

TOWARDS A DATA VISION AND STRATEGY @SCIENSANO

**Une évaluation de la maturité des données
et plan de mise en œuvre**

—

Groupe de travail sur la stratégie de données

QUI SOMMES-NOUS ?

Sciensano, ce sont plus de 900 collaborateurs qui s'engagent chaque jour au service de notre devise « Toute une vie en bonne santé ».

Comme son nom l'indique, la science et la santé sont au cœur de la mission de cette organisation. Sciensano puise sa force et sa spécificité dans une approche holistique et multidisciplinaire de la santé. Plus spécifiquement, nos activités sont guidées par l'interconnexion étroite et indissoluble entre la santé humaine et animale et leur environnement (le concept « One health » ou « Une seule santé »). Dans cette optique, en combinant plusieurs angles de recherche, Sciensano contribue d'une manière unique à la santé de tous.

Sciensano s'appuie sur plus de 100 ans d'expertise scientifique.

Sciensano

Services de la direction générale - Stratégie et positionnement externe

Stratégie de données

Rédigé Avril 2022 • Bruxelles • Belgique

Adapté : Septembre 2022 • Bruxelles • Belgique

Approuvé par : Conseil d'administration



Stratégie et positionnement externe

Personne de contact Sofie De Broe • T+32 478 843194 • sofie.debroe@sciensano.be

TABLE DES MATIÈRES

ABRÉVIATIONS	6
INTRODUCTION	9
1. Qu'est-ce qu'une stratégie de données et pourquoi est-ce important ?	9
1.1. Motivations pour une stratégie de données de Sciensano.....	10
1.2. Initiatives favorables préexistantes	11
2. Visualisation des données de Sciensano	13
2.1. Cadres légaux.....	14
2.2. Utiliser les données internes et externes à Sciensano	16
COMPOSANTES D'UNE STRATÉGIE DE DONNÉES	20
POUR SCIENSANO	20
1. Interopérabilité organisationnelle	21
1.1. Objectifs stratégiques	21
1.2. Leadership et gestion.....	21
1.3. Supervision et responsabilité	21
1.4. Respect de la législation	22
1.5. Éthique des données	22
1.6. Approvisionnement	23
1.7. Liens avec un écosystème de données plus large	23
2. Interopérabilité humaine	25
2.1. Administration des données.....	25
2.2. Protection des données à caractère personnel et préservation de la confidentialité	26
2.3. Connaissances et compétences du personnel	27
2.4. Communication interne et externe	27
2.5. Adaptabilité	28
3. Interopérabilité des données	28
3.1. Capacité de modélisation des données et des métadonnées.....	28
3.2. Capacité d'organisation et de classification des données.....	29
3.3. Accès, ouverture et partage des données	29
3.4. Analyse et automatisation des données	32
3.5. Protection et stockage des données	32
4. Interopérabilité technologique	33
4.1. Infrastructure des données	33
4.2. Cybersécurité et réaction aux incidents	34
FUTUR	35
1. Interopérabilité organisationnelle	35
1.1. Objectifs stratégiques	35
1.2. Leadership et gestion.....	35
1.3. Supervision et responsabilité	35
1.4. Respect de la législation	36
1.5. Éthique des données	36
1.6. Approvisionnement	36
1.7. Liens avec un écosystème de données plus large	37
2. Interopérabilité humaine	37
2.1. Administration des données.....	37
2.2. Protection des données à caractère personnel et préservation de la confidentialité	37
2.3. Connaissances et compétences du personnel	38
2.4. Communication interne et externe	38
2.5. Adaptabilité	39

3. Interopérabilité des données	39
3.1. Capacité de modélisation des données et des métadonnées	39
3.2. Capacité d'organisation et de classification des données.....	43
3.3. Accès, ouverture et partage des données	43
3.4. Analyse et automatisation des données	43
3.5. Protection et stockage des données	44
4. Interopérabilité technologique	45
4.1. Infrastructure des données	45
4.2. Cybersécurité et réaction aux incidents	45
5. Instituts de santé publique orientés sur les données	45
5.1. Le rôle des instituts de santé publique dans la société	45
5.2. UE et exemples inspirants	46
CONCLUSION	48
RÉFÉRENCES	50
ANNEXES	51

ABRÉVIATIONS

AHEAD	Vers le développement d'une plateforme nationale de données sur la santé
AHLICS	Système de centralisation de l'information des laboratoires de santé animale
IA	Intelligence artificielle
AirAllergy	Réseau belge de surveillance aérobiologique
BeBOD	Étude nationale belge sur le fardeau de la maladie
CESSDA	Réseau européen d'archives de données en sciences sociales
CINECA	Infrastructure commune pour les cohortes nationales en Europe, au Canada et en Afrique
CoZo	Plateforme de soins collaborative
DIPoH	Infrastructure décentralisée sur la santé de la population
PGD	Plan de gestion des données
DPD	Délégué à la protection des données
DXP	Plateforme d'expérience numérique
EAN	Réseau européen des aéroallergènes
ECMWF	Centre européen pour les prévisions météorologiques à moyen terme
EEDS	Espace européen des données de santé
EJP RD	Programme européen conjoint sur les maladies rares
IELS	Implications éthiques, légales et sociales
EOSC	Cloud européen pour la science ouverte
Epistat	Épidémiologie des maladies infectieuses
HERA	Autorité européenne de préparation et de réaction en cas d'urgence sanitaire
EUMETNET AutoPollen	Programme européen AutoPollen
Principes FAIR	Facile à trouver, Accessible, Interopérable, Réutilisable
FDP	FAIR Data Point
RGPD	Règlement général sur la protection des données
ADS	Autorité des données de (soins de) santé
HISIA	Analyse interactive de l'enquête de santé belge par interview
HRIC	Portail de données pour la recherche et l'innovation dans le domaine de la santé
IANPHI	Association internationale des instituts nationaux de santé publique
IMI	Initiative en matière de médicaments innovants
InfAct	Action commune sur l'information sanitaire
SGSI	Système de gestion de la sécurité de l'information
SGIL	Système de gestion de l'information des laboratoires
INSP	Instituts nationaux de santé publique
ISP	Instituts de santé publique
PHIRI	Infrastructure de recherche sur les informations de santé de la population
GDR	Gestion des données de la recherche

REDCap	Research Electronic Data Capture
SCRA	Analyse des petites cellules
POS	Procédure opérationnelle standard
SPMA	Procédures standardisées d'analyses de la mortalité
TDI	Indicateur de demande de traitement
TEHDAS	Vers l'espace européen des données de santé
TOP	Technologie, organisation, procédure
TDC	Tiers de confiance
UXP	Plateforme d'expérience utilisateur
IBV	Infrastructure de bureau virtuel

INTRODUCTION

1. Qu'est-ce qu'une stratégie de données et pourquoi est-ce important ?

Les données jouent un rôle de plus en plus important dans la gestion de notre vie quotidienne, du secteur public et des entreprises. Le potentiel des données pour fournir de l'information pertinente sur le plan politique est énorme. En parallèle, la quantité de données non pertinentes augmente également. Il est donc d'autant plus difficile de protéger correctement les données utiles, de les gérer et de les distinguer des données « parasites ». Les institutions et entreprises publiques se voient donc dans l'obligation de faire des choix quant aux données qu'elles gèrent et analysent (ou du moins qu'elles veulent gérer et analyser), à leur manière de protéger et partager ces données et aux informations qu'elles veulent rendre disponibles plus tard.

Une stratégie de données est un processus très dynamique utilisé afin de soutenir l'acquisition, l'organisation, l'analyse et la transmission de données en faveur d'objectifs commerciaux. La stratégie de données spécifique à Sciensano est un plan sur le long terme qui décrit la manière dont Sciensano va appliquer les principes de données FAIR, qui garantissent que nos données sont faciles à trouver, accessibles, interopérables et réutilisables. Ce plan définit également d'une part les mesures que Sciensano prendra pour rendre ses données disponibles et contrôlables, et d'autre part la manière dont Sciensano générera et utilisera à l'avenir les nouveaux flux de données et les informations afin de remplir la mission de l'organisation. Tout dépend du niveau de maturité des données de l'organisation par rapport à divers indicateurs d'interopérabilité (1) : l'objectif de nos données, la vision future de Sciensano quant au rôle qu'il entend jouer dans le monde de la gestion des données et la mesure dans laquelle il compte innover dans le cadre de sa mission future.

Le développement d'une stratégie de données est donc essentiel. Étant donné que Sciensano génère et utilise une quantité toujours plus grande de données et que celles-ci sont son plus grand atout, une stratégie de données est une nécessité. Le projet de stratégie de données a commencé dans notre institut à la fin de l'année 2021 et sera mis en œuvre en 2022. Il suit l'évolution présente en Europe, Europe qui a adopté une position forte dans le monde de la gestion des données et qui possède sa propre stratégie de données.

Tout d'abord, ne pas avoir de stratégie de données risque d'ébranler la réputation d'une organisation en termes de fiabilité, de transparence, de qualité et de sécurité en matière de données (à caractère personnel) de santé. De plus, cela reflète aussi une attention insuffisante aux principes FAIR. Enfin, cela revient à manquer des opportunités en ne prenant pas suffisamment en compte le potentiel futur des données pour les objectifs et la vision de l'organisation (2). Une stratégie de données efficace et bien élaborée garantit que les données que Sciensano utilise, génère et conserve sont valorisées de manière optimale pendant leur cycle de vie. Sans stratégie de données et sans gouvernance des données, celles-ci sont éparpillées, non identifiées, invisibles pour la majorité des personnes à cause de l'accès unilatéral : l'accès est difficile, aléatoire, non documenté. La stratégie de données de Sciensano doit inclure une série de décisions qui élaborent un cadre de haut niveau qui facilite l'exploitation des données d'un point de vue holistique : stockage, accès, sécurité, environnement de traitement, sauvegarde, etc. Elle doit adhérer au rôle légal de Sciensano et correspondre aux besoins de l'utilisateur : permettre l'utilisation des données dans un écosystème opérationnel et optimisé. Elle doit assurer la visibilité, l'accessibilité et la traçabilité ainsi que la qualité et la description des données

afin de garantir leur valorisation-réutilisation. La stratégie de données n'inclut pas les données relatives par exemple aux indicateurs de performance financière ou aux ressources humaines.

1.1. MOTIVATIONS POUR UNE STRATÉGIE DE DONNÉES DE SCIENSANO

Beaucoup de bonnes raisons justifient le besoin de Sciensano d'avoir une vision et une stratégie de données. Outre les arguments énoncés ci-dessus, les raisons suivantes peuvent être ajoutées :

- L'offre et la demande des données sont plus importantes que jamais auparavant, mais les citoyens / chercheurs sont confrontés à des obstacles pour accéder aux données dont ils ont besoin pour la recherche et les conseils politiques ;
- L'expertise de Sciensano repose sur la surveillance, le contrôle et la recherche en matière de santé humaine, de santé animale et d'environnement depuis une perspective One Health qui requiert la collecte et la mise en relation des données ;
- Le besoin d'informations et de conseils transparents, factuels, en temps réel, détaillés et fondés sur des données probantes en matière de santé (publique) se fait de plus en plus sentir. Afin de répondre à ce besoin d'information, Sciensano doit être capable de collecter et/ou recevoir / partager des données qui satisfassent nos finalités de production (par exemple, les données doivent être accessibles dans les meilleurs délais) ;
- De nombreux chercheurs et organisations veulent travailler avec Sciensano, donc il doit pouvoir être capable de partager ses bases de données et de relier les différentes bases de données provenant de chercheurs internes et externes ;
- Une stratégie de données efficace donne également à Sciensano de la crédibilité dans sa position d'expert au regard des données qu'il génère. Les bases de données bien gérées amènent souvent les fournisseurs de données au centre des projets / réseaux de collaboration scientifique ;
- En Belgique, les données sont éparpillées entre différentes organisations dans le système de santé, et une méthode de partage des données plus rentable est nécessaire. Les données de Sciensano devront s'intégrer dans une future structure « belge » de gouvernance des données de santé ;
- Il est de plus en plus nécessaire d'intégrer la science et l'ingénierie des données dans les organisations afin de traiter des quantités croissantes de (grands) ensembles de données ;
- Le COVID a mis au jour les problèmes liés à la gouvernance et à la gestion des données et la quantité de données a explosé en à peine un an et demi. Le COVID-19 a accéléré non seulement la santé numérique, mais aussi la sensibilité concernant le partage et la réutilisation des données.

Il existe également des tendances claires dans le monde qui nous entoure :

- La création d'un espace européen des données est l'une des priorités de la Commission pour 2019-2025, notamment dans le secteur de la santé (3). Un espace européen commun des données de santé (EEDS) permettra un meilleur accès à différents types de données de santé et facilitera leur échange (dossiers médicaux électroniques, données génomiques, données des registres des patients, etc.), non seulement pour soutenir la prestation des soins de santé (fonction première des données), mais aussi pour la recherche médicale et l'élaboration de politiques de santé (fonction secondaire des données). C'est dans le cadre de l'EEDS que la demande d'une Autorité des données de (soins de) santé (ADS) a été formulée. L'entièreté du système de données sera construite de manière transparente afin de protéger les données des citoyens et de renforcer la portabilité de leurs données de santé, comme stipulé à l'article 20 du règlement général sur la protection des données ;
- L'acte européen sur la gouvernance des données (DGA) vise à renforcer plusieurs mécanismes de partage des données et à faciliter l'accès aux données qui peuvent être utilisées pour alimenter des applications et des solutions avancées dans les domaines de l'intelligence artificielle (IA), de la médecine personnalisée, de la mobilité verte, de la fabrication intelligente et bien d'autres ;

- Il y a actuellement une tendance très marquée pour l'informatique en nuage (dans le cloud). L'UE a ses propres politiques en ce qui concerne l'informatique en nuage comme nous pouvons le lire sur [ce site](#) ;
- La science ouverte est devenue la norme pour tous les subsides et subventions aux échelles fédérale et européenne ;
- Le calcul haute performance, l'intelligence artificielle ainsi que les dernières technologies et méthodes en matière de techniques de protection de la vie privée sont pertinents pour les activités de recherche de Sciensano ;
- L'UE a accordé de nombreux financements pour la numérisation de l'Europe afin d'avoir davantage recours aux sources de données massives et aux nouvelles méthodes et techniques. Certains des sujets sont d'une grande importance pour les activités de recherche sur la santé (publique) de Sciensano et font partie de la stratégie de données et du plan de gestion de Sciensano ;
- Avec la mise en application du RGPD, de nombreuses organisations évoluent vers une analyse fédérée ou décentralisée. Cela signifie que les données ne quittent plus l'institut qui les gère, mais elles deviennent accessibles localement et peuvent être analysées localement également dans un espace sécurisé pour le chercheur. Cela requiert des développements technologiques et un changement dans les procédures de traitement des données.

1.2. INITIATIVES FAVORABLES PRÉEXISTANTES

Certaines des initiatives existantes, des projets que Sciensano coordonne ou auxquels il participe, et des situations propices pour la stratégie de données chez Sciensano sont citées ci-dessous. Le développement ultérieur de ces initiatives s'appuiera sur la vision / stratégie principale de Sciensano et servira les besoins de la partie prenante externe pour la recherche et l'élaboration de politiques de santé (publique). Cependant, Sciensano est (partiellement) dépendant de partenaires collaboratifs pour le partage ou la transmission de données et pour le développement de la recherche.

- Sciensano joue un rôle clé à l'échelle internationale en influençant le futur de l'Europe au regard de la fonction secondaire des données de santé par des activités de l'unité Système d'information sanitaire de l'UE. Sciensano est un partenaire actif dans [l'espace européen des données de santé](#), [le nuage européen pour la science ouverte](#) et [le portail de données pour la recherche et l'innovation dans le domaine de la santé](#). Sciensano facilite l'échange d'informations sur la santé à travers l'Europe en dirigeant et en participant à sept grands projets européens dans le domaine de l'information sur la santé. Sciensano a construit son expertise dans les mécanismes européens de gouvernance des données et a créé un vaste réseau à travers l'Europe par le biais de projets tels que l'Action commune sur l'information sanitaire ([InfAct](#)) (2018-2021), l'Infrastructure de recherche sur les informations de santé de la population ([PHIRI](#)), le portail européen d'information sur la santé, le DIPoH (Infrastructure décentralisée sur la santé de la population) et l'action conjointe Vers l'espace européen des données de santé ([TEHDAS](#)), qui soutient la mise en place de [l'espace européen des données de santé](#) (EEDS).
- Le projet HealthyCloud étudie la faisabilité du partage de données de santé à caractère personnel entre différents pays européens à des fins de recherche et identifie les défis ainsi que les solutions pour les relever. Il nous apporte une connaissance approfondie des principes FAIR afin d'identifier le meilleur moyen d'évaluer le degré FAIR d'une infrastructure de données ;
- Le projet EGI-ACE vise à permettre aux chercheurs de toutes les disciplines de collaborer à des recherches nécessitant beaucoup de données et de calculs grâce à des services gratuits à l'usage. Le projet EGI-ACE collabore étroitement avec le cloud européen pour la science ouverte (EOSC) afin de proposer des installations de stockage, des services techniques, des outils d'analyse ainsi qu'une assistance ;

- Le projet BY-COVID vise à mobiliser et à connecter les ressources de données bien établies conformément aux principes FAIR et à donner un accès à des données hétérogènes mais interconnectées et organisées dans les domaines scientifique, médical, de la santé publique et de la politique. Cela se fera par le biais du portail de données sur le COVID-19 tout en collaborant avec le cloud européen pour la science ouverte (EOSC) d'Horizon Europe et avec l'espace européen des données de santé (EEDS) ;
- Le Réseau belge de surveillance aérobiologique (AirAllergy), géré par Sciensano, a accepté une coopération européenne avec plusieurs réseaux nationaux aérobiologiques afin de partager les données issues de la recherche scientifique sur la présence d'allergènes dans l'air. Historiquement, la Belgique est l'un des membres fondateurs notables de cet accord qui a donné naissance au Réseau européen des aéroallergènes (EAN) en 1987 ;
- Sciensano prend part au consortium européen sur l'aérobiologie (EUMETNET AutoPollen), qui vise à définir dès le départ les normes principales de la nouvelle ère de la mesure automatique en temps réel des bioaérosols ;
- Le consortium BelCoVac : dans le contexte des vaccins contre le COVID-19 qui sont actuellement mis en vente ou le seront bientôt en Belgique, les institutions du consortium souhaitent participer à un programme du consortium en coopération avec et sous les auspices de Sciensano, avec comme objectif d'organiser plusieurs essais cliniques non commerciaux et des projets de recherche complémentaires en ce qui concerne lesdits vaccins contre le COVID-19 afin d'obtenir des informations et des données qui pourraient être utiles pour les patients, pour la pratique clinique et/ou pour les décideurs politiques ;
- Sciensano travaille à la consolidation de son Système actuel de gestion de l'information des laboratoires (SGIL) et à plusieurs projets visant à améliorer l'échange des données du SGIL avec des partenaires et réseaux externes et ainsi rendre ces données plus facilement accessibles aux utilisateurs internes ;
- Sciensano est un acteur central dans le paysage belge de l'information sur la santé :
 1. Nombreuses enquêtes, notamment l'enquête de santé belge par interview, l'enquête de consommation alimentaire, etc.
 2. Suivi actif de la pandémie du COVID-19 grâce à des sources de données existantes et nouvelles
 3. Coordination du Point focal de collecte de données internationales
 4. Plateforme Healthdata.be pour la facilitation technique des projets scientifiques de santé et soins de santé (plus d'informations ci-dessous)
 5. Gestion du projet HERA-BE-Incubator afin d'améliorer l'infrastructure nationale pour l'échange de données génomiques et épidémiologiques en Belgique
- Sciensano entend clairement soutenir la valorisation scientifique du paysage belge de l'information sur la santé :
 1. BELCOHORT a exploré activement les possibilités de mettre en place une cohorte en Belgique et a formulé une série de recommandations conjointement avec les parties impliquées.
 2. AHEAD améliorera de façon proactive la visibilité du paysage belge de l'information sur la santé et explorera les conditions techniques, légales et éthiques afin d'être capable de travailler en vue du développement d'une plateforme nationale de données sur la santé pour faciliter l'accès à différentes sources de données et pour étudier des questions plus complexes (par exemple la relation entre la santé et les indicateurs socioéconomiques).
 3. Sciensano joue un rôle crucial dans la création d'une Autorité des données de soins de santé (ADS).

- Sciensano se concentre sur la création de liens entre différentes sources :
 1. Le projet HISLink étudie la qualité des données de soins de santé en mettant en relation l'Enquête de santé par interview (HIS) et les données de soins de santé de l'Agence inter-mutualiste (AIM).
 2. HELICON étudie les inégalités sociales face au COVID-19 et l'impact à long terme et indirect de la crise du coronavirus en faisant des liens entre les données COVID de Sciensano et les données démographiques de Statbel, ainsi que les données de soins de santé de l'AIM.
 3. TDI-IMA étudie l'utilisation des soins de santé et des médicaments chez les personnes suivant un traitement pour usage de drogue en mettant en relation le registre d'indicateur de demande de traitement (TDI) et les données de soins de santé de l'AIM.
 4. Le projet SPADIS étudiera l'emploi et la participation sociale des personnes atteintes de maladies chroniques. Pour ce faire, le projet reliera des enquêtes et registres de Sciensano préexistants à la Banque-Carrefour et les données de soins de santé de l'AIM. Une étude similaire des patients atteints d'un cancer (Fondation Registre du Cancer) a déjà été menée.
 5. Le projet LINK-VACC met en relation la base de données des vaccinations COVID-19 réalisées (VACCINNET+) avec les données des hôpitaux, les données d'analyse, ainsi qu'avec les sources de données externes telles que CoBRHA.

- Sciensano soutient et dirige plusieurs initiatives visant à rendre les données Faciles à trouver et Accessibles (le F et le A des principes FAIR) :
 1. Healthdata.be gère le portail fair.healthdata.be, qui offre un aperçu des données relatives à la santé et aux soins de santé. Ce projet est né du point d'action 18 du plan eSanté 2013-2018 : un inventaire de toutes les données, disponibles via les registres et la surveillance, relatives à la santé et aux soins de santé en Belgique. Selon la source de données, il peut y avoir des informations sur l'objectif, la licence, les variables, etc. Outre la possibilité de rechercher des données, ce portail offre également une API pour intégrer les informations disponibles dans d'autres systèmes. Ajouter et entretenir les informations relève de la responsabilité du propriétaire des données.
 2. <https://data.gov.be> fait partie de la stratégie fédérale de données ouvertes, qui a pour but l'ouverture des données publiques et qui entend les rendre accessibles à la réutilisation par les citoyens, les chercheurs, les entreprises et les gouvernements. Le site Internet lui-même réutilise les métadonnées d'une série de portails liés à des domaines spécifiques. Pour les données relatives à la santé, les informations sont pour le moment relativement limitées (notamment les sources de données sur le COVID-19 via fair.healthdata.be, les sources de données de Statbel, etc.).

2. Visualisation des données de Sciensano

Une stratégie de données devrait être précédée d'une visualisation des données. Ci-dessous se trouve une brève visualisation des données de Sciensano. Elle se compose des cadres légaux au sein desquels Sciensano doit fonctionner et de sa vision du futur pour la gouvernance des données. Ces données peuvent être la propriété de Sciensano ou d'autres parties et servent à remplir le rôle de Sciensano d'institut scientifique en menant à bien sa mission (4). La stratégie de données concerne *toutes* les données présentes chez Sciensano (y compris les données sans lien avec la santé).

En exécutant ses tâches, Sciensano voit dans le XXI^e siècle l'émergence de (nouvelles et grandes) sources de données et le développement rapide de nouveaux outils, méthodes et technologies tels que l'IA pour utiliser tant les sources de données préexistantes (enquête, registre et données de recensement) que les nouveaux types (données massives, données du monde réel, données ouvertes). Cela implique inévitablement que ces sources de données nouvelles et préexistantes sont disséminées dans différents systèmes de (soins de) santé et dans différentes organisations et différents pays. En parallèle, les décideurs politiques, les chercheurs, les entrepreneurs et le grand public veulent obtenir des informations de haute qualité sur les questions de santé (publique) et donc de grandes initiatives sont en train d'être lancées (comme l'EEDS). Les citoyens ont besoin d'informations transparentes et

factuelles, les gouvernements et les entreprises cherchent des informations à jour, précises et détaillées pour qu'ils puissent définir leurs politiques et leur gouvernance en s'appuyant sur des données probantes, et les scientifiques utilisent ces données comme fondement de leur recherche.

Traditionnellement, Sciensano s'acquitte de ses tâches selon deux axes. Premièrement, Sciensano surveille la santé par ses réseaux, plateformes et laboratoires de contrôle et de surveillance, et il donne des conseils aux décideurs politiques sur les questions de santé (publique). Deuxièmement, Sciensano entreprend des recherches afin de donner des conseils sur la santé (publique) fondés sur des données probantes. Afin de permettre à Sciensano de prodiguer ces conseils et de mener sa recherche, les informations sanitaires sont collectées au travers d'enquêtes, de systèmes de surveillance et de l'accès à certains registres. Obtenir l'accès aux données constitue souvent une activité complexe et chronophage, à cause de l'absence de procédures transparentes et standardisées, où il n'existe aucune unité de saisie, de validation et de collecte des données. La réutilisation des données de (soins de) santé collectées par Sciensano et d'autres agences gouvernementales est sans aucun doute une méthode de collecte des données qui devrait être davantage priorisée, structurée et contrôlée. Il convient de faire pression pour mettre en place des changements législatifs lois en termes d'accès aux données et de protection, d'administration des données et de séparation des rôles dans la gestion des données. Nous avons aussi beaucoup à y gagner si le contrôle des données des hôpitaux, des laboratoires et des médecins généralistes pouvait être un réseau structuré de transmission des données qui soit adapté aux conditions de production de Sciensano (en ce qui concerne la charge des fournisseurs de données, par exemple la surveillance syndromique). Enfin, les rôles des différents détenteurs belges de données de santé (healthdata.be, Statbel, Sciensano, Inami, etc.) devraient être spécifiés. La création d'une Autorité belge des données de soins de santé (ADS) offrirait l'opportunité idéale à cet égard.

En cette nouvelle époque, Sciensano souhaite (re)définir son rôle et sa position dans une société de plus en plus orientée sur les données et la technologie au regard de la réglementation déontologique et éthique (voir les objectifs stratégiques du plan de gestion de Sciensano). Dans son plan de gestion, Sciensano a exprimé son ambition non seulement d'accroître son influence sur les décideurs politiques pour la santé en utilisant les données (ouvertes) disponibles, mais aussi d'adopter une Politique de science ouverte et de traiter les données suivant les principes FAIR (Facile à trouver, Accessible, Interopérable, Réutilisable). Le respect des principes FAIR commence dès le stade de collecte des données, et les principes FAIR sont inclus dans les accords et contrats avec les fournisseurs de données (Les données peuvent-elles être réutilisées ? Les données peuvent-elles être ouvertes ?) et les collaborateurs. Avec l'essor de l'échange de données, Sciensano souhaite faciliter la visibilité et l'accessibilité non seulement de ses propres données, mais aussi des données de Sciensano en combinaison avec les données d'autres départements / organisations. C'est la capacité à combiner et à relier des données qui est fortement demandée et qui crée la valeur ajoutée des données de Sciensano pour la recherche et la prise de décisions. En parallèle, Sciensano entend rendre ses propres procédures plus efficaces et transparentes, comme indiqué dans son plan de gestion.

Cette section présente la visualisation des données de Sciensano, qui décrit plus en détail le rôle que Sciensano souhaite jouer dans un écosystème de données national et international.

2.1. CADRES LÉGAUX

Sciensano a été créé par la loi du 25 février 2018. Parmi les tâches juridiques conférées par cette loi se trouve le traitement des données tel que décrit à l'article 4, § 4 et § 4/1 de la loi :

« § 4. Sciensano assure le traitement, en ce compris la collecte, la validation, l'analyse, le rapportage et l'archivage, des données à caractère personnel notamment relatives à la santé publique ou en un lien avec la santé et d'autres informations scientifiques relatives à la politique de santé, dans le respect des lois applicables en la matière. À cette fin, Sciensano

réalise des analyses scientifiques quantitatives et qualitatives sur base des informations traitées en vue de soutenir la politique de santé. Sciensano peut également mettre des données et des informations traitées à disposition, moyennant les autorisations des comités sectoriels compétents.

[1 § 4/1. Dans le cadre de la gestion des crises touchant la santé publique, Sciensano a pour mission de coordonner et d'implémenter les aspects scientifiques qui y sont liés, de surveiller et d'évaluer les risques au moyen d'analyses spécifiques des données collectées, de fournir des avis et recommandations aux différentes autorités de santé du pays et d'organiser la communication au profit des autorités, des prestataires des soins de santé et du public.] »

En tant qu'établissement de droit public, Sciensano est sujet à diverses lois concernant les données :

- Proposition de RÈGLEMENT DU PARLEMENT EUROPÉEN ET DU CONSEIL sur la gouvernance européenne des données (acte sur la gouvernance des données) :
- Règlement (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016 relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, et abrogeant la directive 95/46/CE (règlement général sur la protection des données) :
- Directive (UE) 2019/1024 du Parlement européen et du Conseil du 20 juin 2019 concernant les données ouvertes et la réutilisation des informations du secteur public. Cette directive est déjà entrée en vigueur et doit être transposée dans la loi nationale (belge) pour le 17/07/2021 au plus tard. Cette échéance n'a pas été respectée.
- 4 mai 2016 – Loi relative à la réutilisation des informations du secteur public
- 5 août 2006 – Loi relative à l'accès du public à l'information en matière d'environnement
- 11 avril 1994 – Loi relative à la publicité de l'administration
- 2 juin 2019 – Arrêté royal relatif à la réutilisation des informations du secteur public

Sciensano relève du champ d'application de ces deux lois qui s'appliquent cumulativement : la loi relative à la publicité de l'administration (11/04/1994) et la loi relative à la réutilisation des informations du secteur public (dites « données ouvertes » (04/05/2016)). Le champ d'application de la loi relative à la publicité de l'administration est très large. Cette loi a été rédigée afin de garantir la transparence des informations détenues par les autorités publiques. Elle stipule un droit à l'information. Les exceptions qui peuvent être invoquées pour refuser une demande d'accès à l'information sont listées de manière exhaustive dans la loi. Il est à noter que quand la loi a été rédigée, il n'existait aucun institut de recherche ayant un statut similaire à celui de Sciensano.

Le champ d'application de la loi relative à la réutilisation des informations du secteur public est différent. En effet, cette loi se concentre sur la réutilisation des informations du secteur public dans un but autre que celui pour lequel les informations avaient été collectées. Les exceptions légales stipulées dans cet acte différent de celles qui existent pour la loi relative à l'accès du public aux informations du gouvernement. L'application de ces deux lois est désavantageuse pour Sciensano. Il suffirait d'invoquer une seule de ces deux lois pour demander des informations. Un avant-projet de loi est actuellement en préparation pour remplacer la loi relative à la réutilisation des informations du secteur public. L'objectif est d'avoir une meilleure protection des données de la recherche. Malheureusement, cette proposition de loi ne permet pas (à ce stade) aux citoyens de se passer d'invoquer la loi de 1994.

La situation légale actuelle est un malheureux concours de circonstances et de nombreuses lois requièrent une mise à jour : la loi relative aux droits du patient ; la loi relative à la qualité de la pratique des soins de santé (les dossiers médicaux personnels devraient être accessibles à tous les prestataires de soins) ; l'accord de coopération entre les régions concernant le partage des données ; la loi sur la protection de la vie privée. Il existe également plusieurs lois à paraître : la loi relative à l'analyse des données (la quantité des données de santé, leur organisation, la procédure, les objectifs pour le futur)

et l'acte sur la gouvernance des données, qui mentionne le besoin de créer un comité de sécurité de l'information.

2.2. UTILISER LES DONNÉES INTERNES ET EXTERNES À SCIENSANO

Les sources de données internes et externes à Sciensano

Sciensano est conscient de l'importance de mettre en relation les données (de soins de santé) préexistantes et les nouvelles sources de données émergentes (massives et ouvertes) avec la mise au point rapide de nouveaux outils, méthodes et technologies pour créer une plus-value pour la recherche sur les soins de santé et pour les résultats pertinents pour la politique. Les données que possède Sciensano sont variées. La stratégie de collecte des données concerne *toutes* les données présentes chez Sciensano (y compris les données sans lien avec la santé). Ci-dessous se trouve un aperçu non exhaustif des sources et systèmes de données qui rendent les données accessibles :

- [Healthdata.be](https://healthdata.be)
- Enquêtes : voir ce [lien](#) pour un aperçu des projets qui collectent les données d'enquête
- Le contrôle et la surveillance du COVID via l'utilisation de données internes et externes à Sciensano :
 1. Données de santé et tests PCR / antigéniques
 2. Collecte de données sur la surveillance de la capacité de pointe des hôpitaux et sur la surveillance clinique
 3. Collecte de données sur le nombre de cas de COVID-19 et sur le nombre de décès dans les maisons de repos
 4. Surveillance par les médecins généralistes
 5. Surveillance de l'absentéisme au travail
 6. Traçage des contacts via l'IFC Testing & Tracing
 7. Détection des clusters à l'aide de données provenant de plusieurs instances
 8. Formulaire de localisation du passager pour les voyageurs revenant de l'étranger
 9. Surveillance des administrations du vaccin et du pourcentage de vaccination
 10. Surveillance de l'efficacité et de la sécurité du vaccin
 11. Surveillance de la production d'anticorps contre le SARS-CoV-2 (études de séroprévalence parmi différents groupes : donneurs de sang, enfants, etc.)
- Autres systèmes de données utilisés par Sciensano :
 1. Système de gestion de l'information des laboratoires (SGIL)
 2. [REDCap](#) (Research Electronic Data Capture)
 3. [SLIMS](#) : système de gestion de l'information des laboratoires (SGIL) combiné avec un carnet de laboratoire électronique (ELN).
 4. [BIOTECH platform](#)
 5. [Tiquesnet](#)
 6. [ORPHANET](#)
 7. Bases de données dotées d'informations sur les produits chimiques : par exemple le projet SCANNER sur les nouvelles substances psychoactives
 8. [EAN](#)
- Données ouvertes :
 1. <https://epistat.wiv-isp.be/covid/covid-19.html>
 2. <https://fair.healthdata.be/sources/covid19>
- Bases de données interactives :
 1. <https://www.belgiqueenbonnesante.be/fr/etat-de-sante/a-propos-du-health-status-report/bases-de-donnees-interactives-sciensano>
 2. <https://hisia.wiv-isp.be/SitePages/Home.aspx>
 3. <https://www.sciensano.be/fr/node/68347/spmah>

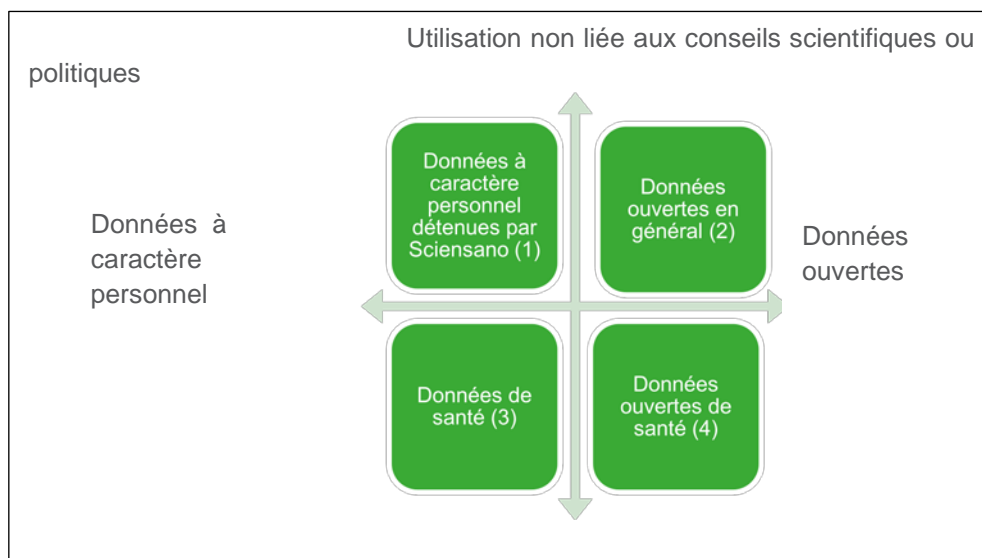
En plus de ses propres données, Sciensano a accès à d'autres données (ouvertes) et les utilise :

- Données sur la biodiversité
- NEVO
- NUBEL
- Milieu alimentaire : Données de Locatus
- Données sur la pollution de l'air
- Données sur la couverture des terres et données de télédétection
- Institut géographique national belge
- Référentiels de données géographiques régionales :
 1. <https://geoportail.wallonie.be/home.html>
 2. <https://environnement.brussels/> et <https://geobru.irisnet.be/>
- <https://www.geopunt.be>
- Données météorologiques

La liste ci-dessus est non exhaustive et fera partie d'une stratégie de données afin d'avoir un inventaire de toutes les données chez Sciensano.

Le rôle que Sciensano souhaite jouer dans un écosystème de données national et international

La figure ci-dessous présente un premier aperçu et une analyse des différents types d'activités que Sciensano réalise en utilisant les données et présente les différents types de données disponibles pour ces activités. Selon la loi, Sciensano n'a pas le droit d'utiliser des microdonnées à caractère personnel à des fins qui ne sont pas liées aux conseils scientifiques ou politiques ; toutes les activités des données qui ne sont **pas** couvertes par le cadre légal dans lequel fonctionne Sciensano relèvent du **1^{er} quadrant**.



Les données citées dans le **2^e quadrant** ne sont pas fournies par Sciensano, mais par d'autres acteurs comme le gouvernement ou les entreprises privées. Les données sont rendues disponibles par d'autres acteurs que Sciensano en tant que données ouvertes et peuvent contenir des données quantitatives, mais aussi d'autres données, par exemple sur les cartes topographiques, etc. Elles peuvent aussi contenir des données sur des entreprises individuelles, sur des personnes ou sur des institutions.

Les activités de Sciensano sont actuellement situées dans les deux quadrants inférieurs. Le résultat des rapports de politique et de recherche de Sciensano (informations sanitaires, conseils et recherche scientifique) est dans le **4^e quadrant**, où les données ouvertes sont dérivées des microdonnées du

3^e quadrant. Les microdonnées (ou parfois données brutes)¹ utilisées pour la recherche et les conseils politiques fondés sur la science se trouvent dans le 3^e quadrant. Une terminologie assez courante de l'utilisation des données est la suivante :

- Usage sécurisé : il s'agit de microdonnées confidentielles qui peuvent être utilisées au sein de Sciensano.
- Usage scientifique : il est possible d'offrir des données à des acteurs externes de manière sécurisée et contrôlée. Il doit être précisé si les données « quittent » l'organisation ou si elles restent en son sein.
- Usage public : il s'agit de données ouvertes, sécurisées et non confidentielles.

Les usages sécurisé et scientifique appartiennent au 3^e quadrant, alors que l'usage public appartient au 4^e quadrant.

Ci-dessous, vous trouverez la description de l'ambition de Sciensano concernant les données et leur utilisation pour chaque quadrant. Les données du **4^e quadrant** sont agrégées et non rattachables à un individu identifiable (dans le cas d'entreprises ou d'institutions, le RGPD ne s'applique pas, à moins que les personnes physiques qui y travaillent ne soient identifiables, mais d'autres lois économiques s'appliquent) et sont rendues accessibles à l'utilisation en tant que données ouvertes. Le contenu de ce quadrant est fourni par Sciensano. Ce contenu consiste en des publications sous forme de rapports, d'outils interactifs ou de tableaux de bord des résultats de l'analyse d'enquêtes, de registres ou de données ouvertes en tant que produits finaux de conseils personnalisés ou de recherches scientifiques pour des parties externes. Les personnes utilisant les données de ce quadrant (décideurs politiques, citoyens, chercheurs) ont parfois (rapidement) besoin d'informations sur une grande variété de sujets en lien avec la santé. L'ambition de Sciensano est de satisfaire ce besoin autant que possible et d'avancer dans cette direction à partir de ses services actuels afin de s'attaquer encore mieux aux questions politiques actuelles en rendant les données largement accessibles.

Le **3^e quadrant**, celui des données de santé, contient des microdonnées qui peuvent être utilisées à des fins consultatives ou scientifiques et sont dès lors accessibles aux scientifiques internes et externes (à Sciensano) dans de strictes conditions. Nous distinguons plusieurs types d'utilisation interne des microdonnées :

- Pour décrire la santé de la population : à ce titre, la collecte des données doit respecter certains aspects tels qu'un échantillon représentatif de la population, c'est le but principal du monitoring et de la surveillance de la santé ;
- Pour examiner les associations entre les risques et l'état de santé : étant donné que la majorité des collectes de données sont transversales, aucune conclusion sur l'interférence causale ne peut être tirée. Les projets décrits ci-dessus (BELCOHORT, AHEAD, etc.) sont des étapes vers une collecte des données et une recherche longitudinales pour permettre une interférence causale ;
- Pour étayer les conseils en matière de santé publique : par exemple l'intégration de facteurs déterminants distaux comme facteurs habilitants ou restrictifs, pour examiner l'impact tout au long de la vie du style de vie et des maladies sur la santé et/ou la participation sociale, pour estimer l'impact de mesures politiques potentielles sur les conséquences en matière de santé ;
- Pour prédire les tendances futures de la santé et des maladies chez les humains et les animaux, etc.

L'ambition de Sciensano est de rendre les données faciles à trouver et accessibles en aidant les chercheurs à mettre en relation les données ouvertes de Sciensano (4^e quadrant) avec les données

¹Les microdonnées sont une série de dossiers contenant des informations sur les individus, les ménages ou les entreprises. Les données brutes (parfois appelées données de base, données atomiques ou données primaires) sont des données qui n'ont pas été traitées pour usage.

ouvertes du 3^e quadrant. Ici, Sciensano peut appliquer sa connaissance de l'infrastructure des données, du couplage des données, de la gestion des métadonnées et de la protection de la vie privée. De cette façon, Sciensano contribuerait à un paysage belge non fragmenté à partir de données bien décrites et utilisables qui ne révèlent aucune information à caractère personnel, en combinant les sources de données de différentes administrations de la santé.

COMPOSANTES D'UNE STRATÉGIE DE DONNÉES POUR SCIENSANO

Pour les différentes composantes de la stratégie de données, nous suivons les subdivisions de l'évaluation de la maturité des données (1). Selon l'approche proposée, quatre types différents d'interopérabilité sont utilisés comme fondement de l'évaluation de maturité, à savoir :

- Organisationnelle : l'interopérabilité organisationnelle peut faire référence à la manière dont les responsabilités de gouvernance et de gestion des données qui touchent à l'interopérabilité sont réparties au sein d'une organisation. Elle peut aussi être définie comme les méthodes avec lesquelles des organisations entières, ou des départements individuels en leur sein, collaborent avec l'écosystème de données plus large pour décider du degré d'interopérabilité qu'elles aimeraient atteindre entre leurs ressources de données collectives.
- Humaine : l'interopérabilité humaine peut faire référence au besoin de s'assurer qu'alors qu'on rend les données interopérables dans les systèmes de données, elles restent tout de même lisibles et utilisables par les utilisateurs humains. Cela fait aussi référence à la capacité des individus, des groupes, des équipes et des départements d'individus au sein d'organisations de communiquer et de collaborer de manière à encourager l'interopérabilité dans l'ensemble de leurs flux de travail et de leurs ressources de données.
- Des données : l'interopérabilité des données est le besoin de s'assurer que les systèmes de données et les ensembles de données sont conçus pour permettre l'interopérabilité en s'assurant de pouvoir collecter, stocker et traiter les données dans des formats lisibles à la machine et de réfléchir aux besoins d'interopérabilité au moment de modéliser et de classer les données.
- Technologique : l'interopérabilité technologique nécessite une infrastructure de données adéquate pour permettre aux trois interopérabilités précitées de fonctionner de manière efficace.

Chacune des dimensions de l'interopérabilité inclut plusieurs indicateurs de maturité et un niveau de maturité est assigné à chacun de ces indicateurs. Les niveaux de maturité possibles sont listés ci-après dans l'ordre croissant : Indéterminé, Émergence, Apprentissage, Construction et Consolidation. Les niveaux de maturité sont bien définis et l'objectif est d'identifier à quel niveau Sciensano se situe.

Pour l'évaluation, un groupe de travail interne a été créé au sein de Sciensano avec des représentants de chaque conseil d'administration scientifique. Chaque indicateur de maturité a été abordé, et le niveau de maturité de Sciensano a été déterminé. Le même exercice a été répété afin d'arriver à un accord sur le niveau de maturité que Sciensano souhaite atteindre au cours des cinq prochaines années. Le résumé et l'explication des niveaux de maturité de Sciensano obtenus par l'évaluation de la maturité des données actuelle et future sont disponibles en Annexe 1 et 2. Chez Sciensano, healthdata.be est pris en charge par le conseil d'administration des risques biologiques pour la santé². La maturité des

² La plateforme healthdata.be, développée par Sciensano (anciennement l'ISP) et financée par l'Institut national d'assurance maladie-invalidité (INAMI), offre de nouvelles perspectives sur l'eSanté en simplifiant l'enregistrement et le stockage des données de santé envoyées par les différents prestataires de soins de santé. Healthdata.be assure une recherche médicale de meilleure qualité. En 2006, le Centre fédéral d'expertise des soins de santé (KCE) avait accès à près de 50 bases de données sur la santé, auxquelles la communauté scientifique avait accès sur le site Internet du KCE. L'objectif à l'époque était de donner aux chercheurs un accès rapide aux informations dont ils avaient besoin pour leur recherche dans divers domaines de la santé publique, des soins de santé et de l'épidémiologie. En unissant leurs forces dans le cadre d'un nouveau projet d'inventaire, le KCE et Sciensano sont parvenus à créer 150 bases de données sur la santé. Healthdata.be est gouverné par le comité de pilotage de la plateforme healthdata, la commission de la protection

données de healthdata.be est en moyenne plus élevée que celle des données du reste de Sciensano (appelé Sciensano dans ce texte) ; pour les besoins de cet exercice, la distinction avec healthdata.be sera faite par moments.

1. Interopérabilité organisationnelle

1.1. OBJECTIFS STRATÉGIQUES

« La gestion des données peut être définie comme le développement, la mise en place et le suivi de stratégies et de plans qui permettent d'exploiter la valeur des données de manière sécurisée... Pour les organisations qui gèrent, collectent ou traitent divers ensembles de données, l'interopérabilité doit être un principe fondamental dans leur stratégie de gestion des données... Au moment de fixer des objectifs stratégiques en lien avec l'interopérabilité, il est important de leur allouer des ressources budgétaires et humaines pour soutenir les plans de mise en œuvre... » (1).

Healthdata.be est pris en charge par le conseil d'administration des risques biologiques pour la santé ; les données du reste de Sciensano sont gérées par les conseils d'administration scientifiques soutenus par des technologies de l'information et par le Système de gestion de l'information des laboratoires (SGIL). Healthdata.be atteint de hauts niveaux d'interopérabilité et une collaboration étroite entre les conseils d'administration scientifiques et fait partie de la stratégie de données de Sciensano. L'ambition de Sciensano est d'adhérer aux principes FAIR.

Pour Sciensano dans son ensemble, au mois d'avril 2022, il n'y avait aucun objectif stratégique lié à l'interopérabilité et aucune ressource budgétaire ou humaine n'y était allouée. Par conséquent, en termes d'objectifs stratégiques, pour healthdata.be, les technologies de l'information et le SGIL, nous estimons que Sciensano est au niveau « Apprentissage ». Pour les données externes à ces systèmes, Sciensano est au niveau « Émergence ».

1.2. LEADERSHIP ET GESTION

« Le leadership est crucial pour la gestion des données et pour s'assurer que l'interopérabilité est considérée comme étant un objectif stratégique. Afin que les organisations puissent exploiter le potentiel de leurs données, les dirigeants doivent s'engager dans des processus stratégiques liés à l'interopérabilité des données ainsi que comprendre et être conscients de la valeur que l'interopérabilité apporte à leurs organisations... » (1).

La crise du COVID a sensibilisé les dirigeants à la valeur de l'interopérabilité des données, à la valeur des données et au besoin d'un plan de gestion des données. En termes de leadership et de gestion, Sciensano se situe entre les niveaux « Émergence » et « Apprentissage ».

1.3. SUPERVISION ET RESPONSABILITÉ

« Les structures de supervision et de responsabilité sont fondamentales pour la gouvernance des données et pour les rôles de gestion liés à la définition d'objectifs stratégiques, au leadership, au management et à l'administration des données. L'intégration d'une surveillance et d'une responsabilité efficaces dans les processus organisationnels permet de garantir la gestion et l'entretien corrects des processus visant à rendre les données interopérables (et donc intégrables, partageables et accessibles à d'autres). Il doit également y avoir une supervision des ensembles de données susceptibles de contenir des données qui pourraient révéler l'identité d'un individu ou toute autre caractéristique sensible sur des individus ou des groupes vulnérables » (1).

de la vie privée, le groupe de travail pour l'architecture de la plateforme eSanté et le comité d'orientation interne de healthdata.be.

Les structures de supervision et de responsabilité telles que le comité de pilotage, les procédures visant à rendre les données interopérables et l'administration des données sont en place sur healthdata.be afin d'assurer la gouvernance des données et la supervision des données qui contiennent des informations sensibles. Pour les données de Sciensano, un inventaire des données a été entrepris dans le cadre de la stratégie de données. Cet inventaire révélera où se trouvent les données, qui en est responsable, qui en est le propriétaire (s'il y en a un) ou le point de contact et si des métadonnées sont disponibles. Des initiatives sont prises dans différentes unités afin de permettre la supervision. Les délégués à la protection des données (DPD) et les coordinateurs de sécurité des informations veillent à la sécurisation des données sensibles.

En termes de supervision et de responsabilité, le niveau de Sciensano se trouve entre « Émergence » et « Apprentissage ».

1.4. RESPECT DE LA LÉGISLATION

« L'interopérabilité peut se heurter à des problèmes juridiques et réglementaires quand il s'agit de partager ou d'intégrer des données entre organisations et au-delà des frontières nationales. Les lois fixent des limites applicables à ce qui constitue une conduite acceptable et à ce qui ne l'est pas. Dans certains cas, elles régissent la manière dont les données peuvent être partagées (par exemple, les lois qui réglementent et fixent les normes pour la communication, la sécurité et la protection des données). Dans d'autres cas, elles définissent quelles sont les données qui peuvent, et surtout celles qui ne peuvent pas, être partagées et intégrées (par exemple, les lois sur la protection des données et sur la vie privée)... » (1).

L'article 4, § 4 de la loi portant création de Sciensano constitue la base juridique grâce à laquelle Sciensano est autorisé à traiter les données à caractère personnel. Sciensano doit par la présente respecter les lois applicables, y compris le principe de traitement minimal des données stipulé dans le RGPD et le titre 4 de la loi du 30 juillet 2018 relative à la protection des personnes physiques à l'égard des traitements de données à caractère personnel. Sciensano peut utiliser des données pseudonymisées seulement quand il est impossible de parvenir à l'objectif scientifique ou statistique avec le traitement de données anonymes ; il peut seulement utiliser des données non pseudonymisées s'il est impossible de parvenir à l'objectif de recherche avec des données pseudonymisées. Le traitement de données non pseudonymisées doit donc toujours être justifiable.

L'article 4, § 4/1 de la loi portant création de Sciensano concerne expressément le rôle de Sciensano (et tout traitement de données à caractère personnel lui étant lié) dans la gestion d'une crise de santé publique. La loi du 4 mai 2016 relative à la réutilisation des informations du secteur public et l'arrêté royal du 2 février 2019 réglementent la réutilisation des informations du secteur public et expliquent la stratégie fédérale de données ouvertes. Sciensano est conforme au RGPD et développe aussi des outils logiciels à cet effet.

Sciensano possède un département juridique qui s'assure que Sciensano respecte les lois auxquelles il est soumis. En termes de respect de la législation, le niveau de Sciensano se trouve entre « Apprentissage » et « Construction ».

1.5. ÉTHIQUE DES DONNÉES

« L'interopérabilité peut soulever des dilemmes éthiques. Ceux-ci diffèrent des problèmes juridiques dans la mesure où ils sont susceptibles de porter sur des questions d'équité ou sur ce qui est considéré comme bien ou mal dans un contexte précis, et non sur ce qui est légal ou illégal. Un problème éthique lié à l'interopérabilité pourrait être de savoir si les entreprises internationales de traitement de données qui obtiennent l'accès à des microdonnées démographiques nationales dans le cadre d'un programme de développement devraient être autorisées à réutiliser ces données et à en tirer profit à l'avenir en les intégrant à d'autres ensembles de données. Selon les lois nationales d'une certaine juridiction, cette

situation pourrait être légale, mais elle soulève tout de même des questions éthiques à propos du partage équitable de la valeur extraite des données nationales... » (1).

Plusieurs actions sont menées chez Sciensano pour s'occuper des questions éthiques concernant la gestion et le traitement des données (à caractère personnel). Dans le cadre de l'action conjointe « Vers l'espace européen des données de santé » (TEHDAS) et du projet Belspo « Vers le développement d'une plateforme nationale de données sur la santé » (AHEAD), une consultation en ligne a été lancée afin d'informer les citoyens des questions éthiques liées à la réutilisation des données de santé et de leur offrir l'occasion de créer de manière conjointe un cadre éthique et sociétal pour les espaces de données sur la santé. Dans le cadre de BY-COVID, un projet d'Horizon Europe, Sciensano dirige l'initiative de participation citoyenne visant à informer les citoyens sur l'utilisation des données de santé pendant la pandémie et sur les plateformes de données de santé qui sont créées pour mieux se préparer à la pandémie.

Le niveau de maturité de Sciensano sur l'éthique des données est « Apprentissage ».

1.6. APPROVISIONNEMENT

« L'approvisionnement en services informatiques, en systèmes de données ou en autres solutions liées au traitement, au stockage ou au partage de données soulève, pour les organisations, d'importantes questions concernant l'interopérabilité... L'enfermement propriétaire apparaît quand des organisations achètent séparément différentes solutions de données et solutions numériques privées dans le cadre de différents programmes de travail ou de projets parrainés par des donateurs sans prendre en considération les architectures de données et l'infrastructure numérique préexistantes. Cela peut mener au cloisonnement des systèmes d'informations, ce qui produit des systèmes ou des ensembles de données qui n'interopèrent pas ou dont les données résultantes ne peuvent être intégrées les unes dans les autres » (1).

Chez Sciensano, les actions scientifiques sont soutenues par une infrastructure informatique avec des installations sur site, un réseau sécurisé, des calculs haute performance et des solutions de stockage judicieusement choisies. En plus de cette infrastructure, une infrastructure virtuelle (dite infrastructure de bureau virtuel ou IBV) est hébergée. Le département informatique peut étendre sa structure avec une capacité de mémoire ou une puissance de calcul supplémentaires selon les besoins des unités individuelles et les exigences de performance, sans être limité par des conditions matérielles. De plus, la plateforme IBV est sécurisée et accessible uniquement aux personnes autorisées et permet une analyse intensive et performante.

L'approvisionnement se fait selon deux objectifs différents. Le premier objectif concerne les besoins communs en matière d'infrastructure informatique et de logiciels. Tous deux font l'objet d'un suivi, et le département informatique prévoit chaque année la croissance, l'évolution et les expansions. Le deuxième objectif concerne les projets scientifiques à la demande assortis de leurs propres exigences. Ils sont analysés au cas par cas pour sélectionner la solution la plus rentable et la plus performante. La mission et la vision d'une infrastructure informatique commune sont évidemment appliquées, mais pour le stockage par exemple, il est important d'investir dans une technologie sur mesure.

Le niveau de maturité de Sciensano pour l'approvisionnement se trouve entre « Indéterminé » et « Émergence ».

1.7. LIENS AVEC UN ÉCOSYSTÈME DE DONNÉES PLUS LARGE

« De nombreux autres départements gouvernementaux, organisations internationales, acteurs du secteur privé et organisations non gouvernementales peuvent être considérés comme faisant partie d'un écosystème de données commun, un réseau d'organisations, de systèmes de données, d'infrastructures numériques et d'applications interconnectés. La manière dont les organisations choisissent de collaborer avec d'autres entités au sein de leurs propres écosystèmes de données, ou

de ceux qui se superposent, est une décision stratégique qui comporte d'importantes considérations en matière d'interopérabilité. [...] Les accords mis en place par les organisations avec d'autres parties de l'écosystème de données sont particulièrement importants pour s'assurer que toute interopérabilité planifiée est pensée et documentée de façon stratégique. Les contrats de licence des données (documents juridiques qui spécifient ce que l'utilisateur peut et ne peut pas faire avec les données) sont l'un des outils disponibles à cette fin » (1). Un autre article (2) souligne les trois conditions préalables pour créer un écosystème robuste de partage de données de santé et de soins de santé : tout d'abord, la familiarité du pays avec la gestion des données, ses accords de partage de données et sa technologie de partage des données ; ensuite, la prévalence de l'utilisation des systèmes de dossiers médicaux personnels ; enfin, la présence d'un organisme national responsable du développement de cet écosystème.

Universités

En juillet 2021, Sciensano a signé un protocole d'accord avec les universités wallonnes (Université catholique de Louvain, Université libre de Bruxelles, Université de Liège, Université de Mons, Université de Namur) et flamandes (Université catholique de Louvain, Université d'Anvers, Université de Gand, Université de Hasselt, Vrije Universiteit Brussel). Ce protocole stipule que les domaines communs de recherche comprennent notamment la recherche sur le bien-être et sur la santé. Les partenaires conviennent d'un échange mutuel des données et des résultats avec l'objectif commun d'avancées scientifiques dans le domaine de la santé et du bien-être humains et animaux.

Healthdata.be

En Belgique, le secteur des soins de santé compte de nombreux acteurs qui appliquent chacun leurs propres méthodes pour la collecte et l'enregistrement des données de santé. Par conséquent, la collecte des données est très hétérogène, tant en termes quantitatifs que qualitatifs. La plateforme healthdata.be permet de standardiser et d'homogénéiser l'enregistrement et la protection des données de santé tout en garantissant une confidentialité absolue dans leur diffusion à des fins de recherche. En offrant une solution concrète au problème de fragmentation des données en différents registres de santé, healthdata.be contribue de manière significative au développement d'une infrastructure du système pour la recherche (sur la santé) en Belgique.

Service MetaHub de la plateforme eSanté

Par leur intégration à la plateforme de soins collaborative CoZo, les résultats des laboratoires médicaux de Sciensano sont partagés sur le service MetaHub de la plateforme eSanté.

Autorité des données de santé

Le besoin d'une initiative fédérale capable d'organiser et de soutenir l'utilisation et l'interconnexion éventuelle des données de santé et de soins de santé à des fins scientifiques et politiques a été souligné à plusieurs reprises par le passé³.

Dans sa note de politique générale Santé publique du 2 novembre 2020, le ministre de la Santé publique propose la création d'une ADS pour répondre à ce besoin. L'unité politique du ministère de la Santé publique a ensuite commissionné l'AFMPS, le SPF SPSCAE, le KCE, l'INAMI et Sciensano pour développer conjointement une proposition pour l'autorité des données de soins de santé, en consultation avec la plateforme eSanté.

Espace européen des données de santé (EEDS)

³Voir par exemple « Inventaire des bases de données de soins de santé » (KCE, rapport 2006, chapitre 5) ; « Le soutien scientifique à la politique de santé fédérale » (Rekenhof, 2010) ; « Data for better health - Conclusion note » (Policy Cell Minister of Health, 2020) ; « Call and advice – Better government data for strong policy research and good governance » (SERV, VLIR, 2021).

Sciensano collabore activement au débat sur l'EEDS. Il s'agit d'un lien important avec des écosystèmes de données plus larges, qui permet aussi à Sciensano de façonner l'EEDS. Le projet TEHDAS permet de mieux comprendre les systèmes de gouvernance et de gestion des données en Belgique et dans les pays européens. En se lançant dans le développement de l'EEDS, Sciensano crée de manière conjointe les normes qui seront développées pour la Facilité à trouver, l'Accessibilité, l'Interopérabilité et la Réutilisation des données. Sciensano participera aussi au développement de la technologie qui permettra de fonctionner dans un écosystème de données plus large, conformément aux exigences des implications éthiques, légales et sociales (IELS).

En termes de liens avec l'écosystème de données plus large, Sciensano se situe au niveau « Apprentissage ».

2. Interopérabilité humaine

2.1. ADMINISTRATION DES DONNÉES

« Dans le domaine de la gestion des données, l'administration des données est la « pratique qui consiste à gérer les données pour le compte d'autrui et dans le meilleur intérêt de l'organisation » (5). Une administration effective et responsable des données est fondamentale pour l'interopérabilité. Souvent, les administrateurs des données sont chargés de mettre en place des procédures et pratiques qui permettent une meilleure interopérabilité. Les administrateurs des données sont souvent des membres du personnel nommés stratégiquement au sein d'une organisation afin de superviser les rôles clés de la gestion des données pour lesquels ils ont une expertise considérable. Par exemple, l'administrateur principal des données pourrait présider des conseils ou des comités de gouvernance des données, alors que l'administrateur technique des données pourrait être un expert informatique qui supervise notamment l'intégration des données dans des départements spécifiques... » (1). Un livre blanc (2) distingue différents rôles relatifs aux données dans l'écosystème de partage des données (générateurs de données de santé, régulateurs de données de santé, instances dirigeantes, partage des données de santé, fournisseurs de services informatiques et utilisateurs des données de santé). Il demande à ce que les « propriétaires » des données soient rebaptisés « administrateurs » des données. Le terme « propriétaire » des données devient moins pertinent quand les données ne sont plus possédées par le générateur de données, mais bien partagées entre des parties prenantes qui sont responsables de protéger et gérer les données.

Healthdata.be joue un rôle important pour l'accessibilité des données de santé en Belgique. L'administration des données peut être améliorée en ce qui concerne la mise en place de procédures, de protocoles et de conditions budgétaires clairs, transparents et standard afin d'avoir accès aux données. Il est nécessaire de délimiter clairement le rôle et la responsabilité des différents comités dans la gestion de l'accès aux données entre healthdata.be et (le reste de) Sciensano.

Chez Sciensano, les chefs de service sont responsables des données utilisées dans leur service. Pour certains départements scientifiques et dans certains services, il y a une personne qui fait office de point de contact pour les ensembles de données utilisés dans le service. Pour les données sur le COVID, des administrateurs de données distincts ont été désignés pour prendre en charge la gestion des données, mais aucun rôle officiel de propriétaire des données n'a été attribué pour les différents ensembles de données (le propriétaire n'est souvent pas Sciensano, mais les régions). En général, les directeurs scientifiques sont chargés de s'assurer que les données présentes au sein de leur conseil d'administration sont conformes aux règles du RGPD. Chez Sciensano, deux rôles ont été officiellement définis : le délégué à la protection des données et le coordinateur de sécurité des informations.

En termes d'administration des données, Sciensano est au niveau « Émergence ». En prenant en compte l'expertise de healthdata.be, Sciensano est, dans son ensemble, au niveau « Apprentissage ».

2.2. PROTECTION DES DONNÉES À CARACTÈRE PERSONNEL ET PRÉSERVATION DE LA CONFIDENTIALITÉ

« Établir l'interopérabilité entre des systèmes requiert une objectivité minutieuse au moment de déterminer quelles données devraient être ouvertes, partagées et intégrées à d'autres ensembles et quelles données ne devraient pas l'être. Nombreuses sont les raisons pour lesquelles une organisation pourrait ne pas vouloir rendre ses données interopérables avec d'autres ensembles de données, mais l'une des raisons principales est le souhait de préserver la confidentialité et de protéger le droit à la vie privée que les sujets des données ont sur leurs informations personnelles et sensibles... » (1).

Protection des données à caractère personnel et évaluation de l'impact

Sciensano a aussi mis en place un plan d'action spécifique pour surveiller le respect du règlement (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016 relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données (loi du 30 juillet 2018). Il est important de mentionner qu'en raison de la nature des tâches de Sciensano, le type de données traitées (données de santé) doit, conformément au règlement, faire l'objet de mesures de sécurité spécifiques, comme les évaluations de l'impact sur la protection des données.

Les projets scientifiques de Sciensano demanderont souvent l'approbation d'un comité éthique. Des informations sur la gestion des données se trouvent dans les protocoles d'étude évalués par les comités éthiques. Les sections qui concernent la gestion des données comportent des informations sur le flux des données et les mesures de protection des données à caractère personnel, dont les procédures d'accès aux données.

En plus de l'approbation du comité éthique, le traitement des données de santé par Sciensano requiert souvent a) des délibérations positives à la chambre sécurité sociale et santé du Comité de sécurité de l'information (CSI) et b) des autorisations du ministre de l'Intérieur concernant l'utilisation des informations du Registre national. Dans le cadre des demandes d'autorisation de ces parties, les employés de Sciensano devront justifier leurs besoins de données et adopter des mesures de sécurité. En outre, les postes et services de Sciensano qui auront accès aux données doivent être spécifiés.

Ces organismes externes d'éthique et de protection des données établiront si, oui ou non, le projet respecte les principes de protection des données tels que la légalité, l'équité et la transparence, la limitation des finalités, la limitation des données, la précision, la limite de stockage, l'intégrité et la confidentialité. Une analyse des petites cellules (SCRA) est souvent demandée par le Comité de sécurité de l'information. Sciensano a entamé une collaboration avec différentes administrations de la santé publique s'occupant des données de santé à caractère personnel afin de procéder à des SCRA avec des bourses et, par conséquent, de standardiser les méthodologies des SCRA et de réduire les coûts.

Spécificité de healthdata.be

Chez healthdata.be, les informations collectées peuvent être transmises à des chercheurs et des médecins superviseurs agréés uniquement pour améliorer la qualité et la gestion du secteur de la santé et des soins de santé. Les données peuvent alors être transmises uniquement à des fins de recherche scientifique, de surveillance de la santé et de promotion des connaissances. Les données ne sont jamais communiquées à des tiers dans un objectif autre que la recherche ou la protection de la santé publique. Par exemple, elles ne sont jamais communiquées à un employeur potentiel dans le cadre d'une procédure de recrutement, à une banque afin d'obtenir un prêt ou à une compagnie d'assurance pour obtenir une police d'assurance. De plus, les données de santé sont systématiquement pseudonymisées avant d'être rendues accessibles aux chercheurs. En d'autres mots, seules les données nécessaires à la recherche seront communiquées, et les informations qui rendent possible l'identification officielle des patients seront toujours codées.

En termes de protection des données à caractère personnel et de préservation de la confidentialité, Sciensano est au niveau « Construction ».

2.3. CONNAISSANCES ET COMPÉTENCES DU PERSONNEL

« Les connaissances et compétences du personnel sont fondamentales pour l'interopérabilité humaine. Pour les organisations qui gèrent divers ensembles de données et veulent améliorer leur interopérabilité, il est impératif qu'au moment de développer leurs stratégies de gestion des données et de définir les objectifs stratégiques, elles prennent en compte les questions de supervision et de responsabilité ainsi que les rôles d'administration des données et les connaissances et compétences dont le personnel aura besoin pour effectuer ses tâches efficacement... » (1).

Le département des Ressources humaines de Sciensano et les différents départements scientifiques sont bien conscients de la nécessité de former le personnel et d'embaucher de nouvelles personnes dotées des compétences nécessaires à l'amélioration de la gestion des données et de compétences en analyse de données. Les gestionnaires et coordinateurs des données sont disséminés chez Sciensano ; les exigences en matière de connaissances et de compétences nécessaires pour relier les données ne sont pas reprises dans les descriptions de poste et restent souvent liées à des projets. Plusieurs initiatives pour répondre à ces besoins en compétences ont été lancées, notamment des séminaires avec des intervenants externes, de nouveaux modules de formation et le recrutement de personnes dotées de compétences en matière d'infrastructure informatique et de science des données. En outre, il est également prévu de former des conseillers en sécurité des informations. Il y a aussi un besoin de compétences en matière de connaissances pour la mise en application des principes FAIR, pour la gestion des métadonnées et pour l'analyse fédérée.

En termes de connaissances et de compétences du personnel, Sciensano se trouve entre les niveaux « Émergence » et « Apprentissage ».

2.4. COMMUNICATION INTERNE ET EXTERNE

« Un autre principe de l'interopérabilité humaine est d'assurer une communication interne et externe efficace autour des questions fondamentales de gestion des données qui affectent la capacité des utilisateurs à utiliser et à intégrer les données. En interne, des canaux de communication officiels doivent être ouverts entre les individus, les équipes et les départements au moment de fixer des objectifs stratégiques dans les stratégies de gestion des données, les postes de direction et de gestion, la supervision et la responsabilité et les postes d'administration des données. En externe, il est essentiel de créer des boucles de feedback utilisateur qui permettent la communication sur la qualité des données, les lacunes, la facilité d'utilisation et d'autres aspects de l'interopérabilité humaine pour y apporter des solutions. De la même façon, pour les publics non spécialisés, il est important que les organisations [...] puissent communiquer efficacement avec les décideurs politiques, qui sont les utilisateurs finaux de leurs données... » (1).

En interne, davantage de canaux de communication officiels devraient être mis en place pour permettre la communication entre les individus, les départements, les points de contact des données, les DPD, les coordinateurs de sécurité des informations et healthdata.be.

En externe, Sciensano communique avec les parties prenantes sur les questions de gestion des données par un contact direct avec les fournisseurs de données ou des contacts avec le Comité de sécurité de l'information via des points de contact uniques (les DPD).

Sciensano a acquis une expertise dans les questions fondamentales de gestion des données de santé à l'échelle européenne par l'intermédiaire de l'unité Système d'information sanitaire de l'UE. Sciensano est considéré comme étant un acteur de poids dans la communauté européenne. Les activités de cette

unité ont renforcé la présence de Sciensano au niveau européen et ont contribué à faire de Sciensano une institution réputée pour son leadership dans les projets européens, pour ses connaissances et son expertise dans le domaine des informations de santé, pour sa vision du développement en matière d'informations de santé et de réutilisation des données. Sciensano est aussi reconnu comme un collaborateur international de confiance produisant des résultats de qualité.

En termes de communication interne et externe, Sciensano est au niveau « Apprentissage ».

2.5. ADAPTABILITÉ

« La capacité d'une organisation à garantir que ses procédures internes de gestion des données sont adaptables est un aspect clé de l'interopérabilité humaine. Alors que les innovations dans le domaine du numérique et des données émergent, les organisations et leur personnel doivent s'assurer que leurs stratégies de gestion des données suivent de près les avancées. L'adaptabilité requiert la communication et l'autonomisation des membres du personnel, ainsi que la flexibilité pour changer de cap à tout moment afin de répondre à l'évolution des besoins des parties prenantes... Un bon exemple de la plus-value que peut apporter l'adaptabilité concerne la nécessité de sensibiliser les individus et les organisations à l'émergence de nouvelles composantes de l'infrastructure des données. Par exemple, les organisations qui cherchent à améliorer l'interopérabilité, l'accessibilité et l'utilisation de leurs données devraient être conscientes des tendances relatives à l'accessibilité et à l'application de l'infrastructure numérique et être capables de s'y adapter ou de les adapter à leurs besoins... » (1).

Afin de répondre aux besoins de données et d'informations des décideurs politiques, des chercheurs et des utilisateurs, les systèmes de gestion des données doivent être standardisés et homogénéisés. Pour ce faire, les responsabilités, les postes et les services responsables doivent être répartis de manière à pouvoir s'adapter aux nouveaux besoins en matière d'information. Actuellement, les procédures de gestion des données chez Sciensano ne sont pas assez simplifiées et restent trop complexes, car les différents départements ont mis en place différents systèmes de gestion des données. Un paysage très diversifié de gestion des données affaiblit l'adaptabilité.

Sciensano existe depuis 2018. Pendant la crise du COVID, Sciensano a pu démontrer, pour certaines actions, une excellente adaptabilité en termes de transmission des informations pour combattre la crise du COVID en prodiguant des conseils politiques en temps réel. Par conséquent, Sciensano est toujours au niveau « Apprentissage », mais a montré des signes de « Construction ».

3. Interopérabilité des données

3.1. CAPACITÉ DE MODÉLISATION DES DONNÉES ET DES MÉTADONNÉES

« La modélisation des données et des métadonnées est le fondement de l'interopérabilité des données. Elle fait partie du processus de conception au cours duquel la structure interne et les corrélations entre les différents ensembles de données sont définies, optimisées et décrites afin de prendre en compte toutes les mesures et dimensions pertinentes pour l'entreprise (6). Le processus de modélisation des données et des métadonnées comprend un certain nombre d'étapes, allant de la découverte, de l'analyse et de la définition des besoins en données jusqu'à leur transmission sous forme de modèle (5)... » (1).

Chaque département de Sciensano consomme et génère très certainement des données. Celles-ci sont utiles, nécessaires à l'acquisition de connaissances, à la prise de décisions, au développement des programmes de santé publique, etc. Les données sont au cœur des activités et du travail scientifique de Sciensano. Le serveur de fichiers de Sciensano héberge plus d'un demi-pétaoctet de données.

Actuellement, les données externes à healthdata.be sont stockées dans des centres de données de Sciensano. Le département informatique distingue le stockage des données actives, passives et

archivées et fournit des solutions techniques pour ces trois types de données. Les catalogues de métadonnées sont importants dans le cadre de la mise en œuvre d'une stratégie de données pour Sciensano, étant donné qu'ils abordent les notions de partage et d'adoption des principes FAIR : Facile à trouver, Accessible, Interopérable et Réutilisable. Ces principes traitent du cycle de vie des données et de leur valorisation avec une devise simple : « créer une fois, utiliser maintes fois » (*create once, use many times*). Il existe cependant de beaux exemples de la manière dont les métadonnées peuvent être décrites.

De quoi a-t-on besoin pour entretenir un catalogue de métadonnées ? Comment est-il en lien avec la stratégie de données ?

Le processus consistant à passer d'une approche cloisonnée à des infrastructures ouvertes, décentralisées et fédérées vise à ajouter de la valeur aux données. Nous devons améliorer nos méthodes de travail et créer de nouvelles approches de l'exploration et de l'analyse des données. La stratégie de données fournit un cadre opérationnel pour la gouvernance des données selon le concept des trois axes Technologie, Organisation et Procédure (TOP) : L'axe de la technologie (« T ») concerne toutes les technologies visant à faciliter la gestion et l'exploitation des données selon les règles de l'entreprise. Le catalogue de métadonnées est un élément essentiel pour garantir la découverte et l'utilisation des données. Le « P » renvoie à l'application des règles visant à augmenter la valeur des données (rendre les données FAIR), et le « O » porte sur la mise en place d'une équipe de gouvernance des données responsable de la mise en œuvre de la stratégie de données. Les principes directeurs FAIR facilitent la valorisation des données en rendant les données faciles à trouver, accessibles, interopérables et réutilisables. L'application de ces principes consiste en la mise en œuvre du catalogue de métadonnées FAIR. Elle aidera les utilisateurs « humains » et les machines dans la découverte, l'accès, l'intégration et l'analyse des données.

En termes de capacité de modélisation des données et des métadonnées, Sciensano *dans son ensemble* est au niveau « Émergence ». Si l'on prend en compte healthdata.be, le niveau de maturité est « Construction ».

3.2. CAPACITÉ D'ORGANISATION ET DE CLASSIFICATION DES DONNÉES

« La coordination au sein des organisations et entre elles, ainsi qu'au sein de systèmes (statistiques) nationaux entiers ou d'écosystèmes de données d'ailleurs, pour décider de classements et d'identifiants communs, est essentielle pour permettre l'interopérabilité des données. Les vocabulaires standard, les classements et les identifiants uniques font partie de l'infrastructure de données de base d'un pays ou d'une organisation. Ils permettent d'améliorer la cohérence et d'éviter les ambiguïtés dans la description d'un ensemble de données, tout en permettant aux utilisateurs de repérer plus facilement des éléments de données connexes et de les lier entre eux... Les classifications standard permettent l'intégration d'ensembles de données multiples ainsi que leur analyse et leur interprétation cohérentes... » (1).

À l'exception de healthdata.be, où des systèmes de classement communs sont en place, les pratiques en matière de classification et d'identification varient entre les différents services, unités et conseils d'administration scientifiques de Sciensano. En termes de capacité d'organisation et de classification des données, Sciensano est au niveau « Apprentissage ».

3.3. ACCÈS, OUVERTURE ET PARTAGE DES DONNÉES

« ...il existe également des considérations techniques au niveau de l'interopérabilité des données. Au niveau de l'interopérabilité des données, des considérations sur la manière de planifier stratégiquement la publication de données ouvertes touchent aussi à la façon de modéliser et de classer les ensembles de données. Dans la mesure du possible, il convient d'utiliser des logiciels open source et des standards de données pour la publication de données ouvertes et de métadonnées, car ils permettront aux utilisateurs des données d'intégrer plus facilement de nouveaux ensembles de données dans leurs systèmes. Les outils et les technologies qui permettent de publier des données de manière à ce que les machines puissent facilement identifier et intégrer des sources d'information connexes directement sur

le web suscitent un intérêt croissant. Cette approche de « données liées » facilite l'interopérabilité et l'intégration des données en connectant les ensembles de données structurés et non structurés avec des éléments de métadonnées spéciaux conçus pour être référencés sur le web... » (1). Pour développer les capacités nécessaires pour créer des écosystèmes de partage de données, il est impératif de mettre en place un cadre de partage de données qui implique plusieurs parties prenantes et qui comprend plusieurs composantes de base : une stratégie de partage des données, des capacités technologiques, une capacité réglementaire et juridique et une approche visant à rendre opérationnel le partage de données (2). Cet article distingue également différents niveaux de sensibilité (accès libre, accès contrôlé restreint et accès contrôlé fermé) qui peuvent aider à définir la stratégie optimale de partage des données.

Sciensano rend ses microdonnées accessibles pour la recherche (statistique). Sciensano a mis en place des services de microdonnées pour les utilisateurs externes par le biais de healthdata.be. Ce service permet de faire le lien avec le numéro de registre national⁴. Les données provenant notamment de l'enquête de santé et de l'enquête de consommation alimentaire sont aussi disponibles via différents services selon les procédures décrites sur le site web. Ces données sont particulièrement sensibles en matière de respect de la vie privée. Pour cette raison, la recherche sur les microdonnées est soumise à des conditions strictes et à des exigences de sécurité. La recherche externe peut également être une recherche multithématique, ce qui nécessite de combiner et de relier les microdonnées de Sciensano avec les données des chercheurs. Pendant la crise du COVID, un groupe de demande de données a été mis en place afin de gérer les différentes demandes de données émanant des universités. Ce groupe se réunit toutes les deux semaines, et les DPD ainsi que le département juridique participent à ces réunions. Sur base d'un avis positif, un accord de transfert de données est conclu entre Sciensano et l'université.

La science ouverte, y compris les données ouvertes, est un moteur de changement considérable en science. Une nouvelle façon de faire de la recherche exige de plus en plus le partage et la publication des données ainsi que le traitement adapté des données de recherche. Selon la Commission européenne, la science ouverte doit devenir la norme pour la recherche financée par les fonds publics⁵.

La procédure de partage des données traitées par Sciensano avec des tiers est décrite dans la procédure opérationnelle standard. Les données ouvertes sont disponibles via le lien suivant :

<https://www.belgiqueenbonnesante.be/fr/etat-de-sante/a-propos-du-health-status-report/bases-de-donnees-interactives-sciensano> Sciensano rend ses données disponibles sur des plateformes :

- <https://data.gov.be/fr> à <https://epistat.sciensano.be/covid/> et
- <https://www.healthinformationportal.eu/> et
- <https://fair.healthdata.be/search/?theme=COVID-19>

Le Moniteur belge a publié une disposition de libre accès dans la législation belge le 5 septembre 2018. Voici l'article XI.196, paragraphe 2/1 du Code de droit économique (belge), qui stipule ce qui suit : « *L'auteur d'un article scientifique issu d'une recherche financée pour au moins la moitié par des fonds*

⁴ La clé est le numéro de registre national, pseudonymisé par un tiers de confiance (TDC) grâce à un algorithme. En outre, toute donnée liée au système d'information géographique peut être reliée (par exemple via Statbel) aux microdonnées basées sur le numéro de registre national. Si aucun numéro de registre national n'est trouvé dans les microdonnées, des informations (comme la pauvreté, la pollution de l'air, etc....) peuvent être reliées grâce à d'autres informations comme le secteur statistique, la commune, etc. tant que ces informations sont définies sur le niveau de secteur statistique ou la commune ou bien peuvent être extrapolées statistiquement d'une quelconque manière jusqu'à ces niveaux.

⁵Le terme « doit » est à nuancer, car les données de santé au niveau des microdonnées brutes ne seront jamais « ouvertes » au sens large des données ouvertes.

publics conserve [...] le droit de mettre le manuscrit gratuitement à la disposition du public en libre accès après un délai de douze mois pour les sciences humaines et sociales et six mois pour les autres sciences, après la première publication, dans un périodique, moyennant mention de la source de la première publication. »

Le contenu de cet article est obligatoire et s'applique, en dépit de la loi choisie par les parties, dès qu'un facteur de rattachement est situé en Belgique. Ce droit ne concerne que les articles de journaux scientifiques, pas les ouvrages (ou contributions à des ouvrages). Ce droit est rétroactif et s'applique donc également aux publications antérieures à 2018. Il est donc tout à fait possible de rendre des publications antérieures disponibles en accès libre grâce à cet article. Sciensano adhère à une politique de données ouvertes et de science ouverte conforme au cadre du RGPD. Pour la science ouverte, Sciensano a publié ses articles scientifiques sur son site Internet.

Healthdata.be

La plateforme technique healthdata.be, créée en 2015, vise à fournir aux décideurs politiques des rapports scientifiques à jour, fiables et valables de manière rapide et autonome pour qu'ils puissent élaborer et (re)diriger leurs actions politiques. Afin de compiler ces rapports, les chercheurs ont besoin d'outils (techniques et administratifs) qu'ils peuvent utiliser de manière rapide, facile et autonome afin de collecter et de traiter les données souhaitées de façon fiable et efficace. La collecte de ces données peut ne pas constituer une charge administrative ou technique pour les prestataires de soins de santé ou les prestataires de services y prenant part et doit garantir une confidentialité et une protection des données optimales pour les prestataires de soins de santé et pour les patients.

En 2021, la collaboration pour le SARS-CoV-2 entre la plateforme healthdata.be et les autorités régionales de santé a conduit au développement d'une plateforme technique, haute performance, en temps réel et évolutive pour l'échange de données, qui peut également soutenir des projets coopératifs futurs.

La plateforme healthdata.be est devenue une référence pour les objectifs nationaux et internationaux en relativement peu de temps. Par exemple, en 2021, la plateforme a été consultée pour avoir recours à ses services pour des initiatives telles que l'ADS belge, l'Espace européen des données de santé (EEDS) et l'autorité européenne de préparation et de réaction en cas d'urgence sanitaire (HERA).

Un changement important concerne la gouvernance de la plateforme healthdata.be : en plus du comité de pilotage déjà opérationnel de la plateforme healthdata.be, un nouveau comité d'orientation a été créé. Ce comité d'orientation, composé des représentants des administrations de santé fédérales et fédérées, a été nommé responsable de la planification et de la priorisation de healthdata.be ainsi que du financement des ressources.

Données ouvertes et COVID

La crise du COVID a clairement démontré l'importance des données ouvertes et FAIR pour la visibilité et la réputation des institutions de santé publique. Les chercheurs externes et les journalistes de données ont découvert Sciensano et ont porté un vif intérêt à la valorisation des données sur le COVID-19 générées, collectées et gérées par Sciensano. Bien que ces demandes aient causé une charge administrative supplémentaire pour Sciensano, la réutilisation des données a conduit à une situation gagnant-gagnant pour l'institut : les chercheurs externes ont validé des idées et en ont développé de nouvelles, les journalistes de données ont aidé à diffuser les messages scientifiques et les organisations internationales ont établi des comparatifs entre les pays. En d'autres termes, la réutilisation des données a conduit à une diffusion plus large et plus efficace des résultats de Sciensano et a permis de générer des idées et des produits pour lesquels Sciensano n'aurait ni les ressources ni le temps. Les instituts de santé publique dotés de plateformes de données ouvertes disponibles ont reçu des louanges, tandis que les autres ont reçu des reproches.

Compte tenu de l'attention portée actuellement aux données ouvertes et FAIR, Sciensano devrait diffuser de manière proactive des résultats pertinents sous forme de données ouvertes et FAIR. Cela permettra de promouvoir une image positive d'ouverture, de transparence et de volonté de collaboration. Une future plateforme de données FAIR doit donc être particulièrement visible, par exemple grâce à un sous-domaine.

Ce portail de données FAIR servira de plus à donner un aperçu des activités relatives aux données, augmentant ainsi la visibilité des différents projets de Sciensano. Enfin, ce portail sera également un atout pour les propositions de projets, étant donné que les chercheurs de Sciensano pourront prouver que les résultats de leurs recherches seront publiés sur le portail de données, ce qui, soit dit en passant, est de plus en plus souvent exigé par les organismes de financement belges et européens.

En plus d'une plateforme de données FAIR, Sciensano doit fournir des outils de visualisation faciles à utiliser pour explorer de manière intuitive les résultats obtenus. Le tableau de bord COVID a démontré l'intérêt et la valeur de tels investissements, mais d'autres exemples existent aussi :

- Étude nationale belge sur le fardeau de la maladie (BeBOD), fardeau non mortel du cancer
- Étude nationale belge sur le fardeau de la maladie (BeBOD), mortalité et années de vie perdues
- Procédures standardisées d'analyses de la mortalité (SPMA)
- Espérance de santé
- Analyse interactive de l'enquête de santé belge par interview (HISIA)
- Épidémiologie des maladies infectieuses (Epistat)

En termes d'accès, d'ouverture et de partage des données, Sciensano est au niveau « Apprentissage » avec des éléments du niveau « Construction ».

3.4. ANALYSE ET AUTOMATISATION DES DONNÉES

« Les révolutions dans le domaine du numérique et des données ont éclaté dans le monde entier, tout comme la capacité d'un plus grand nombre d'entités et d'individus à collecter, à organiser, à structurer et à analyser des types de données plus diversifiés. À mesure que davantage de données et types de données sont rassemblés, l'interopérabilité gagne en importance pour garantir la modélisation des données de manière à permettre le traitement et l'analyse, y compris par voie automatisée... » (1).

Une première condition pour l'analyse de données est d'avoir une vue d'ensemble de toutes les données dont Sciensano (hormis healthdata.be) dispose en interne et de leurs métadonnées afin de pouvoir identifier le potentiel en vue de la valorisation des données. Sciensano a pris des mesures relatives à l'analyse des données en reliant les sources de données d'enquêtes et administratives et en menant des recherches à l'aide des données environnementales et des informations géospatiales pour la cartographie. Cependant, l'infrastructure actuelle ne permet pas d'effectuer ces recherches de manière automatisée sans copier ou déplacer les données entre les institutions.

En termes d'analyse et d'automatisation des données, Sciensano est au niveau « Émergence ».

3.5. PROTECTION ET STOCKAGE DES DONNÉES

« À son niveau le plus basique, la protection des données est la réglementation de la manière de contrôler l'accessibilité et l'utilisation des données stockées dans les ordinateurs, les appareils numériques et les dossiers au format papier. La protection des données peut s'envisager sous trois angles. Tout d'abord, il s'agit d'une série d'obligations dans les juridictions où les lois relatives à la protection des données ont été votées pour établir des règles quant à la manière dont les données devraient être protégées. Le règlement général sur la protection des données (EU 2016) de l'Union européenne en est un bon exemple. Ensuite, la protection des données peut se référer à la protection des données à caractère personnel des citoyens et est intrinsèquement liée à la préservation de la vie privée et de la confidentialité. Enfin, elle peut renvoyer à un mélange de considérations techniques et de considérations liées aux données. Les premières concernent la cybersécurité et la réaction aux

incidents, alors que les secondes concernent la sécurité des données, etc. » (1). Un livre blanc (2) distingue plusieurs façons de stocker des données (entrepôts de données, coffres-forts numériques, lacs de données, fabriques de données). On constate que de nombreuses parties prenantes détenant une grande quantité de données disposeraient d'un entrepôt de données stockant les données de manière sûre et efficace. Néanmoins, quand les conditions de sécurité et de protection se superposent au fil du temps et que les organisations font preuve de réticence à partager des données puisque considérées sensibles, ces entrepôts de données peuvent devenir des coffres-forts de données. La transition devrait alors se faire d'un entrepôt de données à un lac ou une fabrique de données, qui sont des infrastructures de partage de données plus progressives.

Protection des données

Afin d'améliorer la sécurité de ses données, Sciensano met au point un système de gestion de la sécurité de l'information (SGSI) selon la norme ISO 27001. Cette approche holistique permet une meilleure gestion de la sécurité en fournissant aux décideurs politiques des données pour justifier leurs décisions. Sciensano utilise des analyses des risques liés à la sécurité de l'information pour développer un plan de gestion des risques identifiés.

Stockage des données

Toutes les données sont stockées dans l'espace de stockage central fourni par Sciensano. De l'espace disque est fourni tant pour les données scientifiques que pour les données administratives. Cet espace de stockage est automatiquement mis à la disposition de tous les employés de Sciensano quand ils utilisent les ordinateurs de Sciensano (ou des volumes spécifiques pour une grande quantité de données). Stocker des données sur n'importe quel support externe (par exemple, des disques durs ou clés USB) n'est pas autorisé. Les données stockées localement sur des ordinateurs peuvent seulement être de nature temporaire, ce qui est autorisé dans les cas suivants :

- Utilisation hors ligne des données centralisées lors d'enquêtes, de recherches sur les missions, d'interventions et d'actions avec des partenaires sans connexion au réseau ;
- Utilisation hors ligne lors de voyages ou de déménagements.

Le principe standard au sein de Sciensano est que les données de l'organisation ne sont pas stockées via un stockage externe dans le cloud. Toutes les données sont stockées, préservées et gérées en interne chez Sciensano. Les traitements de données sont décrits au stade de la demande de projet (pré-cycle PMO). Ce traitement comprend la nature des données qui doivent être produites / collectées, l'importation et l'exportation des données ainsi que le stockage final des données ou des résultats.

En termes de protection et de stockage des données, Sciensano se situe entre les niveaux « Apprentissage » et « Construction ».

4. Interopérabilité technologique

4.1. INFRASTRUCTURE DES DONNÉES

« Bien que cela semble évident, il est utile d'affirmer de manière explicite que toutes les entités et organisations souhaitant collaborer avec l'écosystème de données plus large et travaillant à l'amélioration de l'accessibilité et de l'utilisation de leurs données doivent avoir une infrastructure numérique de base, à savoir des ordinateurs, une capacité de base de données (sur serveur ou dans le cloud) et une connexion à Internet. Toutefois, le concept d'infrastructure numérique est beaucoup plus large que ce qu'il n'était il y a à peine quelques années. Il existe maintenant de nombreuses catégories d'infrastructures qui dépendent toutes de l'interopérabilité et l'influencent. Voici quelques exemples de composantes infrastructurelles interconnectées :

- Infrastructure de stockage des données (sur serveurs locaux ou dans le cloud) ;
- Infrastructure de gestion des données, comme les systèmes de gestion de l'information sur mesure ;

- Infrastructure de traitement des données (de plus en plus souvent dans le cloud pour les grands ensembles de données) ;
- Infrastructure de diffusion des données composée de plateformes de données, plateformes communautaires, etc. » (1).

Sciensano et healthdata.be fournissent séparément un environnement virtuel pour l'analyse de leurs données dans un environnement sécurisé avec différents flux de travail. Actuellement, le partage de microdonnées à des fins relatives à la santé se limite aux microdonnées situées sur healthdata.be. Ce flux de travail comprend les approbations de healthdata.be et les autorisations du Comité de sécurité de l'information ou du Comité éthique. L'analyse est menée avec un ensemble de logiciels et de politiques par défaut proposés par l'infrastructure informatique.

L'infrastructure des données pour les données qui ne sont pas gérées par healthdata.be se trouve dans les dossiers du serveur de fichiers avec un accès restreint et contrôlé via l'application de gestion de l'identité Sciensorg. Les droits de délégation, le traçage et l'automatisation sont combinés pour assurer une solution sécurisée pour tous les accès aux données. L'accès anonyme aux données de Sciensano se fait via le site web public ou les sites des projets de Sciensano. Toutes les données hébergées par Sciensano (hormis healthdata.be) sont uniquement accessibles avec un compte Sciensano personnel et sont limitées aux fonctions que remplit le titulaire du compte (Sciensorg). Les DPD analysent la demande d'accès aux données et donnent un avis ou une approbation.

En termes d'infrastructure des données, Sciensano est au niveau « Apprentissage ».

4.2. CYBERSÉCURITÉ ET RÉACTION AUX INCIDENTS

« Alors que de plus en plus de fonctions primordiales de gestion des données se font en ligne et que le télétravail devient la règle et non plus l'exception, la cybersécurité et les protocoles de réaction aux incidents prennent de l'importance. Prenons l'exemple de l'informatique dans le cloud Basée sur la technologie de virtualisation matérielle, l'informatique dans le cloud permet aux organisations de sauvegarder rapidement des données, des applications ou même des systèmes d'exploitation sur un centre de données à distance et de les envoyer à plusieurs utilisateurs en de nombreux endroits différents. Cependant, toutes ces transmissions de données en ligne les exposent à des menaces de cybersécurité... » (1).

L'infrastructure informatique de Sciensano est protégée par plusieurs niveaux de sécurité, avec la classification de réseaux, la double authentification, les solutions sur mesure, la gestion de l'identité, les antivirus, les logiciels ciblant les programmes malveillants, les filtres antispam et le suivi intelligent des données. La correction des bugs et la maintenance d'une infrastructure informatique sont essentielles et constituent à ce titre les premières défenses contre les cyberincidents. Par conséquent, un suivi journalier et un plan de maintenance sont obligatoires chez Sciensano. L'enregistrement des incidents et les flux de travail sont hébergés sur une plateforme sécurisée distincte, et le système interne de tickets d'assistance technique permet d'enregistrer les infractions sur les systèmes. La réaction aux incidents et la communication seront améliorées et optimisées pour sensibiliser les employés, les partenaires et les responsables. La formation, les principes et les pratiques exemplaires en matière de sécurité sont mis en application dans tous les services informatiques (développement, centre de support, etc.) et sont renouvelés ou optimisés si nécessaire. Les exigences de la norme ISO 27001 sont en train d'être mises en pratique.

En termes de cybersécurité et de réaction aux incidents, Sciensano a le niveau « Émergence ».

FUTUR

1. Interopérabilité organisationnelle

1.1. OBJECTIFS STRATÉGIQUES

Dans le cadre de son plan de gestion, Sciensano aura une stratégie de données d'ici la fin de l'année 2022. Sciensano est un fournisseur de données de healthdata.be, mais aussi un client avec qui healthdata.be a des accords de service. La stratégie de données déterminera clairement la position et la structure de healthdata.be au sein de Sciensano pour la stratégie de données. Selon les résultats de l'analyse des lacunes, la mise en place de la stratégie de données est censée prendre au moins cinq ans. La stratégie de données rendra possible l'interopérabilité avec d'autres producteurs et utilisateurs de données pour créer de la valeur ajoutée pour la politique de santé et la recherche des données de Sciensano. De plus, elle sera en coordination avec les initiatives de l'Autorité des données de santé.

En termes d'objectifs stratégiques, Sciensano souhaite atteindre un niveau de maturité intermédiaire entre « Construction » et « Consolidation » au cours des cinq prochaines années.

Dans le cadre du plan de mise en œuvre, Sciensano prévoit deux actions :

1. La stratégie de données fait partie du plan de gestion en tant que cinquième mission de Sciensano ;
2. La position et la structure de healthdata.be doivent être déterminées au sein de Sciensano pour la stratégie de données.

1.2. LEADERSHIP ET GESTION

Sciensano mettra sur pied un comité de gouvernance des données ayant pour mission de prendre en charge les quatre dimensions de l'interopérabilité (des données, humaine, technologique et stratégique). La direction soutiendra l'ajout de nouveaux postes dans la gestion des données. Dans les descriptions de poste des analystes et scientifiques de données, l'interopérabilité sera désignée comme une exigence. Les exigences et règles en matière de gestion des données pourraient faire partie intégrante de la réglementation du travail.

En termes de leadership et de gestion, Sciensano souhaite atteindre le niveau « Consolidation » au cours des cinq prochaines années.

Dans le cadre du plan de mise en œuvre, Sciensano prévoit deux actions :

1. La création d'un comité de gouvernance des données au sein de Sciensano ;
2. La reconnaissance par la direction des quatre dimensions de l'interopérabilité comme nécessité (allocation de postes, descriptions de poste, réglementation du travail).

1.3. SUPERVISION ET RESPONSABILITÉ

Sciensano souhaite assurer la supervision et la responsabilité en créant un nouveau rôle central pour un responsable de la gouvernance des données, en allouant des ressources afin de désigner des administrateurs des données pour les différents conseils d'administration scientifiques, des coordinateurs des données pour les différents services et des points de contact des données pour les différents ensembles de données. En termes d'accès aux données et de collecte, Sciensano reconnaît la valeur d'une unité distincte pour la saisie, la validation et l'assurance qualité des données. Cela apporterait également une meilleure transparence, professionnaliserait le contrôle de la qualité et la validation des données et permettrait à Sciensano d'être reconnu comme tiers de confiance grâce à la

séparation de trois rôles : l'assurance qualité / validation des données, leur pseudonymisation et leur analyse.

La collecte et la gestion des données sont des activités coûteuses. Par conséquent, il est aussi important d'allouer suffisamment de ressources à ces activités dans les demandes de financement de la recherche.

Sciensano entend atteindre le niveau de maturité « Construction » au cours des cinq prochaines années.

Dans le cadre du plan de mise en œuvre, Sciensano prévoit deux actions :

1. La mise en place de structures de supervision pour les données de Sciensano qui ne sont pas sur healthdata.be à l'aide d'une stratégie claire pour les données personnelles et non personnelles. Un groupe (ou unité) distinct pour la saisie, la collecte, l'évaluation de la qualité et la validation des données, permettant ainsi la séparation de ces trois rôles : la saisie et la validation des données, la pseudonymisation des données et la recherche ;
2. L'allocation de ressources suffisantes pour les tâches de collecte et de gestion des données au moment d'analyser les demandes de financement de la recherche.

1.4. RESPECT DE LA LÉGISLATION

Sciensano souhaite renforcer et offrir son expertise juridique en ce qui concerne la protection des données, l'acte sur la gouvernance des données et le RGPD. Sciensano joue un rôle primordial en apportant des retours sur la reformulation et la mise à jour des lois relatives aux données.

En termes de respect de la législation, Sciensano entend se hisser au niveau « Consolidation » au cours des cinq prochaines années.

Dans le cadre du plan de mise en œuvre, Sciensano prévoit deux actions :

1. L'utilisation des connaissances juridiques et de l'expertise internes sur l'accès aux données, leur traitement et leur partage afin que Sciensano joue un rôle actif dans la mise à jour des lois actuelles relatives aux données ;
2. L'explicitation des principes des données FAIR dans les contrats et accords signés avec les fournisseurs de données et les partenaires.

1.5. ÉTHIQUE DES DONNÉES

Sciensano n'a qu'une expertise partielle en matière d'éthique des données. Au vu de l'importance de l'éthique des données, il serait judicieux de créer un poste d'éthicien des données au sein de Sciensano. Le personnel en place est éligible à ce poste.

En termes d'éthique des données, Sciensano entend atteindre le niveau de maturité « Construction » au cours des cinq prochaines années.

Dans le cadre du plan de mise en œuvre, Sciensano prévoit une action :

La création d'un comité d'éthique des données et d'un poste d'éthicien des données chez Sciensano.

1.6. APPROVISIONNEMENT

Sciensano cherche à coordonner les solutions informatiques dans l'intérêt de l'ensemble de l'organisation. En termes d'approvisionnement, Sciensano souhaite atteindre le niveau de maturité « Construction » au cours des cinq prochaines années.

Dans le cadre du plan de mise en œuvre, Sciensano prévoit une action :

La coordination de tous les approvisionnements en solutions informatiques et techniques dans l'ensemble de l'organisation.

1.7. LIENS AVEC UN ÉCOSYSTÈME DE DONNÉES PLUS LARGE

En ce qui concerne ses interactions avec un écosystème de données plus large, Sciensano vise à établir des réseaux officiels non seulement avec les universités nationales grâce à des protocoles d'accord et à des accords stratégiques, mais aussi avec d'autres organes administratifs par le biais de l'ADS et des projets de l'UE tels que PHIRI, TEHDAS et l'EEDS.

En ce qui concerne l'écosystème de données plus large, Sciensano entend atteindre un niveau intermédiaire entre « Construction » et « Consolidation » au cours des cinq prochaines années.

Dans le cadre du plan de mise en œuvre, Sciensano prévoit deux actions :

1. La formalisation des interactions au sein de l'écosystème de données grâce à des protocoles d'accord et des accords stratégiques avec les universités et les institutions (scientifiques) fédérales / régionales sur les échanges de données et les collaborations ;
2. Le statut de Sciensano comme acteur clé de la création de l'ADS et de l'EEDS.

2. Interopérabilité humaine

2.1. ADMINISTRATION DES DONNÉES

Au sein de Sciensano, deux postes relatifs à la sécurité des données ont été officiellement définis : le délégué à la protection des données et le coordinateur de sécurité des informations. En outre, il est également prévu de former des conseillers en sécurité des informations. De nouveaux postes seront alloués à la gestion des données. Des champions des données ont été désignés pour remplir un rôle moins officiel comme points de référence spécialisés sur les questions relatives aux données (visibilité des données, métadonnées, principes FAIR, questions juridiques, plan de gestion des données en ligne).

Pour chacun des conseils d'administration scientifiques, les rôles des administrateurs des données seront clairement définis, et il y aura des coordinateurs des données pour chaque service et des points de contact pour les différentes bases de données gérées dans ce service ou cette unité. Le coordinateur des données et les points de contact doivent s'assurer que les données sont faciles à trouver et accessibles, qu'il y a un journal de bord des variables dans les ensembles de données, que les métadonnées sont disponibles et que les autres départements peuvent utiliser les données. Des règles claires seront fixées pour l'accès aux données et leur utilisation ainsi que pour les règles de collaboration pour les résultats scientifiques.

Pour l'administration des données, Sciensano compte atteindre un niveau de maturité intermédiaire entre « Construction » et « Consolidation » au cours des cinq prochaines années.

Dans le cadre du plan de mise en œuvre, Sciensano prévoit une action :

La définition des rôles pour les postes d'administration des données pour permettre la gestion des données FAIR : désigner un directeur en chef des données et désigner des administrateurs des données par un conseil d'administration scientifique, des coordinateurs des données par service et des points de contact pour chaque ensemble de données.

2.2. PROTECTION DES DONNÉES À CARACTÈRE PERSONNEL ET PRÉSERVATION DE LA CONFIDENTIALITÉ

Sciensano souhaite renforcer son expertise et sa connaissance de la préservation de la confidentialité et de la vie privée afin d'aborder des problématiques telles que la limitation et la proportionnalité des données (le contrôleur des données devrait limiter la collecte d'informations personnelles aux données directement pertinentes et nécessaires pour accomplir un objectif spécifique). Afin d'améliorer sa transparence et sa fiabilité, une organisation se doit de séparer les rôles et les postes relatifs aux données (la saisie, la validation et l'évaluation de la qualité des données ; le tiers de confiance et la

pseudonymisation des données ; la recherche et l'analyse). Sciensano a déjà mis en place certaines de ces unités (le délégué à la protection des données (DPD) et le coordinateur de sécurité des informations). Cette unité pourrait être davantage renforcée par des méthodologues spécialisés dans le contrôle de la divulgation de statistiques et les techniques de préservation de la vie privée. Cette séparation des fonctions sera également davantage développée. Une unité distincte (plus technique), par laquelle les données entrent chez Sciensano, affectée à un rôle de validation et de contrôle de la qualité et un personnel doté d'une expertise méthodologique dans le domaine de l'assurance qualité et de la validation des données, permettent de professionnaliser la gestion des données au point d'entrée.

En raison des deux postes officiels qui existent pour gérer les questions de vie privée et de confidentialité, Sciensano a un plus haut niveau de maturité pour cet indicateur. Sciensano entend à cet égard atteindre le niveau de maturité « Consolidation » au cours des cinq prochaines années.

Dans le cadre du plan de mise en œuvre, Sciensano prévoit une action :

Le renforcement de la protection des données à caractère personnel et de la sécurité des informations grâce à une expertise méthodologique du contrôle de la divulgation de statistiques, des techniques de préservation de la vie privée et de l'analyse des risques.

2.3. CONNAISSANCES ET COMPÉTENCES DU PERSONNEL

Sciensano prévoit plusieurs projets de formation pour sensibiliser les chercheurs aux principes FAIR, pour approfondir ses connaissances sur la gestion et la protection des données et pour comprendre les implications pour ses chercheurs des lois et règles relatives aux données. Les chercheurs de Sciensano seront formés pour les sensibiliser sur les problèmes de protection des données et pour prévoir ces problèmes. Afin de développer davantage l'innovation et la science des données, le recrutement de scientifiques de données est considéré comme une priorité.

Pour les connaissances et compétences du personnel, Sciensano souhaite atteindre le niveau de maturité « Construction » au cours des cinq prochaines années.

Dans le cadre du plan de mise en œuvre, Sciensano prévoit une action :

La création d'une Académie Sciensano pour améliorer la formation grâce aux actions suivantes :

- Un kit de démarrage pour les chercheurs sur la gestion des données et l'analyse des risques en matière de protection de la vie privée ;
- La sensibilisation juridique aux problèmes liés aux données ;
- Une formation méthodologique (contrôle de la divulgation de statistiques, qualité et validation des données) ;
- Une formation sur la science des données.

2.4. COMMUNICATION INTERNE ET EXTERNE

Pour la communication interne et externe, Sciensano souhaite atteindre le niveau de maturité intermédiaire entre « Construction » et « Consolidation » au cours des cinq prochaines années.

Sciensano prévoit d'améliorer sa communication interne sur les principes et problématiques de données convenus et de donner l'occasion de partager les meilleures pratiques et des exemples de création de valeur par le biais de canaux de communication modernes et d'une communauté de science de données. Il y a également les « mardis EPI » et les séminaires SEP qui améliorent les connaissances et l'échange d'informations sur les principes FAIR et la maturité des données.

Par les projets de l'UE tels que PHIRI, TEHDAS et l'EEDS, Sciensano communique et partage ses expériences et exemples avec les autres dans le contexte national et international en s'engageant auprès des parties prenantes. Par ces projets de l'UE, des actions de formation sont organisées. Elles sont consacrées notamment aux principes FAIR, aux catalogues de métadonnées, à la gouvernance des données, au respect du RGPD, etc. Ces événements sont ouverts à tout le personnel de Sciensano

et sont communiqués en interne. Il s'agit d'un cadre idéal pour échanger des bonnes pratiques au sein de Sciensano, mais aussi pour attirer une expertise externe à Sciensano grâce au réseau international de l'unité HIS de l'UE.

Dans le cadre du plan de mise en œuvre, Sciensano prévoit une action :

Le développement du plan de communication interne et externe sur la stratégie de données de Sciensano par le biais de séminaires, de la communauté de science de données et des projets nationaux et européens.

2.5. ADAPTABILITÉ

La méthode Agile est une approche itérative et incrémentielle de la gestion de projets. Elle porte sur la flexibilité, la transparence, la qualité et l'amélioration continue. Elle permet de se concentrer davantage sur les besoins spécifiques et de réduire les déchets en limitant les ressources. La flexibilité accrue permet aux équipes de s'adapter facilement au changement et offre un meilleur contrôle des projets. Sciensano envisage d'adopter une méthode Agile via une approche multidisciplinaire avec une bonne vue d'ensemble dans l'équipe centrale de collecte de données. Une unité distincte de saisie / collecte des données dotée d'une expertise professionnelle et spécialisée et de méthodologues à consulter pour les méthodes génériques augmente aussi l'adaptabilité de l'organisation par le biais de l'harmonisation en cas de nouvelle crise.

Au vu des expériences vécues pendant la crise du COVID, Sciensano souhaite atteindre un niveau de maturité intermédiaire entre « Construction » et « Consolidation » pour l'adaptabilité au cours des cinq prochaines années.

Dans le cadre du plan de mise en œuvre, Sciensano prévoit une action :

S'appuyer sur les preuves et résultats et continuer à les développer et innover par l'application d'une méthode Agile.

3. Interopérabilité des données

3.1. CAPACITÉ DE MODÉLISATION DES DONNÉES ET DES MÉTADONNÉES

Pour les données absentes de healthdata.be, un inventaire exhaustif est nécessaire, incluant les données que Sciensano (hormis healthdata.be) possède et génère, qu'elles soient à caractère personnel ou non, où qu'elles soient sauvegardées, que des métadonnées soient disponibles ou non et quelle que soit l'identité de la personne de contact pour cet ensemble de données. En ce qui concerne la virtualisation des données, le département informatique a obtenu la certification ISO 27001. Le recours à des langages de programmation tels que SAS, R et Python pour l'analyse des données et l'utilisation de sources de données ouvertes devraient être encouragés pour poursuivre l'amélioration des résultats scientifiques et des services.

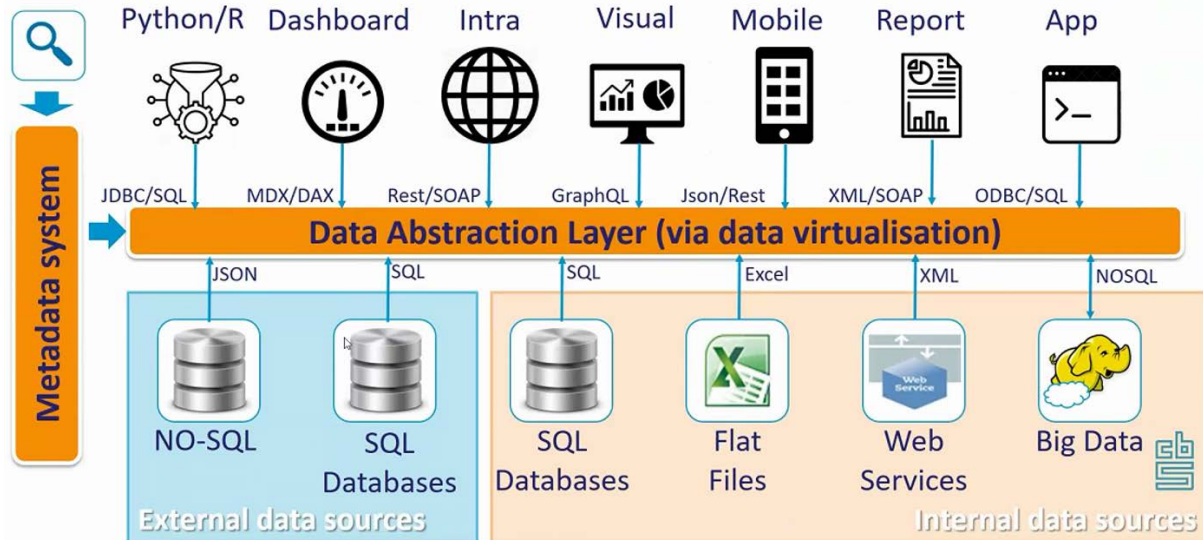
Qu'est-ce que la virtualisation des données ?

La virtualisation des données consiste à fournir une interface de service qui masque les détails techniques relatifs aux données, comme leur localisation dans la base de données, leur structure de stockage, leur accès API (point terminaux), le langage de leur requête, etc. L'interface de virtualisation répond aux besoins essentiels des utilisateurs en simplifiant l'« accès aux données » sur une plateforme Internet, la rendant efficace et entièrement opérationnelle. Tous les services indépendants sont intégrés dans une seule plateforme d'expérience utilisateur.

Une plateforme d'expérience utilisateur (UXP) est un ensemble intégré de technologies utilisées pour fournir une interaction entre un utilisateur et un ensemble d'applications, de processus, de contenus, de services ou d'autres utilisateurs. Une UXP compte plusieurs composantes, notamment les portails, la gestion de contenu, la recherche, les outils d'application Internet riche (RIA), l'analyse, la collaboration,

les outils sociaux et mobiles. Elle peut être fournie sous la forme d'une suite de produits ou d'un produit unique. Le diagramme proposé par le bureau central de la stratégie de données statistiques offre un bon aperçu du principe de virtualisation des données.

'Data at the source' as the guiding principle



*Au-dessus se trouvent toutes les solutions d'accès aux données proposées aux utilisateurs.

Comment concevoir la couche de virtualisation des données, un « magasin de données Sciensano » ?

Le magasin de données Sciensano peut être conçu comme une plateforme d'expérience numérique (DXP) où les données, les informations, le code, l'utilisation des données et le support utilisateur sont tous organisés, gérés et optimisés afin de créer une expérience utilisateur « réussie et satisfaisante ». Un service de virtualisation des données permet de créer une seule passerelle intégrée, un seul point d'entrée, un « magasin de données Sciensano » : un emplacement central où les données et les métadonnées sont disponibles, un espace de science ouverte, de données ouvertes et de collaboration. Il s'agit d'un guichet unique qui compile des services afin de faciliter l'accès aux données grâce à un traitement efficace du langage naturel et à un accès amélioré aux données affichant certaines mesures telles que les valeurs moyennes, les écarts types, le nombre d'échantillons, etc., mais aussi des options de prévisualisation telles que l'affichage des données pour certaines régions et la possibilité de télécharger des cohortes spécifiques.

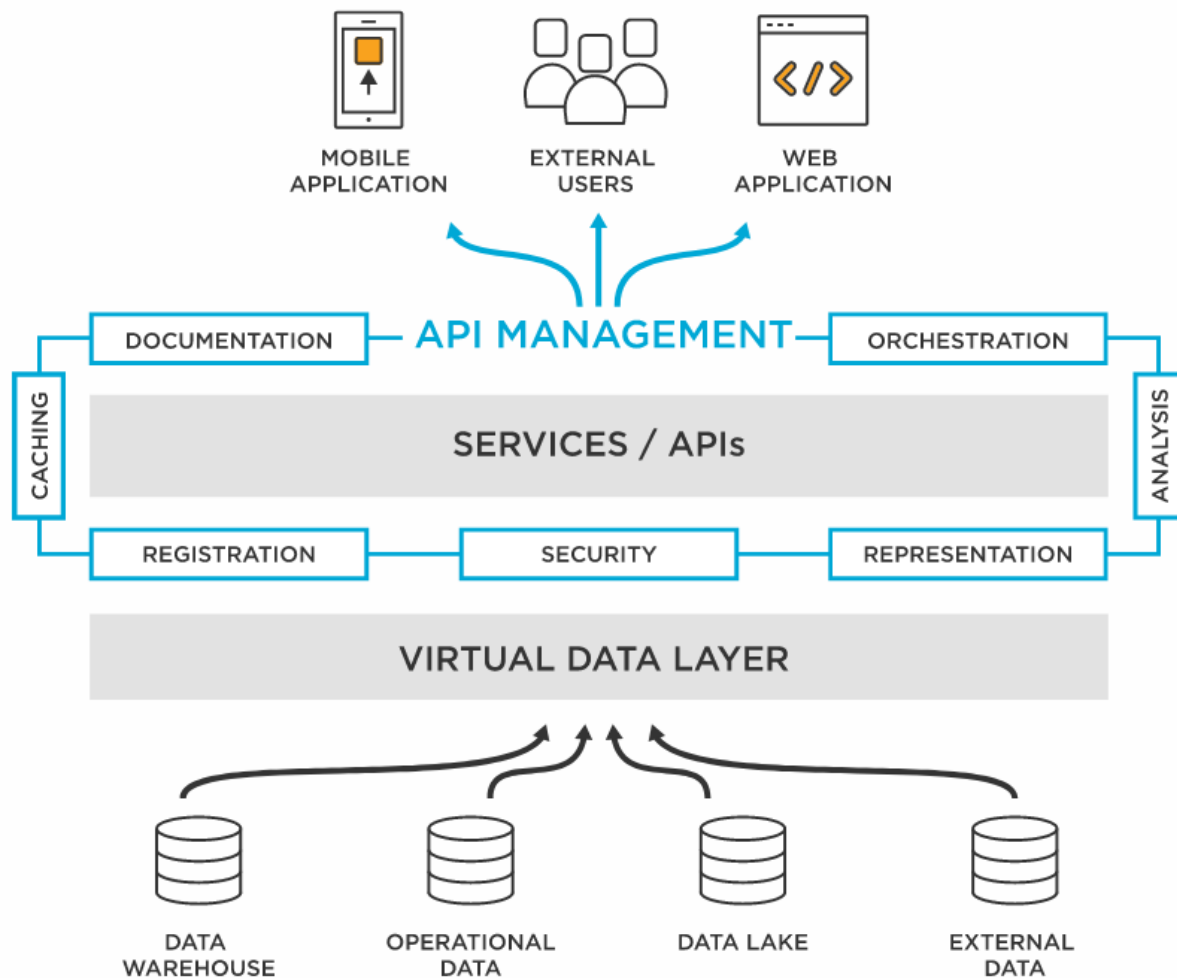
Découvrabilité (sur le diagramme : « metadata system ») : un FAIR DATA POINT (FDP)

« Le FDP est un référentiel de métadonnées qui donne accès aux métadonnées en suivant les principes FAIR (Facile à trouver, Accessible, Interopérable et Réutilisable) pour la publication des données / métadonnées. Le FDP est un logiciel qui, d'une part, permet aux propriétaires / éditeurs d'objets numériques d'exposer les métadonnées de leurs objets numériques en respectant les principes FAIR et, d'autre part, permet aux consommateurs des objets numériques de découvrir des informations (métadonnées) sur les objets numériques proposés. En général, le FAIR Data Point est utilisé pour exposer les métadonnées d'ensembles de données, mais les métadonnées d'autres types d'objets numériques peuvent aussi être exposées, par exemple les ontologies, les référentiels, les algorithmes d'analyse, les sites web, etc. »

Traitement (sur le diagramme : « Python/R » / ainsi que « dashboard » ou « report ») : JUPYTERHUB

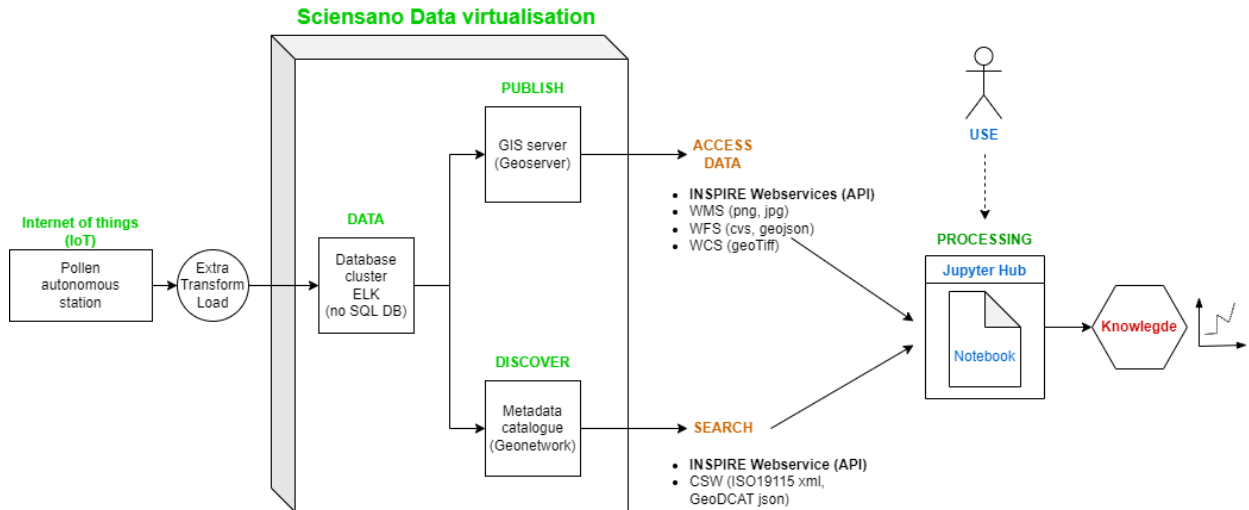
L'approche orientée utilisateur (DXP) : les utilisateurs veulent utiliser les outils avec lesquels ils sont le plus à l'aise.

« JupyterHub » met la puissance des ordinateurs portables à la portée des groupes d'utilisateurs. Il donne aux utilisateurs l'accès à des environnements et des ressources informatiques sans les encombrer par des tâches d'installation et de maintenance. Les utilisateurs, notamment les étudiants, chercheurs et scientifiques de données, peuvent effectuer leur travail dans leurs propres espaces de travail sur des ressources partagées, qui peuvent être gérées efficacement par des administrateurs du système.



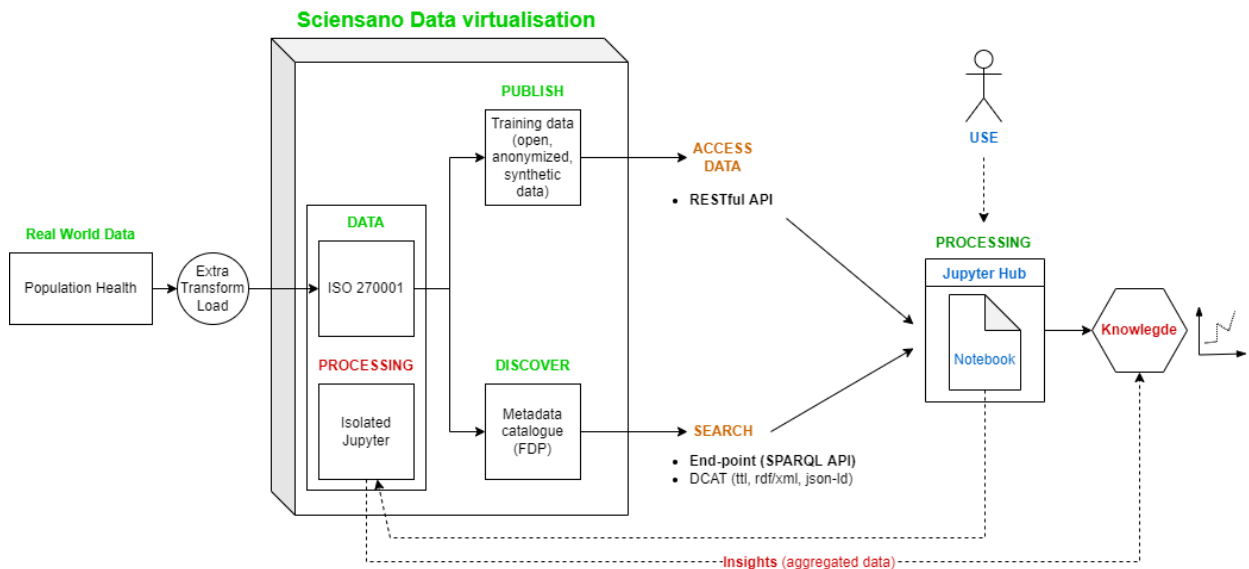
Les cas d'utilisation 1 et 2 ci-dessous servent d'exemples pour fournir des informations sur le rôle de l'équipe de virtualisation des données de Sciensano. Le premier cas d'utilisation vient du service d'aérobiologie et de mycologie. Le diagramme du cas d'utilisation 1 reflète le flux de données pour une station d'enregistrement autonome du pollen (longitude / latitude). Il montre la quantité de technologies impliquées dans le cycle de vie des données depuis la curation des données jusqu'à leur traitement et à la production d'indicateurs. Les données sont rendues accessibles par les services web de données géospatiales d'INSPIRE permettant l'interopérabilité avec d'autres infrastructures de données géospatiales et l'accès à des données environnementales.

Cas d'utilisation 1 : « Station d'enregistrement du pollen in situ »



Le second cas d'utilisation présente le principe d'un « centre de connaissances » soutenant l'accès à des données de santé sensibles. Il permet à l'utilisateur de chercher et de trouver les données grâce à un catalogue de métadonnées liées (c'est-à-dire une requête SPARQL). Il fournit un environnement informatique (à savoir JupyterHub) et donne accès à des échantillons d'entraînement (c'est-à-dire des données ouvertes, anonymisées ou synthétiques) pour programmer des analyses de données sous la forme de carnets informatiques. Un cadre d'interopérabilité technique est créé pour permettre une analyse des données (à savoir un carnet) pour être reproduit sur support informatique dans un environnement de traitement isolé et sécurisé en accédant aux données sensibles.

Cas d'utilisation 2 :



En termes de capacité de modélisation des données et des métadonnées, Sciensano compte atteindre le niveau de maturité intermédiaire entre « Construction » et « Consolidation » au cours des cinq prochaines années.

Dans le cadre du plan de mise en œuvre, Sciensano prévoit cinq actions pour un magasin de données DXP :

1. La création d'un inventaire des données ;

2. La disponibilité des métadonnées pour toutes les données de Sciensano (hormis healthdata.be) ;
3. La virtualisation des données comme infrastructure centrale pour permettre la mise en relation des données avec celles de healthdata.be et des partenaires externes ;
4. La promotion de langages de programmation comme SAS, R et Python ;
5. L'autorisation d'accéder à des ensembles de données complexes.

3.2. CAPACITÉ D'ORGANISATION ET DE CLASSIFICATION DES DONNÉES

Pour le moment, il n'existe aucun système, règlement ou procédure commun de classification. En termes de capacité d'organisation et de classification des données, Sciensano entend atteindre le niveau de maturité intermédiaire entre « Construction » et « Consolidation » au cours des cinq prochaines années.

Dans le cadre du plan de mise en œuvre, Sciensano prévoit une action :

La mise en place de systèmes de classification communs pour toutes les données de Sciensano.

3.3. ACCÈS, OUVERTURE ET PARTAGE DES DONNÉES

Pendant la crise du COVID, Sciensano a joué un rôle important pour la publication de certaines données en temps réel ainsi que pour permettre la mise en relation des données provenant de différentes institutions. Il semble également que Sciensano était très dépendant d'autres organisations pour l'accès aux données. Dans le cadre de l'ADS, Sciensano souhaite continuer à jouer un rôle primordial en rendant accessibles des données en temps réel et en s'attaquant à la liaison de différents ensembles de données entre différentes institutions. Au vu de l'augmentation du nombre de données ouvertes disponibles, il serait judicieux d'envisager la désignation d'un découvreur de données qui ferait pression de manière proactive pour obtenir de nouveaux ensembles de données gratuits ou abordables. Parallèlement, des actions juridiques devraient être prises pour changer la loi afin de faciliter l'accès aux données pour la recherche et les résultats de Sciensano, comme c'est le cas pour d'autres institutions de santé.

En termes d'accès, d'ouverture et de partage des données, Sciensano entend atteindre le niveau de maturité intermédiaire entre « Construction » et « Consolidation » au cours des cinq prochaines années.

Dans le cadre du plan de mise en œuvre, Sciensano prévoit cinq actions :

1. Des données ouvertes grâce à une plateforme de données FAIR pour Sciensano ;
2. Une procédure standard pour les accords de transfert de données ;
3. Le calcul standard des coûts par activité pour les demandes de données, le stockage, l'archivage et la sauvegarde des données privées et publiques ;
4. La désignation d'un découvreur de données ;
5. Le lancement d'une procédure pour changer la loi afin de faciliter l'accès aux données pour la recherche et les résultats de Sciensano.

3.4. ANALYSE ET AUTOMATISATION DES DONNÉES

L'ambition de Sciensano est clairement d'améliorer l'utilisation des outils et méthodes de science des données (le webscraping, la fouille de textes ou de données, l'apprentissage machine et l'apprentissage profond, l'intelligence artificielle, les techniques de science des données visant à préserver la vie privée) et d'exploiter les données (ouvertes) pertinentes préexistantes pour encore mieux remplir la mission de Sciensano en rendant l'information et la recherche politiques plus opportunes et plus détaillées afin d'examiner des questions politiques complexes.

Sciensano a accès à de vastes registres, à des enquêtes et à d'autres données massives (visuelles, d'échantillons, génomiques, chimiques) qui pourraient servir à concevoir des modèles et à calculer des probabilités (à développer une maladie par exemple), à développer des modèles d'alerte précoce ou à lier de grands ensembles de données grâce à l'analyse de réseau et des plateformes pour la détection précoce des menaces pour la santé.

De nouvelles (sources de) données ouvertes sont générées chaque jour (par exemple, les données environnementales). Les hôpitaux s'améliorent dans l'enregistrement de leurs données (dans l'objectif

de développer la surveillance syndromique) en raison des exigences relatives aux dossiers médicaux électroniques. De plus, de nouveaux systèmes qui collectent les données des pharmacies et des médecins généralistes ont été mis en place pendant la crise du COVID. Ces systèmes permettront de rendre disponibles les données en temps réel, ce qui offre une nouvelle possibilité de mise en relation des données et davantage d'études approfondies et complexes sur la santé publique.

Afin de favoriser l'innovation, Sciensano créera la communauté de science des données en 2022. Grâce à cette communauté, Sciensano réunira des experts en épidémiologie et (bio)statistique dotés de l'expertise nécessaire en science des données pour développer et promouvoir une approche méthodologique hybride à l'aide de cas d'utilisation. La communauté de la science des données facilitera également le renforcement des capacités de la communauté de recherche au sein de Sciensano. En outre, la communauté de la science des données pourrait aider à identifier des sources de données (ouvertes) (la découverte de données) et des compétences parmi nos partenaires régionaux (communauté flamande de l'IA) et nationaux pour atteindre une approche holistique One Health dans le domaine de la recherche. Le recrutement de personnes dotées des compétences adéquates et le financement actuel de la recherche en matière d'innovation aux niveaux européen et national devraient concrétiser cette innovation.

Vers un jumeau numérique de la population pour la santé

Un jumeau numérique est, en substance, une infrastructure numérique qui utilise les données du monde réel pour créer des simulations capables de prédire comment un produit, un processus ou un système va évoluer. Cette infrastructure peut intégrer l'Internet des objets (web sémantique), l'intelligence artificielle et l'analyse de logiciels pour obtenir plus de résultats.

Avec les avancées de l'apprentissage machine et des facteurs tels que les technologies de données massives, ces modèles virtuels sont devenus un élément de base de l'analyse moderne pour soutenir l'innovation et approfondir les connaissances. La création d'un jumeau numérique pour la santé peut soutenir la recherche en utilisant des capacités d'analyse, de contrôle et de prédiction avancées, des processus d'essai et des solutions.

En termes d'analyse et d'automatisation des données, Sciensano compte atteindre le niveau de maturité « Construction » au cours des cinq prochaines années.

Dans le cadre du plan de mise en œuvre, Sciensano prévoit deux actions :

1. La création d'une communauté de la science des données ;
2. L'inclusion des compétences en science des données et en ingénierie des données dans les offres d'emploi.

3.5. PROTECTION ET STOCKAGE DES DONNÉES

Le département informatique soutient plusieurs technologies de stockage et d'archivage. Les données actives sont hébergées sur la technologie de disque la plus performante (principe de la mémoire flash) et sont utilisées pour les activités les plus performantes. Les données passives sont hébergées sur un support économique plus lent. L'archivage, qui équivaut aux données hors ligne, s'effectue sur des bandes ou sur des disques à faible consommation d'énergie. La protection des données est assurée par les processus de gestion de l'identité, bien que les données en elles-mêmes peuvent être protégées par d'autres mesures de sécurité telles que le chiffrement, les fichiers protégés par mot de passe, les rôles d'application intégrés et les ordinateurs mis en quarantaine ou bien les PC virtuels. Le niveau supérieur de protection des données à atteindre est celui des réseaux classifiés séparés où les données ne peuvent pas être transférées.

Sciensano pourrait prévoir un fournisseur de services (dans le cloud) certifié ISO 27001 pour sécuriser et gérer les données de santé à caractère personnel qui répondent aux besoins des chercheurs. Sciensano pourrait également prévoir un serveur géospatial pour le stockage et la gestion

de toutes les données (non personnelles) de santé des systèmes d'information géographiques (SIG). En termes de protection et de stockage des données, Sciensano entend atteindre le niveau de maturité « Consolidation » au cours des cinq prochaines années.

Dans le cadre du plan de mise en œuvre, Sciensano prévoit deux actions :

1. La hausse du niveau de protection des données à l'aide de réseaux classifiés séparés ;
2. La prise en compte de la possibilité d'un fournisseur de serveur dans le cloud pour sécuriser les données à caractère personnel.

4. Interopérabilité technologique

4.1. INFRASTRUCTURE DES DONNÉES

En interne, le département informatique évoluera vers le niveau suivant des offres de services cloud, à savoir l'infrastructure en tant que service (IaaS), le logiciel en tant que service (SaaS) et la sécurité en tant que service (SECaaS). Le département informatique soutiendra les missions et les visions de Sciensano grâce à son statut de fournisseur légal, accepté et approuvé de données.

En termes d'infrastructure des données, Sciensano souhaite atteindre le niveau de maturité « Construction » au cours des cinq prochaines années.

Dans le cadre du plan de mise en œuvre, Sciensano prévoit une action :

Le développement des technologies de l'information au niveau supérieur de l'IaaS, du SaaS et du SECaaS en interne.

4.2. CYBERSÉCURITÉ ET RÉACTION AUX INCIDENTS

En termes de cybersécurité et de réaction aux incidents, Sciensano compte atteindre le niveau de maturité « Construction » au cours des cinq prochaines années.

Dans le cadre du plan de mise en œuvre, Sciensano prévoit une action :

Amener Sciensano au niveau supérieur de la gestion de la cybersécurité et des réactions aux incidents.

5. Instituts de santé publique orientés sur les données

5.1. LE RÔLE DES INSTITUTS DE SANTÉ PUBLIQUE DANS LA SOCIÉTÉ

Le site web de l'Association internationale des instituts nationaux de santé publique (IANPHI) cite ce qui suit au sujet du rôle des ISP et des avantages de consolider ces fonctions au sein d'un institut national de santé publique (INSP). « Pour faire le travail de santé publique, c'est-à-dire détecter, mesurer et relever les défis en matière de santé grâce à des interventions basées sur la population, chaque pays doit remplir un ensemble de fonctions qui sont la pierre angulaire de systèmes de santé publique robustes. Parmi ceux-ci, on peut citer :

- Évaluation de la santé de la population (évaluation de l'état de santé de la population) ;
- Protection de la santé (surveillance et réaction) ;
- Recherche (preuves pour éclairer les politiques et les programmes). »

De nombreux pays consolident ces fonctions sur le plan organisationnel dans un « institut national de santé publique » : une organisation scientifique (ou un réseau d'organisations) qui assure le leadership et la coordination de la santé publique au niveau national. Dans la plupart des cas, les INSP font partie du gouvernement (généralement sous le ministère de la Santé) ou y sont étroitement liées.

Les INSP du monde varient considérablement. Pourtant, malgré leurs différences dans l'histoire, la portée et les ressources, les INSP fournissent toutes des fonctions essentielles de santé publique qui améliorent les efforts de leur pays pour relever les défis de santé publique à l'intérieur et à l'extérieur de leurs frontières. La consolidation de ces fonctions ainsi que des compétences, des disciplines, de l'expérience et de l'expertise associées au sein d'un INPH offre de nombreux avantages :

- Amélioration de la prestation des **services** de santé publique ;
- Utilisation plus **efficace** des fonds ;
- Capacité de générer et de **partager des connaissances**, des données et des données probantes pour éclairer les décisions et les politiques en matière de santé publique ;
- Capacité accrue à mettre en place une **réponse rapide, décisive et coordonnée** lors d'une urgence de santé publique ;
- **Leadership national** visible pour les questions de **santé publique** ;
- Capacité d'élaborer des **programmes stratégiques en matière de santé publique** et d'allocation des ressources conformément aux propres priorités du pays ;
- **Politiques** cohérentes et **procédures harmonisées** ;
- **Liens** entre tous ceux qui œuvrent à l'amélioration de la santé publique dans le pays.

Les INSP constituent également un point central pour les ressources humaines de la santé publique et offrent un plan de carrière et un noyau de professionnels de la santé publique afin de mener à bien les fonctions essentielles de la santé publique. »

Les données sont devenues l'atout principal de nombreuses organisations. Étant donné que la plupart des institutions possèdent ou traitent des données et que les ISP ont accès aux données les plus sensibles disponibles (vu qu'elles concernent la santé), on valorise énormément leur expertise en matière non seulement de protection, de sécurité, de traitement et d'analyse des données, mais aussi de transformation des données en informations (de santé publique), de transmission aux décideurs politiques et au grand public, de technologies de l'information et d'analyse des données, ou encore d'intervention rapide et de systèmes de surveillance (des données). Ils pourraient encore mieux remplir leur rôle en devenant des instituts de recherche orientés données, en introduisant des outils modernes pour améliorer le contrôle qualité des données, en jouant le rôle de tiers de confiance, en utilisant de nouvelles sources de données comme les données (massives) ouvertes, en encourageant l'utilisation de méthodes et de techniques dans le domaine de la science des données et de l'analyse (apprentissage machine, intelligence artificielle) et en entreprenant des recherches sur des questions de santé complexes qui n'ont pas pu être abordées auparavant.

Les fonctions identifiées par l'IANPHI sont toujours pertinentes, mais il est possible de les renforcer au cours des cinq prochaines années. Compte tenu de l'évolution actuelle du paysage des données de l'UE, ces fonctions pourraient être : anticiper et identifier les questions de politiques de santé grâce à un système de recherche sur la santé, entreprendre des recherches de manière proactive et encourager la recherche et l'innovation orientées données. Les INSP pourraient en outre bénéficier de l'harmonisation des expériences, de la mobilisation de nouvelles compétences et de l'adoption de stratégies de données afin d'avoir un impact plus important sur la société.

À l'échelle nationale aussi, les ISP pourraient participer à la gestion nationale des données de santé en ayant une stratégie de données et en montrant l'exemple aux autres organisations (de la santé). Le gouvernement belge a lancé une initiative de création d'une ADS où toutes les organisations de la santé s'engagent à respecter de nouvelles réglementations et procédures quant à la mise en relation et le partage des données de santé. Les INSP devraient jouer un rôle proactif dans le soutien de ce type d'initiatives.

5.2. UE ET EXEMPLES INSPIRANTS

Il existe certains beaux exemples d'ADS dans d'autres pays de l'UE, en voici quelques-uns.

Findata est l'autorité finlandaise d'octroi des permis d'accès aux données sociales et sanitaires (*Health and Social Data Permit Authority*) et est l'autorité principale permettant l'utilisation secondaire des données sociales et sanitaires. Findata assure la gestion centralisée des traitements d'accès et de la découvrabilité des collectes de données et ensembles de données provenant de différentes sources de données par le biais d'un catalogue de métadonnées accessible au public.

L'autorité danoise des données de santé fait partie du ministère danois de la Santé. Elle est responsable de la maintenance des registres nationaux de santé, qui sont au nombre de 40. Ils contiennent des données relatives à la santé de la population danoise et aux services de soins de santé. L'autorité danoise des données de santé a créé des bibliothèques décrivant les collectes de données qu'elle gère.

Le Health Data Hub français est une structure publique dont l'objectif est de permettre aux porteurs de projet d'accéder facilement à des données non nominatives hébergées sur une plateforme sécurisée. La plateforme rassemble les copies de bases de données préexistantes dans lesquelles les données sont pseudonymisées ou anonymisées. Les données seront accessibles aux coordinateurs de projet contribuant à l'intérêt public sous réserve d'une procédure d'approbation.

CONCLUSION

Sciensano est pleinement conscient de la valeur des données de santé publique et donc de leur protection. Sciensano est parvenu à publier des données sur l'épidémie de COVID et était (et est encore) le point de référence et d'information pour les décideurs politiques et le grand public pendant toute la durée de la crise. C'est au lendemain de la crise et au lancement de l'initiative d'ADS du ministre de la Santé Frank Vandenbroucke que Sciensano a mis au point sa stratégie de données en entreprenant une évaluation de la maturité des données de la situation actuelle, en formulant un objectif de maturité des données et en élaborant un plan de mise en œuvre correspondant. Sciensano entend montrer l'exemple aux autres organisations de la santé en Belgique en disposant d'un système et une stratégie professionnels de gestion des données où toutes les composantes de l'interopérabilité des données, humaine, technologique et organisationnelle sont traitées.

L'évaluation de la maturité des données montre clairement que Sciensano est toujours aux niveaux « Émergence » et « Apprentissage » de la maturité des données et de l'interopérabilité, sauf pour certains indicateurs relatifs à la protection des données et aux aspects juridiques, où la maturité est plus élevée. Il est important de faire la distinction entre Sciensano en général d'une part et healthdata.be d'autre part. En raison de son rôle, healthdata.be a généralement une maturité des données plus élevée. Au cours des cinq prochaines années, Sciensano souhaite évoluer jusqu'aux niveaux « Construction » et « Consolidation », ce qui signifie que pour tous les indicateurs d'interopérabilité, l'organisation tout entière sera organisée de manière à ce que Sciensano puisse adhérer aux principes FAIR et puisse réutiliser ainsi que valoriser les sources de données nouvelles et préexistantes afin de fournir des informations de santé et de remplir sa mission en démontrant sa fiabilité, sa transparence, sa qualité et sa sécurité au moment de traiter les données de santé à caractère personnel.

Le plan de mise en œuvre esquisse les premières étapes pour atteindre des niveaux plus élevés de maturité des données. D'importantes actions de mise en œuvre sont la désignation de rôles de gestion des données dans l'organisation (responsable de la gouvernance des données et administrateur des données), la création d'un inventaire exhaustif des données, la garantie de la description des métadonnées de toutes les données de Sciensano, l'augmentation de la capacité informatique pour la virtualisation des données et l'insertion de la stratégie de données dans le plan de gestion comme cinquième mission stratégique.

Depuis sa création en 2018 et grâce à son rôle (pro)actif pendant la crise du COVID, Sciensano a reçu beaucoup d'attention et de questions sur son rôle dans la société. Sciensano est un institut de santé publique qui, non seulement fournit des conseils et des services au gouvernement et au grand public, mais entreprend aussi des recherches sur la santé, constituant le fondement de ses autres activités. De par ces activités, Sciensano entretient un vaste réseau de collaborations avec des universités, des instituts de connaissances et d'autres organismes de santé par lesquelles les données sont échangées et partagées. C'est dans ce rôle que Sciensano souhaite adopter une approche innovante plus proactive pour devenir une organisation orientée données qui, par son expertise sur le stockage, le traitement, la protection et l'analyse des données pour assurer un suivi de la santé et fournir des informations sur la santé, souhaite montrer l'exemple dans le paysage belge des données et accroître son influence sur la société.

RÉFÉRENCES

- (1) https://www.data4sdgs.org/sites/default/files/file_uploads/Joined_Up_Data_Maturity_Assessment_draft5.pdf
- (2) Saw Swee Hock School of Public Health et Asia Society. *White Paper on Responsible Data Sharing in Health and Healthcare*, 2022.
- (3) Commission européenne. *Bâtir l'avenir numérique de l'Europe*, <https://digital-strategy.ec.europa.eu/fr/policies/strategy-data>, 2022.
- (4) Sciensano. <https://covid-19.sciensano.be/fr/missions-de-sciensano>, 2022.
- (5) DAMA International. *Data Management Body of Knowledge (2nd Edition)*, New Jersey: Technics Publications, 2017.
- (6) Robert Chang. *A Beginner's Guide to Data Engineering — Part II*, medium.com/@rchang/a-beginners-guideto-data-engineering-part-ii-47c4e7cbda71, 2018.

ANNEXES

Annexe 1 • Évaluation de la maturité des données de Sciensano 2022

Critère d'opérabilité	Niveau de maturité des données de Sciensano	Explication
INTEROPÉRABILITÉ ORGANISATIONNELLE		
Objectifs stratégiques	Émergence / Apprentissage	<p>La capacité de mettre en relation les données est reconnue, mais n'est pas explicitement identifiée en tant qu'objectif stratégique.</p> <p>L'interopérabilité est identifiée comme objectif stratégique dans les unités techniques de l'organisation, mais pas en dehors.</p>
Leadership et gestion	Émergence / Apprentissage	<p>Un leadership ad hoc sur les problèmes d'interopérabilité émerge organiquement, mais n'est pas coordonné.</p> <p>Un leadership sur l'interopérabilité émerge dans différentes unités techniques, mais reste fragmenté.</p>
Supervision et responsabilité	Émergence / Apprentissage	<p>L'organisation est consciente de la nécessité de créer des chaînes de responsabilité pour superviser la manière dont les données sont mises en relation, mais n'a pas encore pris de mesures pour les créer.</p> <p>Les structures de supervision émergent dans différentes unités techniques, mais ne sont ni coordonnées ni harmonisées. La responsabilité quant à la manière dont les efforts d'interopérabilité sont entrepris est fragmentée et manque de clarté.</p>
Respect de la législation	Apprentissage / Construction	<p>Des mesures actives sont prises pour mieux comprendre les exigences du respect de la législation relative à la conservation, la transmission, le partage et l'utilisation des données et pour rendre les données accessibles pour les utilisateurs.</p> <p>Le respect des lois applicables relatives au partage, à la transmission et à l'utilisation des données est intégré aux fonctions de supervision et de responsabilité, et est reflété dans la stratégie de données de l'organisation, qui est publiée en ligne.</p>
Éthique des données	Apprentissage	<p>Des mesures actives sont prises pour mieux comprendre les répercussions éthiques que la mise en relation des données peut avoir et pour comprendre comment elle peut causer des dommages sans le vouloir. Des évaluations rudimentaires des répercussions éthiques sont entreprises, mais uniquement pour répondre à des besoins spécifiques.</p>
Approvisionnement	Indéterminé / Émergence	<p>L'organisation n'est pas consciente de l'impact de l'approvisionnement en solutions techniques et de données sur l'interopérabilité.</p>

		L'organisation se rend de plus en plus compte de la nécessité de mettre en relation les données entre les activités d'approvisionnement, mais il n'y a pas d'approche cohérente. Il y a un recours fréquent aux contractants externes pour combler les lacunes, mais aucune coordination entre les différentes unités sur la manière de le faire, résultant souvent en l'approvisionnement de solutions de données incompatibles.
Liens avec un écosystème de données plus large	Apprentissage	L'organisation commence à assortir les données qu'elle publie, transmet ou partage de conditions de licence pro forma comportant des dispositions relatives à l'intégration des données, mais ne surveille pas les utilisateurs des données et ne collabore pas avec eux. L'organisation commence à enregistrer les données qu'elle reçoit des autres organisations. L'organisation commence à enregistrer et coordonner ses collaborations avec d'autres parties de l'écosystème de données.
INTEROPÉRABILITÉ HUMAINE		
Administration des données	Émergence / Apprentissage	La mise en relation des données fait partie des fonctions de certains membres du personnel, mais elle n'est pas reprise dans la description de leur poste et est ad hoc. Le personnel des différentes unités a des tâches reconnues liées à l'interopérabilité des données, mais il y a peu ou pas de coordination centrale.
Protection des données à caractère personnel et préservation de la confidentialité	Construction	Le droit applicable, les principes, les meilleures pratiques et les conseils sont régulièrement pris en compte. L'organisation entreprend des évaluations d'impact sur la vie privée avant et pendant les projets relatifs aux données et ces évaluations incluent des considérations quant aux risques liés aux données.
Connaissances et compétences du personnel	Émergence / Apprentissage	Certains employés ont les connaissances et compétences requises pour mettre en relation des données, mais elles ne sont pas reprises dans la description de leur poste et sont secondaires par rapport à leurs tâches principales. Les connaissances et compétences pertinentes pour l'interopérabilité commencent à être reconnues comme partie intégrante de la description des postes dans certaines unités, mais l'approche est fragmentée.
Communication interne et externe	Apprentissage	Le personnel et les unités commencent à partager entre eux des exemples de bonnes pratiques, mais la démarche n'est pas coordonnée. La valeur de l'interopérabilité commence à être comprise par le personnel non technique, mais n'est pas encore communiquée en externe.
Adaptabilité	Apprentissage / Construction	Des unités disséminées dans l'organisation commencent à reconnaître officiellement la nécessité de veiller à ce que les fonctions du personnel et la supervision des systèmes de données soient adaptables afin de garantir que leurs données continuent de générer de la valeur.

		L'organisation reconnaît et indique dans sa stratégie de données la valeur de la responsabilisation des employés pour qu'ils s'adaptent dans leur façon d'utiliser les données et de les mettre en relation avec d'autres. Le personnel est en mesure d'adapter ses méthodes de travail et de superviser les données de l'organisation de manière à en accroître la valeur, y compris en les mettant en relation.
INTEROPÉRABILITÉ DES DONNÉES		
Capacité de modélisation des données et des métadonnées	Émergence / Construction	L'organisation se rend de plus en plus compte de la valeur que la modélisation des données et métadonnées peut conférer aux données, mais la modélisation des données n'est pas une priorité pour les unités techniques. Les unités techniques coordonnent leur approche de la modélisation des données et métadonnées et s'efforcent de modéliser les données de manière cohérente en fonction des besoins de l'organisation. Les besoins internes sont encore prioritaires par rapport aux groupes externes, mais les données sont modélisées de manière cohérente.
Capacité d'organisation et de classification des données	Apprentissage	Certaines unités essaient de manière non officielle d'utiliser des classifications communes, mais celles-ci ne sont ni officialisées ni coordonnées entre toutes les unités concernées. Des classifications communes sont parfois utilisées dans l'organisation, mais ce n'est pas fait de manière cohérente.
Accès, ouverture et partage des données	Apprentissage / Construction	Des efforts coordonnés sont déployés dans les unités techniques pour veiller à ce que les données soient accessibles et partagées de manière responsable. Il s'agit notamment d'autorisations de licence ou de limitations pertinentes pour l'intégration et l'utilisation futures des données. Certaines données sont publiées sur une plateforme de l'organisation, mais les ensembles de données sont incomplets, inopportuns ou ne sont pas passés par un contrôle qualité. Les données sont partagées de manière responsable de façon à protéger les droits que les tiers peuvent avoir sur ces données. Les données publiées ouvertement le sont dans des formats lisibles par une machine sous une licence de données ouvertes claire avec des conditions d'utilisation. Ces données sont aussi dépourvues de caractéristiques pouvant servir à la réidentification d'individus ou de groupes vulnérables. Les portails de données ouvertes sont dotés d'informations contextuelles pertinentes et sont visualisés de manière à encourager de nombreux publics à les utiliser.
Analyse et automatisation des données	Émergence	La compréhension du rôle de l'interopérabilité dans l'analyse automatisée des données varie d'une unité à l'autre dans l'organisation. La compréhension de la manière dont les données interopérables devraient être utilisées pour former les algorithmes est limitée.

Protection et stockage des données	Apprentissage / Construction	<p>Des unités distinctes appliquent régulièrement des techniques de protection des données appropriées à leurs ensembles de données avant l'intégration des données, mais il n'y a que peu ou pas de cohérence dans la manière de les appliquer. Les risques de réidentification inhérents aux données interopérables font l'objet d'une certaine compréhension, mais celle-ci est limitée.</p> <p>Les données personnelles, sensibles et de groupes vulnérables font l'objet de protections adéquates avant d'être intégrées, partagées ou traitées à l'aide d'outils d'analyse automatisés. Les risques de réidentification inhérents aux données interopérables sont compris et appliqués, mais pas de manière systématique.</p>
INTEROPÉRABILITÉ TECHNOLOGIQUE		
Infrastructure des données	Apprentissage	Tous les membres du personnel concernés ont accès à des outils matériels et logiciels adéquats, ainsi qu'à une connexion réseau. Il existe des serveurs et des référentiels de données sécurisés, mais ils sont utilisés de manière incohérente par le personnel et les unités de l'organisation ; l'infrastructure numérique est peu surveillée.
Cybersécurité et réaction aux incidents	Émergence	Les différents membres du personnel et les différentes unités de l'organisation sont conscients des risques que représentent leurs données réutilisables en cas de cyberattaque ou d'autre violation de données, ou du moins s'en inquiètent. Des défenseurs émergent et font pression en faveur d'un protocole ou d'une politique de lutte contre la violation de données.

Annexe 2 • Maturité des données de Sciensano : niveaux de maturité pour le futur

Critère d'opérabilité	Niveau de maturité des données de Sciensano	Explication
INTEROPÉRABILITÉ ORGANISATIONNELLE		
Objectifs stratégiques	Construction / Consolidation	<p>La nécessité de mettre en relation les données entre les systèmes est reconnue comme objectif stratégique dans la stratégie de données de l'organisation. La valeur des normes et d'une solide gouvernance des données est reconnue.</p> <p>La valeur stratégique que des données mises en relation peuvent apporter à la prise de décision est reconnue dans les stratégies de l'organisation. L'interopérabilité fait partie de la stratégie de collaboration externe de l'organisation avec d'autres producteurs et utilisateurs de données.</p> <p><i>Actions de mise en œuvre :</i></p> <ol style="list-style-type: none"> 1. La stratégie de données fait partie du plan de gestion en tant que cinquième mission de Sciensano ; 2. La position et la structure de healthdata.be doivent être déterminées au sein de Sciensano pour la stratégie de données.
Leadership et gestion	Consolidation	<p>Il existe un comité ou conseil de gouvernance des données et il a pour mission explicite de prendre en charge les questions d'interopérabilité. La valeur de la mise en relation des données est comprise par les dirigeants et les responsables de l'organisation et est clairement identifiée comme tâche dans les descriptions de poste concernées.</p> <p><i>Actions de mise en œuvre :</i></p> <ol style="list-style-type: none"> 1. La création d'un comité de gouvernance des données au sein de Sciensano ; 2. La reconnaissance par la direction des quatre dimensions de l'interopérabilité comme nécessité (allocation de postes, descriptions de poste, réglementation du travail).
Supervision et responsabilité	Construction	<p>Les missions de contrôle et de responsabilité sont intégrées dans la stratégie de l'organisation et se reflètent dans la structure de direction. Les chaînes fragmentées commencent à se rejoindre et des procédures opérationnelles standardisées communes apparaissent.</p> <p><i>Actions de mise en œuvre :</i></p> <ol style="list-style-type: none"> 1. La mise en place de structures de supervision pour les données de Sciensano qui ne sont pas sur healthdata.be à l'aide d'une stratégie claire pour les données à caractère personnel ou non : un groupe (ou unité) distinct responsable de la saisie, de la collecte, de l'évaluation de la qualité et de la validation des données, permettant ainsi la séparation des trois rôles, à savoir la saisie et la validation des données, la pseudonymisation des données et la recherche ; 2. L'allocation de ressources suffisantes pour les tâches de collecte et de gestion des données au moment d'analyser les demandes de financement de la recherche.
Respect de la législation	Consolidation	<p>Les activités de transmission, de partage et d'utilisation des données de l'organisation sont pleinement conformes aux lois applicables et surpassent parfois les normes juridiques.</p>

		<p><i>Actions de mise en œuvre :</i></p> <ol style="list-style-type: none"> 1. L'utilisation des connaissances juridiques et de l'expertise internes sur l'accès aux données, leur traitement et leur partage afin que Sciensano joue un rôle actif dans la mise à jour des lois actuelles relatives aux données ; 2. L'explicitation des principes des données FAIR dans les contrats et accords signés avec les fournisseurs de données et les partenaires.
Éthique des données	Construction	<p>Les questions éthiques que la mise en relation des données pourrait soulever sont comprises et classées de manière appropriée. Des mesures adaptées sont prises pour réduire les dommages causés par une violation des normes éthiques. Des évaluations de l'impact éthique sont régulièrement effectuées.</p> <p><i>Action de mise en œuvre :</i> La création d'un comité d'éthique des données et d'un poste d'éthicien des données chez Sciensano.</p>
Approvisionnement	Construction	<p>L'approvisionnement en systèmes de données compatibles et interopérables dans l'organisation est officialisé et coordonné. Il existe une politique commune d'approvisionnement au sein de l'organisation qui nécessite que le personnel prenne en compte les questions d'interopérabilité au moment d'acheter de nouveaux systèmes. Le recours à des contractants externes est stratégique et coordonné.</p> <p><i>Action de mise en œuvre :</i> La coordination de tous les approvisionnements en solutions informatiques et techniques dans l'ensemble de l'organisation.</p>
Liens avec un écosystème de données plus large	Construction / Consolidation	<p>L'organisation classe efficacement ses données et les soumet à des licences d'utilisation appropriées. L'organisation répertorie toutes les données qui sont partagées avec elle et a une compréhension générale de ce qu'elle peut et ne peut pas faire avec ces données. L'organisation collabore avec d'autres parties de l'écosystème des données de manière coordonnée, conformément à sa stratégie de données.</p> <p>Il existe un ensemble de licences bien établies et sur mesure qui définissent des paramètres clairs pour l'utilisation, notamment pour l'intégration en fonction de la catégorie de données concernée. L'organisation répertorie toutes les données qui sont partagées avec elle et dispose d'orientations et de procédures claires qui déterminent si et comment ces données peuvent être associées à d'autres ensembles de données qu'elle contrôle. L'organisation fait de la collaboration avec d'autres parties de l'écosystème des données une priorité stratégique et dispose d'une approche bien coordonnée avec des procédures claires pour mettre en relation ses données avec des données externes.</p> <p><i>Actions de mise en œuvre :</i></p> <ol style="list-style-type: none"> 1. La formalisation des interactions au sein de l'écosystème de données grâce à des protocoles d'accord et des accords stratégiques avec les universités et les institutions (scientifiques) fédérales / régionales sur les échanges de données et les collaborations ; 2. Le statut de Sciensano comme acteur clé de la création de l'ADS et de l'EEDS.
INTEROPÉRABILITÉ HUMAINE		

Administration des données	Construction / Consolidation	<p>Il existe un plan de recrutement coordonné qui reflète les différentes dimensions et rôles liés à l'administration des données, notamment l'interopérabilité, dans l'ensemble de l'organisation. Les unités sont coordonnées et communiquent entre elles au sujet des données qu'elles intègrent et de la manière et de la raison de le faire.</p> <p>Un plan stratégique de gestion des données est supervisé par le conseil ou le comité de gouvernance des données de l'organisation et comprend un plan clair d'administration des données, incluant les missions relatives à l'interopérabilité des données. Le personnel de l'organisation est conscient de la manière dont les données sont utilisées, mises en relation et partagées avec d'autres entités.</p> <p><i>Action de mise en œuvre :</i> La définition des rôles pour les postes d'administration des données pour permettre la gestion des données FAIR : désigner un directeur de la gouvernance des données et désigner des administrateurs des données par un conseil d'administration scientifique, des coordinateurs des données par service et des points de contact pour chaque ensemble de données.</p>
Protection des données à caractère personnel et préservation de la confidentialité	Consolidation	<p>La préservation de la vie privée et de la confidentialité des données fait partie de l'examen juridique et éthique de l'organisation et est intégrée dans le cycle de vie des données. L'organisation adhère aux normes applicables les plus élevées en matière de préservation de la vie privée et de la confidentialité. L'organisation intègre la préservation de la vie privée et de la confidentialité dans le cadre de sa stratégie de données et prodigue explicitement des conseils sur les risques liés aux données interopérables, comme l'effet mosaïque. L'organisation est tournée vers l'avenir et est consciente des risques potentiels pour la vie privée inhérents aux données interopérables et résultant des technologies émergentes telles que l'Internet des objets, l'authentification biométrique ou les processus automatisés en général.</p> <p><i>Action de mise en œuvre :</i> Le renforcement de la protection des données à caractère personnel et de la sécurité des informations grâce à une expertise méthodologique du contrôle de la divulgation de statistiques, des techniques de préservation de la vie privée, des SCRA et de l'analyse des risques.</p>
Connaissances et compétences du personnel	Construction	<p>Il existe une approche coordonnée du renforcement des connaissances et des compétences au sein de l'organisation, qui reconnaît explicitement les besoins en matière d'interopérabilité et y répond.</p> <p><i>Actions de mise en œuvre :</i> La création d'une Académie Sciensano pour offrir une formation grâce aux actions suivantes :</p> <ol style="list-style-type: none"> 1. Un kit de démarrage pour les chercheurs sur la gestion des données et l'analyse des risques en matière de protection de la vie privée ; 2. La sensibilisation juridique aux problèmes liés aux données ; 3. Une formation méthodologique (contrôle de la divulgation de statistiques, qualité et validation des données) ; 4. Une formation sur la science des données.
Communication interne et externe	Construction / Consolidation	<p>Les mécanismes destinés à faciliter la communication interne et le partage des bonnes pratiques en matière d'interopérabilité font partie de la stratégie de données de l'organisation. La communication entre</p>

		<p>les unités permet de traduire les bonnes pratiques et les exemples de création de valeur pour les publics externes.</p> <p>L'organisation dispose de divers canaux de communication interne coordonnés entre les unités et le personnel, permettant le partage des bonnes pratiques et des exemples de création de valeur.</p> <p>L'organisation se fait le champion de la valeur des données mises en relation avec les écosystèmes de données et communique activement ses expériences et ses exemples à d'autres de manière convaincante et efficace, notamment en collaborant avec des journalistes de données et des écrivains.</p> <p><i>Action de mise en œuvre :</i> Le développement du plan de communication interne et externe sur la stratégie de données de Sciensano par le biais de la communauté de science de données et des projets nationaux et européens.</p>
Adaptabilité	Construction / Consolidation	<p>L'organisation reconnaît et indique dans sa stratégie de données la valeur de la responsabilisation des employés pour qu'ils s'adaptent dans leur façon d'utiliser les données et de les mettre en relation avec d'autres. Le personnel est en mesure d'adapter ses méthodes de travail et de superviser les données de l'organisation de manière à en accroître la valeur, y compris en les mettant en relation.</p> <p>L'organisation devient un leader en matière de gestion adaptable et le personnel se sent responsabilisé et croit en sa capacité à adapter sa supervision des systèmes de données en fonction des besoins, notamment sa méthode de mise en relation des données, afin d'en maximiser la valeur.</p> <p><i>Action de mise en œuvre :</i> S'appuyer sur les preuves et résultats et continuer à les développer et innover en utilisant une méthode Agile.</p>
INTEROPÉRABILITÉ DES DONNÉES		
Capacité de modélisation des données et des métadonnées	Construction / Consolidation	<p>Les unités techniques coordonnent leur approche de la modélisation des données et métadonnées et s'efforcent de modéliser les données de manière cohérente en fonction des besoins de l'organisation. Les besoins internes sont encore prioritaires par rapport aux groupes externes, mais les données sont modélisées de manière cohérente.</p> <p>L'organisation utilise régulièrement des modèles de données et de métadonnées conformes aux modèles standard, rendant leur réutilisation possible et permettant le partage de données. La sélection et l'utilisation de modèles standard sont soigneusement planifiées, notamment par la collaboration avec des utilisateurs de données et d'autres entités dans l'écosystème de données.</p> <p><i>Actions de mise en œuvre pour un magasin de données DXP :</i></p> <ol style="list-style-type: none"> 1. La création d'un inventaire des données 2. La disponibilité des métadonnées pour toutes les données de Sciensano (hormis healthdata.be) ; 3. La virtualisation des données comme infrastructure centrale pour permettre la mise en relation des données avec celles de healthdata.be et des partenaires externes ; 4. La promotion de langages de programmation comme R et Python ; 5. L'autorisation d'accéder à des ensembles de données complexes.
Capacité d'organisation et	Construction / Consolidation	Il existe une approche coordonnée de l'utilisation des classifications de données dans l'organisation. Les unités collaborent pour identifier

de classification des données		<p>les classifications les plus adéquates pour leurs données et s'assurent que les données dont elles sont responsables sont classifiées de manière appropriée.</p> <p>L'organisation non seulement utilise des classifications de données de manière régulière et appropriée, mais elle produit aussi ses propres classifications pour combler les lacunes et assurer la cohérence. L'organisation collabore activement avec d'autres entités de l'écosystème de données afin d'améliorer les systèmes de classification utilisés en général et d'en établir de nouveaux si nécessaire. L'organisation communique de manière efficace la valeur pour l'interopérabilité d'une classification cohérente des données.</p> <p><i>Action de mise en œuvre :</i> La mise en place de systèmes de classification communs pour toutes les données de Sciensano.</p>
Accès, ouverture et partage des données	Construction / Consolidation	<p>Les données sont partagées de manière responsable de façon à protéger les droits que les tiers peuvent avoir sur ces données. Les données publiées ouvertement le sont dans des formats lisibles par une machine sous une licence de données ouvertes claire avec des conditions d'utilisation. Ces données sont aussi dépourvues de caractéristiques pouvant servir à la réidentification d'individus ou de groupes vulnérables. Les portails de données ouvertes sont dotés d'informations contextuelles pertinentes et sont visualisés de manière à encourager de nombreux publics à les utiliser.</p> <p>L'organisation dispose d'une politique de partage des données efficace qui fournit des conseils sur les différentes manières dont le partage des données devrait avoir lieu, de la publication sous licence libre à l'utilisation d'accords de partage ou de traitement des données. Des conseils juridiques sont disponibles pour le personnel souhaitant partager des données qui seront intégrées par des tiers à d'autres ensembles de données. Les données ouvertes ne sont pas seulement publiées dans des formats lisibles par la machine et par l'homme, mais sont également mises à disposition sous forme de données liées sur le web sémantique. Il existe des boucles de feedback avec les principaux groupes d'utilisateurs, et l'organisation est réactive aux besoins des utilisateurs.</p> <p><i>Actions de mise en œuvre :</i></p> <ol style="list-style-type: none"> 1. Des données ouvertes grâce à des portails et une plateforme de données FAIR pour Sciensano ; 2. Des procédures standard pour les accords de transfert de données ; 3. Le calcul standard des coûts par activité pour les demandes de données, le stockage, l'archivage et la sauvegarde des données privées et publiques ; 4. La désignation d'un découvreur de données ; 5. Le lancement d'une procédure pour changer la loi afin de faciliter l'accès aux données pour la recherche et les résultats de Sciensano.
Analyse et automatisation des données	Construction	<p>Les tâches d'analyse des données et d'apprentissage machine sont reprises dans la stratégie de données de l'organisation. Les avantages et les risques relatifs de l'exécution d'analyses automatisées sur des données interopérables, ou de leur utilisation pour former des algorithmes, sont généralement compris, mais il n'existe pas encore d'approche cohérente au sein de l'organisation.</p> <p><i>Actions de mise en œuvre :</i></p>

		<ol style="list-style-type: none"> 1. La création d'une communauté de la science des données ; 2. L'inclusion des compétences en science des données et en ingénierie des données dans les offres d'emploi.
Protection et stockage des données	Consolidation	<p>Toutes les données sont protégées à l'aide des techniques appropriées et sont soit archivées de manière responsable soit définitivement supprimées à la fin de leur cycle de vie prévu. L'accès aux ensembles de données sensibles est contrôlé et enregistré afin de garantir la responsabilité en matière de protection des données. Avant l'intégration, le partage ou le traitement des données au moyen d'analyses automatisées, les risques de réidentification ou d'autres dommages sont évalués et les données ne sont utilisées qu'en cas de forte certitude qu'elles resteront protégées après leur réutilisation. L'organisation contribue à établir des normes de protection des données au sein de l'écosystème de données plus large et défend l'idée d'une utilisation responsable des données.</p> <p><i>Actions de mise en œuvre :</i></p> <ol style="list-style-type: none"> 1. Hausse du niveau de protection des données à l'aide de réseaux classifiés séparés où les données ne peuvent pas être transférées ; 2. La prise en compte de la possibilité d'avoir un fournisseur de serveur dans le cloud pour sécuriser les données à caractère personnel.
INTEROPÉRABILITÉ TECHNOLOGIQUE		
Infrastructure des données	Construction	<p>Tous les membres du personnel concernés savent comment utiliser les systèmes de gestion et de traitement des données de l'organisation et y sont formés. Les membres du personnel utilisent régulièrement des serveurs et des référentiels de données sécurisés, et la supervision de l'infrastructure numérique fait partie de la stratégie de données de l'organisation.</p> <p><i>Action de mise en œuvre :</i> Le développement des technologies de l'information au niveau supérieur de l'IaaS, du SaaS et du SECaaS en interne.</p>
Cybersécurité et réaction aux incidents	Construction	<p>Une politique claire en matière de violation des données est établie, définissant les étapes à suivre et les responsabilités. Le personnel reçoit une formation sur ce qu'il doit faire en cas de violation des données et est informé des risques associés à la réutilisation de données interopérables volées.</p> <p><i>Action de mise en œuvre :</i> Amener Sciensano au niveau supérieur de la gestion de la cybersécurité et des réactions aux incidents.</p>

CONTACT

Sofie De Broe • T+32 2 478 843194 • sofie.debroe@sciensano.be

POUR PLUS D'INFORMATIONS

—
Rendez-vous sur notre site
Internet www.sciensano.be
ou contactez-nous à l'adresse