Prof. Dr. David Basin: Évaluation des risques de l'utilisation du numéro AVS

5 Analyse des risques

Nous allons commencer par mettre en évidence de grandes catégories d'attaquants motivés et qualifiés. Nous examinerons ensuite les caractéristiques des systèmes pertinentes pour nos scénarios et leurs effets sur les risques en matière de sécurité et de protection des données. Enfin, nous analyserons les risques pour la protection des données associés aux différents scénarios.

5.1 Qui pourrait attaquer les registres suisses?

Comme il a été expliqué au chapitre 4.1, les risques pour la protection des données résident dans les actions entreprises par des attaquants. Ce chapitre montre qu'il existe effectivement des attaquants qui ont des motifs suffisants et les capacités nécessaires pour accéder aux données conservées dans différents registres suisses.

Les actions problématiques portant sur les données conservées dans les registres peuvent être le fait d'*utilisateurs internes*. Les attaques et méfaits commis par les utilisateurs internes constituent généralement un grave problème, qui est souvent sous-estimé [14]. Un utilisateur interne peut utiliser son droit d'accès à des fins non autorisées. Il peut par exemple être curieux de connaître le statut d'une personne donnée ou vouloir aider un collègue travaillant dans un autre office. Des employés mécontents peuvent chercher à nuire, par exemple en portant atteinte à la réputation de leur employeur pour se venger. Les utilisateurs internes aux intentions criminelles peuvent profiter de leur droit d'accès pour collecter des données personnelles et les vendre sur le marché noir à des personnes qui les utiliseront par exemple à des fins d'usurpation d'identité ou de publicité ciblée. Enfin, les utilisateurs internes peuvent être négligents. Si négliger de sécuriser les systèmes de manière appropriée, de réduire au minimum la conservation de données sensibles, d'effacer certaines données, etc. ne constitue pas forcément un problème en soi, cela peut cependant permettre aux attaquants externes d'accéder plus facilement aux données.

Les systèmes peuvent également être attaqués de l'extérieur. Les pirates informatiques sont intéressés par le défi consistant à s'introduire dans des systèmes informatiques, et la publication en ligne de données sensibles peut être vue comme un trophée. Les pirates motivés par des considérations politiques (« hacktivistes ») peuvent effectuer des opérations problématiques afin d'humilier leurs adversaires. Les délinquants, quant à eux, s'introduisent dans des systèmes dans le but de voler des données et de les vendre. Les États attaquants peuvent vouloir accéder à des données sensibles (données financières ou médicales, antécédents judiciaires, etc.) permettant d'obtenir des renseignements sur les individus, afin, par exemple, d'arrêter les personnes coupables de fraude fiscale ou

d'exercer une influence sur des élections. Le rapport MELANI sur l'affaire d'espionnage qui a touché RUAG [8] illustre l'habileté et les efforts que peuvent déployer des États attaquants pour s'introduire dans des systèmes hautement sécurisés.

Ci-dessous nous donnons deux exemples concrets des motifs poussant à collecter et à agréger des données personnelles (premier exemple) et des préjudices que peut causer la compromission d'un seul système (deuxième exemple).

Moneyhouse.ch Le premier exemple, tiré du rapport [5], illustre la manière dont des données personnelles peuvent être monnayées. Il montre également comment des données provenant de différentes sources peuvent être associées les unes aux autres afin d'établir des profils révélant une grande partie de la vie des personnes concernées. Enfin, il montre qu'il n'est pas nécessaire, pour que de tels risques surviennent, de craquer des systèmes, qu'il suffit de relier entre elles des données non pseudonymisées accessibles publiquement.

L'entreprise suisse itonex AG gérait la plateforme moneyhouse.ch, servant à collecter et à vendre des données concernant des entreprises et des particuliers. À l'époque où le rapport susmentionné a été établi, itonex offrait un large éventail de services d'information, payants ou gratuits, dont la valeur provenait du croisement de différentes sources de données, dont des registres administratifs tels que le registre du commerce. Ce croisement de données permettait à itonex d'établir des profils d'individus auxquels il était possible d'accéder depuis la plateforme ou en faisant des recherches dans Google. Les clients payants pouvaient accéder à des données personnelles telles que le prénom, le nom, le domicile, le code postal, la date de naissance, la profession, les membres du ménage (enfants compris) et les voisins. Les données relatives au logement comprenaient des vues aériennes et des images Street View de l'immeuble, des informations sur le type de bâtiment, le nombre de ménages habitant dans le bâtiment, les coûts et la période de construction ainsi que le nombre d'étages. L'ensemble de ces données fournissait une image globale des conditions de vie de personnes identifiables, conditions formant une partie importante et sensible de leur profil.

Equifax Le deuxième exemple concerne les États-Unis. En septembre de cette année, Equifax, l'une des plus grandes sociétés américaines d'informations sur la solvabilité, a fait état d'une fuite de données consécutive à une faille dans une application web [19, 10]. Des pirates ont exploité cette faille pour accéder aux noms, adresses, numéros de sécurité sociale, numéros de cartes de crédit et numéros de permis de conduire de 143 millions d'Américains, ce qui représente près de la moitié de la population des États-Unis, ainsi qu'à des documents relatifs à des litiges et contenant des informations permettant une identification personnelle.

Cette faille aura de lourdes conséquences. Les données vont circuler sur le marché noir pendant des décennies et permettre différentes formes d'usurpation d'identité. Comme l'explique Brian Womack [19], les fraudeurs combinent généralement les numéros de sécurité sociale volés et éventuellement d'autres données sur les personnes concernées avec des adresses d'emprunt et demandent de nouvelles cartes de crédit, sur lesquelles ils auront le contrôle. Selon le même auteur, certains escrocs persévérants utilisent ces nouvelles identités pour demander des cartes de crédit supplémentaires ou des prêts, moyens qu'ils exploitent ensuite simultanément au maximum, ce qui leur permet de constituer un butin pouvant s'élever à plusieurs dizaines de milliers de dollars.

Cet exemple illustre lui aussi le risque découlant de l'agrégation d'informations permettant une identification personnelle. Certains attaquants sont fortement motivés pour voler et monnayer des données. De plus, la sécurité des systèmes est imparfaite et le contrôle des accès et d'autres mécanismes peuvent être déjoués. L'entreprise Equifax se veut un organisme fournissant des informations fiables en matière de crédit à la consommation et est présumée appliquer des normes de sécurité élevées. Enfin, ce piratage aura des conséquences à long terme, étant donné que les numéros de sécurité sociale (comme le NAVS13) sont attribués une fois pour toutes et que les autres attributs personnels sont difficiles à modifier. Il constituera donc une source durable de problèmes pour les personnes concernées.

5.2 Caractéristiques des systèmes influant sur les risques pour la protection des données

Nous allons faire ici quelques remarques générales qui seront utiles lors de la comparaison des différents scénarios.

5.2.1 La sécurité parfaite est un leurre

Les systèmes informatiques ne sont jamais sûrs à 100 %. La sécurité est une condition de la protection des données, étant donné qu'il faut empêcher les accès non autorisés à ces dernières. Malheureusement, il y a toujours un risque que des attaquants parviennent à exploiter des failles du matériel, du système d'exploitation, d'une application, d'un processus, etc. ou des failles humaines. Des attaquants sont donc susceptibles de compromettre des systèmes et d'accéder aux données qu'ils contiennent, même si les systèmes en question sont protégés par des mesures de sécurité de pointe.

5.2.2 Les systèmes ne sont pas tous également sûrs

La sécurité d'un système dépend des efforts mis en œuvre pour le sécuriser et de l'efficacité des mesures prises sur les plans technique et organisationnel. Il est également important de mettre en œuvre des processus d'assurance sécurité afin de vérifier que ces mesures ont les effets prévus. Les systèmes de la Confédération

(voir chap. 2.2.2) offrent généralement un niveau de sécurité relativement élevé, grâce aux hautes exigences en matière de sécurité et de processus d'assurance des organes responsables de ces systèmes, tels que l'Unité de pilotage informatique de la Confédération.

Il en va autrement des systèmes gérés par les organismes cantonaux, communaux ou privés (par ex. hôpitaux, écoles) autorisés à utiliser systématiquement le NAVS13. Les exigences de la législation fédérale relatives à la sécurité et aux processus d'assurance sont moins élevées pour ces systèmes non fédéraux [16] (exigences auxquelles peuvent cependant s'ajouter les exigences d'autres administrations). Ces systèmes sont généralement nettement moins sûrs.

5.2.3 Traitement des informations relatives à l'identité

Comme il a été expliqué au chapitre 3.1, les données contenues dans un registre peuvent être organisées de différentes manières. Copier localement des parties du registre UPI constitue généralement un risque pour la protection des données. La compromission d'un certain nombre de systèmes (moins bien protégés) suffit pour qu'un attaquant puisse reconstituer une grande partie de la base de données UPI. Par ailleurs, veiller à ce que des copies restent cohérentes est astreignant.

Un autre risque pour la protection des données découle du fait que les données issues du registre UPI sont généralement conservées dans les registres administratifs avec des attributs relatifs à l'identité (voir chap. 2.2.3) et que ces derniers suffisent à rattacher les données aux personnes concernées et à relier les tables avec une haute précision (voir chap. 2.2.3). Supprimer uniquement les attributs liés au NAVS13 ou uniquement les quasi-identifiants des registres administratifs ne réduit donc guère les risques pour la protection des données. **Pour renforcer celle-ci, il faut dissocier les données des registres de services administratifs ou d'autres organismes *à la fois* du NAVS13 *et* des quasi-identifiants permettant d'identifier une grande partie de la population.** À noter que cela implique un mode d'organisation des données différent de celui qui est actuellement recommandé aux organismes autorisés à utiliser systématiquement le NAVS13 (voir chap. 3.1.1).

5.2.4 Lieu et mode de conservation des données

Ce point est étroitement lié au précédent. Non seulement les données contenues dans un système peuvent être *organisées* de différentes manières, mais en outre les tables de données peuvent être *réparties* entre plusieurs bases de données et plateformes afin de constituer un système d'information distribué. La répartition, si elle est effectuée de manière appropriée, peut offrir les mêmes fonctionnalités qu'un système centralisé (par ex. la possibilité de relier les enregistrements de la manière requise pour l'exécution de certaines tâches), mais garantit une plus grande sécurité et une meilleure protection des données.

Nous allons illustrer cela au moyen d'un exemple basé sur la figure 4. Considérons les deux scénarios suivants :

Scénario 1 : chaque registre sectoriel est conservé dans la *même* base de données que la table de liaison du secteur.

Scénario 2 : ces deux tables sont conservées dans des bases de données distinctes, enregistrées sur des serveurs différents. Pour assister les organismes concernés dans l'exécution de leurs tâches, des requêtes distribuées sont possibles, c'est-à-dire qu'on peut par exemple établir une correspondance entre un identifiant sectoriel et un NAVS13 en combinant les données issues des deux bases de données, si cela est (absolument) indispensable pour que les organismes puissent s'acquitter de leur mission. Pour des raisons de sécurité, toute communication implique une authentification appropriée.

Dans le premier scénario, un attaquant pourrait exploiter une faille logicielle dans la base de données dans laquelle est conservé le registre d'un secteur pour compromettre *les deux* tables de données de ce secteur. Cela lui permettrait d'accéder à tout le contenu de ces tables et de relier ces dernières, autrement dit d'associer le registre sectoriel aux NAVS13. Dans le deuxième scénario, la compromission du registre sectoriel ne permettrait pas d'associer les données aux personnes concernées ou à des données extérieures au secteur. Pour que cela soit possible, il faudrait que l'attaquant compromette, *outre* le registre sectoriel, (i) soit la base de données contenant la table de liaison, (ii) soit la voie de communication nécessitant une authentification et servant au traitement des requêtes distribuées, afin de pouvoir envoyer de fausses requêtes dans le système distribué. Ainsi, dans le deuxième scénario, il est plus difficile de rattacher les données aux personnes concernées ou de combiner les données provenant de différents registres, et les risques en matière de sécurité et de protection des données sont donc plus faibles.

À noter que la différence entre les deux scénarios se révèle encore plus grande si des mesures de sécurité supplémentaires sont prises dans le deuxième cas, telles qu'un renforcement des systèmes, le chiffrement des données, une gestion des clés fondée sur des modules de sécurité matériels, etc. Dans un système conçu de manière appropriée, de telles mesures peuvent réduire le risque de certains types d'attaques et donc les risques pour la protection des données. Le revers de la médaille consiste dans une augmentation de la complexité et du coût du système.

5.3 Comparaison des scénarios

Nous allons maintenant revenir aux scénarios mentionnés au chapitre 1.2 et comparer les risques associés à chacun d'entre eux.

5.3.1 Extension de l'utilisation du NAVS13

À l'heure actuelle, plus de 14 000 organismes sont autorisés à utiliser systématiquement le NAVS13, dont beaucoup emploient des systèmes relativement peu sûrs. Il existe donc déjà un risque élevé que des attaquants compromettent des systèmes et en extraient des données comprenant tant les NAVS13 que les attributs personnels. L'intrusion dans un seul registre peut conduire à la fuite de données sensibles permettant une identification personnelle. Par ailleurs, si plusieurs registres sont compromis, l'utilisation du NAVS13 permet de relier facilement les données contenues dans ces derniers.

Plus le nombre d'organismes utilisant le NAVS13 dans leurs systèmes d'information est important, plus ces risques sont élevés, et ce pour deux raisons :

1. Plus le nombre d'organismes qui collectent, conservent et traitent des données personnelles associées au NAVS13 augmente, plus la probabilité que des systèmes soient compromis s'accroît.

2. Plus le nombre d'organismes extérieurs à l'administration fédérale qui sont autorisés à utiliser systématiquement le NAVS13 augmente, plus le volume de données qui peuvent être reliées de manière directe et univoque aux personnes concernées grâce au NAVS13 et qui sont collectées, conservées et traitées dans des systèmes informatiques non fédéraux relativement peu sûrs croît. En effet, les exigences en matière de sécurité et de processus d'assurance sécurité applicables à ces systèmes sont moins élevées et la sécurité n'était généralement pas un objectif prioritaire lors de la conception de ces derniers (voir chap. 5.2.2).

Les données des organismes qui, à l'heure actuelle, collectent et traitent des données personnelles qu'ils conservent avec des attributs relatifs à l'identité constituant des quasi-identifiants de haute qualité (par ex. le prénom, le nom, la date de naissance) risquent déjà d'être reliées aux personnes concernées ou à d'autres données au moyen des quasi-identifiants. Le risque supplémentaire découlant du fait d'enregistrer *en outre* les NAVS13 est marginal (voir chap. 2.2.3 et 5.2.3), étant donné que l'établissement de liens est possible sans cela.

En suivant cette logique, on peut faire valoir qu'il n'y a aucune raison de *ne pas* conserver les NAVS13 dans les registres, étant donné qu'ils n'ont qu'une faible influence sur les risques pour la protection des données. Cet argument a été avancé pour justifier la pratique qui a été adoptée et son extension. Cette conclusion est correcte, *mais seulement dans le contexte actuel*, dans lequel les NAVS13 sont

toujours accompagnés d'attributs relatifs à l'identité. Or, conserver ces derniers dans le même registre n'est pas une obligation. L'enregistrement redondant des informations relatives à l'identité est étranger tant à une base de données bien conçue qu'à une bonne pratique en matière de protection des données.

La question fondamentale qui se pose alors du point de vue de la protection des données est celle de la comparaison entre, d'une part, la pratique actuelle et son extension et, d'autre part, des scénarios susceptibles de réduire les risques grâce au renforcement des exigences en matière de sécurité (chap. 5.2.2) et à la réorganisation, à la réduction au minimum (chap. 5.2.3) et à la répartition (chap. 5.2.4) de la conservation des attributs relatifs à l'identité et des informations permettant une identification personnelle qui leur sont associées. Nous allons étudier cette question en considérant les deux cas suivants.

5.3.2 Autres identifiants non signifiants et sectoriels

Le NAVS13 est un identifiant non signifiant utilisé dans différents secteurs. Par comparaison, les identifiants sectoriels sont aussi des identifiants non signifiants, mais leur utilisation est limitée *localement* à un secteur. Quand on utilise un identifiant sectoriel, les registres de ce secteur ne peuvent être reliés directement aux registres d'autres secteurs. En conséquence, si on recourt à des identifiants sectoriels en respectant certains principes, on réduit le risque qu'un attaquant agrège les grandes quantités de données personnelles existant dans différents secteurs, révélant ainsi des pans importants de la vie des individus concernés.

Nous soulignons une fois de plus qu'au vu de la *pratique actuelle* consistant à conserver les données des registres avec des attributs relatifs à l'identité, remplacer simplement le NAVS13 par des identifiants sectoriels ne réduirait pas substantiellement les risques pesant sur la protection des données. Les données des registres peuvent toujours être reliées aux personnes concernées en utilisant les attributs relatifs à l'identité comme quasi-identifiants. Ces attributs permettent aussi de relier des tables de données avec une grande précision.

Il est possible et souhaitable d'introduire des identifiants sectoriels en tant que solution de rechange au NAVS13, mais il faut le faire en respectant certains principes. Il faut en particulier à la fois réduire au minimum la conservation d'attributs relatifs à l'identité et protéger adéquatement toutes les tables de liaison (voir chap. 5.2.2 à 5.2.4). **Introduire de cette manière des identifiants sectoriels réduirait les risques en matière de sécurité dus à l'utilisation qui est faite aujourd'hui du NAVS13 et à son extension continuelle.**

Pour ce faire, il peut être nécessaire de changer les architectures logicielles existantes. Le scénario 2 présenté au chapitre 5.2.4 donne un exemple, qui suggère que le traitement de requêtes distribuées peut être nécessaire dans certains cas. Autre exemple : le domaine de la santé, discuté au chapitre 3.2.3. Les attributs relatifs à l'identité ne devraient pas être conservés dans l'index principal des patients (Master Patient Index) des communautés de services de soins et leur conservation devrait être réduite au minimum dans les bases de données des

fournisseurs de soins et dans les dossiers des patients. Il faudrait, en lieu et place, travailler en priorité avec le pseudonyme associé. Les patients, à leur tour, auraient à s'identifier avec leur pseudonyme (par ex. via une carte santé électronique) ; les autres attributs relatifs à l'identité pourraient aussi être présentés, mais ils ne devraient pas être conservés.

Lorsque des identifiants sectoriels doivent être reliés à des informations relatives à l'identité, dans des processus administratifs, il reste possible de le faire en procédant à des requêtes distribuées utilisant les tables de liaison. Ces tables devraient répondre à des exigences accrues en termes de protection et des mesures supplémentaires devraient être prises pour les sécuriser et assurer qu'elles ne sont utilisées qu'à des fins appropriées. Plusieurs options sont ici possibles. Les tables de liaison peuvent par exemple être conservées auprès de la Centrale de compensation, à l'intérieur d'un service administratif spécifique, comme l'Office fédéral du registre du commerce, ou encore au sein d'un organisme privé comme une communauté de services de soins. Il est possible de réaliser toutes ces options avec des risques considérablement plus faibles pour la protection des données qu'en recourant directement au NAVS13, à condition que les tables de liaison soient protégées correctement.

5.3.3 Combinaison interne du NAVS13 avec des identifiants utilisés à l'externe

Le registre du commerce est un exemple d'une telle combinaison. En principe, la seule différence entre un identifiant sectoriel utilisé à l'interne et un identifiant sectoriel utilisé à l'externe réside dans le fait que les attaquants ne doivent fournir aucun effort pour accéder aux données disponibles à l'externe. Pour des données moins sensibles, telles que les données figurant dans le registre du commerce, ce type de combinaison peut être pertinent, en particulier lorsque les données devraient être publiques. **Dans ce cas, l'utilisation d'un identifiant sectoriel externe permet, par rapport à l'utilisation directe du NAVS13, de réduire le risque que des attaquants agrègent des données provenant de registres extérieurs au secteur.** La mesure dans laquelle ce risque diminue dépend cependant de la facilité et de la précision avec laquelle les personnes concernées peuvent être identifiées au moyen des autres attributs relatifs à l'identité également publics.

27 septembre 2017