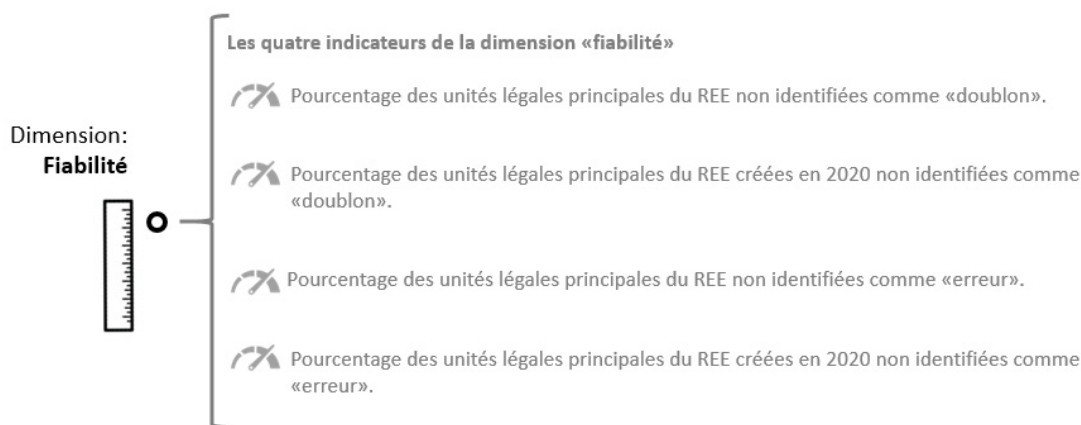


Figure 5: Dans le REE, plusieurs indicateurs servent à mesurer l'une des dimensions de la qualité (voir à ce sujet l'annexe technique B).

Dans un premier temps, **21 indicateurs** ont été retenus, dont la liste complète figure à l'annexe technique B. Ce travail est toutefois encore en cours de finalisation. Outre l'échelle de qualité proprement dite, les limites de certains indicateurs relatifs au contenu des données sont encore à l'examen. La liste existante permet néanmoins de se faire une bonne idée de la solution qui sera adoptée.

Tous les indicateurs figurent actuellement dans un **«tableau de bord»** du REE.⁵ Ces indicateurs avaient déjà été soumis à différents services concernés, de sorte que leur version actuelle tient compte des premières suggestions et remarques reçues.

Les **indicateurs du REE** sont par ailleurs déjà en cours de publication. Dans ce cadre, un lien a par exemple été mis en place entre la plateforme I14Y et le REE, à l'instar du lien qui existait déjà vers les normes eCH (création d'une nouvelle rubrique «indicateurs de qualité du REE»).

Compte tenu du succès obtenu lors de l'élaboration des premiers indicateurs et des réactions reçues jusqu'ici – tant à l'interne qu'à l'externe – les responsables du projet prévoient de déterminer s'il ne vaudrait pas la peine d'appliquer la même solution au **registre des bâtiments et des logements** (RegBL). Ce registre pourrait également inclure des indicateurs de qualité, qui seraient mis à la disposition de ses utilisateurs et utilisatrices dans un tableau de bord.

⁵ « Tableau de bord » du REE : <https://www.bfs.admin.ch/bfs/fr/home/registres/registre-entreprises/registre-entreprises-etablissements/indicateurs-qualite.html>

Qualité de l'authentification des entreprises

Le **processus d'authentification** des entreprises dans le cadre de leurs échanges avec l'administration publique constitue à lui seul un pilier de la qualité des données. Voilà pourquoi le Conseil fédéral a explicitement demandé quels éléments du jeu d'indicateurs utilisés pour évaluer la qualité de l'authentification de personnes physiques dans le cadre de la gestion des identités et des accès (Identity Access Management IAM) pourraient également s'appliquer aux entreprises.

L'association eCH encourage, développe et adopte des normes dans le domaine de la cyberadministration pour une collaboration électronique efficace entre les autorités, les entreprises et les personnes privées.⁶ Après les analyses menées l'an dernier, l'OFS a conclu que la **norme eCH-0170** contient un modèle de qualité exhaustif pour l'authentification de sujets, de sorte qu'aucun travail supplémentaire ne s'impose en la matière. Les indicateurs requis sont déjà entièrement définis et harmonisés avec ceux appliqués aux personnes physiques. Il suffit de les appliquer rigoureusement au sein de l'administration publique et de les inclure dans des solutions logicielles.

Les responsables du projet ont ensuite également déterminé dans quelle mesure l'OFS pouvait définir ses propres mesures. Après avoir étudié les options les plus variées, ils ont conclu qu'il convient de toute évidence de privilégier les solutions communes : du point de vue de l'office, l'option la plus judicieuse réside dans une **solution centralisée pour toute l'administration publique suisse**.

À l'heure actuelle, l'**Administration numérique suisse (ANS)**, le secteur **Transformation numérique et gouvernance de l'informatique de la Chancellerie fédérale** (secteur TNI) et d'autres services administratifs mènent des réflexions qui visent à mettre en place une solution globale. Une première étude doit servir à élaborer une définition de l'IAM qui soit commune aux trois niveaux de l'État fédéral et à identifier les synergies potentielles ainsi que les possibilités d'harmonisation entre ces trois niveaux.

⁶ eCH est une association à but non lucratif dont l'objectif est de promouvoir des normes uniformes dans le domaine de la cyberadministration. Ses membres sont la Confédération, tous les cantons, diverses communes, environ 120 entreprises ainsi que différentes hautes écoles, associations et personnes privées. Ses recommandations sont d'autant plus importantes que les normes eCH sont largement utilisées en Suisse. <http://www.ech.ch>

Annexe technique A: Questionnaire qualité pour la plateforme I14Y – projet

Questionnaire qualité des données des entreprises (QSUD)

Informations sur la source (du jeu de données)

Quelles clés primaires utilisez-vous comme identificateurs pour les données? Quel type utilisez-vous ?

Un suivi des modifications est-il assuré? Autrement dit, en cas de modification d'une valeur, l'ancienne valeur reste-t-elle disponible?

Oui

Non

Non pertinent

Est-il possible de générer un extrait des valeurs à une date donnée? Autrement dit, en cas de modification d'une valeur, la date de la modification est-elle enregistrée?

Oui

Non

Non pertinent

Quand le contenu des modifications est-il disponible ? Existe-t-il un processus régulier et défini de disponibilité des données ? Dans quel délai les données sont-elles introduites dans le système ?

Immédiatement

Au cas par cas

Régulièrement

Annexe technique B: Tableau de bord des indicateurs de qualité du REE

Dimension: complétude	
Indicateur	Échelle <u>provisoire</u>
Pourcentage des unités légales principales du REE pour lesquelles un «NOM/RAISON SOCIALE» est spécifié.	0 à 100% <i>Exemples:</i> 100%: entièrement complet 95% à 99%: partiellement complet < 95%: incomplet
Pourcentage des unités légales principales du REE pour lesquelles une adresse principale (SIÈGE/DOMICILE) entièrement géocodée est spécifiée.	
Pourcentage des unités légales principales du REE pour lesquelles un «NUMÉRO DE COMMUNE» est spécifié.	
Pourcentages des unités légales principales du REE pour lesquelles une «ACTIVITÉ ÉCONOMIQUE» spécifiée.	
Pourcentage des unités légales principales du REE pour lesquelles une «FORME JURIDIQUE» est spécifiée.	
Comparaison du nombre d'unités légales principales créées en 2020 et du nombre de créations d'entreprises transmises en 2020 au REE par les autres registres administratifs.	Nombre d'annonces reçues (#1) et celui d'unités légales créées (#2) <i>Exemples:</i> #1 = #2: les sources administratives couvrent à 100% le périmètre du REE. #1 > #2: les sources administratives sont redondantes. # 1 < # 2: les sources administratives ne suffisent pas à couvrir tout le périmètre du REE.

Dimension: précision

Indicateur	Échelle <u>provisoire</u>
Pourcentage des unités légales du REE ayant le statut «actif» qui figurent dans au moins un autre registre administratif.	0 à 100% <i>Exemples:</i> 100%: excellent 95% à 99%: satisfaisant < 95%: insatisfaisant

Dimension: cohérence

Indicateur	Échelle <u>provisoire</u>
Pourcentage des unités légales du REE ayant le statut «radié» qui ne figurent effectivement plus dans aucun autre registre administratif.	0 à 100% <i>Exemples:</i> 100%: excellent 95% à 99%: satisfaisant < 95%: insatisfaisant

Dimension: fiabilité

Indicateur	Échelle <u>provisoire</u>
Pourcentage des unités légales principales du REE non identifiées comme «doublon».	
Pourcentage des unités légales principales du REE créées en 2020 non identifiées comme «doublon».	0 à 100% <i>Exemples:</i>
Pourcentage des unités légales principales du REE non identifiées comme «erreur».	100%: excellent 95% à 99%: satisfaisant < 95%: insatisfaisant
Pourcentage des unités légales principales du REE créées en 2020 non identifiées comme «erreur».	

Dimension: rapidité d'exécution

Indicateur	Échelle <u>provisoire</u>
Temps moyen (en heures) du traitement des annonces de mutation reçues en 2020 du registre du commerce (FOSC).	Nombre d'heures <i>Exemples:</i> < 24 h: excellent 24 h à 72 h: satisfaisant > 72 h: insatisfaisant
Temps moyen (en heures) du traitement des annonces de mutation reçues en 2020 du registre TVA.	
Temps moyen (en heures) du traitement des annonces de mutation reçues en 2020 du service de l'agriculture (AGR).	
Temps moyen (en heures) du traitement des annonces de mutation reçues en 2020 du registre IDE.	
Temps moyen (en heures) du traitement des annonces de mutation reçues en 2020 des caisses AVS.	
Pourcentage des annonces de mutation «TVA» (concernant des unités légales) qui ont été reçues en 2020 et qui ont été traitées la même année.	0 à 100% <i>Exemples:</i> 100%: excellent 95% à 99%: satisfaisant < 95%: insatisfaisant
Pourcentage des annonces de mutation «AGR» (concernant des unités légales, «personnes») qui ont été reçues en 2020 et qui ont été traitées la même année.	
Pourcentage des annonces de mutation «IDE» (concernant des unités légales) qui ont été reçues en 2020 et qui ont été traitées la même année.	
Pourcentage des annonces de mutation «AVS» (concernant des unités légales) qui ont été reçues en 2020 et qui ont été traitées la même année.	

Références

- 1) Conseil fédéral : **décision** du 27 septembre 2019, concernant l'utilisation multiple des données (mise en œuvre du principe *once-only*, système statistique suisse et gestion des données de l'administration fédérale)
- 2) Office fédéral de la statistique : **Programm Nationale Datenbewirtschaftung: Projekt Qualitätssicherung Unternehmensdaten**, rapport adressé au Conseil fédéral, 25 novembre 2020
- 3) Office fédéral de la statistique : **Plateforme d'interopérabilité**, <https://www.i14y.admin.ch>, consulté le 3.3.2022
- 4) Statistics Austria: **Quality assessment of administrative data**, Documentation of Methods V2.0, 13 novembre 2019
- 5) Office fédéral de la statistique : **Registre des entreprises et des établissements**, <https://www.bfs.admin.ch/bfs/fr/home/registres/registre-entreprises/registre-entreprises-etablisements.html>, était du 8 avril 2022
- 6) Ehling M. et Körner T. (éd.): **Handbook on Data Quality Assessment Methods and Tools**, <https://ec.europa.eu/eurostat/documents/64157/4373903/05-Handbook-on-data-quality-assessment-methods-and-tools.pdf/c8bbb146-4d59-4a69-b7c4-218c43952214>
- 7) Otto B. et Oesterle H.: **Corporate Data Quality: Voraussetzung erfolgreicher Geschäftsmodelle**, Springer Verlag, 2015
- 8) Batini C. et Scannapieca M.: **Data Quality: Concepts, Methodologies and Techniques**, Springer Verlag, 2006