



Berne, le 8 novembre 2023

Science des données pour le bien commun – cas d'application dans l'élaboration des politiques

Dans le cadre de la mise en œuvre de la stratégie de la
Confédération en matière de science des données

Table des matières

Résumé	3
1 Introduction	5
2 Observations confirmées	6
3 Documentation des cas d’application.....	10
3.1 Vue d’ensemble.....	10
3.2 Utilisation de renseignements d’origine sources ouverts (<i>open source intelligence</i> , ROSO).....	10
3.3 Tableaux globaux de la situation en cas de crise	17
3.4 Aide à la surveillance basée sur les données.....	24
3.5 Aide à la décision basée sur des systèmes de modèles holistiques	31
3.6 Contrôles de plausibilité destinés à garantir la qualité des données	37
3.7 Jumeaux numériques	43
4 Annexe.....	52
4.1 La science des données dans l’élaboration des politiques.....	52
4.2 Élaboration du rapport.....	53
4.2.1 Démarche.....	53
4.2.2 Caractéristiques d’un cas d’application	54
4.2.3 Membres du groupe d’accompagnement et des groupes de travail.....	55

Résumé

Le Département fédéral de l’intérieur (DFI) a été chargé de rédiger, en collaboration avec les autres départements et le domaine des EPF, un rapport sur l’application de la science des données pour le bien commun en prenant des cas d’application propres à l’administration fédérale et s’inscrivant dans le processus d’élaboration des politiques. Le rapport doit servir, d’une part, à montrer le potentiel de la science des données au sein de l’administration fédérale et, d’autre part, de déterminer les conditions appropriées à son utilisation transversale à grande échelle.

Lors de plusieurs séances de travail réunissant des représentants de l’administration fédérale et du domaine des EPF, les six cas d’application suivants ont été sélectionnés et analysés: 1) utilisation de renseignements d’origine sources ouvertes (ROSO); 2) tableaux globaux de la situation en cas de crise; 3) aide à la surveillance basée sur les données; 4) aide à la décision basée sur des systèmes de modèles holistiques; 5) contrôles de plausibilité destinés à garantir la qualité des données; 6) jumeaux numériques. Les cas sélectionnés ne prétendent nullement à l’exhaustivité ou à un ordre des priorités. Ils ont surtout un caractère exemplaire et couvrent un éventail aussi large que possible des méthodes de la science des données et des domaines d’activités de l’administration fédérale.

L’élaboration de ces différents cas a non seulement permis d’acquérir de nouvelles connaissances, mais a aussi confirmé les mesures identifiées dans la stratégie de la Confédération en matière de science des données (DSStB). Les cas d’application soulignent en particulier le grand potentiel de la science des données pour l’administration fédérale. Si certaines unités administratives recourent d’ores et déjà à la science des données, son vaste potentiel demeure dans l’ensemble inexploité au sein de l’administration fédérale. Pour étendre son utilisation, également au niveau transversal, il importe de créer ou d’améliorer les conditions appropriées dans différents champs d’action. Les principales conditions appropriées pour recourir à la science des données comprennent notamment les éléments suivants: la volonté et l’appui de la hiérarchie, la formation et la formation continue en science des données, le recrutement de personnel qualifié, la disponibilité et l’interopérabilité des données (en particulier grâce à la promotion d’espaces de données dignes de confiance et interopérables), une législation qui facilite l’utilisation de l’aide à la décision basée sur les données et la prise en compte de méthodes de l’intelligence artificielle (IA) parmi les principaux outils dans l’application de la science des données. Le pilotage et la coordination (gouvernance) de la science des données au sein de l’administration fédérale revêtent une importance particulière.

La création du cadre nécessaire à l’utilisation de la science des données, telles qu’identifiées ci-dessus, a déjà commencé au sein de l’administration fédérale dans le cadre de projets déjà en cours, tels les mandats destinés à suivre la mise en œuvre de la DSStB. En toute logique, les résultats du présent rapport sont dès lors pris en compte dans leur traitement. Ces projets comptent notamment l’évaluation du Centre de compétence en science des données (DSCC) et celle du Réseau de compétence en intelligence artificielle (CNAI), l’élaboration d’un code de bonnes pratiques pour une science des données centrée sur l’être humain et digne de confiance ainsi que la mise au point d’une «plateforme collaborative consacrée à la science des données».

Ces différents éléments sont présentés dans le graphique ci-dessous.

La science des données recèle un potentiel énorme, qui reste inexploité

Si certaines unités administratives recourent d’ores et déjà avec beaucoup de compétence à la science des données, son vaste potentiel demeure largement inexploité au sein de l’administration fédérale.

Pour exploiter le potentiel existant, il faut créer les conditions appropriées

	Volonté et appui de la hiérarchie
	Formation, formation continue et recrutement
	Outils, plateformes et puissance de calcul
	Disponibilité et interopérabilité des données
	Collaboration transversale et interdisciplinaire
	Modèles d’exploitation et de financement durables
	Ancrage des principes fondamentaux régissant le recours à la science des données
	Législation qui facilite l’utilisation de l’aide à la décision basée sur les données
	Emploi de l’intelligence artificielle comme outil important
	Pilotage dans certains domaines, coordination dans nombre d’entre eux

De nombreuses activités sont déjà en cours, mais beaucoup reste à faire

Nombre des actions requises présentées ci-dessus désignent des projets ou des évaluations déjà en cours. Elles soulignent leur importance et confirment qu’il importe de les réaliser de manière systématique. D’autres mesures sont ponctuellement nécessaires.

Illustration 1: Aperçu des besoins identifiés

Il existe une condition appropriée qui n’est pas abordée explicitement dans d’autres mandats, mais qui revêt néanmoins une importance capitale: l’ancrage stratégique de la science des données dans les départements et les offices grâce à la volonté des responsables politiques et administratifs. Il importe par ailleurs de définir les prochaines étapes et les compétences en matière de formation et de perfectionnement en science des données au sein de l’administration fédérale, de faire de même pour ce qui est du recrutement de talents externes et d’appliquer ensuite rigoureusement ces décisions dans le cadre de la stratégie numérique de la Confédération (IS 6).

1 Introduction

Après l’adoption de la Stratégie de la Confédération en matière de science des données (DSStB)¹, le Conseil fédéral a attribué le 2 décembre 2022 différents mandats pour favoriser son déploiement.

Le Département fédéral de l’intérieur (DFI), plus précisément l’Office fédéral de la statistique (OFS), a été chargé de collaborer avec les autres départements et le domaine des EPF afin de rédiger un rapport sur l’utilisation de la science des données pour le bien commun en prenant des cas d’application qui sont propres à l’administration fédérale et qui s’inscrivent dans le processus d’élaboration des politiques. Le rapport poursuit deux objectifs.

Il sert d’une part à inciter les diverses unités administratives à prendre conscience du potentiel de la science des données et à encourager son utilisation. Il permet par ailleurs d’identifier des cas où la science des données est appliquée de manière exemplaire et transversale dans l’ensemble de l’administration fédérale. À cet effet, le rapport explique comment diverses unités administratives de la Confédération recourent à la science des données dans le cadre de leur travail quotidien. Il illustre de manière simple et pratique comment les responsables politiques et administratifs ainsi que les spécialistes peuvent utiliser la science des données au quotidien pour accroître l’efficacité et l’efficience de l’État dans ses actions². L’accent n’est par conséquent pas mis sur les aspects techniques de la mise en œuvre, mais sur les activités de l’administration fédérale. Le rapport vise d’autre part à mettre en évidence les conditions appropriées pour utiliser la science des données au sein de l’administration fédérale et les défis à relever dans ce sens. Ce faisant, il cerne les champs d’action en vue de promouvoir le recours futur à la science des données dans l’administration fédérale.

Au vu de ces objectifs, les cas d’application présentés résultent d’une sélection qui ne prétend nullement à l’exhaustivité et ne correspond pas à un ordre des priorités. Leur description ne constitue pas non plus une proposition ou une planification de projet. De caractère exemplaire, ces cas d’application couvrent un éventail aussi large que possible des méthodes de la science des données et des domaines d’activités de l’administration fédérale. Ils ont été sélectionnés par un groupe d’accompagnement interdépartemental (voir chapitre 4.2).

Le présent rapport s’articule comme suit:

- Le chapitre 2 expose les principales observations provenant de l’élaboration des cas d’application. Il présente de plus la suite des travaux.
- Le chapitre 3 décrit les six cas d’application sélectionnés et les principales exigences propres à chacun d’entre eux.
- L’annexe résume les bases stratégiques et théoriques issues de la DSStB. Elle présente également la démarche suivie.

¹ <https://www.news.admin.ch/news/message/attachments/74271.pdf> et <https://www.bfs.admin.ch/bfs/fr/home/dscc/dscc.gnpdetail.2022-0738.html>

² Dans ce contexte, l’ambition de la DSStB est formulée comme suit: «Aide à la décision basée sur les données au sein de l’administration fédérale: recourir à la science des données pour accroître l’efficacité et l’efficience des politiques publiques».

2 Observations confirmées

L'élaboration des cas d'application ainsi que les discussions menées avec des représentants de l'administration fédérale et du domaine des EPF ont permis à la fois de faire de nouvelles observations et de confirmer les actions requises déjà identifiées dans la DSStB.

La science des données recèle un grand potentiel pour l'administration fédérale.

Allant de l'établissement de tableaux globaux de la situation en cas de crise à la simulation de la sécurité alimentaire en Suisse sur le long terme, en passant par la planification d'interventions humanitaires et l'automatisation (partielle) d'activités de surveillance, les applications de la science des données peuvent faciliter le travail de l'administration fédérale dans de nombreux domaines politiques et secteurs d'activités. De plus, la science des données est aujourd'hui «prête à l'emploi» et il serait dommage de rater le train.

Si certaines unités administratives recourent d'ores et déjà à la science des données, son vaste potentiel demeure largement inexploité au sein de l'administration fédérale.

Nombre des cas d'application transversale présentés ici se fondent sur des projets ou des réflexions déjà en cours dans les unités administratives concernées. Les collaborateurs et les collaboratrices impliqués sont motivés et disposent des compétences nécessaires. Mais il s'agit d'une niche réservée aux spécialistes. Le potentiel disponible demeure hélas méconnu en dehors de certains services administratifs, disciplines et niveaux hiérarchiques.

Pour utiliser la science des données à grande échelle, il importe de créer ou d'améliorer les conditions appropriées dans différents domaines. La DSStB a identifié des champs d'intervention pour promouvoir la science des données au sein de l'administration fédérale. Ces champs d'intervention ont été confirmés et précisés. Voici les principaux d'entre eux:

- **Volonté et appui de la hiérarchie:** il incombe aux responsables politiques et administratifs d'encourager, voire d'imposer, le recours à la science des données dans leurs départements ou offices. Ils doivent être conscients de son potentiel, connaître les défis à relever dans leurs domaines politiques et secteurs d'activités, et bénéficier de l'appui nécessaire pour traduire les besoins (d'information) politiques et organisationnels en exigences concrètes relevant de la science des données («ingénierie des exigences»). Ils doivent surtout bien comprendre à quoi sert la science des données et être prêts à lui consacrer des ressources.
- **Formation, formation continue et recrutement:** l'administration fédérale doit investir dans son personnel et adopter des profils professionnels clairement définis dans le domaine de la science des données. Outre des spécialistes en science des données et des professionnels qualifiés dans ce domaine, des cadres et du personnel spécialisé doivent être formés en fonction de leur niveau hiérarchique. Ce champ d'intervention est déjà traité dans la stratégie numérique de la Confédération (IS 6)³. Il reste à définir les compétences et la suite des travaux, puis à progresser résolument dans la bonne direction.

³ <https://www.bk.admin.ch/bk/fr/home/digitale-transformation-ikt-lenkung/ikt-vorgaben/strategien-teilstrategie/sb000-ikt-strategie-des-bundes.html>

- **Outils, plateformes et puissance de calcul:** au sein de l'administration fédérale, la disponibilité d'outils et de plateformes de science des données n'est aujourd'hui garantie que de manière sporadique et s'avère en règle générale insuffisante. Voilà pourquoi un mandat de suivi de la DSSStB vise déjà à concevoir une «plateforme collaborative en science des données». Par ailleurs, pour recourir en permanence à des applications de la science des données, il faut disposer de la puissance de calcul correspondante. L'administration fédérale devra par conséquent renforcer sa coordination en matière d'acquisition de services informatiques en nuage (*cloud*) pour ce qui est de la puissance de calcul (*compute*) et du stockage (*storage*). Elle devra également évaluer la durabilité environnementale des solutions envisagées.
- **Disponibilité et interopérabilité des données:** l'utilité de la science des données repose pour l'essentiel sur la disponibilité de données pertinentes et sur la possibilité de les échanger facilement. Cette condition vaut en particulier pour les cas d'application transversale, qui font par exemple intervenir différents offices fédéraux et niveaux étatiques et qui exigent une compilation rapide des données (tels les tableaux de la situation lors de catastrophes naturelles). Plusieurs projets de la Confédération visent d'ores et déjà à améliorer la disponibilité et l'interopérabilité des données. La disponibilité des géodonnées est aujourd'hui déjà assurée en grande partie à l'échelon fédéral grâce à l'infrastructure fédérale de données géographiques (IFDG)⁴, tandis que le projet d'une infrastructure nationale de données géographiques (INDG)⁵ est en cours d'élaboration au niveau fédéral. Sur la base du rapport du DETEC (OFCOM) et du DFAE (DDIP) intitulé «Création d'espaces de données fiables, sur la base de l'autodétermination numérique», des mandats de suivi ont été attribués en vue de promouvoir l'exploitation d'espaces de données fiables et interopérables⁶. Ces mandats ont pour but de mieux exploiter à l'avenir le potentiel que recèle l'utilisation des données et de garantir que les individus et les entreprises puissent disposer de leurs données de manière autonome. Le programme de gestion nationale des données (NaDB) encourage quant à lui l'harmonisation des bases de données générales et publie des normes communes sur la plateforme d'interopérabilité I14Y⁷. Quels que soient leurs objectifs, les activités doivent respecter en tout temps la protection des données et le principe de minimisation des données. À cet effet, l'OFS élabore actuellement un concept sur l'application de la science des données afin de préserver la sphère privée lors du partage de données au sein de l'administration fédérale.
- **Collaboration transversale et interdisciplinaire:** l'utilisation de la science des données passe par l'interaction entre différents domaines, disciplines et organisations au sein de l'administration fédérale, de même qu'avec des tiers (notamment la science et la recherche, les cantons et l'économie). De telles interactions nécessitent, d'une part, une culture de collaboration ouverte et, d'autre part, des structures et des processus de collaboration.
- **Modèles d'exploitation et de financement durables:** il importe d'assurer à long terme le financement, l'exploitation, le développement et l'évaluation des applica-

⁴ [Infrastructure fédérale de données géographiques \(admin.ch\)](https://www.admin.ch/gov/fr/accueil/documentation/communiques.msg-id-87780.html)

⁵ [Infrastructure nationale de données géographiques \(INDG\): e-geo.ch - Canton de Lucerne](https://www.admin.ch/gov/fr/accueil/documentation/communiques.msg-id-87780.html)

⁶ <https://www.admin.ch/gov/fr/accueil/documentation/communiques.msg-id-87780.html>

⁷ <https://www.bfs.admin.ch/bfs/fr/home/nadb/nadb.html>

tions de la science des données. Il convient d’élaborer une étude de rentabilité (*business case*). Lorsque le projet visant à développer une application est terminé, il convient de mettre les ressources correspondantes à disposition et de définir les structures, les processus et les rôles requis. Même après la mise en service de l’application, un suivi est nécessaire pour assurer son bon fonctionnement. Il est essentiel de créer les conditions appropriées, en particulier pour les applications utilisées en cas de crise, comme celles qui servent à établir des tableaux de la situation.

- **Ancrage des principes fondamentaux régissant le recours à la science des données:** afin de garantir la qualité et l’éthique des applications de la science des données, il faut respecter rigoureusement les principes fondamentaux définis dans la DSSStB de même que les règles de l’État de droit. L’avancement des travaux doit par ailleurs être transparent pour le grand public. C’est le seul moyen d’instaurer la confiance. L’élaboration d’un code de bonnes pratiques pour une science des données centrée sur l’être humain et digne de confiance constitue un premier pas dans cette direction. D’autres étapes devront suivre pour ancrer les principes fondamentaux dans la culture et les institutions.
- **Législation qui facilite l’utilisation de l’aide à la décision basée sur les données:** les lois, les ordonnances et les directives doivent prendre en compte et anticiper la numérisation croissante du travail de l’administration publique et l’intensification du recours aux données. Il faut par exemple prévoir des consignes destinées à des tiers (tels des établissements de droit public) pour régir la production, la qualité et la transmission de données. Mentionnons à titre d’exemple la loi sur le système d’information dans le domaine de la sécurité sociale, actuellement en préparation.
- **Rôle essentiel de l’intelligence artificielle:** l’intelligence artificielle (IA) est un outil essentiel pour exploiter pleinement le potentiel de la science des données. Il incombe à l’administration fédérale de piloter et de coordonner l’évolution de son utilisation de manière globale, c’est-à-dire pour la grande variété des méthodes envisageables (telles les méthodes issues des statistiques avancées, de l’apprentissage automatique et du domaine de l’IA).
- **Pilotage dans certains domaines, coordination dans beaucoup (gouvernance):** différents domaines présentent un besoin de pilotage et de coordination. Comme pour la transformation numérique et l’informatique, il est nécessaire de créer des synergies grâce à une approche commune et concertée, d’assurer l’interopérabilité au sein de l’administration fédérale et avec les tiers et de renforcer la confiance du public dans l’utilisation de la science des données. Cette remarque s’applique à plusieurs des conditions appropriées pour utiliser la science des données (plateforme et outils, normalisation, éthique, législation, formation et formation continue, etc.). Ces points seront abordés et approfondis dans le cadre de l’évaluation du Centre de compétences en science des données (DSCC) (EXE 2020.0572) et de l’évaluation du Réseau de compétences en IA (CNAI) (EXE 2021.1863).

Comme relevé plus haut, nombre des actions requises présentées ici désignent des projets ou des évaluations déjà en cours. Une seule d’entre elles n’est pas encore traitée.

tée: **l'ancrage stratégique de la science des données dans les départements et les offices grâce à la volonté des dirigeants politiques et administratifs.** La décision d'utiliser la science des données relève de la compétence des responsables des services spécialisés. Des unités administratives comme le CNAI et le DSCC peuvent certes créer des liens, conseiller et soutenir. Il incombe toutefois aux responsables des départements et des offices de veiller à inclure et à utiliser la science des données de manière systématique dans leurs processus spécialisés et dans les bases légales correspondantes. Il convient en outre de définir les compétences et la suite des opérations en matière de formation et de formation continue en science des données au sein de l'administration fédérale ainsi qu'en ce qui concerne le recrutement de talents externes.

La question des ressources nécessaires est forcément étroitement liée à ces activités. Pour atteindre les objectifs fixés dans la DSStB, les différentes unités administratives doivent investir dans des projets concrets de science des données, comparables aux cas d'application présentés ici. Des fonds sont par ailleurs nécessaires pour réaliser les projets prioritaires mentionnés ci-dessus et déjà en cours.

3 Documentation des cas d’application

3.1 Vue d’ensemble

Ce chapitre décrit six cas d’application⁸ : 1) utilisation de renseignements d’origine sources ouvertes (*open source intelligence*, ROSO); 2) tableaux globaux de la situation en cas de crise; 3) aide à la surveillance basée sur les données; 4) aide à la décision fondée sur des systèmes de modèles holistiques; 5) contrôles de plausibilité destinés à garantir la qualité des données; 6) jumeaux numériques.

La description s’effectue en quatre temps. Nous commençons par décrire de manière générale les méthodes sous-jacentes de la science des données, leur valeur ajoutée et leur pertinence. Nous donnons ensuite un exemple d’application de la science des données dans une unité administrative de la Confédération, pour un domaine d’activité ou un domaine politique concret. Cet exemple se fonde d’une part sur des projets ou des considérations en cours dans l’unité administrative en question, d’autre part sur les développements possibles envisagés avec des spécialistes de l’administration fédérale et du domaine des EPF. Nous examinons ensuite les autres domaines de l’administration fédérale dans lesquels le cas considéré pourrait s’appliquer. Enfin, nous nous intéressons aux exigences qui doivent être remplies pour qu’il puisse être mis en œuvre de manière transversale et à grande échelle au sein de l’administration fédérale. Nous avons pour ce faire considéré sept dimensions (voir illustration 2).



Illustration 2: Dimensions considérées pour les exigences

Les cas d’application ont été documentés par un groupe de travail dédié, composé de spécialistes de l’administration fédérale et du domaine des EPF. La procédure d’élaboration des cas d’application et la composition des groupes de travail sont décrites au chapitre 4.2.

⁸ Les cas d’application peuvent être regroupés dans le cadre de l’évaluation du CNAI à des fins de pilotage stratégique de l’application de la science des données au sein de l’administration fédérale.

Utilisation de renseignements d'origine sources ouverts (*open source intelligence*, ROSO)

Exemple: aide humanitaire en situation de crise

Généralités

Description

L'utilisation de renseignements d'origine sources ouvertes (ROSO) repose sur la collecte, l'analyse et l'évaluation systématiques d'informations publiquement accessibles. Ces informations peuvent provenir de différentes sources: médias d'information, réseaux sociaux, blogs, forums, rapports publiés, documents ou données de gouvernements, d'entreprises privées ou d'organisations non gouvernementales (ONG). La clé du traitement et de l'utilisation de ces grands volumes d'informations réside dans l'application de méthodes et d'algorithmes issus de la science des données.

L'intelligence artificielle (IA) – l'apprentissage automatique notamment – permet d'analyser et de traiter automatiquement d'importants volumes de données, y compris des données non structurées (données textuelles et données de télédétection, p. ex.). Les méthodes et les algorithmes de la science des données peuvent aider les décideurs à appréhender les ressentis au sein de la population, à évaluer l'impact de mesures politiques et à identifier les nouvelles thématiques sur lesquelles il conviendrait de s'appesantir.

Une fois les données analysées, différentes approches permettent de rendre les valeurs et les résultats agrégés compréhensibles et accessibles aux utilisateurs. On peut ainsi envisager de développer des systèmes d'information basés sur l'IA qui aident à trouver intuitivement les documents pertinents dans une base de données interne, grâce au traitement du langage naturel (*natural language processing*, NLP).

Valeur ajoutée et pertinence

L'utilisation de ROSO apporte une valeur ajoutée significative au processus d'élaboration des politiques. En substance, il s'agit d'exploiter tout le potentiel des informations publiques pour prendre des décisions plus éclairées et plus efficaces.

Jusqu'ici, l'analyse et l'évaluation de données issues de sources publiques s'effectuaient le plus souvent manuellement. Cela représente un investissement considérable, surtout pour les données textuelles non structurées. Grâce à l'utilisation de méthodes et d'algorithmes de la science des données, de l'apprentissage automatique et du domaine de l'IA notamment (traitement du langage naturel, p. ex.), l'efficacité de ce processus peut être notablement améliorée. L'analyse, le traitement et la structuration automatisés des données permettent une analyse ciblée et systématique de grandes quantités de données, ce qui allège la charge de travail du personnel à qui ces tâches sont habituellement

confiées. Les processus décisionnels de l’élaboration des politiques s’en trouvent accélérés et leur efficacité renforcée, sans que l’on se prive pour autant des compétences humaines en matière d’analyse et de décision.

Une autre valeur ajoutée réside dans le développement d’interfaces conviviales comme point d’accès aux ROSO, qui permettent aux utilisateurs d’accéder rapidement et de manière ciblée aux informations qui les intéressent, sans recherche fastidieuse.

Le recours à des méthodes et des algorithmes de la science des données est également intéressant pour identifier certains modèles et tendances au niveau des données. C’est particulièrement utile pour établir des prévisions et d’un soutien précieux quand on a affaire à d’importants volumes de données.

Les ROSO et les méthodes de la science des données permettent d’améliorer sensiblement l’ensemble du processus de décision politique – de l’identification et de la compréhension du problème à son évaluation, en passant par l’élaboration d’options.

Exemple d’application

Évaluation des besoins fondée sur les ROSO dans une situation de crise humanitaire

Contexte

Les situations de crise, en particulier celles induites par les catastrophes naturelles (séisme, inondations, etc.), nécessitent une réponse humanitaire rapide et efficace. Dans de tels cas, il est essentiel d’évaluer aussi rapidement que possible les besoins de la population concernée afin de fournir l’aide nécessaire, et de déterminer l’ampleur et le calendrier d’une assistance plus poussée. Dans les heures et les jours qui suivent la survenance d’une crise, le chaos règne de sorte qu’il est souvent difficile de se faire une idée claire des conditions sur place. Une multitude d’informations provenant de différentes sources doivent être analysées, alors que l’accès physique à la région concernée est partiellement, voire totalement impossible. Ces informations sont dispersées et se présentent généralement sous la forme de données non structurées: publications sur les réseaux sociaux susceptibles de fournir des témoignages directs de la population touchée, communiqués de presse officiels et rapports des autorités et des ONG, articles de la presse locale, données issues de la télédétection lorsqu’elles sont publiques (données de mesure des capteurs des satellites, des avions et des drones, notamment), etc.

Défis

L’analyse des informations disponibles et l’extraction de données pertinentes pour évaluer les besoins représentent un défi de taille. Passer en revue et analyser manuellement une grande quantité de données non structurées prend beaucoup de temps. Compte tenu de l’évolution rapide de la situation pendant une crise humanitaire, ce processus peut ne pas être achevé ou se révéler rapidement obsolète. À cela s’ajoute la difficulté à extraire les informations pertinentes des données obtenues et à déterminer les besoins précis. Notons par

ailleurs qu’à l’heure actuelle, la pratique est essentiellement réactive: on manque en effet souvent d’informations solides sur lesquelles établir des scénarios et des tendances qui permettraient de se préparer à la survenance de crises.

Solution proposée

Les spécialistes du Département fédéral des affaires étrangères (DFAE) qui fournissent une aide humanitaire à l’étranger en coordination avec d’autres services fédéraux et des tiers recourent à une application ROSO qui les aide à déterminer rapidement et de manière exhaustive les besoins en lien avec la crise. Cette application analyse en continu les données pertinentes sur la base de méthodes et d’algorithmes de la science des données (apprentissage automatique et IA, p. ex. *computer vision* et NLP). L’objectif est de permettre aux spécialistes et aux décideurs d’évaluer la situation en temps réel, sur la base de données, et de prendre des décisions en conséquence.

Les données ainsi traitées sont accessibles au personnel du DFAE en cas de crise humanitaire. Un tableau de bord présente les résultats (et les tendances) les plus importants nécessaires pour planifier l’aide humanitaire sur la base des données ROSO (approvisionnement en eau potable, localisation de la population civile en fuite, p. ex.). Les indicateurs peuvent être adaptés en fonction de la crise et de la région. Les données prises en compte comprennent entre autres les médias locaux, les publications des réseaux sociaux et les rapports des organisations d’aide nationales et internationales dans les langues pertinentes. Pour localiser des groupes de personnes en fuite, on peut utiliser des informations géoréférencées sur les connexions aux réseaux sociaux et des images satellite récentes.

La traçabilité des informations présentées est garantie par l’indication de la source principale des données (rapport d’une ONG, p. ex.) ou de la méthode (analyse des réseaux sociaux, p. ex.). L’attention des utilisateurs est par ailleurs attirée en cas de doute quant à l’exactitude d’une information ou d’informations contradictoires (données divergentes sur les réfugiés entre les rapports de plusieurs ONG, p. ex.).

Les fausses informations (liées p. ex. à des campagnes de propagande) représentant un grand risque pour les personnes concernées, tout particulièrement dans les situations de crise. De ce fait, des mécanismes de la science des données à même de reconnaître certains modèles et donc ce type de campagne sont intégrés (analyses de topographies de réseaux, p. ex.).

Le tableau de bord a ceci d’intéressant qu’il permet d’ajouter rapidement, et lorsque c’est possible en temps réel, de nouveaux points de données au jeu de données existant. Cela garantit une mise à jour rapide de l’analyse des besoins, ce qui est particulièrement important dans une situation de crise susceptible d’évoluer rapidement.

La qualité de l'analyse des données est ici d'une importance capitale. Le système fondé sur l'IA combiné à l'expertise humaine garantit la fiabilité et la précision des résultats obtenus. Grâce à l'apport humain, le système peut également apprendre à mieux identifier les données et les sources de données problématiques (défaut de qualité, p. ex.).

Sur la base des informations mises à disposition, le personnel responsable peut élaborer des scénarios et, dans le meilleur des cas, des prévisions sur les conséquences immédiates de la crise humanitaire. Cette méthode, appelée *nowcasting*, permet par exemple d'estimer les besoins en eau pour les jours à venir sur la base d'informations sur les mouvements de fuite et les réserves d'eau disponibles dans une région. Les mesures nécessaires peuvent être définies sur cette base. On peut ainsi passer d'une approche purement réactive dans la gestion des crises à une approche proactive.

Transférabilité

L'aide apportée aux spécialistes et aux décideurs par le traitement et l'analyse des données ROSO est pertinente pour tout domaine politique dans lequel des avis doivent être recueillis à grande échelle concernant une situation sur le territoire national ou à l'étranger. Ce type d'évaluation pourrait servir de base à certains briefings, par exemple sur des questions de politique de sécurité. Une application serait également possible dès lors qu'il s'agit d'identifier des tendances sur le plan international (évolution de la politique climatique dans d'autres pays et au niveau mondial, p. ex.).

Les représentations suisses à l'étranger pourraient également utiliser le tableau de bord, mais celui-ci devrait être adapté à leurs besoins. La communauté humanitaire, notamment la Genève internationale, pourrait également en bénéficier.

Un autre avantage réside dans la possibilité d'apparier les données avec des données internes, le cas échéant avec des données confidentielles. Dans ce cas, les données devraient être accessibles à certains utilisateurs seulement afin de garantir la confidentialité tant des données que des informations.

Exigences

Données

- Accéder aux sources de données publiques pertinentes.
- Prendre en compte une multitude de langues, styles, terminologies et abréviations dans les documents, pour l'analyse des données et l'interface utilisateur.
- Mettre à jour et étoffer les jeux de données en continu.
- Assurer la qualité des données et la gestion de cette qualité.
- Traiter et stocker efficacement de grands volumes de données.
- Assurer l'actualité des données collectées.

- Assurer la traçabilité des métadonnées (géolocalisation, horodatage, etc.).

Réglementation et éthique (sur la base des principes fondamentaux de la DSStB)

- Protéger les données et les informations ainsi que la sécurité de l’information (restriction des droits d’accès, p. ex., notamment en cas d’utilisation de données personnelles).
- Clarifier les droits d’accès pour l’application.
- Assurer la traçabilité du traitement des données.
- Assurer la transparence des algorithmes utilisés et l’explicabilité des résultats de l’analyse éventuels.

Technologie et infrastructure

- Fournir une architecture évolutive, souple et performante, capable de gérer un volume croissant de données.
- Intégrer techniquement l’application dans l’infrastructure informatique existante conformément à la réglementation.
- Fournir une puissance de calcul suffisante et évolutive, pour l’entraînement des modèles et l’utilisation du NLP, p. ex.
- Mettre à disposition une plateforme cloud interne à l’administration pour l’application (*storage* et *computing*).
- Accorder l’accès à l’application à tous les utilisateurs désignés, même en dehors de l’administration fédérale (Corps suisse d’aide humanitaire, p. ex.).

Gestion

- Mettre en place des structures de gestion de projet pour l’implémentation de l’application, y compris le suivi et l’évaluation de la mise en œuvre du projet.
- Assurer la maintenance et le développement à long terme de l’application.
- Encourager la collaboration interdisciplinaire dans le cadre du développement de l’application (utilisateurs, spécialistes, experts informatiques, scientifiques des données).
- Mettre à disposition des ressources financières nécessaires à long terme pour la mise au point et la maintenance de l’application.
- Organiser le soutien nécessaire et contribuer à l’acceptation du projet au sein de l’organisation.

Connaissances et compétences

- Développer les compétences nécessaires dans le domaine de l’IA et de l’apprentissage automatique.

- Encourager la formation initiale et continue chez les scientifiques des données et les spécialistes pour leur permettre de développer les compétences nécessaires liées à l’application.
- Prévoir des formations et des ateliers pour sensibiliser aux possibilités et aux défis de l’application au sein de l’organisation.
- Développer des connaissances approfondies dans le domaine de l’évaluation de la qualité des données.
- Développer la capacité à évaluer les possibilités techniques et analytiques ou les conséquences à partir de la qualité des données.
- Intégrer les connaissances nécessaires issues de la recherche, en particulier pour les projets ayant une forte pertinence pour la recherche fondamentale.
- Développer l’attitude critique des collaborateurs et des décideurs vis-à-vis des informations (qualité des données et des sources de données, désinformation, informations manquantes ou rares, etc.).
- Former le personnel (sur le terrain et au siège) pour garantir l’utilisation adéquate du tableau de bord, même en cas de stress extrême.

Parties prenantes

- Impliquer les parties prenantes, les spécialistes de différents domaines (humanaire, recherche, etc.), les utilisateurs et les décideurs dans tous les processus de développement de l’application (identification des principaux besoins d’information / exigences du produit, p. ex.).
- Communiquer avec le grand public et les parties concernées pour répondre aux préoccupations / questions et assurer la transparence.
- Échanger avec les parties prenantes concernées sur les avantages et les utilisations possibles de l’application, ainsi que sur les risques potentiels.

Durabilité écologique

- Définir et surveiller la performance du modèle et du budget ressources.
- Définir des mesures pour compenser les émissions de CO₂ générées.
- Identifier, analyser et, le cas échéant, quantifier les potentiels de durabilité positifs et négatifs de l’application.
- Déterminer les critères permettant de s’assurer de la durabilité du matériel et des centres de calcul utilisés.
- Recourir à des pratiques durables de collecte de données, de stockage et d’entraînement des modèles, y compris en réduisant autant que possible la redondance des données et l’utilisation de techniques de compression des modèles.

3.2 Tableaux globaux de la situation en cas de crise

À l’exemple d’un hypothétique séisme dévastateur et de grande ampleur survenu en Suisse

Généralités

Description

Dans les situations de crise, par exemple en cas de catastrophe naturelle ou de pénurie d’énergie, des décisions rapides et solidement étayées s’imposent. Des tableaux globaux de la situation revêtent alors une importance décisive pour garantir l’accès à toutes les informations pertinentes.

Ces tableaux réunissent et analysent différentes sources de données habituellement isolées. Ils servent à comprendre l’ensemble de la situation et facilitent ainsi la prise de décision en cas de crise. Un tableau global regroupe différents tableaux spécifiques, telles des informations sur la situation météorologique, l’état des routes, l’approvisionnement en eau et en électricité, la protection contre les catastrophes et la disponibilité de ressources humaines.

Le traitement et l’analyse de ces données font appel à des méthodes de la science des données, y compris des méthodes issues des statistiques avancées, de l’apprentissage automatique et de l’IA. Grâce à ces méthodes, les données sont saisies, réunies, agrégées et analysées pour permettre aux responsables de prendre des mesures ciblées.

L’interface utilisateur d’un tableau de la situation se présente sous la forme d’un tableau de bord basé sur un modèle présentant une structure spatiale et temporelle.

Pour dresser efficacement des tableaux globaux de la situation, il est nécessaire d’avoir un plan qui spécifie les données requises, les endroits où les trouver, leurs formats et la manière de les rassembler. Il est alors souvent nécessaire d’assurer une coordination et une collaboration par-delà les frontières organisationnelles, afin de fournir une représentation aussi complète et précise que possible de la situation.

Valeur ajoutée et pertinence

Le recours à des tableaux globaux de la situation permet aux décideurs d’être mieux informés pour agir en cas de crise. Cette forme d’aide à la décision basée sur des données accroît l’efficacité de la gestion des crises, augmente la sécurité de la population et réduit les dommages potentiels.

La science des données peut contribuer à exploiter le potentiel de tableaux globaux de la situation en cas de crise, en particulier lorsqu’ils ne se limitent pas à certains domaines thématiques et qu’ils concernent tous les niveaux de l’État.

Elle facilite la saisie et la collecte rapides de grandes quantités de données, leur conversion dans les formats et la qualité nécessaires, leur regroupement et, enfin, leur présentation simple et compréhensible. Des méthodes comme le traitement automatique du langage naturel (NLP) et la vision par ordinateur permettent en particulier de traiter rapidement des données non structurées qui sont souvent difficiles à analyser, tels des textes et des images, et de les rendre accessibles pour la prise de décision. Cette manière de procéder s’avère particulièrement précieuse, car il n’est souvent pas possible de prévoir exactement les informations qui seront nécessaires en cas de crise.

Un autre potentiel réside dans la possibilité de combiner des informations centralisées «descendantes» (*top-down*) issues notamment de systèmes d’information géographique avec des informations décentralisées «ascendantes» (*bottom-up*) provenant par exemple des organes civils de conduite et des forces d’intervention sur place.

Sur la base de tels tableaux de la situation, il est en outre possible d’élaborer des scénarios et, dans l’idéal, de prévoir les conséquences immédiates de la crise (*nowcasting*).

Les approches basées sur l’apprentissage automatique, à même de recommander de manière autonome des actions à entreprendre, conviennent moins bien en situation de crise. D’une part, une traçabilité des prises de décision est nécessaire spécialement en cas de crise et les systèmes d’apprentissage automatique ne sont pas toujours à même de la garantir.

D’autre part, ces systèmes ne sont pas en mesure de gérer des situations nouvelles et, pour eux, uniques. Or les situations de crise se distinguent en général par leur caractère unique (caractéristique que l’on appelle la «stationnarité»).

Dans un monde de plus en plus imprévisible et marqué par les crises, où les incidents deviennent toujours plus fréquents, il est essentiel pour les responsables de disposer des meilleures informations disponibles pour mener des actions rapides et efficaces.

Exemple d’application

Tableau global de la situation lors d’un séisme majeur en Suisse

Contexte

Un séisme majeur (soit un tremblement de terre qui entraîne des dommages aux bâtiments et aux infrastructures) a touché une grande partie de la Suisse, provoquant des destructions à grande échelle et d’importants dommages matériels, tout en faisant de nombreux blessés. Les communes situées dans la principale zone sinistrée sont en grande partie coupées du monde extérieur, les routes d’accès étant bloquées et les lignes de communication interrompues, ce qui complique la situation. Des précipitations persistantes ne font qu’aggraver les choses, car elles peuvent provoquer des inondations et des glissements de terrain et détériorer encore la situation déjà précaire des habitants. De plus,

on ne sait pas à ce stade dans quelle mesure les infrastructures critiques (barages, centrales électriques, centres de calcul, etc.) ont subi des dégâts.

La priorité immédiate consiste à assurer la sécurité de la population touchée et à éviter autant que possible d'autres dégâts. À cet effet, il est crucial de dresser un tableau détaillé et actualisé de la situation afin de coordonner les mesures d'urgence et d'utiliser au mieux les ressources disponibles.

Il est par ailleurs nécessaire d'élaborer sans attendre des scénarios et, idéalement, de formuler des prévisions quant aux conséquences immédiates du séisme (dégâts subis par des infrastructures critiques, glissements de terrain, etc.). Une situation de ce genre implique un scénario complexe et dynamique, qui nécessite une approche globale et axée sur les données afin d'obtenir des solutions efficaces.

Défis

Les tableaux globaux de la situation établis lors d'une crise, tel un séisme majeur, placent les autorités face à toute une série de défis.

Premièrement, les informations nécessaires varient en fonction de la phase de la crise et peuvent évoluer au fil du temps. Au départ, il importe en priorité de disposer par exemple d'informations géoréférencées détaillées sur le système de santé, l'approvisionnement en énergie ou les infrastructures (avant et après l'événement). Dans les phases ultérieures, d'autres domaines peuvent passer au premier plan, comme la reconstruction d'infrastructures ou le rétablissement de la circulation des biens et des personnes.

Deuxièmement, la gestion des crises implique de nombreux acteurs à tous les niveaux de l'État. Ils ont tous des ressources, des capacités et des besoins différents, ce qui complique la coordination et la communication. De plus, les données nécessaires pour dresser un tableau global de la situation sont souvent réparties entre plusieurs organismes et institutions, présentent des qualités et des formats différents et sont même parfois non structurées.

Troisièmement, selon la complexité et la dynamique de la situation de crise, il s'avère souvent difficile d'entreprendre une préparation et une planification prospectives. La situation peut en effet évoluer rapidement et de manière imprévisible, de sorte que bon nombre des mesures et solutions nécessaires ne pourront être mises au point et appliquées que pendant la crise.

Solution proposée

En réponse aux défis identifiés, une «plateforme de suivi de la situation» est créée, qui sera à même d'établir rapidement le tableau nécessaire de la situation si un séisme majeur devait se produire en Suisse. Cette plateforme définira les données à réunir ainsi que les bases techniques et organisationnelles.

Chaque crise étant unique, tout comme chaque tremblement de terre et ses conséquences, il est impossible de prévoir de manière absolue les informations qui seront nécessaires lors d’une crise. Voilà pourquoi cette plateforme de suivi reposera sur la modularité. Dans la mesure du possible, l’échange et la préparation des données seront prévus à l’avance, de telle sorte que les procédures en question puissent s’appliquer selon les besoins de chaque crise (données générales d’information géographique ou données météorologiques, p. ex.). Le moment venu, il sera possible d’activer les différentes couches d’information en fonction de la phase et le type de crise, ou en fonction de l’utilisateur ou de l’utilisatrice (maire ou présidente de la commune, commandant des sapeurs-pompiers, etc.). Un tel système exige de distinguer les informations urgentes des informations nécessaires à moyen et à long terme et de définir les sources de données pertinentes pour chaque phase.

Aux couches d’information centralisées mises à disposition «du haut vers le bas» (*top-down*), telles les données des systèmes d’information géographique, il est possible d’ajouter systématiquement les informations fournies par les forces d’intervention et les particuliers sur place, ainsi que les informations librement disponibles sur Internet, par exemple sur les réseaux sociaux (ce que l’on appelle des «renseignements d’origine sources ouvertes», ROSO).

En cas de crise, les ressources et les infrastructures seront mises à disposition pour intégrer rapidement, aux données identifiées à l’avance, d’autres informations, dont le besoin n’est par exemple pas prévisible, qui sont très spécifiques à la crise ou ne figurent dans aucune base de données générale. Des scientifiques des données accompagneront la gestion de crise. Ils contribueront à identifier les sources de données nécessaires ainsi qu’à préparer les données, à les plausibiliser et à les intégrer dans le tableau de la situation. À cet effet, ils utiliseront des applications de la science des données (p. ex. pour structurer les données ou les plausibiliser). À titre d’exemple, les informations requises d’urgence en cas de crise peuvent concerner l’impact que la destruction d’une infrastructure de transport engendrée par un séisme survenu dans un pays voisin peut avoir sur l’accès aux soins de santé dans un canton proche de la frontière. Lorsque des «données effectives» font défaut, les méthodes de la science des données (y c. les méthodes issues des statistiques avancées, de l’apprentissage automatique et du domaine de l’IA) peuvent contribuer à combler ces lacunes par des simulations (dans le cas, p. ex., d’infrastructures complexes avec des effets en cascade) ou à générer des données synthétiques. Lors du recours à des données synthétiques, il importe toutefois de garantir la traçabilité et l’explicabilité de la manière dont elles ont été générées.

La «plateforme de suivi de la situation» offrira un tableau de bord convivial qu’il sera possible de personnaliser, de sorte que chaque utilisateur et utilisatrice puisse l’adapter à ses propres besoins, à la crise spécifique ou à la phase de crise en cours. Un commandant des sapeurs-pompiers intervenant au niveau local ou une cellule de crise travaillant à l’échelon national pourront activer les couches d’information qui les intéressent et adapter la présentation à leurs besoins (p. ex. à l’aide de «vignettes»). Les données intégrées étant associées à un modèle qui fonctionne dans le temps et l’espace, leur visualisation sous la

forme d’un tableau de bord fournira une représentation simple et compréhensible d’informations complexes.

En situation de crise, la traçabilité et la fiabilité des informations présentées revêtent une grande importance. Elles seront garanties par l’indication de la source principale des données (statistiques publiques ou rapport de police, p. ex.) ou de la méthode appliquée (p. ex. analyse des réseaux sociaux). Par souci de transparence, le système signalera l’incertitude d’une information, de même que les informations contradictoires (p. ex. en cas d’écart important entre les données, provenant de différentes sources, sur l’étendue des dégâts aux bâtiments).

La plateforme ne se contentera pas de présenter la situation actuelle, mais permettra aussi d’élaborer des scénarios et, idéalement, des prévisions quant aux conséquences immédiates d’un événement. Cette méthode dite de «prévision immédiate» (*nowcasting*) permet par exemple d’obtenir une représentation détaillée des dommages probables à partir des informations sur l’intensité et la propagation du séisme et de modèles géoréférencés de bâtiments et d’infrastructures. Sur cette base, les tableaux de la situation peuvent fournir des informations en temps réel aux responsables (pour faire p. ex. fermer les conduites de gaz dans certaines régions).

Transférabilité

Il est possible de transposer largement cette approche à d’autres situations de crise, à condition que le cadre juridique permettant de rassembler et d’utiliser les données nécessaires soit en place.

Une autre condition de cette transférabilité réside dans l’existence d’une solution universelle, généralement utilisable. Celle-ci devrait être compatible avec diverses interfaces, car une situation de crise implique un grand nombre d’acteurs, tels les cantons, les communes et les organismes d’entraide, qui sont aussi bien des fournisseurs que des utilisateurs de données. Elle doit aussi être suffisamment souple pour pouvoir s’adapter à des exigences spécifiques, de sorte qu’il ne soit pas nécessaire de développer une solution entièrement nouvelle pour chaque situation de crise. La carte numérique de la Suisse (<https://test.map.geo.admin.ch>) pourrait servir de base de départ, puisqu’elle permet de superposer, en ligne et à volonté, toutes les couches géoréférencées requises.

Cette approche conviendrait en particulier pour les situations où des tableaux globaux de la situation doivent être mis à disposition rapidement, par exemple en cas de pandémie, de guerre, d’attaque terroriste ou d’urgence due à des phénomènes météorologiques extrêmes. Or, les changements géopolitiques et politiques pourraient à l’avenir accroître tant la fréquence et l’intensité de tels événements. Dans tous ces cas, la prise en compte et l’analyse de différentes sources de données peuvent servir à prendre des décisions plus éclairées et d’agir avec une plus grande efficacité.

Exigences

Données

- Regrouper différentes sources de données, telles que les données d'imagerie (emplacement), les informations météorologiques (danger, aléa), les données concernant le trafic, les infrastructures (exposition au danger), la protection contre les catastrophes et les ressources humaines (mesures) ainsi que les informations sur la vulnérabilité spécifique (vulnérabilité).
- Garantir l'actualité, l'exactitude, l'exhaustivité et la traçabilité des données afin de fournir des bases de décision précises, fiables et vérifiables.
- Préparer les données potentiellement nécessaires (la création de la base de données constituant un élément crucial de la préparation aux situations de crise, qui pourrait p. ex. relever de la cellule de crise).
- Mettre à disposition des métadonnées complètes.
- Présenter un aperçu des données existantes (c.-à-d. un catalogue de métadonnées).

Réglementation et éthique (sur la base des principes fondamentaux de la DSStB)

- Définir clairement les règles et les compétences régissant l'accès aux données et leur utilisation en cas de crise.
- Garantir le respect de la sphère privée et la sécurité des données et des informations, notamment en ce qui concerne les informations personnelles et les données sensibles sur les infrastructures.
- Assurer un traitement équitable et non discriminatoire des données en cas de crise.
- Garantir la transparence et la traçabilité des algorithmes et des résultats.

Technologie et infrastructure

- Développer une plateforme centrale, conviviale et sécurisée pour mettre à disposition et visualiser les données et les analyses de données (y c. d'éventuelles modélisations).
- Créer une infrastructure à la fois robuste et souple qui puisse être adaptée rapidement à l'évolution de la situation en cas de crise.
- Mettre en place des mesures de sécurité élevée pour protéger la plateforme et les données qu'elle contient contre les accès non autorisés et les cyberattaques.
- Créer des canaux, dans le système informatique, qui soient à même de regrouper rapidement les données en période de crise, mais aussi de les isoler en temps normal.
- Assurer aux différents acteurs (services de secours sur place, Conseil d'État, Conseil fédéral, etc.) un accès aux informations pertinentes, par exemple via des interfaces présentes dans des systèmes existants ou des applications dédiées.

- Garantir des interfaces de données lisibles par ordinateur (API, *application programming interfaces*). De telles interfaces assurent aussi l’intégration dans d’autres systèmes (complémentaires), tout en favorisant l’interopérabilité.
- Mettre à disposition des instruments propres à la science des données afin de pouvoir intégrer rapidement les données requises (disponibles souvent sous différents formats) dans le tableau de la situation en cas de crise.

Gestion

- Instaurer des procédures claires (y c. la définition de tâches, de compétences et de responsabilités) à appliquer en cas de crise (p. ex. matrices TCR ou RACI⁹) ainsi que d’éventuels modèles d’utilisation et d’imputation, de sorte que les différents acteurs et sources de données soient pilotés et coordonnés efficacement.
- Communiquer les informations et les décisions de manière claire, opportune et adaptée aux groupes cibles.
- Mettre en place un système de feed-back et d’apprentissage afin de tirer les enseignements des situations de crise, d’optimiser leur approche en permanence et de réduire les incertitudes dans le tableau de la situation.
- Assurer l’harmonisation et l’intégration complètes avec les structures et les processus de gestion de crise de la Confédération, des cantons et des communes ainsi que d’autres acteurs importants.
- Prévoir un mécanisme capable, en situation de crise, de réunir les bonnes personnes autour de la table, notamment dans le domaine de la science des données, afin de construire et de développer de manière flexible les modèles nécessaires, couche par couche, de créer des interfaces et de procéder aux analyses de données.
- Fournir les ressources humaines et financières nécessaires dans les domaines de la science des données, de l’analyse des problèmes et de la gestion des crises, tant pour mettre le système en place que pour l’utiliser en période de crise.
- Adopter un modèle de pensée qui reconnaît que seule une approche interdisciplinaire – qui dépasse les limites des domaines politiques traditionnels ainsi que des disciplines et de la science des données – est à même de relever les défis actuels.
- Prévoir des procédures et des collaborateurs capables de traduire rapidement le besoin d’informations des responsables politiques en exigences et en tâches concrètes relevant de la science des données (ingénierie des exigences).

Connaissances et compétences

- Améliorer la capacité à gérer les situations imprévisibles, où il n’est pas possible de cerner d’emblée les données requises (telle la gestion de jeux de données incomplètes grâce à l’utilisation de proxys).

⁹ TCR = tâches, compétences et responsabilité; RACI = réalisateur, approbateur, consulté et informé

- Assurer la capacité des collaborateurs à gérer des informations incertaines, notamment en ce qui concerne la qualité des données et des sources de données (l’évaluation des informations des réseaux sociaux, p. ex.), y compris à concilier rapidité d’exécution et qualité requise pour la prise de décision.
- Promouvoir les connaissances et les compétences interdisciplinaires (notamment pour ce qui est de la science des données et de la science de l’environnement) afin de relever plus efficacement les défis qui touchent plusieurs domaines (tels qu’ils se présentent souvent dans les situations de crise).
- Organiser régulièrement des cours de formation et des exercices de gestion de crise pour les responsables et le personnel spécialisé, afin de leur apprendre à utiliser la plateforme et les données (y c. l’amélioration des compétences numériques en général).
- Développer les connaissances et le savoir-faire pour intégrer rapidement des données de différents formats dans le tableau de la situation en cas de crise.

Parties prenantes

- Impliquer activement les parties prenantes telles que les autorités, les distributeurs d’énergie, les sociétés de transport et les services de secours, ainsi que les éventuels acteurs concernés au sein de la population et de l’économie, à la mise au point et à l’application de la solution (co-conception et co-développement). Les exercices de gestion de crise (voir sous «Connaissances et compétences») conviennent tout spécialement pour apporter des améliorations et tester des prototypes en «conditions quasi réelles».
- Amener tous les acteurs concernés à réaliser et à comprendre le rôle de l’intégration et de l’analyse des données en situation de crise.

Durabilité écologique

- Opter pour des serveurs et des centres de calcul économes en énergie afin de réduire la consommation énergétique et l’empreinte carbone.
- Appliquer des pratiques durables de collecte et de stockage des données, y compris les techniques de compression et les solutions visant à éviter les redondances.

3.3 Aide à la surveillance basée sur les données

Exemple du secteur aérien

Généralités

Description

Dans le cadre de ses missions et de ses attributions en matière de surveillance, l’administration fédérale mène un certain nombre d’activités (p. ex. élaboration, contrôle et application des lois et des ordonnances, délivrance d’autorisations

et de licences). La réalisation de ces missions implique de créer, collecter, regrouper, traiter et analyser de nombreuses données et informations provenant des sources les plus diverses. Les méthodes de la science des données peuvent être utilisées à différentes étapes de ce processus. Elles facilitent la gestion de volumes importants et hétérogènes de données et améliorent l'efficacité et l'efficacité des activités de surveillance.

Valeur ajoutée et pertinence

La science des données peut être un soutien précieux pour le personnel de l'administration fédérale assumant des tâches de surveillance. Ce potentiel est d'autant plus grand que les technologies et la société ne cessent d'évoluer (avec pour conséquence un accroissement des volumes de données disponibles).

Premièrement, les méthodes de la science des données peuvent en particulier être utilisées lors de la collecte et la validation des données: la qualité des données exploitées pour les activités de surveillance s'en trouve alors améliorée et la charge de travail des personnes assumant ces tâches allégée. Ces méthodes peuvent par exemple servir à convertir des documents électroniques non structurés (documents PDF, XLS, Word, etc.) en formats structurés ou à automatiser partiellement la collecte de données. Elles peuvent aussi être utilisées pour valider les données générées ou collectées, ce qui permet de détecter avant leur exploitation d'éventuelles anomalies (erreurs dans les données ou dues à des modifications opérées dans le domaine de la surveillance, p. ex.).

La science des données est par ailleurs d'une grande aide pour la mise en œuvre de la surveillance. Grâce à l'analyse et à la combinaison d'un grand nombre d'indicateurs, elle permet une évaluation fondée des risques et de la qualité de la performance des organisations ou domaines soumis à surveillance. L'accent peut ainsi être mis sur les risques ou sur la performance pour la planification de la surveillance. Cela permet de prioriser des domaines particulièrement à risque, et ainsi d'optimiser la surveillance et l'utilisation des ressources. Dans les cas moins complexes ou ceux comportant peu de risques, les systèmes basés sur l'IA peuvent même prendre en charge automatiquement certaines étapes de la surveillance (prolongation d'une autorisation, p. ex.). Ils permettent également d'automatiser et d'améliorer les étapes du processus de planification des activités de surveillance.

Enfin, la science des données sert également à améliorer les bases réglementaires des autorités. L'évaluation de grands volumes de données et de valeurs empiriques dans le cadre de l'apprentissage automatique peut ainsi contribuer à l'amélioration continue des prescriptions légales et à leur adaptation dans un contexte cohérent en constante évolution.

Surveillance de la sécurité aérienne

Contexte

L'Office fédéral de l'aviation civile (OFAC) est l'organe compétent en matière de surveillance de l'aviation civile et de développement durable de la navigation aérienne en Suisse. Il contribue notablement à la sécurité dans le domaine de l'aviation civile et coordonne étroitement ses activités avec l'Autorité de l'aviation militaire (MAA).

Pour satisfaire à son obligation de surveillance, l'OFAC doit pouvoir s'appuyer sur des données fiables et de qualité provenant de différentes sources: fournisseurs de services d'information aéronautique, de navigation aérienne et de trafic aérien (Skyguide, MétéoSuisse, Engadin Airport), aérodrômes, compagnies aériennes, pilotes, autorités étrangères, organisations internationales, etc. Indispensables pour la navigation aérienne et les instructions pour les pilotes, ces données sont essentielles pour assurer la sécurité et l'efficacité de l'aviation.

Pour garantir le respect de standards élevés en matière de sécurité aérienne, l'OFAC doit utiliser ses ressources de manière ciblée et efficiente, notamment dans les domaines où un potentiel d'amélioration de la sécurité existe.

Défis

Plusieurs défis doivent être relevés dans le cadre de la surveillance de la sécurité aérienne.

L'une des difficultés réside dans l'hétérogénéité et la qualité des données obtenues. Garantir la constance de la qualité des données collectées auprès de diverses sources et dans divers formats impose de remédier aux incohérences et lacunes constatées. Cette tâche mobilise de précieuses ressources. Une trop grande hétérogénéité des données peut notamment induire des erreurs dans le processus de surveillance, avec les conséquences négatives que l'on peut imaginer.

Aujourd'hui, les indicateurs sont collectés et intégrés manuellement dans des tableaux Excel et des bases de données Access, qui sont mis à jour manuellement sous forme de listes de priorités à partir desquelles l'activité de surveillance est déterminée. Cette façon de faire présente plusieurs points faibles. D'une part, elle est extrêmement chronophage, car chaque information doit être saisie et recoupée manuellement. D'autre part, elle résiste mal aux erreurs humaines (mauvais recoupements de données, saisies incorrectes, etc.). Ces lacunes peuvent affecter à la fois la qualité des données collectées et les décisions prises sur la base de ces données. Le traitement manuel des données rend par ailleurs difficile l'identification de modèles et de corrélations (problèmes d'entretien récurrents, manquements systématiques à l'application des procédures obligatoires, p. ex.). Le recours à des outils et des méthodes d'analyse de données appropriés s'avère décisif pour l'amélioration continue de la

sécurité aérienne et assure que ces constatations précieuses soient identifiées et exploitées.

Approche de solution

L’application de la science des données dans chaque aspect de l’utilisation des données peut contribuer significativement à l’amélioration des processus de surveillance de la sécurité aérienne. La condition préalable est de veiller à ce que la qualité des données, des sources de données et des méthodes de la science des données utilisées soit suffisamment élevée.

L’OFAC a reconnu le potentiel que recèle la science des données et développé sous le nom «Data Collection Service» (DCS) une plateforme de collecte de données aéronautiques. S’inscrivant dans le cadre de la transformation numérique, la création de cette plateforme vise à garantir que les données sont collectées numériquement dans la qualité requise et à valider les données aéronautiques par une mise en relation avec d’autres données et informations, et les exigences réglementaires. Des contrôles automatiques relatifs aux exigences réglementaires et techniques garantissent que les données sont transmises sans rupture de média et mises à disposition efficacement, et permettent de regrouper des évaluations de différents secteurs de l’aviation. Enfin, le système doit servir de référence pour les décisions réglementaires (octroi d’autorisations, p. ex.) et pour la publication des données. Une attention particulière est accordée à la simplification et à l’automatisation des processus. L’idée est notamment que les processus de travail de tous les acteurs puissent être exécutés sans rupture de média et que les exigences en matière de qualité des données soient garanties dans tous les cas.

D’autres applications de la science des données peuvent être initiées sur la base du DCS pour améliorer la surveillance et la qualité des données aéronautiques. La détection d’anomalies potentielles peut par exemple être d’une aide précieuse avant la validation des données par le personnel de surveillance. Les systèmes d’IA peuvent par ailleurs aider à identifier les domaines dans lesquels une qualité de données insuffisante pourrait être observée. Les personnes qui génèrent les données peuvent par exemple être alertées des erreurs ou des sources d’erreur détectées, et ainsi prêter une plus grande attention à ces domaines, ce qui permet d’accroître la qualité et la fiabilité des données.

Le deuxième domaine d’application est l’analyse des risques et l’évaluation de la qualité des prestations des organismes producteurs de données. Ces opérations se fondent sur deux principes couramment appliqués pour la surveillance dans le secteur de l’aviation: la «surveillance basée sur les risques» (*risk based oversight*, RBO) et la «surveillance basée sur les performances» (*performance based oversight*, PBO). Les activités de surveillance sont dimensionnées en fonction du risque que les règles ne soient pas suffisamment respectées. Dans le cadre de la surveillance fondée sur les risques, les activités de surveillance se concentrent sur les domaines à haut risque. La surveillance basée sur les performances, elle, se fonde sur des indicateurs tels que l’évaluation et la comparaison de la performance de différentes organisations en

matière de livraison de données, la gestion des risques de sécurité ou les temps de réaction pour la mise en œuvre de mesures correctives avant l'évaluation finale. Des analyses aléatoires sont par ailleurs réalisées en complément de ces deux approches et pour prévenir l'instauration de routines. Grâce à l'analyse d'un grand nombre d'indicateurs, les applications de la science des données contribuent à une évaluation fondée et complexe des risques et de la qualité des prestations des organisations ou des secteurs soumis à surveillance. La planification de la surveillance peut être ainsi mieux ciblée et l'affectation des ressources optimisée.

Troisièmement, la science des données peut aussi être utilisée pour élaborer un plan de surveillance. Pour une planification efficace et efficiente, de nombreux critères sont pris en compte, tels que la disponibilité et les compétences du personnel de surveillance, la complexité des mesures de surveillance ainsi que la situation géographique de l'objet concerné. Il est également possible de procéder à des ajustements de dernière minute dans le système de planification pour réagir à des incidents ou des événements imprévus. L'utilisation de systèmes basés sur l'apprentissage permet de tirer des enseignements des activités de surveillance passées et d'adapter la planification en fonction des connaissances acquises.

Enfin, la science des données peut contribuer au développement continu de prescriptions, notamment à la définition optimale de valeurs limites. L'exploitation de grands volumes de données et des valeurs empiriques du personnel de surveillance sur la base de l'apprentissage automatique permet par exemple d'examiner si les prescriptions définies produisent les effets escomptés (adaptation de la hauteur maximale autorisée des obstacles à la navigation aérienne, obligation d'autorisation moins poussée (et donc d'examen individuel) pour les objets d'une classe de risque inférieure pour l'aviation, etc.).

Transférabilité

Les possibilités de transfert de cette approche de la science des données à d'autres domaines politiques sont nombreuses et prometteuses. Dans presque tous les domaines d'intervention de l'État, il y a une nécessité d'améliorer la prise de décision et l'exécution en matière de surveillance, et d'optimiser la législation de manière rationnelle en se fondant sur des données probantes. Les méthodes de la science des données sont essentielles pour le traitement rationnel des informations, en particulier dans les domaines où une analyse de données exhaustive et complexe est nécessaire pour répondre aux exigences réglementaires dans le but de valider ces informations. Une telle approche fondée sur la science des données pourrait par exemple être utilisée dans le domaine de l'énergie pour tester et approuver de nouvelles technologies et procédures avant leur mise en œuvre.

Exigences

Données

- Avoir accès à des sources de données pertinentes et de qualité.
- Établir des normes et des protocoles de données en vue de la collecte cohérente et, si possible, automatisée des données dans un format simple et lisible par machine.

Réglementation et éthique (sur la base des principes fondamentaux de la DSStB)

- Assurer la protection des données et des informations, ainsi que la sécurité de l’information.
- Disposer d’une base légale pour une prise de décision autonome fondée sur des algorithmes et des méthodes de la science des données issues de l’apprentissage automatique et du domaine de l’IA.
- Assurer la transparence des algorithmes et des méthodes de la science des données utilisés dans l’analyse, et vérifier leur caractère non discriminatoire afin d’instaurer la confiance chez les parties prenantes concernées.
- Élaborer des mécanismes permettant de comprendre et d’expliquer comment les outputs du système d’analyse sont produits (transparence au niveau des processus, p. ex.).
- Disposer d’une base légale relative aux instructions destinées aux producteurs / fournisseurs de données concernant les types de données, leur format, la fréquence et le canal de transmission, les normes techniques, les processus d’assurance qualité et de plausibilité, etc. (soit dans les lois spécialisées soit dans des lois dédiées).
- Établir des responsabilités claires pour l’ensemble du processus de contrôle de plausibilité.

Technologie et infrastructure

- Élaborer une architecture évolutive, souple et performante, capable de gérer un volume croissant de données.
- Mettre à disposition l’infrastructure cloud nécessaire.
- Mettre à disposition une plateforme de données centralisée pour la saisie, le stockage et le traitement conviviaux des données, y compris une infrastructure de partage de données dotée d’interfaces définies et conviviales.
- Disposer d’une puissance de calcul suffisante et évolutive, pour l’entraînement des algorithmes et des méthodes de la science des données.
- Concevoir de nouveaux processus dans le cadre des activités de surveillance, en tenant compte des bases de données existantes et potentielles ainsi que des capacités pertinentes (puissance de calcul, p. ex.), en vue d’une exploitation optimale du potentiel des données et de la science des données.
- Développer des outils et interfaces permettant d’assister les producteurs et fournisseurs de données dans leurs tâches.

Gestion

- Décrire précisément la finalité et les tâches dans le cadre de l'activité de surveillance pour laquelle les données sont collectées et traitées (y compris la communication aux instances ou personnes soumises à surveillance).
- Établir une structure de gouvernance efficace pour la mise en œuvre et l'utilisation de la science des données dans l'activité de surveillance.
- Favoriser la compréhension et l'acceptation des méthodes de la science des données en vue de l'établissement d'une gestion administrative moderne axée sur les besoins changeants de la société.
- Créer des synergies et des coopérations entre les différentes autorités et domaines de surveillance.
- Développer des stratégies d'amélioration et d'adaptation continues des algorithmes et des méthodes de la science des données.

Connaissances et compétences

- Développer des compétences et une expertise dans le domaine de la science des données et de l'IA au niveau des autorités de surveillance (mesures de formation et de perfectionnement en lien avec l'utilisation de la science des données et de l'IA, p. ex.).
- Encourager le partage de connaissances et la coopération entre les scientifiques des données et les spécialistes.
- Développer et renforcer les capacités en vue d'accompagner les transformations numériques (compétences en matière d'innovation et de gestion du changement, p. ex.).

Parties prenantes

- Impliquer les parties prenantes de l'industrie, de la communauté scientifique et de l'aviation militaire (forces aériennes et *Military Aviation Authority*, MAA) dans le développement et l'application d'approches de surveillance fondées sur la science des données pour l'aviation civile (société).
- Favoriser la compréhension et l'acceptation des méthodes de surveillance fondées sur la science des données auprès des parties prenantes.

Durabilité écologique

- Utiliser des technologies et des infrastructures énergétiquement efficaces et ménageant les ressources afin de réduire au maximum la consommation d'énergie et l'empreinte écologique.
- Utiliser des solutions de stockage respectueuses de l'environnement, p. ex. des services cloud exploitant les énergies renouvelables ou des supports de stockage physiques respectueux de l'environnement.

3.4 Aide à la décision basée sur des systèmes de modèles holistiques

À l’exemple de la sécurité alimentaire

Généralités

Description

Les modèles holistiques informatisés sont des systèmes logiciels intégrés qui collectent, réunissent, analysent et représentent une multitude d’informations. Ils associent des données provenant de capteurs, de bases de données historiques et de connaissances spécialisées afin d’appréhender dans son intégralité le système étudié. Pour refléter la complexité de ce système, ils prennent en compte ses aspects économiques, écologiques, sociaux et techniques ainsi que leurs interactions.

Ces systèmes sont en mesure de fournir des prévisions, des simulations et des analyses reposant sur plusieurs sources de données de différents niveaux. Capables d’évaluer un grand nombre de scénarios et d’actions possibles, ils aident les décideurs à évaluer les conséquences possibles sur les différents aspects d’un système.

De tels systèmes se prêtent à de nombreux cas d’application, notamment le management environnemental et l’évaluation des risques. Précisons que les systèmes de modèles holistiques se distinguent des jumeaux numériques et des tableaux globaux de la situation. Loin de se limiter à la représentation virtuelle d’objets réels ou à la représentation de la situation actuelle, elles intègrent aussi des éléments abstraits – tels que des modèles mathématiques théoriques, des facteurs socioéconomiques et des aspects culturels – dans le but de fournir une compréhension plus globale de systèmes complexes et de réaliser des simulations dynamiques. Ces systèmes de modèles peuvent aussi recourir à des algorithmes et à des méthodes issus de l’apprentissage automatique et de l’IA. Le comportement d’un jumeau numérique peut par exemple constituer l’un des multiples facteurs d’une modélisation globale.

Valeur ajoutée et pertinence

Le recours à des systèmes de modèles holistiques offre des avantages considérables dans l’élaboration des politiques, en particulier dans le contexte de défis complexes à relever à long terme ou posés dans une situation exceptionnelle.

Le premier avantage de ces systèmes est de pouvoir fonder des processus décisionnels sur une base de données solide. Plutôt que de considérer des points de données isolés, les systèmes de modèles holistiques intègrent une multitude d’informations pertinentes provenant de différentes sources. Ils extraient et analysent ces données, permettant ainsi une approche globale de la problématique.

Deuxième avantage, les systèmes de modèles holistiques permettent de mieux comprendre les relations complexes, illustrant en particulier les interactions entre les différents facteurs du système considéré. Le décideur obtient une compréhension détaillée de ces interactions, ce qui lui permet de prendre des mesures appropriées tout en réduisant autant que faire se peut les effets secondaires indésirables.

Troisième avantage des systèmes de modèles holistiques, ils sont à même de simuler différents scénarios de façon à aider les décideurs à évaluer au préalable l'efficacité de différentes mesures. Plutôt que de présenter uniquement des données et des expériences historiques et de les enrichir de prévisions extrapolées, les systèmes de modèles holistiques sont en mesure de présenter des scénarios futurs possibles et d'évaluer l'impact de différentes options politiques. En identifiant à temps les risques ainsi que les effets potentiels, voire inattendus, des mesures politiques, ces systèmes offrent une base de décision importante.

Enfin, les systèmes de modèles holistiques contribuent à adapter et à améliorer en continu les processus de décision. Ils sont capables d'apprendre, d'intégrer les retours d'information – notamment dans le cadre de feedbacks humains – et les expériences tirées de décisions antérieures, puis d'intégrer toutes ces informations dans des simulations et des analyses subséquentes. Ce faisant, les systèmes de modèles holistiques contribuent à améliorer sans cesse la qualité des décisions tout en aidant les décideurs à mieux répondre aux exigences et aux conditions appropriées en constante évolution.

Exemple d'application

Systèmes de modèles holistiques en matière de sécurité alimentaire

Contexte

Déterminer le cadre politique permettant d'assurer la sécurité alimentaire de la population fait partie des tâches fondamentales de tout État. De nos jours, il faut d'une part garantir l'approvisionnement alimentaire de la population à tout moment, même si les rendements sont en baisse ou les possibilités d'importer limitées. D'autre part, il est décisif d'assurer la durabilité à long terme du secteur agroalimentaire afin de couvrir les besoins alimentaires actuels et futurs, tout en permettant une alimentation saine et en réduisant autant que possible l'impact négatif sur l'environnement et la société. Par «sécurité alimentaire», nous entendons précisément cet équilibre entre garantie de l'approvisionnement alimentaire à court et à moyen terme d'une part, et durabilité écologique, économique et sociale à long terme du secteur agroalimentaire d'autre part.

Pour contribuer à la sécurité alimentaire, l'Office fédéral pour l'approvisionnement économique du pays (OFAE) et Agroscope (centre de compétence de la Confédération pour la recherche dans les domaines de l'agriculture et de la filière alimentaire) ont mis au point un système de modèles holistiques du secteur agroalimentaire. Ce système représente, du point de vue production, les diverses interactions entre les surfaces agricoles, les effectifs d'animaux de

rente, la transformation des denrées alimentaires, le stockage, le commerce extérieur et les besoins alimentaires. Son objectif est de déterminer, pour certains scénarios de pénurie grave, le type de production alimentaire suisse qui serait le mieux à même de couvrir les besoins énergétiques et nutritionnels de la population.

Le système de modèles holistiques en question est en voie d’être élargi pour devenir un modèle de durabilité global intégrant des indicateurs économiques, écologiques et sociaux. Ce modèle élargi fait partie du projet *Agri-Food Economy* d’Agroscope, qui vise à fournir une approche encore plus globale du secteur agroalimentaire afin d’évaluer la résilience et la durabilité des futurs systèmes alimentaires suisses.

Défis

En raison de l’évolution des habitudes alimentaires, des préoccupations environnementales, des restrictions commerciales, de la volatilité des chaînes d’approvisionnement (les engrais provenant de Russie, p. ex.) et des conséquences du changement climatique, le secteur agroalimentaire est devenu plus complexe, plus dynamique et plus vulnérable.

Dans un système aussi dynamique et complexe, il n’est que difficilement possible de garantir la sécurité alimentaire en recourant aux modélisations actuelles. Il faut une vision plus globale du secteur agroalimentaire, qui tienne compte de l’ensemble des éléments susceptibles d’influer sur les aspects de durabilité et sur leurs interactions au sein du système.

D’une part, la complexification des processus du secteur agroalimentaire – car il faut considérer aussi les chaînes internationales d’approvisionnement en denrées alimentaires et en moyens de production – exige de nouvelles sources de données et augmente donc le travail de collecte et d’intégration de données dans le système de modèles. Face à l’hétérogénéité des données et de leur qualité, cette phase peut s’avérer chronophage tout en comportant un risque d’erreurs susceptibles de fausser les résultats des simulations inhérentes, le tout augmentant l’incertitude des données. À quoi s’ajoute que les décideurs ont besoin de types de données différents en fonction de l’utilisation prévue (court ou long terme). La modélisation doit être à même de répondre aux exigences des différents utilisateurs tout en étant suffisamment flexible pour être adaptée à des besoins en constante évolution.

D’autre part, avec l’accroissement de la complexité et de la dynamique, il devient aussi nécessaire de mettre à jour périodiquement tant les logiques du système de modèles que les données sous-jacentes et d’en vérifier l’exactitude. Voilà les conditions à remplir pour que la modélisation reste en mesure de fournir des affirmations et des prévisions valables et continue donc de remplir sa fonction d’instrument d’aide à la décision.

En raison de la grande complexité du système de modèles, il est tout aussi important que les décideurs soient à même de comprendre son utilisation et

l’interprétation de ses résultats, rapidement et sans grands efforts. Faute de quoi, ils risquent de ne guère exploiter, voire d’ignorer, les conclusions du système de modèles pour prendre des décisions, ce qui limiterait considérablement l’efficacité de ladite modélisation comme outil d’aide à la décision.

Solution proposée

Le système de modèles existant doit être développé et affiné pour prendre en considération la complexité croissante du système alimentaire et fournir une aide à la décision basée sur des données en matière de sécurité alimentaire. Le système de modèles reproduit le système alimentaire dans toute sa complexité, offrant ainsi une base fiable pour la prise de décisions stratégiques et opérationnelles, tant dans le but d’optimiser l’approvisionnement en denrées alimentaires à court et à moyen terme que d’assurer la sécurité et la durabilité de l’approvisionnement à long terme.

Il intègre de nombreuses informations afin de donner une vision complète du système alimentaire. En font partie les besoins en moyens de production dans le secteur agroalimentaire (machines, engrais minéraux, capacités de transport, etc.). Sont également prises en compte les fonctions et les alternatives de production (si certains moyens de production devaient s’avérer indisponibles), les données relatives à l’aptitude des sols, les informations sur la situation actuelle de l’approvisionnement et sur les menaces potentielles, ainsi que les coûts et les effets des mesures opérationnelles et politiques.

Le système de modèles est par ailleurs en mesure d’évaluer la sécurité d’approvisionnement et la durabilité des futurs systèmes alimentaires. Pour ce faire, il doit être complété par des informations sur les conditions appropriées futures (méthodes de production, rendements, etc.), les tendances et les comportements alimentaires, les effets des régimes alimentaires sur la santé et les indicateurs de durabilité de l’alimentation.

Pour que le système de modèles existant reste gérable, en dépit de ces extensions, il importe de définir les fonctions souhaitées au début de la phase de développement ou d’extension. La simulation de pénuries d’approvisionnement à court terme exige par exemple d’autres bases de données et des algorithmes différents que ceux nécessaires à la prévision et à l’évaluation des systèmes alimentaires futurs. Dès le départ, il s’agit de clairement définir à quelles questions le système de modèles pourra répondre, et auxquelles il ne le pourra pas. En parallèle, il doit être conçu de manière modulaire afin de permettre une mise à l’échelle et une adaptation du système dès lors que d’autres aspects (telle la sécurité alimentaire) doivent être pris en compte dans le processus de prise de décision.

Le défi consiste à réunir et à intégrer toutes ces informations, de manière efficace et sans erreur. C’est précisément ici que les méthodes de la science des données issues des statistiques avancées, de l’apprentissage automatique et du domaine de l’IA s’avèrent particulièrement utiles, par exemple pour automa-

tiser et perfectionner les processus de collecte et de plausibilisation des données. Ainsi, les procédures de détection d’anomalies – basées sur des méthodes de la science des données – servent à déceler des comportements anormaux dans des jeux de données. Elles permettent d’identifier de manière (partiellement) automatique les valeurs aberrantes parmi les données et de les signaler rapidement aux experts ou aux fournisseurs de données afin de vérifier si elles sont effectivement erronées.

Tout comme le traitement des données, la présentation des résultats joue elle aussi un rôle décisif. Le système de modèles présente les résultats des analyses de données, des prévisions et des simulations de manière compréhensible et attrayante dans le but d’offrir aux décideurs un aperçu rapide de la situation actuelle et des options envisageables. Des interfaces et des fonctions conviviales sont alors d’une grande aide, par exemple des recommandations générées de manière (partiellement) automatique. Le modèle indique alors en toute transparence la qualité de ses chiffres, de ses affirmations et de ses recommandations. Si l’incertitude de l’output de la modélisation est par exemple élevée (mesurée au moyen de méthodes de la science des données, notamment des analyses de sensibilité ou des simulations de Monte Carlo), autrement dit s’il est probable que la valeur réelle s’écarte significativement de la valeur représentée, celle-ci est signalée comme telle.

De par l’utilisation de la science des données, la modélisation est en outre capable d’apprendre. Sur la base des connaissances acquises lors d’événements passés et du feed-back des experts ou des utilisateurs, elle est ainsi en mesure de mieux prédire ou simuler l’impact de diverses mesures politiques sur la sécurité de l’alimentation en Suisse à long terme.

Transférabilité

Il est possible de recourir à un système de modèles holistiques de ce genre, conçu de façon modulaire, dans de nombreux autres domaines politiques où les interactions complexes jouent un rôle de premier plan. Il pourrait par exemple être utilisé en politique de la santé pour évaluer les mesures de lutte contre les épidémies, voire en politique énergétique pour soutenir la planification des ressources et la mise en œuvre de technologies durables.

Exigences

Données

- Assurer la qualité des données, notamment en élaborant des normes et des protocoles qui permettent d’uniformiser les différents formats et structures de données provenant de diverses sources.
- Assurer l’actualité des données, notamment en mettant à disposition les interfaces nécessaires.
- Élaborer des processus validés pour collecter, préparer, traiter et stocker les données afin de garantir leur intégrité et leur accessibilité.

- Assurer la disponibilité des données essentielles (données sur l'aptitude à la culture des terres agricoles, données sur les processus de production alternatifs, p. ex.).
- Quantifier la variabilité des données comme base pour les simulations.

Réglementation et éthique (sur la base des principes fondamentaux de la DSStB)

- Garantir la protection des données et de l'information ainsi que la sécurité de l'information, par exemple en limitant l'accès ou en prévoyant des mécanismes d'anonymisation des données à caractère personnel afin de préserver la vie privée des personnes concernées.
- Assurer la transparence des algorithmes et des méthodes issus de la science des données utilisés pour renforcer la confiance des parties prenantes et éviter toute discrimination.
- Préserver les secrets d'affaires des acteurs privés ainsi que les droits de la personnalité des particuliers (flux de mouvements p. ex.).
- Mettre au point des mécanismes qui permettent de comprendre et d'expliquer comment on obtient les outputs du système d'analyse des données.

Technologie et infrastructure

- Mettre en place une infrastructure informatique performante, flexible et évolutive pour faire face à la quantité croissante de données et permettre des simulations et des analyses de données exigeant une grande puissance de calcul.
- Assurer un stockage efficace et à long terme des données (*inputs* et *outputs*) afin de pouvoir contrôler les versions et de reproduire les modélisations.
- Garantir la compatibilité des systèmes avec les technologies et les normes existantes, et anticiper les changements technologiques à venir.
- Disposer d'une puissance de calcul évolutive suffisante permettant d'entraîner les algorithmes et les méthodes issus de la science des données (surtout lors de l'utilisation d'éléments de l'apprentissage automatique et de l'IA).
- Mettre à disposition une plateforme cloud interne à l'administration pour l'application (*storage* et *computing*).
- Mettre en place des mesures de sécurité solides pour protéger les systèmes contre toute cyberattaque ou fuite de données.

Gestion

- Mettre en place une gestion de projet efficace pour piloter la mise en œuvre et le développement des systèmes de modèles holistiques.
- Examiner périodiquement les performances et les résultats des systèmes pour valider les algorithmes et les méthodes issus de la science des données, si nécessaire les corriger.

- Identifier et atténuer les risques potentiels liés à la mise en œuvre et à l’utilisation des systèmes de modèles holistiques, notamment en contrôlant constamment la qualité des modélisations obtenues au moyen d’un système de suivi.

Connaissances et compétences

- Assurer la disponibilité d’experts de divers domaines afin de reproduire la complexité du système ou du processus examiné.
- Prévoir des formations (continues) périodiques adaptées aux différents niveaux, tant pour les utilisateurs du système que pour les décideurs. Le but est de permettre une utilisation optimale des modélisations, par exemple des formations sur la gestion de l’incertitude dans la modélisation, sur l’interprétation des résultats moyennant des probabilités et des intervalles de confiance. De même, mettre à disposition les informations nécessaires en matière de qualité des données et des analyses de la modélisation.
- Mettre en place des processus de collecte, de stockage et de transmission des connaissances et des expériences afin d’assurer une amélioration et une innovation continues.

Parties prenantes

- Impliquer activement les parties prenantes dans le processus de développement et de prise de décision afin de garantir la transparence et de renforcer la confiance dans les systèmes de modèles holistiques.
- Mettre en place des canaux et des procédures pour recueillir les feedbacks et les suggestions des parties prenantes, puis les intégrer dans le développement des systèmes de modèles holistiques.

Durabilité écologique

- Opter pour des serveurs et des centres de calcul économes en énergie afin de réduire la consommation énergétique et l’empreinte carbone.
- Pratiquer des méthodes durables en matière de collecte, de stockage et d’entraînement des modèles, y compris la réduction de la redondance des données au niveau requis pour assurer la sécurité des données et le recours à des techniques de compression des données et des modèles.

3.5 Contrôles de plausibilité destinés à garantir la qualité des données

À l’exemple des assurances sociales

Généralités

Description

L’assurance qualité des données constitue un élément essentiel pour per-

mettre des décisions politiques fondées ainsi qu’une exécution et une surveillance réglementaires. Dans le contexte de l’augmentation constante de la quantité de données à disposition (et devant être surveillées), les méthodes et les algorithmes de «détection d’anomalies» issus de la science des données peuvent aider à repérer des schémas inhabituels dans la foison de données. Ces schémas peuvent révéler des erreurs involontaires (ou parfois aussi des informations délibérément fausses), par exemple lors de la création ou de la transmission des données ainsi que lors de leur traitement.

Aux fins d’identifier ces erreurs, les méthodes et les algorithmes de la science des données (issus notamment de l’apprentissage automatique et du domaine de l’IA) analysent automatiquement les données reçues. Ces systèmes peuvent marquer et corriger automatiquement les erreurs évidentes. Les incertitudes, elles, sont transmises ou signalées à des spécialistes pour examen. Ainsi par exemple la déclaration d’un congé maternité pour une personne de sexe masculin est indéniablement une erreur et peut être corrigée automatiquement. En revanche, l’immatriculation d’un grand nombre d’étudiants quinquagénaires dans le courant d’un même semestre est certes improbable, mais pas nécessairement une erreur. Un tel cas devrait par conséquent être signalé au fournisseur de données, en l’occurrence une haute école, afin d’en vérifier la plausibilité.

Pour détecter ce type d’«anomalies» dans les données, les modèles sont entraînés au «comportement normal» des données et à l’identification de comportements qui s’en écartent. Une vérification et une mise à jour continues du modèle permettent d’intégrer les retours et donc d’améliorer la précision.

Garantir une assurance qualité efficace des données par le biais de contrôles de la plausibilité revêt une grande importance surtout pour les données générées lors de l’exécution et qui jouent ensuite un rôle majeur dans la surveillance et le pilotage de l’administration fédérale. Des données erronées qui ne sont pas reconnues comme telles peuvent déboucher notamment sur des conséquences financières pour les personnes concernées (dans le domaine des subventions, p. ex.) ou conduire à des conclusions incorrectes en matière de réglementation (dans la planification hospitalière, p. ex.).

Valeur ajoutée et pertinence

Le recours à des méthodes et des algorithmes de «détection d’anomalies» issus de la science des données (en particulier l’apprentissage automatique et l’IA) pour assurer la qualité des données présente de nombreux avantages, l’un des plus décisifs étant d’accroître l’efficacité de l’assurance qualité. Ces modèles sont en mesure de reconnaître automatiquement des schémas dans une foison de données, singularités qui pourraient passer inaperçues lors d’un contrôle manuel ou dont l’identification manuelle demanderait énormément de temps. Ils permettent donc une vérification systématique, complète et précise des données. Une autre valeur ajoutée réside dans la capacité de ces méthodes et algorithmes (issus notamment de l’apprentissage automatique et de l’IA) de repérer des erreurs inconnues. En effet, les procédés de plausibilisation

usuels, qu'ils soient manuels ou automatisés, ne localisent que les erreurs définies par avance, alors que les méthodes et les algorithmes de «détection d'anomalies» trouvent également des erreurs jusque-là inconnues. Ces systèmes reposent sur la probabilité d'anomalies et non pas uniquement sur les erreurs et les règles prédéfinies (p. ex. «congé maternité pour les personnes de sexe féminin seulement»).

Autre avantage: les processus sont optimisés grâce à l'automatisation des tâches routinières. Ainsi, les experts ne doivent plus intervenir que pour les questions complexes. Il est donc possible d'utiliser plus efficacement l'expertise disponible et les spécialistes éprouvent une plus grande satisfaction au travail, car ils peuvent se concentrer sur des tâches complexes et stimulantes.

Enfin, les données plausibilisées fournissent une base pour approfondir, sans distorsion, l'analyse dans le domaine de la «détection d'anomalies» et soutenir ainsi les employés spécialisés et les décideurs dans l'identification de cas de fraude ou d'escroquerie (p. ex. dans les déclarations d'impôt ou les demandes de prestations sociales).

La «détection d'anomalies» visant à garantir la qualité des données revêt une importance considérable à tous les stades du processus de décision politique. L'assurance qualité par le biais de contrôles de plausibilité permet de garantir que les décisions politiques, tout comme l'exécution et la surveillance, reposent à toutes les étapes sur des données solides et de grande qualité.

Exemple d'application

Réformes politiques dans le domaine des assurances sociales

Contexte

L'Office fédéral des assurances sociales (OFAS) gère de grandes quantités de données personnelles qui sont nécessaires pour la surveillance, le pilotage, les rapports statistiques et les analyses ainsi que pour le développement de la sécurité sociale. Différents organes d'exécution, dont les caisses de compensation et les offices AI, transfèrent en continu d'importants flux de données à l'OFAS. Bien que la fréquence de ces transmissions soit variable, l'OFAS reçoit d'habitude les informations nécessaires rapidement et régulièrement.

Selon le processus de flux de données, les informations parviennent des organes d'exécution à la Centrale de compensation (CdC), où elles sont vérifiées, soumises à un contrôle de plausibilité et préparées avant d'être mises à la disposition de l'OFAS.

Jusqu'ici, cette vérification des données est effectuée, d'une part, dans les organes d'exécution sous la forme de contrôles par échantillonnage et, d'autre part, à l'aide de tests de plausibilité automatisés qui ont lieu lors de la transmission et qui sont régis par des directives ad hoc. Ces contrôles de plausibilité comprennent un aspect technique (vérification que la transmission des données est conforme au format prescrit) et portent également sur le contenu. Ce

dernier contrôle se concentre essentiellement au niveau des personnes assurées. Le système vérifie donc notamment si une personne a touché des versements de plusieurs organes d’exécution, si le début du droit à la prestation est antérieur à la fin de ce droit ou si le nombre de jours indemnisés dépasse le maximum légal.

Défis

Au lendemain de réformes politiques, on note un empressement à voir une première évaluation des effets de ces changements. Mais il faut évidemment attendre quelque temps avant de disposer de premiers chiffres, car des assurés doivent commencer par faire valoir des droits aux prestations tombant sous le coup de la nouvelle législation, puis les organes d’exécution doivent les vérifier et, en cas d’approbation, les montants doivent être versés. Ensuite seulement, les informations seront transmises par les canaux prévus. De plus, il y a une vulnérabilité accrue aux erreurs de collecte et de transmission des données durant les phases de transition et lors du lancement d’un nouveau système. Ces incertitudes peuvent entraîner une baisse temporaire de la qualité des données, ce qui, à son tour, retarde ou fausse l’évaluation des effets de la réforme.

Pendant cette phase sensible, il est primordial d’assurer une vérification des données complète et efficace. Le contrôle (partiellement) automatisé réalisé jusqu’ici se limite avant tout à la plausibilisation de données individuelles, alors que la surveillance systématique des évolutions est souvent négligée. Les irrégularités au niveau du système peuvent de ce fait passer inaperçues, par exemple une diminution soudaine du nombre d’assurés déclarés par un organe d’exécution ou une augmentation / baisse inattendue du montant total des paiements. Ce type d’«anomalies» peuvent indiquer des erreurs de données, surtout pendant une phase de transformation, et ne devraient par conséquent pas passer sous le radar.

Solution proposée

Un contrôle de plausibilité complet des données entrantes dans le domaine des assurances sociales permet de relever les défis énumérés, tant dans le cadre des activités opérationnelles que dans le contexte de réformes de grande envergure. Il est réalisé à l’aide d’un modèle de «détection d’anomalies» basé sur des méthodes et des algorithmes de la science des données, y compris des méthodes de la statistique traditionnelle et avancée, de l’apprentissage automatique et de l’IA. Un tel modèle permet de déceler à la fois les erreurs qui sont déjà connues et celles qui sont nouvelles (p. ex. qui apparaissent pour la première fois à la suite d’une réforme). Il permet en particulier de procéder à un contrôle de plausibilité pour le système dans son ensemble et non seulement au niveau du cas individuel. Ainsi, lorsque le nombre d’ajournements de l’AVS augmente après une réforme, la probabilité qu’une nouvelle rentière ou un nouveau rentier ait 65 ans diminue. Si une caisse de compensation devait

s'écarter de ce schéma, le modèle le reconnaîtrait. Une telle divergence pourrait indiquer une erreur de programmation, qui devrait alors être vérifiée par un spécialiste.

D'une manière générale, les contrôles ont lieu automatiquement après réception des données des organismes d'exécution. Selon la gravité des «anomalies» détectées, un retour automatique est envoyé aux services concernés, avec demande de corriger ou de plausibiliser les données. Des cycles d'apprentissage sont ainsi établis, qui améliorent à la fois la saisie des données par les organismes d'exécution et le modèle de «détection d'anomalies».

Il est possible ensuite d'examiner les «anomalies» dans la base de données plausibilisée ainsi que les effets des réformes ou d'autres mesures.

L'efficacité et la fiabilité de la «détection d'anomalies» supposent une surveillance et une mise à jour des algorithmes en continu. Dans ce contexte, il est essentiel que les «anomalies» potentielles soient interprétées par des spécialistes afin de s'assurer qu'elles indiquent bel et bien des erreurs.

Outre le recours systématique aux modèles de «détection d'anomalies», une mesure additionnelle pour améliorer la qualité des données pourrait consister à mettre en place une plateforme informatique destinée aux personnes qui produisent des données. Cette plateforme pourrait aider ces dernières à transmettre correctement les données. Les méthodes et les algorithmes issus du domaine de l'IA pourraient contribuer à garantir une transmission des données dans le bon format et sans erreur.

Transférabilité

Ce cas de figure est spécialement important pour les services fédéraux qui reçoivent différentes données, les traitent, les analysent et prennent des décisions sur cette base. Dans ce type de situation, les unités administratives n'ont souvent aucun contrôle sur la collecte et le transfert des données puisqu'elles reçoivent les données de producteurs de données externes et ne peuvent donc pas garantir elles-mêmes leur qualité.

Cette affirmation est vraie pour la plupart des domaines de la surveillance et de la statistique, qui dépendent de données de tiers pour accomplir leurs tâches. D'autres exemples se trouvent dans le domaine de l'exécution cantonale (p. ex. déclarations de déchets). Les données dépendent en l'occurrence de services externes et il est essentiel qu'elles soient précises et fiables.

Exigences

Données

- Intégrer les sources de données nécessaires à l'entraînement des méthodes et des algorithmes, en tenant compte des normes sur la protection et la sécurité des données.

- Assurer une transmission standardisée des données à l’administration fédérale dans un format facilement lisible par machine.

Réglementation et éthique (sur la base des principes fondamentaux de la DSStB)

- Disposer d’une base légale pour les directives destinées aux producteurs / fournisseurs de données concernant le format de données, la fréquence et le canal de transmission, les normes techniques, les processus d’assurance qualité et de plausibilité, etc. (soit dans les lois spécialisées soit dans des lois dédiées telles que la loi sur le système d’information dans le domaine des assurances sociales [en cours d’élaboration]).
- Garantir la protection des données et des informations ainsi que la sécurité de l’information, par exemple en limitant l’accès aux méthodes et aux algorithmes.
- Déclarer les méthodes et les algorithmes de science des données utilisés afin de s’assurer la confiance des parties prenantes et de garantir un comportement responsable.
- Éviter la discrimination et les distorsions dans les méthodes et algorithmes de la science des données, par exemple en assurant un «étiquetage» précis des données.
- Établir des responsabilités claires pour l’ensemble du processus de contrôle de plausibilité.

Technologie et infrastructure

- Mettre à disposition les interfaces conviviales nécessaires pour la transmission des données.
- Mettre à disposition des ressources de calcul et de stockage suffisantes pour le traitement et l’analyse de grandes quantités de données.
- Implémenter des procédures automatisées de préparation, de modélisation et de mise à jour des données afin d’améliorer l’efficacité et la précision.
- Actualiser régulièrement la pondération des modèles.
- Mettre en place les infrastructures nécessaires à la gestion des données et aux accès aux données.
- Mettre à disposition des outils et des plateformes de science des données appropriés.
- Assurer l’interopérabilité à l’échelon des données ainsi que des méthodes, des outils et des plateformes de science des données.

Gestion

- Mettre à disposition les ressources financières et humaines nécessaires à la mise en œuvre.
- Mettre en place des méthodes efficaces de gestion de projet afin d’assurer une planification, une mise en œuvre et un suivi structurés du projet de détection d’anomalies.

- Encourager la collaboration entre les scientifiques des données et les experts afin de permettre le transfert de connaissances dans le cadre du développement (continu) du système.
- Mettre en place des processus d’entretien et d’amélioration continus par le biais de révisions, de mises à jour et d’intégrations des retours régulières.
- Mettre en place des structures pour mesurer l’effet de l’application.
- Attester la volonté de tous les décideurs de l’administration fédérale et de tous les niveaux étatiques de soutenir les solutions et les structures numériques communes, et de mettre à disposition les ressources nécessaires pour les utiliser.

Connaissances et compétences

- Veiller à ce que les scientifiques des données et les experts disposent des connaissances et des compétences nécessaires pour mener à bien les projets de «détection d’anomalies».
- Fournir aux collaborateurs des formations et des formations continues leur permettant d’approfondir et de maintenir à jour leurs connaissances en matière de science des données ainsi que dans leur propre domaine. Il s’agit en particulier de connaissances de base en science des données à l’intention des experts et des personnes occupant des fonctions dirigeantes afin de renforcer la compréhension du processus de science des données dans l’ensemble de l’organisation.

Parties prenantes

- Instaurer un dialogue permanent avec les producteurs de données afin d’améliorer la qualité des données d’entrée.
- Assurer une base commune de données plausibilisées également en cas de collaborations.
- Associer les parties prenantes importantes dans le processus de développement de l’application (lorsque cela est possible) afin de s’assurer leur acceptation et leur soutien.

Durabilité écologique

- Utiliser des technologies et des infrastructures efficaces sur le plan énergétique afin de réduire la consommation d’énergie et l’empreinte carbone.

3.6 Jumeaux numériques

À l’exemple d’un jumeau numérique de l’agriculture suisse

Généralités

Description

Un jumeau numérique est un modèle virtuel d'un objet ou d'un système physique dynamique. Il est essentiellement utilisé pour des objets et des systèmes de grande complexité pour lesquels il est difficile de comprendre les corrélations et de déduire les effets ou pour lesquels la collecte de données par d'autres moyens (des capteurs, p. ex.) est trop coûteuse ou trop fastidieuse. Les bâtiments agricoles destinés à l'élevage et équipés de systèmes complets de chauffage et de ventilation en sont un exemple. La gestion de tels systèmes doit également tenir compte des besoins biologiques complexes des animaux.

Le jumeau numérique collecte et traite des données et informations sur son jumeau physique, en développant ainsi une représentation numérique. La transmission de ces données et de ces informations peut se faire aussi bien en temps réel qu'en différé. Le jumeau numérique reflète tous les aspects pertinents du système réel, de la forme et des mouvements aux matériaux et aux conditions environnementales, en passant par les indicateurs de performance.

Pour l'utilisateur, le jumeau numérique se présente sous la forme d'un tableau de bord qui est associé à un modèle organisé dans le temps et dans l'espace. Souvent, une telle interface utilisateur est dotée en plus de champs de saisie qui permettent de modifier certains paramètres et d'intégrer de nouveaux inputs et à l'aide desquels les utilisateurs peuvent simuler et analyser les effets des changements apportés. Cela permet de tester différentes options ou de jouer des scénarios de manière rapide et flexible.

Déjà utilisé dans différents domaines de la production industrielle, de la logistique ou de la médecine, le concept de jumeau numérique sert, par exemple, à modéliser des trains et des systèmes ferroviaires, notamment des infrastructures associées comme les rails, la signalisation et les gares. Des simulations permettent d'analyser les flux de trafic et d'optimiser les horaires. De même, il est possible de tester et d'analyser le comportement du système ferroviaire dans différentes situations, par exemple lors de changements de météo, de pannes techniques ou encore d'évolution du nombre de passagers.

Autre exemple, la modélisation des denrées alimentaires et des vaccins dans les chaînes d'approvisionnement, y compris de l'infrastructure associée (chaînes du froid, stockage et moyens de transport). Les jumeaux numériques facilitent en l'occurrence l'analyse de la durée de conservation des produits dans différentes situations. Il peut s'agir par exemple de variations de température, de problèmes logistiques ou d'évolutions de la demande. Les chaînes d'approvisionnement et les processus de vente s'en trouvent améliorés.

Valeur ajoutée et pertinence

Les jumeaux numériques apportent également une valeur ajoutée significative dans l'élaboration des politiques. D'une part, ils modélisent des systèmes et des objets complexes et permettent, grâce à la visualisation et la simplification, de mieux comprendre leurs éléments et leurs interactions dynamiques.

D'autre part, ils permettent une surveillance complète de tels objets ou de tels systèmes très complexes, et dans certains cas même en temps réel. Cela permet d'acquérir des connaissances servant à améliorer les contrôles et la maintenance des objets ou des systèmes, tout en ayant un impact positif sur la qualité des résultats.

Les jumeaux numériques peuvent par ailleurs fournir des données détaillées et actuelles sur l'état et l'efficacité des infrastructures, des ressources et des processus. On peut imaginer, à titre d'exemple, des exploitations agricoles où la consommation de pesticides et d'eau est enregistrée avec précision.

La possibilité de simuler un ensemble de scénarios et de mesures politiques avant leur mise en œuvre est un aspect central supplémentaire des jumeaux numériques. On peut ainsi évaluer l'impact des décisions politiques, et savoir, par exemple, quelles sont les retombées d'un nouveau système de péage routier dans différents domaines comme l'environnement, l'économie ou les transports. Les données ainsi générées sont particulièrement utiles lorsque leur collecte par le biais d'autres méthodes demande un gros investissement sur les plans humain ou financier ou d'importantes ressources naturelles (p. ex. des expériences).

Face à la complexité croissante des systèmes économiques et sociaux et aux exigences accrues en matière de traçabilité des décisions politiques, l'utilisation de jumeaux numériques devient cruciale. Compte tenu des données et de la capacité de calcul dont on dispose actuellement, il est possible de se servir du potentiel des jumeaux numériques pour relever ces défis.

Exemple d'application

Jumeau numérique dans le système agricole suisse

Contexte

En Suisse, les jumeaux numériques sont déjà utilisés avec succès dans certains domaines de l'agriculture et dans l'élaboration de la politique agricole. Ils contribuent à la conception et à la mise en œuvre de plans d'action nationaux et sont par exemple utilisés pour réduire l'utilisation des produits phytosanitaires et pour contrôler les flux d'éléments nutritifs. Leurs résultats peuvent de surcroît être mis à profit dans l'élaboration de modèles de politique environnementale et dans la réalisation d'écobilans. Ces applications permettent une planification et une optimisation plus précises et plus efficaces des processus de production et contribuent à minimiser l'utilisation des ressources et l'impact environnemental des activités agricoles ainsi qu'à optimiser le rendement. En général, chacun de ces jumeaux numériques prend en considération une seule exploitation agricole. Ils sont adaptés aux conditions et aux exigences spécifiques de chaque exploitation et assurent une planification et une optimisation ciblées des processus de production. Jusqu'à présent, le degré d'automatisation et de connexion à d'autres modèles agricoles, systèmes de modèles et

ensembles de données issus de l'économie agricole n'a cependant été mis en œuvre que de manière limitée.

Défis

Les défis à relever en Suisse dans le domaine des processus de production agricole et de la science du travail, en particulier, consistent avant tout à surmonter la complexité systémique. Les outils de planification et d'optimisation du travail actuellement disponibles couvrent la complexité des exploitations individuelles et se concentrent essentiellement sur l'amélioration de l'efficacité au niveau des exploitations individuelles. L'optimisation de l'interaction entre homme et machine est au cœur des enjeux.

D'une part, il manque une approche qui puisse rendre compte des effets et des corrélations tant au niveau régional que national. En outre, d'autres facteurs doivent être intégrés si l'on veut comprendre le système agricole plutôt que de se focaliser uniquement sur l'efficacité, comme jusqu'ici. L'accent est alors mis sur des facteurs écologiques et sociaux.

Par exemple, la planification et la mise en œuvre de mesures visant à diminuer l'impact environnemental et à répondre aux besoins des individus (p. ex. la main-d'œuvre) dans les processus agricoles nécessitent une analyse détaillée à un niveau supérieur. Il est essentiel de comprendre les interactions entre les différents éléments du système comme les résidus de pesticides, la consommation d'eau ou le bien-être des exploitants agricoles, mais aussi des éléments qui n'ont pas encore été pris en compte dans les logiques économiques traditionnelles, comme la durabilité environnementale et les besoins des exploitants et du personnel agricoles.

Une telle vision globale permet par exemple d'évaluer exactement quelles sont les répercussions des mesures politiques prises: qu'advient-il si la part de l'agriculture biologique augmente, si des subventions sont octroyées pour certains types de culture ou encore si l'élevage est soumis à davantage de contraintes? Les effets de ces mesures sur le système dans son ensemble peuvent par ce moyen être observés et analysés précisément et rapidement.

Lorsqu'une décision politique est prise en se basant sur des modèles qui ne tiennent pas compte d'une situation dans sa globalité, des effets non désirés peuvent apparaître. La conception de politiques efficaces passe nécessairement par l'adoption d'une approche plus globale qui est basée sur les données et qui tient compte de l'interaction de tous les aspects de l'agriculture.

Solution proposée

Le jumeau numérique de tout le système agricole suisse consiste en une représentation numérique complexe, mais claire, du système agricole réel et de ses différents processus de production. Les décideurs peuvent accéder aux

principaux paramètres et informations du jumeau numérique à partir d'une interface conviviale qui se présente sous la forme d'un tableau de bord. On peut non seulement y obtenir des informations actuelles sur le système, mais aussi y présenter de manière simple et claire les répercussions de scénarios qui simulent les besoins futurs en main-d'œuvre, en eau ou en engrais ou les modifications de l'utilisation des terres suite à telle ou telle décision politique. Les changements opérés dans les processus de travail actuels et futurs, en raison par exemple des progrès de la numérisation et de la mécanisation, peuvent être analysés par cette technique et être mis en relation avec d'autres données et modèles. Il en va de même pour les changements opérés dans la gestion des cultures, en cas par exemple de rotation des cultures, de changement de variétés cultivées et de réduction des pesticides.

Le jumeau numérique reprend à cet effet des données statistiques existantes sur l'agriculture et sur l'environnement ainsi que d'autres sources comme la géoinformation. À cela peuvent s'ajouter des modèles de la science des données et des calculs physiques qui viennent compléter les données du jumeau numérique. On peut donc appliquer des méthodes issues de la science des données à des fins de simulations ou de création de données synthétiques, lorsque l'accès à certaines données est restreint pour des raisons de protection des données ou lorsque les systèmes réels ne fournissent aucune donnée de mesure. Prenons l'exemple de l'agriculture de précision. Ce domaine a besoin de données de mesure détaillées sur la taille des champs, le type de plantations et l'irrigation. Avec un véhicule agricole équipé de capteurs, ces données seraient accessibles. Sans capteur, il est toutefois possible de générer des valeurs, sous une forme synthétisée, en ayant recours à des méthodes de science des données.

Le jumeau numérique contribue à une meilleure compréhension du système agricole. Il donne en effet une représentation transparente de toutes les caractéristiques pertinentes, telles que la production, les outils de travail, les chaînes et les itinéraires d'approvisionnement, la consommation de ressources (p. ex. le personnel et les engrais), les caractéristiques du sol, la croissance des plantes, l'utilisation de nutriments et de produits phytosanitaires, les conditions météorologiques, la consommation d'eau et l'efficacité énergétique, ainsi que les aspects sociaux tels que les conditions de travail et leurs interactions.

Le jumeau numérique facilite grandement la surveillance et le pilotage du système. Les décideurs politiques peuvent surveiller tout le système agricole suisse ou restreindre leur surveillance à certaines régions ou à certaines surfaces cultivées, dans l'objectif de mettre en œuvre des mesures efficaces. On pourrait, par exemple, avoir besoin d'adapter l'utilisation de produits phytosanitaires dans une zone de culture donnée. Avec le jumeau numérique, il est possible de faire des simulations pour voir si de telles mesures produisent l'effet escompté.

Les données collectées et générées dans le jumeau numérique peuvent servir à rédiger des rapports sur l'agriculture et à apporter une plus grande transpa-

rence dans le système. Par ce biais, on peut par exemple comprendre comment a évolué l'utilisation de pesticides après l'introduction d'une mesure politique. Dans une telle situation, le jumeau numérique présente un atout majeur: il permet de contrôler et de mesurer les paramètres de manière centralisée ou descendante («top-down»), par décision politique par exemple, mais aussi de manière décentralisée ou ascendante («bottom-up»), par des agriculteurs par exemple.

Le jumeau numérique de tout le système agricole suisse permet en plus de simuler différents scénarios et de développer des solutions optimisées qui profitent à l'ensemble du système de production agricole suisse. Avec une simulation, on pourrait par exemple savoir quelles seraient les répercussions d'un été plus chaud et plus sec sur la consommation d'eau ou savoir quels seraient les effets de tel ou tel choix des agriculteurs en matière de techniques de culture et de rotation des cultures et des céréales. De même, on pourrait voir quels effets se produiraient sur l'équilibre des nutriments dans les sols ou sur la production agricole en cas de modification de l'utilisation des engrais. Avec de tels scénarios, on peut prévoir l'impact des changements sur des périodes plus longues, simuler les répercussions à long terme et quantifier les conséquences des effets négatifs possibles.

Lorsqu'il est intégré dans les modèles correspondants, le jumeau numérique peut aussi évaluer les effets possibles de mesures agricoles dans d'autres domaines. En cas d'utilisation de cultures résilientes, comment réagissent les compagnies d'assurance agricole et que se passe-t-il au niveau de la protection et de l'information des consommateurs (p. ex. les labels)? En cas de changement des surfaces agricoles, quelles sont les répercussions au niveau du système de géoinformation? Un ajustement de la politique appliquée en matière de subventions a-t-il une incidence sur l'AVS et l'AI?

Le jumeau numérique du système agricole suisse présente un autre atout: il permet de mieux comprendre les systèmes agricoles des autres pays et de les comparer à celui de l'agriculture suisse. Pour ce faire, il faut appliquer la structure du jumeau numérique du système agricole suisse au système d'un autre pays. Cette possibilité de comparaison fournit de précieuses informations pour notre propre politique agricole et aide à mieux comprendre les systèmes agricoles des autres pays. De même, ces informations peuvent être utiles dans l'élaboration de politiques ou dans le cadre d'activités opérationnelles menées par d'autres unités de l'administration fédérale. C'est le cas par exemple lorsque des mesures de développement sont conçues et mises en œuvre dans d'autres pays.

La possibilité d'étendre le jumeau numérique en le reliant à d'autres sources de données et à d'autres jumeaux numériques, par exemple à d'autres sites de production tout au long de la chaîne de valeur comme des abattoirs et des usines de production laitière, offre un potentiel pour une modélisation et une analyse encore plus précises.

Transférabilité

Les jumeaux numériques sont pertinents dans tous les domaines où une vision globale d'un système physique et/ou la simulation d'évolutions hypothétiques au niveau physique du système sont nécessaires, à condition que les données requises soient disponibles. Les domaines politiques dans lesquels on peut envisager l'usage de jumeaux numériques ne manquent pas: l'urbanisme, la planification des transports et de l'espace, la protection de l'environnement, l'approvisionnement en énergie, l'approvisionnement alimentaire et les soins médicaux. S'agissant de l'approvisionnement en énergie, les modèles de processus dynamiques, basés sur des méthodes de la science des données, peuvent par exemple contribuer à optimiser les réseaux électriques et à détecter les points faibles dans l'approvisionnement.

Exigences

Données

- Mettre à jour régulièrement, si possible de manière automatisée, les sources de données afin de garantir que les jumeaux numériques reflètent les informations actuelles et permettent ainsi des analyses de données (p. ex. des prévisions précises). Il s'agit notamment d'informations spécifiques sur les cultures et les animaux, ainsi que de données sur les distances.
- Assurer la qualité des données, notamment en vérifiant périodiquement l'exactitude, l'exhaustivité et la cohérence des données utilisées.
- Mettre à disposition des données détaillées et précises (et pas seulement des données agrégées comme des moyennes) afin de pouvoir représenter avec précision les différents aspects de l'objet ou du système réel.
- Créer un pipeline de collecte de données au sein d'un système de gestion de base de données centralisé ou décentralisé afin de consolider les différentes sources de données et de faciliter l'accès aux données pour toutes les parties prenantes concernées.
- Assurer l'interopérabilité sémantique entre les données consolidées.
- Satisfaire aux autres exigences en matière de données, notamment mettre à disposition des valeurs absolues (p. ex. les variations absolues dans le temps) et des données de séries chronologiques sur plusieurs années.

Réglementation et éthique (sur la base des principes fondamentaux de la DSStB)

- Élaborer des directives de protection des données afin de garantir la confidentialité des données collectées et la protection de la vie privée des personnes concernées.
- Mettre en œuvre des lignes directrices éthiques et des évaluations des risques pour l'application de la science des données dans l'élaboration des politiques, afin de prévenir les abus potentiels, de garantir des décisions équitables et d'éviter les effets négatifs non souhaités des analyses de données et des décisions.
- Garantir la transparence des algorithmes et des méthodes de la science des données utilisés dans les jumeaux numériques, vérifier leur caractère

non discriminatoire afin d’instaurer la confiance chez les parties prenantes concernées et éviter toute discrimination.

- Préserver les secrets commerciaux des acteurs privés et des droits de la personnalité des particuliers (p. ex. flux de mouvements).
- Aligner la conception du modèle sur la garantie de la reproductibilité des résultats, par exemple par la mise en place d’un «provenance system» pour le suivi des réglages et des valeurs du modèle entrés précédemment.
- Mettre au point des mécanismes permettant de comprendre et d’expliquer comment les résultats des jumeaux numériques sont obtenus (p. ex. transparence des processus et des différentes étapes du processus).
- Développer et exploiter le jumeau numérique selon les normes «open source» et «open data».

Technologie et infrastructure

- Mettre en place une infrastructure informatique performante pour permettre le traitement et le stockage des grandes quantités de données nécessaires à la création de jumeaux numériques.
- Mettre à disposition une puissance de calcul suffisante et évolutive pour les simulations de modèles et l’application des algorithmes et des méthodes de la science des données.
- Fournir une infrastructure cloud adéquate, servant en premier lieu à la puissance de calcul (*computing*), mais aussi au stockage des données.
- Mettre à disposition des outils d’analyse et de simulation avancés pour pouvoir modéliser et optimiser des scénarios complexes avec efficacité.
- Créer une plateforme sécurisée et conviviale avec des tableaux de bord permettant d’accéder aux jumeaux numériques et à leurs fonctions d’analyse et de simulation.
- Garantir un environnement système IdO suffisant comme source de données.
- Mettre à disposition un «*institutional repository*» (dépôt institutionnel) pour faciliter l’échange de métadonnées.

Gestion

- Introduire une gestion de projet interdisciplinaire afin d’encourager la collaboration entre les scientifiques des données, les ingénieurs des données, les analystes des données, les experts techniques, les représentants des processus de décision politique (notamment au sein du Secrétariat général) et les autres parties prenantes.
- Créer des structures et des ressources permettant d’associer toutes les expertises pertinentes au développement et à l’exploitation du modèle et de promouvoir des échanges réguliers visant à perfectionner le modèle.
- Fixer des objectifs et des indicateurs de performance clairs pour l’utilisation du jumeau numérique.
- Mettre en place des processus de suivi et d’évaluation afin de vérifier en continu l’efficacité et la valeur ajoutée des jumeaux numériques et d’apporter des améliorations.

- Garantir à long terme les ressources financières et humaines nécessaires à la mise en place et à la maintenance du jumeau numérique.

Connaissances et compétences

- Promouvoir la formation (continue) des scientifiques des données et des experts afin de développer et de poursuivre l’acquisition des compétences nécessaires à cette application.
- Créer un système interne de gestion des connaissances pour favoriser l’échange d’informations et la coopération entre les acteurs concernés.
- Organiser des formations et des ateliers pour sensibiliser aux possibilités et aux défis de l’application au sein de chaque organisation.

Parties prenantes

- Impliquer les parties prenantes concernées dans le processus de développement des jumeaux numériques afin de répondre aux besoins et aux préoccupations de celles-ci.
- Créer un forum de discussion pour favoriser l’échange d’expériences, d’idées et de bonnes pratiques entre les parties prenantes et accroître l’acceptation des jumeaux numériques et des décisions politiques qui en découlent.
- Communiquer régulièrement au sujet des progrès et des réalisations de cette application de la science des données dans l’élaboration des politiques pour assurer la transparence et renforcer la confiance des parties prenantes.

Durabilité écologique

- Utiliser du matériel et des solutions cloud à bon rendement énergétique pour réduire autant que possible la consommation d’énergie dans la mise en œuvre de jumeaux numériques.
- Utiliser des pratiques durables en matière de collecte, de stockage et de modélisation des données, notamment la minimisation de la redondance des données et l’utilisation de techniques de compression des données et des modèles (surtout pertinent lors de stockage de grands volumes de données, p. ex. lors de recours à de nombreux capteurs).

4 Annexe

4.1 La science des données dans l’élaboration des politiques

L’élaboration des politiques (*policy making*) est un processus cyclique qui comprend les étapes suivantes: conception, mise en œuvre et évaluation. La science des données y joue un rôle décisif.

Le processus commence par l’identification d’un problème et du besoin d’agir en conséquence. À ce stade, il est possible de recourir à la science des données pour analyser et évaluer les données pertinentes afin d’obtenir une meilleure compréhension de la problématique ainsi que du contexte spécifique et des diverses relations et interactions.

Une fois la problématique bien comprise, on passe à l’étape suivante du processus: l’identification de solutions envisageables, tels un projet de planification des transports ou une modification de la législation. Des simulations basées sur les données peuvent alors servir à affiner le choix des variantes possibles et à faciliter la prise de décision.

Dans la suite du processus, les connaissances issues de l’analyse des données alimentent le débat politique tout en contribuant à développer ou à préciser les solutions envisageables. Elles permettent de prendre des décisions éclairées sur la base de données et de connaissances fiables.

Une fois la décision prise, la science des données peut servir à analyser la planification nécessaire des ressources et à simuler l’impact potentiel de la décision. Elle contribue ainsi à l’efficacité et à l’efficience de l’application des politiques.

Durant la phase de mise en œuvre, la science des données peut contribuer à traiter des informations et à automatiser des processus, et apporter de la sorte des améliorations considérables.

Enfin, lors de l’étape d’évaluation et d’éventuelle adaptation, l’analyse des données collectées permet d’apprécier le succès de la politique appliquée. Cette analyse contribue à améliorer l’application en continu et permet de mieux comprendre le problème initial.

En résumé, la science des données est à même de renforcer et d’améliorer l’ensemble du processus d’élaboration des politiques, en complétant et en élargissant les outils qui existent déjà dans ce domaine. L’illustration 3 met en évidence le rôle de la science des données à chaque étape de ce processus.

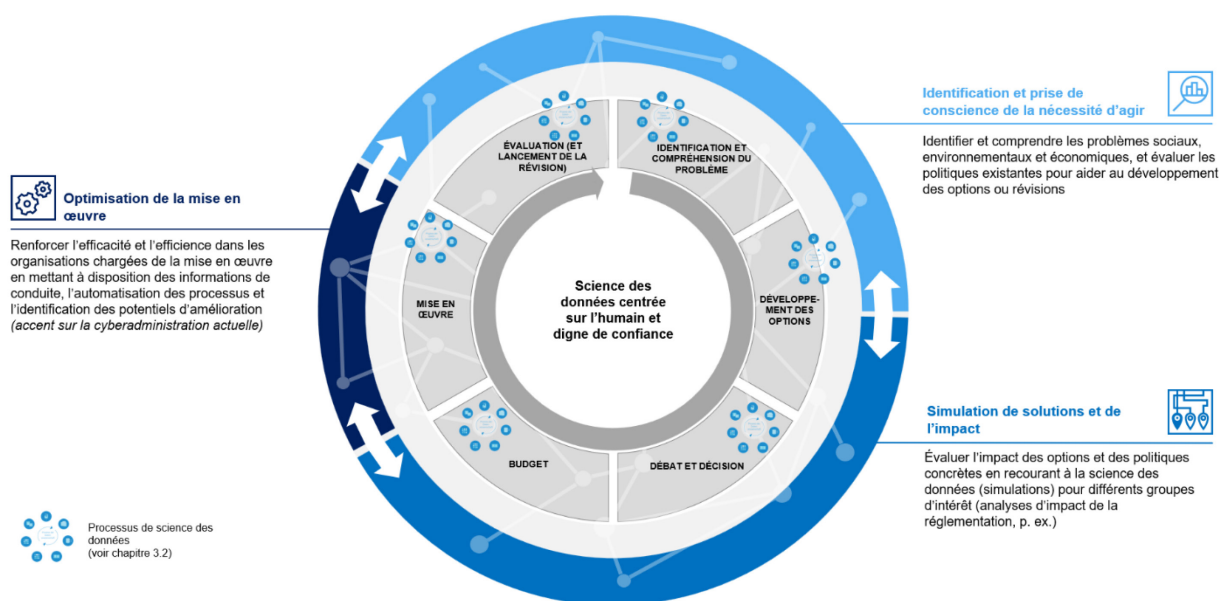


Illustration 3: La science des données dans l’élaboration des politiques (*policy making*).
(Source: DSStB)

4.2 Élaboration du rapport

4.2.1 Démarche

Le présent rapport a été élaboré sur la base d’une collaboration étroite entre des expertes et des experts de l’administration fédérale et du domaine des EPF. Rédigé de mars à août 2023, il a été élaboré au fil des étapes de travail décrites ci-après.

Pour commencer, le groupe d’accompagnement interdépartemental «Science des données» s’est réuni avec des représentants du domaine des EPF. Cette séance a servi à adopter une définition commune d’un cas d’application, des critères pour sélectionner des cas appropriés et une structure sommaire pour les décrire. Dans un deuxième temps, le groupe d’accompagnement et les représentants du domaine des EPF ont recueilli les premières idées de cas d’application à présenter, dont six ont été retenus et précisés lors d’une séance commune. Les responsables ont ensuite identifié, au sein de l’administration et du domaine des EPF, des spécialistes des cas sélectionnés et un groupe de travail a été formé pour chacun d’entre eux. Les groupes de travail ont élaboré leurs cas d’application de manière itérative au cours de trois séances successives. À titre indicatif, l’OFS leur avait soumis un cas d’application hypothétique.

Le présent rapport a été rédigé sur la base des cas d’application élaborés et des discussions approfondies au sein des groupes de travail. Il étaye les principales conclusions ainsi que les exigences, et les met en relation avec les projets et les évaluations en cours. Le rapport a été soumis au groupe d’accompagnement pour validation et précision.

4.2.2 Caractéristiques d’un cas d’application

En science des données, on définit un cas d’application comme un scénario spécifique dans lequel la science des données est utilisée pour exécuter une certaine tâche ou réaliser un certain objectif. Dans le cadre de l’administration fédérale et du présent rapport, les cas d’application identifiés doivent présenter certaines caractéristiques. La description de ces cas doit en tout temps tenir compte des aspects suivants:

1. **Pertinence:** un cas d’application devrait être pertinent pour le travail de l’administration fédérale et, le cas échéant, aussi pour les deux autres niveaux de l’État. Il doit aborder un défi ou un problème réel qu’il incombe à l’administration publique de résoudre.
2. **Transférabilité:** un cas d’application doit être transférable, c’est-à-dire pouvoir s’appliquer dans différents domaines politiques. La transférabilité augmente les possibilités de l’appliquer et permet de démontrer l’utilité de la science des données dans une grande variété de contextes.
3. **Identification des bases nécessaires:** un cas d’application devrait contribuer à identifier les bases nécessaires et à justifier leur création. Il peut s’agir de bases légales, de ressources techniques ou de structures organisationnelles.
4. **Science des données centrée sur l’être humain et digne de confiance:** un cas d’application doit se baser sur l’application d’une science des données digne de confiance et apporter un avantage (direct ou indirect) pour les personnes («centré sur l’être humain»). Autrement dit, les méthodes, les techniques et les pratiques de la science des données doivent respecter les normes éthiques et légales et garantir la qualité et la sécurité des données.
5. **Description:** un cas d’application doit être décrit de manière suffisamment concrète et illustré d’exemples pratiques pour que les responsables politiques et les unités administratives puissent le transposer dans leur travail. Une telle description facilite la compréhension de la science des données et son application dans l’élaboration pratique des politiques.

La prise en compte de ces caractéristiques lors de l’identification et de la description des cas d’application garantit que ceux-ci sont à la fois pertinents et utiles pour l’administration fédérale et qu’ils contribuent à promouvoir l’utilisation de la science des données dans l’élaboration des politiques.

4.2.3 Membres du groupe d’accompagnement et des groupes de travail

Groupe d’accompagnement

Nom (organisation)

Baeriswyl Dario (SG-DFF)	Fraefel Marianne (ChF)	Mielewczik Michael (Agroscope)
Blarer Albert (armasuisse)	Hagi Ivan (CSI-DFJP)	Moret Lionel (MétéoSuisse)
Bremer Walter (ChF)	Johansson Baker Kerstin (OFS)	Roth Anja (OFAS)
<i>Brüning Jakob (OFIT)*</i>	<i>Kaydul Alan (DFAE)*</i>	Schütz Samuel (OFS)
Chatagny Florian (AFF)	Kellenberger Tobias (swiss-topo)	Siegenthaler Claudia (SG-DDPS)
Choirat Christine (OFS)	Keller Eva (DFAE)	Singler Sebastian (PwC)
Cochard Jean-Luc (AFS)	Klopfenstein Hans-Peter (OFIT)	Stephan Thomas (AFC)
Droz-Georget Serge (DFAE)	Krebs Daniel (SG-DETEC)	Vancauwenberghe Laure (SDSC)
Dumont Georges (OFDF)	Kuonen Diego (OFS)	Willi Lukas (PwC)
Egloff Florian (SG-DDPS)	Loison Bertrand (OFS; présidence du groupe d’accompagnement)	
<i>Flury Marco (OFIT)**</i>	Michlig Roger (SG-DDPS)	

* A quitté le groupe d’accompagnement en cours de projet.

** A assuré la suppléance de Hans-Peter Klopfenstein lors de la 2^e séance de coordination (14 juin 2023).

Groupes de travail

Cas d’application	Administration fédérale	Domaine des EPF
Utilisation de renseignements de sources ouvertes (ROSO)	Eva Keller, Alan Kaydul, Mary Brown, Sibylle Obrist, Serge Droz-Georget (DFAE); Albert Blarer (armasuisse); Diego Kuonen (OFS)	Andreas Wenger, Sascha Langenbach, Corinne Bara (Center for Security Studies, EPF Zurich); Elliott Ash (Center for Law & Economics, EPF Zurich)
Tableaux globaux de la situation en cas de crise	Markus Aeschlimann, Markus Böhm, Markus Müller, Philipp Angehrn (OFEV); Samuel Schütz (OFS)	David Bresch (Dép. des sciences de l’environnement, EPF Zurich); Gianfranco Guidati (Energy Science Center, EPF Zurich)
Aide à la surveillance basée sur les données	Gianpaolo Cecchin, Markus Luginbühl, Pascal Imoberdorf, Caroline Gerber (OFAC); Bertrand Loison (OFS)	Giovanni Sansavini (Dép. de génie mécanique et des procédés, EPF Zurich)
Aide à la décision basée sur des modèles holistiques	Albert von Ow, Mélanie Douziech (Agroscope); Diego Kuonen (OFS)	Stephan Pfister (groupe pour la conception d’un système écologique (ESD), EPF Zurich)
Contrôles de plausibilité destinés à garantir la qualité des données	Anja Roth (OSAV); Diego Kuonen (OFS)	Matthias Bannert, Samad Sarferaz (KOF, EPF Zurich); Sergio Galletta (Gruppe für Ökonomie, Recht und Datenwissenschaften, EPF Zurich)
Jumeaux numériques	Michael Mielewczik (Agroscope)	Thijs Defraeye (LFEM, EPF Zurich)