

Attribution intégrée du nouvel identificateur de personnes

Auteurs

Jérôme Magnin, Dr Sc.

Chef de la section « Statistique et registres centraux », Centrale de compensation AVS/AI, Genève.

Patrick Kummer

Chef de projet « Harmonisation des Registres », Office fédéral de la statistique, Neuchâtel.

Abstract

En 2009, le numéro AVS à 13 chiffres a été attribué comme identificateur de personnes dans les registres officiels, ainsi que dans les bases de données des caisses maladie. Ce projet ambitieux a été couronné de succès.

Le numéro AVS à 13 chiffres a été attribué comme identificateur de personnes dans les registres officiels de la Confédération, des cantons et communes, ainsi que dans les bases de données des caisses maladie. Ce projet ambitieux a été couronné de succès. Deux des responsables du projet exposent la méthode choisie, et dressent le bilan de cette entreprise.

Origines du projet

La perspective d'un numéro unique et multisectoriel pour l'identification administrative des personnes physiques en Suisse s'est définitivement cristallisée en 2005-2006 au travers de deux décisions successives du Conseil fédéral : le choix du nouveau numéro AVS à 13 chiffres comme identificateur de personnes dans les nouveaux cadres légaux de l'Harmonisation des registres (HR) d'une part, et de la carte d'assuré de l'assurance obligatoire de soins (CA) d'autre part.

Ces décisions sont intervenues alors que le projet national « NNSS » (introduction du nouveau numéro d'assuré AVS NAVS13 au sein des assurances sociales fédérales, 2003-2008) était déjà en pleine phase de réalisation. Les délais impartis aux projets nationaux adossés à ces nouveaux cadres légaux étaient serrés (janvier 2009 pour l'introduction de la carte d'assuré, janvier 2010 pour le recensement de la population basé sur la collecte de registres).

La transition du numéro AVS à 11 chiffres existant vers un NAVS13 unique pour chaque personne physique représentait déjà en soi, pour la seule institution AVS/AI, un défi organisationnel et technique. Ceci bien que les organes d'application de l'assurance aient couplé depuis longtemps leurs systèmes d'information au registre central des assurés de la CdC, point de référence pour la gestion du numéro AVS. Venait désormais s'ajouter la nécessité d'introduire rapidement, mais surtout de façon la plus *fiable* possible, un nouvel identificateur dans un nombre considérable de registres administratifs de collectivités publiques et organisations privées présentant les deux caractéristiques suivantes : a) une gestion mutuellement indépendante, et b) aucun lien préexistant avec le registre de référence gérant le numéro AVS. La tâche s'avérait donc d'une ampleur conséquente, en complexité comme en volume.

La concomitance des délais requis pour l'introduction du NAVS13 dans ces deux cadres a rapidement rendu manifeste la nécessité d'une gestion de projet unique, englobant l'ensemble des objectifs à atteindre au sein d'une démarche consolidée s'attendant notamment à la coordination des diverses parties prenantes de ce processus. Sur mandat du Département fédéral de l'intérieur, cette responsabilité de coordination a été endossée par l'Office fédéral de la statistique, au travers de son projet « Numéro d'Assuré » (2006-2009), sous-projet du projet d'« Harmonisation des Registres » (2005-2011).

Une Solution : méthode intégrative

Un objectif ambitieux

La tâche à accomplir se caractérisait ainsi : insérer de façon « exacte » le NAVS13 dans environ 2'700 registres administratifs de personnes présentant des propriétés hétérogènes en matière de contenu (nombre et type d'attributs gérés) comme de mode de gestion (fréquence et règles d'actualisation). Dans le présent contexte, le caractère d'« exactitude » de l'attribution signifie que quels que soient les registres dans lesquels une personne donnée figure – et surtout *quelle que soit la façon dont elle s'y trouve représentée* – elle doit se voir attribuer un seul et même NAVS13 dans chacun d'entre ces registres. Qui plus est, celui qu'elle possède déjà le cas échéant.

La difficulté principale provient des inévitables variations qui affectent des représentations mutuellement indépendantes de l'identité administrative d'une même personne. Elles sont imputables aux facteurs suivants :

- a)** Caractère plus ou moins précis de la définition même de certains attributs, qui peut donner lieu à interprétations et fluctuations « naturelles » (« prénom » = « prénom usuel » ou « prénom selon l'état-civil » ?).
- b)** Temporalité de la valeur objective des attributs, en relation avec le mode de mise à jour du registre (exemple : changement de nom suite à un événement d'état-civil, sans répercussion systématique sur de nombreux registres).
- c)** Erreurs objectives (« bruit ») issue de saisies erronées, d'un déchiffrement humain défaillant, etc.

L'atteinte de l'objectif rendait ainsi incontournable une *réconciliation globale* de l'ensemble des fichiers administratifs concernés.

Cette évidence nous a conduit à opter pour une approche dite « intégrative » du problème, qui consiste en les étapes suivantes :

- a)** *Gathering* : réunir d'une façon ponctuelle la *totalité* des registres destinés à recevoir le NAVS13, dans leur état figé à une date précise préétablie.
- b)** *Records linkage* : opérer une comparaison deux à deux quasi-systématique de tous les enregistrements présents (42 millions). A l'aide d'un algorithme finement paramétré et adapté aux spécificités des données à traiter, mesurer de façon pertinente le degré de similarité entre chacune des paires possibles d'enregistrements.
- c)** *Clustering* : grouper les enregistrements caractérisés par un niveau suffisamment élevé de similarité mutuelle pour constituer avec une probabilité non négligeable des représentations d'une seule et même personne physique.
- d)** *Annotation* : sur la base de la connaissance préalable des registres dont on dispose, classifier les clusters obtenus en « réguliers » et « irréguliers ». Les premiers peuvent se voir assigner un NAVS13 de façon sûre. Les seconds doivent faire l'objet d'une opération de
- e)** *Clearing* manuel des clusters irréguliers : ces derniers sont « régularisés » par un opérateur humain, en obtenant de l'information additionnelle au besoin.

Cette procédure permet d'atteindre un excellent niveau de fiabilité du résultat durant un laps de temps contrôlé, tout en maîtrisant l'ampleur des moyens engagés. Elle implique toutefois de devoir s'assurer de disposer de tous les registres à un instant donné, et que ceux-ci soient dans un état de qualité suffisant pour être compatible avec les délais impartis à l'opération. Un risque important affecte donc l'étape initiale du processus (*gathering*). Les moyens requis en aval, également facteurs critiques de succès, sont :

- a)** Une connaissance fine des registres

impliqués.

b) Un système de « records linkage » performant, modulable et adapté !

c) Une équipe de gestionnaires de clearing proprement formés, disposant d'un catalogue exhaustif de processus préalablement répertoriés et d'une infrastructure adaptée à ses activités (moyens de communication, outil informatique, etc.)

Préparation et mise en place (2007-2008)

Lors de cette phase préparatoire de l'opération désormais qualifiée de « 1^{ère} attribution intégrée du NAVS13 hors l'AVS », le gros de notre effort conjoint a porté sur les fronts suivants :

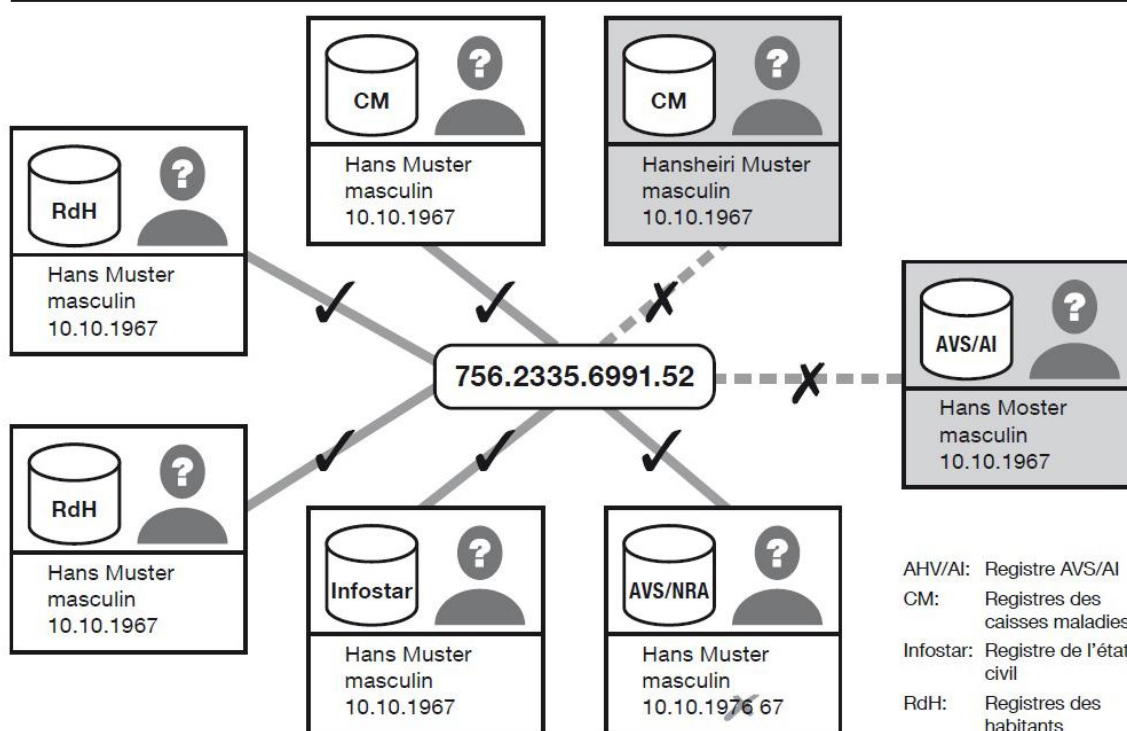
1. Consolidation de la collaboration avec les maîtres de fichiers.

Objectif : livraison en qualité et en heure de chacun des 2'700 registres concernés par l'opération.

Un comité de projet a été constitué et s'est réuni à cadence régulière, dans lequel étaient représentés les registres fédéraux impliqués, les caisses maladie (via santésuisse) ainsi que les registres cantonaux et communaux d'habitants (via l'OFS et les services cantonaux de

coordination pour l'harmonisation des registres d'habitants). Des conventions de collaboration tripartites (OFS – CdC – partie X) ont été élaborées et conclues avec chaque participant, explicitant : l'objectif visé, les tâches à accomplir, les modalités d'exécution particulières, ainsi que les responsabilités propres à chacun. Une analyse détaillée de la situation existante a été conduite : capacité des systèmes d'information à exporter leurs données dans des formats exploitables, données effectivement disponibles dans chaque type de registre, et leur qualité. Des standards spécifiques ont été définis pour l'échange des données (formats eCH). Des séances de formation ont été délivrées aux cantons afin de les familiariser avec le processus et répondre à leurs questions. Des guidelines pratiques de préparation des données (détection et élimination des doublons, des erreurs flagrantes, normalisation) ont été rédigées et distribuées à tous les acteurs. Plusieurs outils et services adaptés à la situation sur le terrain, facilitant la préparation des données et l'amélioration préventive de leur qualité, ont été développés et mis à disposition par l'OFS et la CdC : convertisseurs de formats, normalisateurs et validateurs de données, service online de détection de (quasi-) doublons.

Fiabilité d'attribution du NAVS13



2. Conception, validation et déploiement de l'infrastructure de traitement (records linkage et clearing) à la Centrale de compensation.

Objectif : se donner les moyens de produire un résultat de qualité très élevée durant un laps de temps restreint à quelques mois.

La CdC a mené de front, durant cette période, les projets suivants :

A. Conception et mise en place d'une infrastructure de clearing dimensionnée pour 60 collaborateurs : catégorisation systématique des cas possibles d'anomalies et élaboration de procédures standard de régularisation, ciblage des profils professionnels requis et recrutement, conception et développement d'une plateforme informatique dédiée aux opérations de clearing.

B. Elaboration, déploiement et validation d'une infrastructure technique permettant la comparaison massive de données (matching). Paramétrisation fine du système générique de records linkage choisi, afin de l'adapter aux spécificités des données à traiter.

C. Conception détaillée et développement des produits « UPI » (Unique Person Identification). La base de données UPI était appelée à recevoir le produit final du traitement de 1^{ère} attribution intégrée.

La mise au point des dispositifs élaborés par les projets A et B a grandement bénéficié de deux phases de tests portant

sur des données réelles :

➤ Printemps 2007 : un test-pilote avec les données des registres fédéraux et de 2 cantons a permis de a) valider définitivement le choix de la méthode intégrative et b) dimensionner les volumes attendus dans les différentes catégories de clusters d'enregistrements.

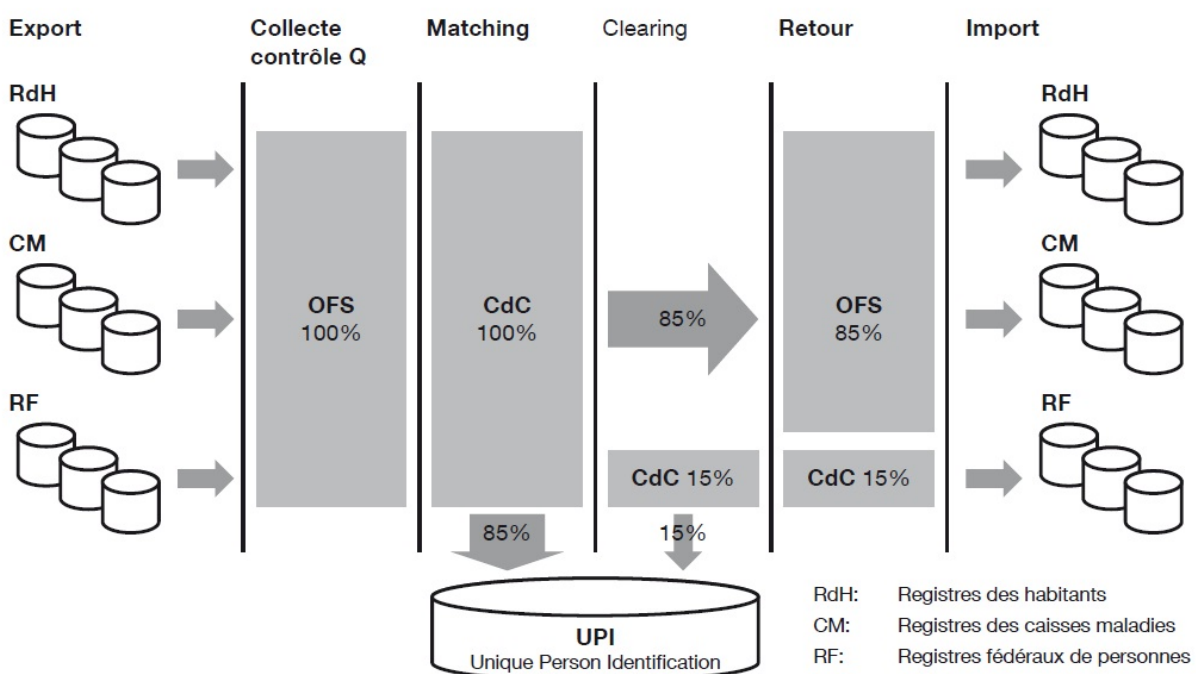
➤ Automne 2008 : une « répétition générale » impliquant la totalité des registres, avec feed-back détaillé aux maîtres de fichiers, a permis (1) d'affiner encore les estimations et (2) de procéder aux derniers ajustements dans les données et les dispositifs d'acheminement et de traitement de celles-ci.

Réalisation (2009)

L'opération productive s'est déroulée de janvier à décembre 2009. Conformément au planning, 100% des registres attendus nous ont été livrés en l'espace de 3 semaines, dans leur état au 15 janvier. La comparaison massive de ces données entre elles (records linkage, clustering et annotation) s'est déroulée sur 12 jours. A mi-février, le résultat final des traitements informatisés était disponible pour la suite des opérations.

Début mars a vu l'entrée en lice de l'équipe de clearing, s'attaquant à un volume d'1 million de clusters d'enregistrements à régulariser. Au même

Processus



moment, les données issues de clusters réguliers ont été insérées dans la base de données UPI, et celle-ci a alors officiellement repris la fonctionnalité d'identification de personnes dans le registre central des assurés AVS/AI.

En mai, les 89% d'enregistrements auxquels un NAVS13 a pu être attribué automatiquement ont été retournés aux maîtres de fichiers pour intégration dans leurs systèmes d'information. Les travaux de clearing, quant à eux, se sont poursuivis sans relâche de mars à novembre, rythmés par les retours mensuels aux maîtres de fichiers de lots de données complétées du NAVS13. En juin, le premier objectif chiffré de l'opération était rempli, avec l'atteinte du seuil de 95% d'attribution du NAVS13 au sein des registres participants. En novembre, ce chiffre était porté à 99.65% et l'objectif final atteint.

Perspectives (2010)

Les lendemains ne chanteront pas trop faux qu'à la stricte condition que le capital investi dans l'opération réalisée en 2009 soit préservé. Pour y parvenir, le concours actif de tous est indispensable, qu'ils soient fournisseurs d'information à UPI (registres fédéraux) ou utilisateurs systématiques du NAVS13 (registres d'habitants, caisse maladie, administrations fiscales etc.). En pratique :

➤ Des annonces électroniques quotidiennes des registres fédéraux à UPI (nouvelles personnes et mutations) assurent désormais la complétude et l'actualité du contenu d'UPI. Elles permettent, entre autres, l'attribution du NAVS13 au plus tôt dans la vie administrative d'une personne, à savoir quelques jours seulement après sa naissance ou son établissement sur le territoire national.

➤ Depuis avril 2009, la CdC permet aux utilisateurs systématiques du NAVS13 d'administrer avec exactitude et facilité, au quotidien, cet identificateur dans leurs systèmes d'information. Cette « mise à jour continue » s'opère à l'aide d'outils

en ligne d'interrogation : UPIViewer (GUI web), UPIServices (bouquet de webservices).

➤ UPI gère désormais en permanence les caractères d'identification personnelle les plus « officiels » disponibles dans des systèmes d'information électroniques fédéraux (services de l'état-civil et/ou du contrôle des migrations). L'association « NAVS13 & caractères d'identification adjoints » qu'UPI offre constitue une implémentation relativement forte et robuste de la notion d'identité administrative. UPI la met désormais à disposition d'un très large cercle d'organisations. A terme, nous espérons que cette plus-value induise, au sein de la communauté *entière* des utilisateurs systématiques du NAVS13 légalement légitimés, l'usage d'un ensemble d'attributs officiels *commun* pour l'identification harmonisée d'une personne physique.

Bilan et conclusion

L'investissement important consenti dans la phase de préparation a, en définitive, porté ses fruits, permettant l'aboutissement de ce que d'aucuns considéraient initialement comme une entreprise hasardeuse. Au travers de cette opération d'envergure, le NAVS13 a été introduit à quasiment 100% dans tous les registres prévus par le législateur. La qualité d'attribution que nous estimons avoir atteinte nous permet d'être aujourd'hui confiants quant à la fiabilité avec laquelle cet identificateur peut dorénavant y être exploité.

Le dispositif « NAVS13 + UPI » mis en place représente sans doute un pas significatif en direction d'une implémentation sensée du concept d'« identité administrative » en Suisse. A noter toutefois que les conditions cadre régissant actuellement les types de noms (harmonisation difficile entre droit suisse et Schengen) rend difficile une telle implémentation pour les ressortissants étrangers.

Remarquons pour conclure que l'établissement d'un lien fort et authentifié entre identité administrative d'une part, et personne physique sous-jacente de l'autre, reste incontournable pour nombre d'applications. L'arrivée récente sur le marché du nouveau passeport biométrique et du produit « SuisseID » fournit deux implémentations intéressantes d'un tel lien. Si le scénario d'une adoption systématique du SuisseID par le citoyen devait se concrétiser, il serait opportun de réfléchir au positionnement du NAVS13 en relation avec ce dernier.
