Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

The Federal Council

Bern, 2 December 2022

Federal data science strategy

The federal administration's common bases, competences and objectives for the application of data science

Contents

Summary

Due to technological change, various issues are becoming increasingly complex, unstructured, data-rich and fast-paced. This makes it increasingly challenging for politics and the government to develop suitable solutions. Data science is an appropriate tool to support addressing these challenges at the operational level and in political decision-making. It helps to create a better understanding of problems and possible solutions in such an unstructured, data-rich and fast-paced environment. Here, data science complements and builds on existing tools that assist in problem solving and decision-making in politics and the government, so that they can fulfil their statutory tasks effectively, efficiently and sovereignly for the community's benefit.

In the Federal Digitalisation Strategy 2020-2023, the core "data science" initiative (SI-6) formulates the objective of developing competencies and successively designing a new discipline for this purpose. The present federal data science strategy takes up this mandate. This shows how the federal administration should use data science in the future, what goals it is pursuing in its application and what measures are required for its targeted and coordinated implementation.

With regard to the federal administration's development in the field of data science, it is essential to clearly define a common language and central terminology. The federal data science strategy therefore defines the term "data science" and describes the importance of data science oriented towards the public good in the context of the policy-making process. This contributes to the federal administration's identity in the field of data science, which in turn guides all data science activities in the federal administration. The identity in the framework of the federal data science strategy is based on a vision statement, core principles and an ambition. These are defined as follows:

Vision statement: "Human-centric and trustworthy data science supports the public good and public policy".

Core principles: Information security, data and information protection, data security, data governance, non-discrimination, explainability, traceability, transparency, reproducibility, neutrality, objectivity and ethical handling of data and results.

Ambition: "Data-driven decision support in the federal administration: Increased effectiveness and efficiency in public policy through data science".

Strategic directions, levers and measures have been formulated to implement the vision statement, core principles and ambition. The first lever and the corresponding concrete measures serve to strategically position data science in the federal administration. The second lever strives to promote the **transversal "data science ecosystem"**. This charts the federal administration's guiding principle as a coordinated value creation system. The ecosystem is the foundation for consolidating innovations, improvements and application fields related to data science and to optimally implement the benefits through five key enablers. These enablers cover talent and personnel development, culture, transversal cooperation, data science tools and platforms, and data stewardship and infrastructure.

A summary of the federal data science strategy's bases and the measures derived from these are presented in the following overview:

Federal data science strategy (DSStB)

Vision statement and core principles

Ambition

"Data-driven decision support in the federal administration: Increased effectiveness and efficiency in public policy through data science"

Strategic directions

- Create trust in data-driven decision support
- Build awareness and competence
- Increase technical accessibility and availability
- Exploiting potential for synergies and tackling challenges together

Levers and measures

A) Strategic anchoring of data science

A.1 Promote, explain and consolidate the core principles
A.2 Clarify and establish governance on the data science strategy
A.3 Identify application cases across the policy-making process

B) Development of the transversal "data science ecosystem"

B.1 Cultivate culture, a common language and knowledge building
B.2 Establish personnel development and promote and strengthen talent
B.3 Facilitate access to data science tools and platforms
B.4 Simplify access to data

1 Introduction

Data science is one of the pioneering areas for digital transformation and it has great potential in the federal administration. For most of the federal administration, data science is already a familiar field as data science applications are already in use in a number of administrative units. Data science can be applied in every field, ranging from epidemiology to spatial planning up to fraud detection. The corresponding examples of data science applications span the entire range of activities covered by the federal government's administrative units.

The federal administration has recognised the relevance of a strategic orientation and the coordination of the application of data science. In the Federal Digitalisation Strategy 2020-2023[1], the corresponding objective is formulated under the strategic "Data Science" core initiative (SI-6) led by the Federal Statistical Office (FSO) and the Federal Chancellery (FC / DTI): "*Information and data at the core of fulfilling administrative tasks*: *Develop competences in data science and successively create a new discipline for this field*". Among other things, the initiative pursues the development of a Federal Data Science Strategy (DSStB) and a data science training concept. This will create further important prerequisites for the federal administration's digital transformation.

In May 2020, the FSO received a mandate from the Federal Council to create a Data Science Competence Center (DSCC).[2] Since January 2021, the DSCC has been providing its data science expertise to Swiss administrations (Confederation, cantons and communes) in the form of "Data Science as a Service". In addition, the Federal Council promotes artificial intelligence (AI) and decided at the end of August 2021 to establish a Competence Network for Artificial Intelligence (CNAI).[3] The associated unit is based in the FSO and has been offering basic services since July 2022. Examples of AI projects can be found in several administrative units.[4] In each department, the use of AI has potential for automating and streamlining data-intensive tasks of a repetitive nature. However, methods from machine learning and AI are only some of the data science methods that are used in the problem-solving process of data science when analysing mass data for specific topics.[5]

The federal data science strategy takes up the mandate from the "data science" (SI-6) strategic core initiative within the "Federal Digitalisation Strategy 2020-2023". The federal data science strategy supports and promotes ongoing inter-departmental projects as well as those of the individual administrative units. This shows why and how the federal administration would like to use data science in the future, what goals it pursues in its application and what measures are required for its accelerated, targeted and coordinated implementation. The two other core objectives of SI-6 "Coordination and support of DSCC and CNAI" and the "data science training concept" are also addressed with concrete measures.

---

[1] "Federal Digitalisation Strategy 2020-2023"
[2] Data Science Competence Center (DSCC)
[3] Competence Network for Artificial Intelligence (CNAI)
[4] The CNAI's public project database provides an overview of AI-relevant projects in the federal administration and creates transparency about existing AI projects in the federal administration.
[5] The following topics are particularly relevant for the application of AI in the federal administration: text recognition, speech recognition, image recognition, fraud detection, plausibility checks and data validation, and predictive maintenance.

The design of data spaces[6] and the further development of "open government data" (OGD)[7] are distinct from the federal data science strategy. Data science applications use the data provided, for example, in data spaces or within the framework of OGD, but do not determine their design. The data policy also needs to be differentiated. It defines the organisational and technical framework conditions of the data stewardship and infrastructure in which data science is applied.

The present federal data science strategy is structured as follows:

- Chapter 2 describes the relevance, purpose and approach of this strategy with an outline of the logic behind the content.

- Chapter 3 defines the process of data science as a discipline and data science in the context of public policy.

- Chapter 4 forms the basis of the data science strategy outlining the federal administration's data science identity as a vision statement, the core principles and its ambition.

- Chapter 5 describes the directions and measures with which the vision statement, core principles and the ambition are to be implemented.

- The interplay of various factors is essential for the coordinated and targeted use of data science. Chapter 6 explains the data science ecosystem in this vision statement.

---

[6] As, for example, discussed in the Federal Council's press release on the report "Creation of trustworthy data spaces based on digital self-determination".

[7] FSO website on OGD.

2 Relevance, purpose and approach of the strategy

Figure 1 offers an overview of the federal data science strategy showing its relevance, purpose and approach.
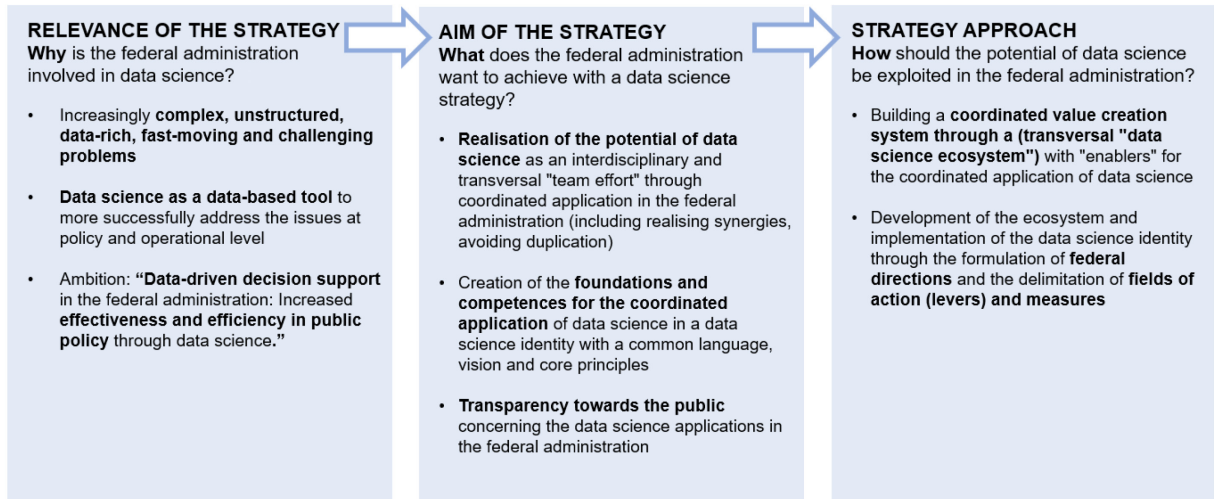

RELEVANCE OF THE STRATEGY
Why is the federal administration involved in data science?

- Increasingly **complex, unstructured, data-rich, fast-moving and challenging problems**

- **Data science as a data-based tool** to more successfully address the issues at policy and operational level

- Ambition: **"Data-driven decision support** in the federal administration: Increased **effectiveness and efficiency in public policy** through data science."**

AIM OF THE STRATEGY
What does the federal administration want to achieve with a data science strategy?

- **Realisation of the potential of data science** as an interdisciplinary and transversal "team effort" through coordinated application in the federal administration (including realising synergies, avoiding duplication)

- Creation of the **foundations and competences for the coordinated application** of data science in a data science identity with a common language, vision and core principles

- **Transparency towards the public** concerning the data science applications in the federal administration

STRATEGY APPROACH
How should the potential of data science be exploited in the federal administration?

- Building a **coordinated value creation system through a (transversal "data science ecosystem")** with "enablers" for the coordinated application of data science

- Development of the ecosystem and implementation of the data science identity through the formulation of **federal directions** and the delimitation of **fields of action (levers)** and measures

Figure 1: Relevance, purpose and approach of the federal data science strategy (*Source*: *own diagram*).

2.1 Relevance of the strategy

Due to technological change, various issues are becoming increasingly complex, unstructured, data-rich and fast-paced. This makes it increasingly challenging for politics and the government to develop suitable solutions. Data science is an appropriate tool to support addressing these challenges at the operational level and in political decision-making. It helps to create a better understanding of problems and possible solutions in such an unstructured, data-rich and fast-paced environment. Here, data science complements and builds on existing tools that assist in problem solving and decision-making in politics and the government, so that they can also fulfil their statutory tasks effectively, efficiently and sovereignly for the community's benefit in the 21st century.

Against this backdrop, the federal data science strategy serves the:

"Data-driven decision support in the federal administration: Increased effectiveness and efficiency in public policy through data science."

2.2 Aim of the strategy

The federal data science strategy lays the foundation for unlocking the potential of data science for all federal administrative units. It promotes the expansion of the use of data science and the development of the necessary organisational, ethical, legal and technical foundations and competences in the federal administration. The focus is on the coordinated use of data science with the aim of exploiting synergies, avoiding duplication and establishing common principles. In addition, the strategy informs the public of how the federal administration uses data science. Against the background of the population's concerns about the government's handling of their data, the federal administration must act as a role model for human-centric and trustworthy data science

(see chapter 3.3.1). The empowerment of data science can only work if its application is viewed, coordinated and practised as an interdisciplinary and transversal common project within the federal administration.

2.3 Strategy approach

Two aspects are key for the development and implementation of the strategy. Firstly, the interplay of various factors is essential for the coordinated and targeted use of data science. The establishment and maintenance of a coordinated value creation system is therefore at the core of the federal data science strategy. Five enablers ("cogs") form the transversal "data science ecosystem" (see chapter 6). They are complemented by the strategic positioning of data science. Together, they form the two central levers of the federal data science strategy, on the basis of which concrete measures are derived. These measures are implemented as follow-up orders.

Secondly, the federal data science strategy forms a framework and creates common ground. It does not define impact or application goals for individual administrative units. This takes into account the different levels of ambition and maturity that prevail between administrative units. The federal data science strategy deliberately allows for development paths determined in line with opportunities to use data science within the federal administrative unit. At the same time, it also calls for the active engagement of administrative units with data science and the definition of an individual development path.

3 Common language and understanding as a basis

3.1 Importance of the common language

Creating a common language can be seen as the heart of the federal data science strategy. A shared understanding of terms and a resulting common language enables the active exchange of experience and knowledge and simplifies communication regarding ideas, projects and services. This is because although considerable experience in data science is already available within the federal administration to some extent, the degree of maturity in the application of data science varies.

Especially in light of the ambitious data science development envisaged for the federal administration, it is essential to prevent misunderstandings right from the start by using a common language and clearly defined core terms. The following shows what data science and its related concepts mean for the federal administration.

3.2 Definition of data science

Data science is the interdisciplinary science of learning from data with the aim of gaining insights from the data in order to facilitate data-based decision-making. It is a problem-solving process based on continuous improvement that aims to solve complex, unstructured and data-rich problems through the application of data science methods (e.g., methods from advanced statistics, machine learning and the field of artificial intelligence), techniques and practices. Data science covers the entire process of problem formulation, data collection, data selection, data preparation, data analysis, evaluation, interpretation, communication and provision of the findings. As Figure 2 illustrates, the process of data science is a cycle of continuous improvement. The cycle can be restarted if it makes sense to do so based on the knowledge gained. Data science is also an interdisciplinary and transversal "team sport".

The fields of data science and statistics are very closely related. In addition to statistics, however, data science also makes use of other disciplines such as data management, engineering, machine learning, AI, algorithms, optimisation and data visualisation. For the purposes of uniform terminology in the field of artificial intelligence at the level of the federal administration, the CNAI has also described the term "data science" as follows: "Data science is also referred to as the child of statistics and computer science. This child metaphor appropriately infers that data science inherits (ideally the best) from both parents, but eventually grows into its own entity. Its focus separates it from its parents.".[8]
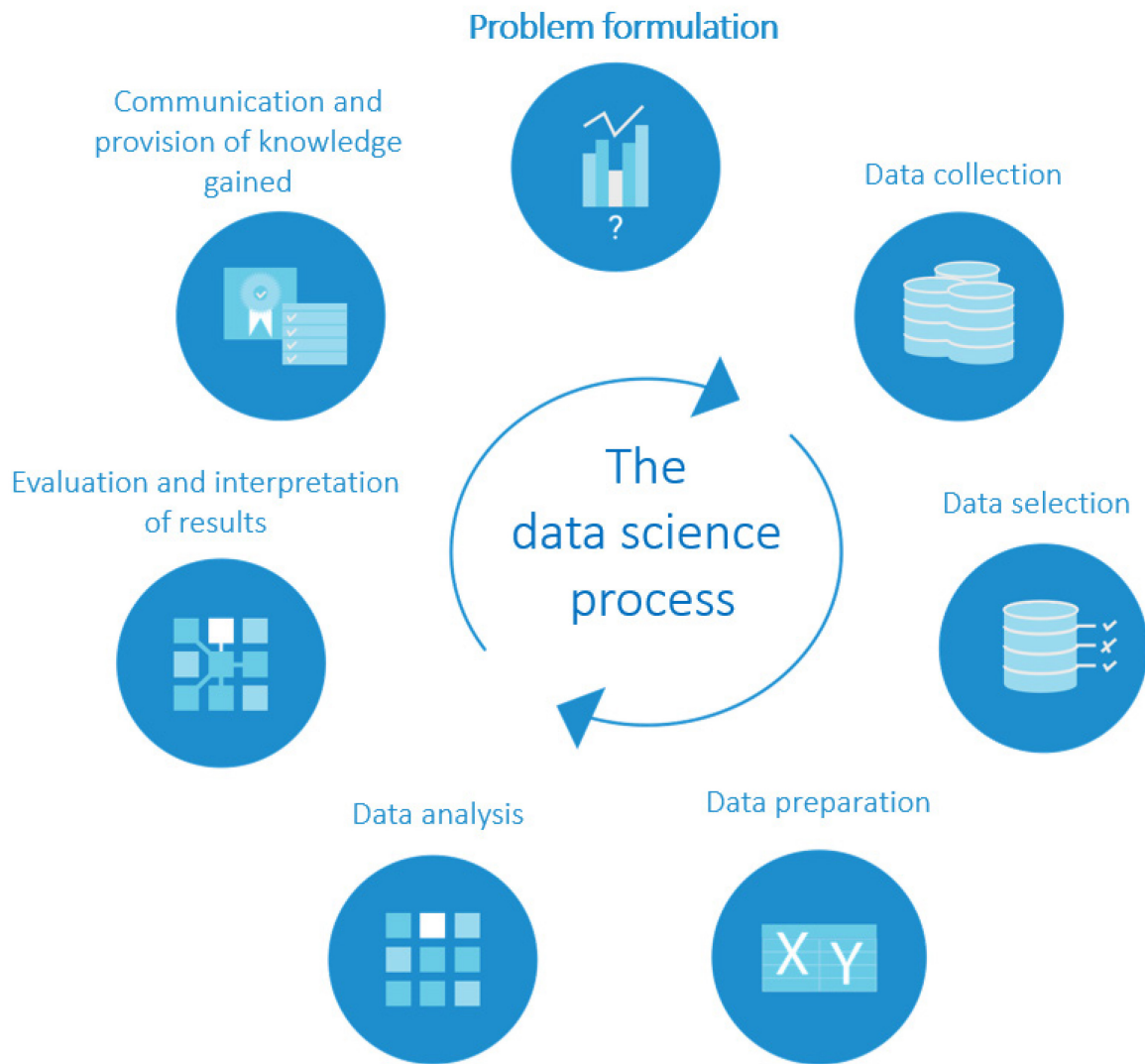
---

[8] Terminology of the CNAI

Figure 2: Interdisciplinary data science problem-solving process. *(Source: Own diagram based on the https://www.bfs.ad-min.ch/bfs/en/home/dscc/dscc.html).*

3.3 Importance of data science in the federal administration

3.3.1 Human-centric & trustworthy data science

The application of data science in the federal administration focuses on humans. The aim is to apply data science in line with humans' current needs and to help them fulfil their tasks. At the same time, people should always remain in control. In this way, applications of data science are to be developed that are oriented towards Switzerland's ethical, legal, democratic and cultural values. This human-centric approach aims to build confidence in the application of data science methods, techniques and practices and the insights gained. At the same time, this offers great potential for new applications and business models.

In addition to the opportunities offered by increasingly data-based decision-making, the federal administration takes the accompanying challenges and risks seriously. In particular, it is important to ensure the protection of fundamental rights when using data

science applications (e.g., ensuring non-discrimination, compliance with data protection). This is why the federal administration must ensure that the application of data science is followed up in a transparent manner at all times. If this is not given, there is a risk that compliance with democratic principles and legal requirements cannot be guaranteed in data-based decision-making. Consequently, trust among the population, businesses and institutions in public policy may decline.

In cooperation with science and research, as well as the economy, it is the task of the state to ensure the regulation of the increasing use of data science and the management of the associated challenges, and to proactively address the concerns and worries of the population, economy and civil society. Because trust is not something that is given, but instead has to be won by proving trustworthiness. To this end, the federal administration must establish clear core principles, ensure compliance with them and work to strengthen citizens' ability to verify the trustworthiness of government agencies when applying data science.

3.3.2 Public good & public policy

Public policy serves the common good. This also applies to the use of data science. Based on the UK Statistics Authority's definition[9], four aspects of data science contribute to the "public good"[10]. Firstly, it provides evidence that supports data-based decision-making in policy and public service delivery. This also includes other decisions that benefit the economy, society and the population's quality of life. Secondly, data science has the effect of replicating, validating, challenging and reviewing existing approaches. Thirdly, the use of data science ensures the improvement of the quality, coverage and presentation of existing applications. Lastly, understanding of social and economic trends or events is supplemented and expanded by improving knowledge.

---

[9] The public good and data science according to the UK Statistics Authority.

[10] In economics, (https://www.investopedia.com/terms/p/public-good.asp), a public good is a "commodity or service that is made available to all members of a society. Typically, these services are administered by governments and paid for collectively through taxation. Examples of public goods include law enforcement, national defense and the rule of law. Public goods also refer to more basic goods, such as access to clean air and drinking water".
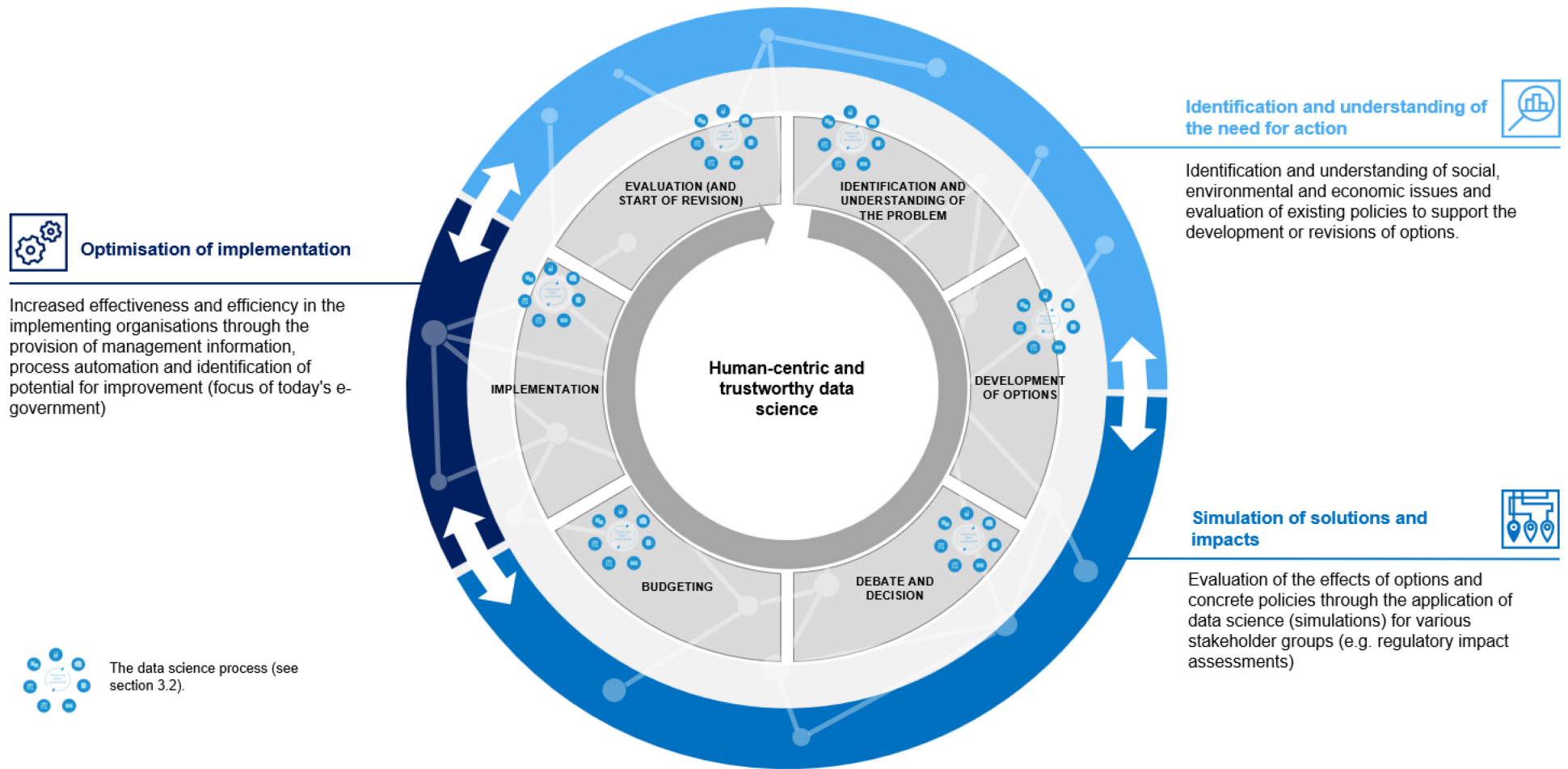
Federal data science strategy

Figure 3: Data science across the policy making process *(Source: own diagram).*

In concrete terms, the contribution of data science to the public good can be demonstrated across the policy-making process. This describes, in a highly simplified form, the cycle of policy making, and its implementation and evaluation. Data science can significantly support the whole process of these activities by complementing and extending existing policy-making and implementation tools. How data science applications are embedded in the process of government activity ("public policy"[11]) is illustrated in Figure 3. The figure makes it clear that data science is relevant to all departments and the Federal Chancellery (e.g., with regard to enquiries from citizens, parliamentary enquiries, problem identification and the evaluation of political measures).

Political decision-making and its implementation first start by identifying the problem at hand and the corresponding need for action. The analysis and evaluation of relevant data lead to a better understanding of the problem and the relevant interrelationships and dependencies. A consolidated understanding is central in the next step, the identification of possible solutions (e.g., an amendment to the law, transport planning concept). For example, data-based simulations can be used to further narrow down the choice of options. The corresponding findings support the political discussion and the further development or specification of options in the further course of the cycle. They thus allow informed decisions to be made. Similarly, in the event of a decision, the allocation of the necessary resources can be analysed and the effects simulated. Finally, improvements can be achieved in implementing decisions through the data science driven preparation of information or process automations. Within the evaluation framework including a possible revision, the collected data are analysed and can provide conclusions on the success of the implemented policies. This in turn helps to continuously improve the clarity of problem understanding. In summary, data science enhances effectiveness and efficiency in line with the federal data science strategy's ambition across the entire policy-making process.

---

[11] Public policy refers to the sum of the substantive decisions, objectives and activities of those involved in the policy-making process. It is thus an interdisciplinary subject of political science that deals with issues of politics, the economy and the society.

4 The federal administration's data science identity

Based on the meaning of data science described in chapter 3.3, a data science identity is defined that guides the actions of all of the federal administration's data science activities. The identity is composed of a vision statement, the core principles and an ambition.

4.1 Vision statement

The vision statement is:

"Human-centric and trustworthy data science for public good and public policy".

Data science applies appropriate methods, techniques and practices to create complementary and extended understanding and to promote data-driven decision support. This creates added value for the public good and public policy and the potential of data science for the federal administration's digital transformation is realised effectively and efficiently. The federal administration coordinates the use of data science across all administrative units and with other government levels in order to create synergies and jointly develop competencies and solutions. The core principles are consistently implemented in the daily work.

4.2 Core principles

Core principles are the basis for human-centric and trustworthy data science. They provide guidance in the application of data science and the implementation of the vision statement and ambition. The federal data science strategy's core principles correspond to the existing key values of the DSCC for the provision of all data science services so that sustainable added value can be created. These are based, among other things, on the UN's Fundamental Principles of Official Statistics and on the European Statistics Code of Practice which also applies to Switzerland.[12] The core principles are accordingly coordinated with existing, internationally recognised principles, but must be further clearly defined for the context of the entire federal administration and underpinned with corresponding content dimensions[13] (see measure A.1 in chapter 5.2). The core principles are:

- Information security
- Data and information protection
- Data security
- Data governance
- Non-discrimination
- Explainability
- Traceability
- Transparency
- Reproducibility

---

[12] The core principles are part of the FSO's quality management.
[13] Such as those based on the Guide for Ethical Data Science from the Royal Statistical Society (RSS) and the Institute and Faculty of Actuaries (IFoA).

- Neutrality
- Objectivity
- Ethical handling of data and results

For example, it should be ensured that the results of the individual projects (as far as permitted by data protection laws and other provisions) are documented and made available in a transparent manner.[14] The relevance and implementation of the core principles are the responsibility of the individual administrative units. Thus, the weighting of the core principles may vary depending on the data science application. On the basis of a code of practice to be drawn up (see chapter 5.2), the core principles are concretised as needed for specific administrative units. The federal data science strategy should also help to ensure longer-term compliance with the core principles.[15]

4.3 Ambition

The previously mentioned ambition renders the vision statement operational and concretises how the federal administration wishes to contribute to the public good with human-centric and trustworthy data science. The goal is to increasingly make decisions based on the insights gained from data science processes. This enables administrative units to deal more effectively and efficiently with the complex problems of today. Humans always remain at the centre of problem-solving using data science. The ambition is:

"Data-driven decision support in the federal administration: Increased effectiveness and efficiency in public policy through data science."

In the following chapter, directions, levers and measures are used to show how the federal administration will implement its data science identity in practice.

---

[14] This can be achieved, for example, through best practices such as the "Reproducible Analytical Pipelines" (RAPs) in the context of operationalising data science applications and projects that can serve as inspiration. The framework developed by Statistics Canada for Reproducible Machine Learning could also provide inspiration.

[15] The governance of these values must be ensured in the longer term, by using, for example, a label such as "Human-centric and trustworthy data science inside!" - in particular with a strong focus on the ethical handling of data and results and public trust. This assurance must also be aligned and coordinated with the DSCC's basic data science services (i.e., "development of quality standards, guidelines for data protection compliance and basic infrastructures ("sandboxes") for data science applications") and the external governance of the DSCC in the federal administration.

5 Directions, levers and measures

Strategic directions, levers and measures have been formulated to implement the vision statement, core principles and ambition. Figure 4 gives an overview of this.
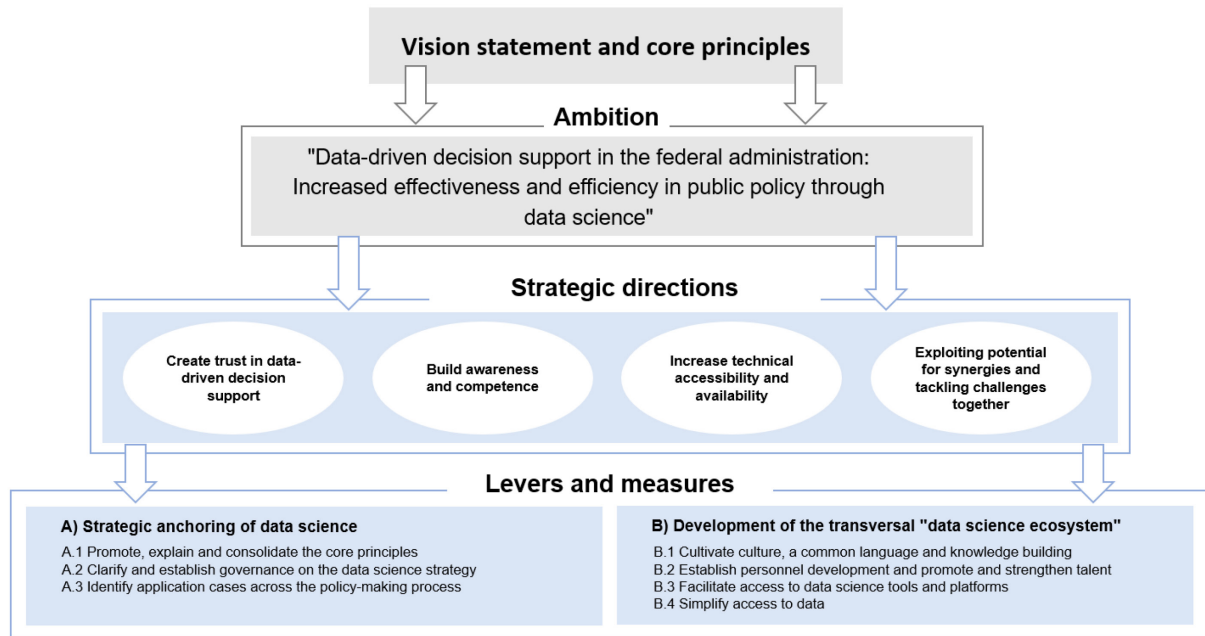

Figure 4 : Structure and hierarchy of the vision statement, directions, levers and measures *(Source: own diagram)*.

5.1 Strategic directions

The strategic directions show which effects are to be achieved with the federal data science strategy. These are equally meaningful and interdependent.

5.1.1 Create trust in data-driven decision support

The trust of stakeholders who are both directly and indirectly involved is a central factor of the federal administration's identity in terms of trustworthy, human-centric and public good-oriented data science for the federal government's activities. The federal data science strategy strives to create this trust in the application of data science methods, practices and techniques in the federal administration itself and also among the public.

5.1.2 Build awareness and competence

In order to realise the potential of data science, this potential must first of all be recognised. The federal data science strategy strives to create awareness in the federal administration and to demonstrate the urgent need to engage more deeply with data science. This includes a greater exchange of data science methods, practices, techniques and their results. It is therefore important to further promote the concrete knowledge and competence required in the federal administration and to consolidate it across the board.

5.1.3 Increase technical accessibility and availability

By providing easy and quick access to data science tools and platforms in a secure environment, the aim is to facilitate experimentation in the exploration of potential solutions to data science related problems and the use of data science applications. For the sustainable implementation of the data science problem-solving process, data stewardship and infrastructure also form a crucial basis. It is possible to do data science with little data (provided the data quality is given), yet access to data should be facilitated technically and organisationally in line with legal requirements.

5.1.4 Exploiting potential for synergies and tackling challenges together

The federal data science strategy contributes to exploiting potential synergies within the federal administration, between the levels of government, with science and research, the economy and other states. This transversal collaboration in data science is central to the federal data science strategy's success, as it allows us to confront the complexity of current problems together. Using various knowledge bases and data science skills, successful approaches to solutions can be developed.

5.2 Levers and measures

To implement the strategic directions described above, the federal data science strategy envisages the following strategic levers "A) Strategic anchoring of data science" and "B) Development of a transversal "data science ecosystem"" with concrete associated measures.

A) Strategic anchoring of data science

The strategic anchoring of the development and planning of data science projects at the management level of all administrative units is critical to the success of the federal data science strategy's implementation and constitutes its first lever. Three concrete measures are at the centre of this.

A.1 Promote, explain and consolidate the core principles

By developing a federal code of practice for human-centric and trustworthy data science for data science teams and decision makers, the core principles are promoted, made understandable and consolidated. The code of practice serves as a guideline for compliance with the core principles in data science projects. The creation and dissemination of a "Data Science Oath"[16] or a "Data Science Code of Conduct"[17] for individuals in data science teams should fall within this framework. The Federal Council's "Guidelines on AI for the Confederation"[18] and the development of the "Code of Conduct for the Operation of Trusted Data Spaces" by the FDFA and DETEC[19] serve as additional guidance.

---

[16] Such as a "Hippocratic Oath" for data science.
[17] Such as those based on the Guide for Ethical Data Science from the Royal Statistical Society (RSS) and the Institute and Faculty of Actuaries (IFoA).
[18] https://www.admin.ch/gov/de/start/dokumentation/medienmitteilungen.msg-id-81319.html
[19] https://www.admin.ch/gov/de/start/dokumentation/medienmitteilungen.msg-id-87780.html

A.2 Clarify and establish governance on the data science strategy

For the successive establishment, implementation, monitoring and further development of the federal data science strategy, it is necessary to define the appropriate governance. Questions such as which organisation assumes what responsibilities and how funding is regulated are clarified. The clear definition of roles also includes the coordination of decision-making bodies to promote transversal cooperation. Governance on the federal data science strategy includes science and research as well as business and international cooperation and takes into account existing bodies (such as the "DSCC Expert Group") to exploit possible synergies. No new decision-making committees are to be created. Governance also sets up the necessary coordination and organisation, as well as relevant processes, to place the requirements from data science in federal data policy committees. This must be done taking into account the clear boundaries between the different strategic initiatives within the federal administration.

A.3 Identify application cases across the policy-making process

Data science can be applied at every step across the policy-making process, with the aim of supporting and promoting more data-based decisions. Currently, however, data science is mainly applied in the context of implementation. In order to increasingly establish transversal cooperation and support the entire policy-making process, overarching application cases are to be identified and actively promoted in the federal administration. In the course of this measure, the legal basis is to be reviewed for each application case and, if necessary, recommendations are to be formulated.

B) Development of a transversal "data science ecosystem"

The effective and efficient application of data science as part of the federal administration's data science identity requires the right framework conditions. The appropriate prerequisites of a legal, ethical, technical and organisational nature - this includes an adapted organisational culture - must be in place. The totality and interplay of these factors together form a transversal "data science ecosystem". This ecosystem is the foundation for consolidating innovations, improvements and application fields related to data science and to optimally implement the benefits. Moreover, such an ecosystem is necessary to implement the vision statement, the core principles and the ambition of the federal data science strategy. Figure 5 shows the ecosystem with its five enablers ("cogs"). In chapter 6 the target image of the transversal "data science ecosystem" is described in detail.
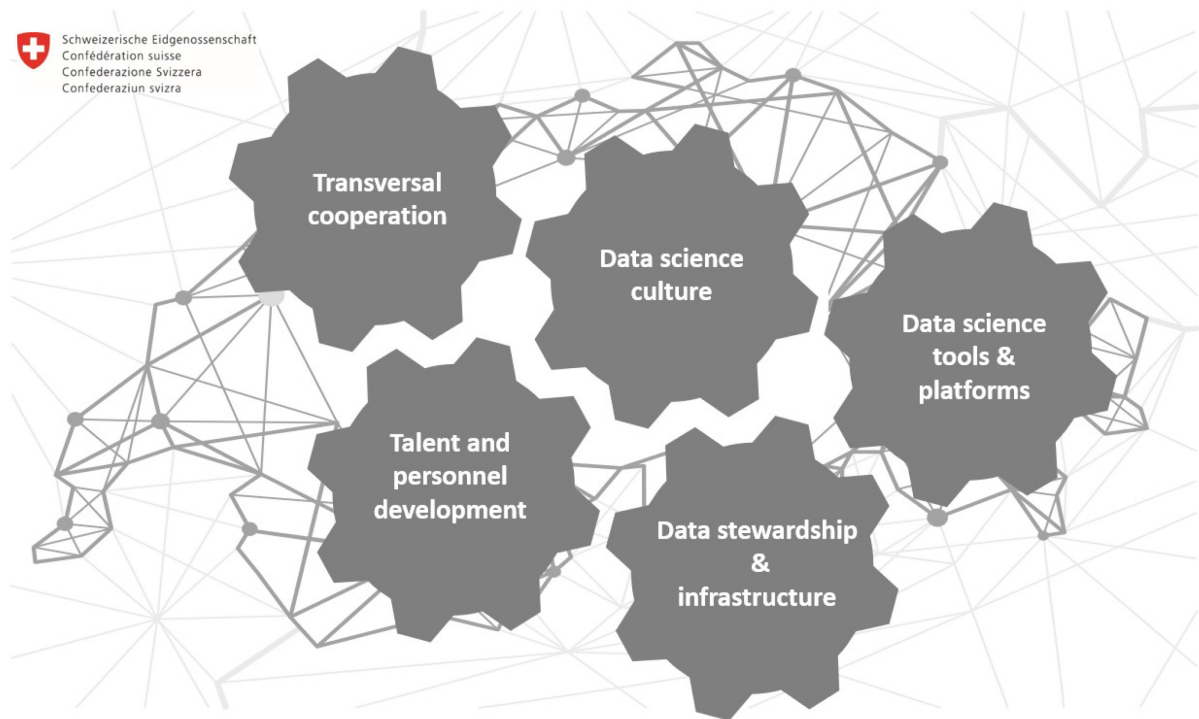
Federal data science strategy (DSStB)

Figure 5: Target image of the transversal "data science ecosystem" with enablers ("cogs") *(Source: own diagram)*.

The development of this ecosystem is the second lever of the federal data science strategy. The roles are described in more depth below.

B.1 Cultivate culture, a common language and knowledge building

The cultivation of a common data science culture and language and collaborative knowledge building are at the heart of the federal data science strategy. To this end, a single common "Community of Practice" (CoP) is to be established. The CoP is a group of people facing similar challenges who can learn from each other. The CoP is intended to cultivate and consolidate the exchange of experience between the federal administrative units and to provide and maintain an overview of data science projects in the federal administration. This is where stakeholders can exchange views on their special topics on a regular basis, including at international level. The CoP is open to anyone with an affinity for data science or who has expertise and interest in data science. It is held together by passion, commitment and identification with the group's specialist knowledge. The CoP meets regularly to share information, build their expertise and actively work to increase general knowledge in this field. To avoid duplication, this CoP is being set up in close cooperation and coordination with the FSO-external community "Data Science for Public Good" from the DSCC[20] and its integrated CNAI "CoP AI"[21]. The focus of this CoP is on the further development of the basic service of the DSCC

---

[20] The DSCC's "Data Science for Public Good" Community
[21] The CNAI's "CoP Artificial Intelligence"

regarding guidelines for the use of data science tools and platforms. In addition, a catalogue is to be compiled of data science tools and platforms available in the federal administration.

B.2 Establish personnel development and promote and strengthen talent

The development of talent and personnel is at the heart of strengthening the federal administration's capacity for data science. To this end, the first element required is a coordinated education and training programme ("up- and reskilling") for the application of data science in the federal administration, based on the experience already acquired.[22] In doing so, the programme aims to strengthen leadership knowledge for the application of data science. Alongside the education and training programme, it is important to recruit external talent that is aligned with clearly defined professional profiles. To achieve this, the federal administration must position itself as an attractive employer in the field of data science. To promote sustainable human resource development, the development of clearly defined job descriptions and profiles in the field of data science needs to be addressed.

B.3 Facilitate access to data science tools and platforms

For the timely application of data science methods, practices and techniques, it is important that access to relevant tools and platforms is as easy as possible (in compliance with the respective protection of the data required for processing). In addition, it must be possible to coordinate the joint use of such data science tools and platforms. In order to further advance this coordination, a standard product (including needs analysis, architecture, operative roles, requirements for security, archiving, performance and availability) for a "collaborative data science platform" is to be established, taking into account existing technical systems. The aim is to enable the dynamic and flexible procurement of the standard product, as tools and platforms in this area are quickly evolving.

Furthermore, it is important to sensitise administrative units and individual employees by communicating guidelines for the use of tools and platforms against the background of the core principles and legal provisions. The latter are coordinated with the corresponding basic service of the DSCC and further developed in a target-oriented manner. In addition, a catalogue of the data science tools and platforms currently available in the federal administration is to be compiled, which includes an assessment of the applicability and challenges of these solutions, taking into account existing experience in the federal administration. This catalogue provides the necessary overview, enables timely reference and ensures that the tools and platforms meet defined quality standards and the extent to which they sustainably promote the core principles. Depending on the needs and requirements arising from the data science problems, the catalogue can range from stand-alone tools to platforms and even hyperscalers.

---

[22] Developed and validated by the FSO, the "data science training concept" has enabled FSO staff since the start of 2022 to acquire and develop data science competences and skills. In this training plan, different target groups, training stages (awareness raising, introduction, consolidation, specialisation) and learning paths were defined. Based on a collaboration with the EPFL Extension School, the training course was divided into different modules and modalities. Initial discussions between the FSO and the FPA (Federal Personnel Office) with a view to a roll-out to the whole federal administration already took place in 2021 and will continue in a targeted manner.

B.4 Simplify access to data

The benefits of data science are enhanced by the availability of relevant data. Due to the amount of data collected by the various administrative units, there is great potential in this regard within the federal administration.[23] In order to exploit this, data science approaches are being tested and further developed that allow efficient shared learning from these data while respecting data protection and the principle of data minimisation. These approaches allow insights to be generated from data on individuals without the need to transmit the data of these individuals. For example, this includes the creation and use of synthetic data and the application of data science to protect privacy[24]. In the course of this measure, the legal bases are to be reviewed and, if necessary, recommendations are to be formulated.

The two levers "A) Strategic anchoring of data science" and "B) Development of a transversal "data science ecosystem"" implement the federal administration's data science identity. In the next chapter, the ecosystem's target image is described in further detail for greater understanding. It represents the foundation pillar of data science identity's implementation.

---

[23] See, for example, measure 4 in the Federal Council report on improving data management in the health sector: https://www.admin.ch/gov/de/start/dokumentation/medienmitteilungen.msg-id-86762.html

[24] Privacy Preserving Techniques

6 Target image of the transversal "data science ecosystem"

6.1 Overview

As described in chapter 5.2, the transversal "data science ecosystem" is driven by five central enablers ("cogs"). Due to their continuous interactions, these can only work and develop optimally when together. In addition, through this, the establishment and maintenance of a coordinated value creation system is promoted.

The individual enablers ("cogs") of the transversal "data science ecosystem" are described in more detail below.

6.2 Description of the enablers of the ecosystem

6.2.1 Data science culture

First of all, for a functioning, transversal data science ecosystem, it is necessary to establish a data science culture that promotes human-centric and trustworthy application. The established core principles (see chapter 4.2) represent the culture's pillars and are to be implemented and ensured through them. Furthermore, compliance with the core principles promotes trust among the population in the federal administration's data science projects.

The targeted culture consists of two levels. Firstly, leadership culture determines how data science projects are tackled. Here leadership should create space to encourage the organisation to experiment with and use data science within its means and to support learning in concrete projects. It also includes an awareness that learning from data adds value and that appropriate data quality (i.e., the assessment of data in terms of its suitability to fulfil a specific purpose) is crucial in this, in addition to a willingness to invest in capturing them correctly and as comprehensively as possible along the defined core principles. The second level is the work culture. This is shaped by the employees who implement data science applications. In their implementation, care should be taken to ensure that the added value relates to a human-centric and trustworthy state and that there is an entrepreneurial climate that strengthens the autonomy of employees and encourages innovation and continuous improvement. For a successfully functioning ecosystem, the culture must in particular create a heightened awareness of the benefits of data science.

6.2.2 Talent and personnel development

The federal administration's employees are both the foundation and a crucial bottleneck. Employees are needed who can ensure the implementation of data science projects with the appropriate knowledge. It is therefore very important to up or reskill personnel through internal and external education and training or recruitment programmes in the federal administration. At the same time, it is important to strengthen leadership knowledge for the application of data science.

Data science talent is in high demand on the job market. Successful recruitment of the appropriate talent is made easier if the administration is seen as an attractive employer in the field of data science. Job profiles with clear training and career paths as well as consistent communication of the diverse fields of activity in data science within all administrative units contribute to its attractiveness. Harmonised data science job descriptions simplify collaboration and clarity.

6.2.3   Transversal cooperation

Data science is an interdisciplinary and transversal "team sport". Its success largely depends on the intensive cooperation of experts from different disciplines, organisations, sectors and countries, who bring different knowledge bases and complement each other in data science skills. Through continuous knowledge exchange and joint projects, valuable insights can be gained and already existing applications can be used for applications in other areas. Cooperation takes place on three levels.

Within the federal administration: Through interdepartmental projects and projects across government levels, potential synergies within the federal administration and between the federal levels are made use of. Ideally, the enablers ("cogs") of the ecosystem form a framework that can also be used by cantons and communes, preferably involving existing platforms such as Digital Public Services Switzerland (DPSS). The aim here is to enable the individual administrative units to solve their own problems through data science applications. It is not about centralising competences, but about empowering each administrative unit to apply data science within its own ambitions and possibilities. To further promote cooperation in the field of data science, legal requirements must also be examined and the need for adaptation evaluated.

With science and research, as well as the economy: Close cooperation between the administration, science and research, as well as the economy enables basic data science research to be established quickly and for novel solution approaches from data science to be shared between stakeholders. A good example is the cooperation with the Swiss Data Science Center (SDSC), run jointly by the Federal Institutes of Technology, and the DSCC to use data science and AI for Switzerland's public good.[25] Similarly, the Data Science research programme from armasuisse Science and Technology S+T ensures that the necessary technological competences are available in the operational units within the armed forces, the Federal Intelligence Service, the Federal Office of Sport and the Federal Office of Police with a corresponding network of experts.[26] Based on such cooperation, valuable new insights are gained from the available data by combining the methodological skills of the researchers with the federal administration's specific expertise. Closer cooperation with decentralised administrative units and private companies could create space to learn from each other, inspire each other and jointly develop new application cases.

In international exchange: Last but not least, international cooperation is also an essential prerequisite for the sustainable development of a data science ecosystem for

---

[25] FSO press release: "Data science and artificial intelligence for the public good"
[26] "Research Programme 3b: Data Science" by armasuisse Science and Technology S+T

the federal administration. As an open and neutral stakeholder, Switzerland can offer a location where researchers and companies can cooperate on an international level.[27] The federal administration actively participates in the development of international and European standards and initiatives at the technical and regulatory level in the field of data science. The Federal Council's acknowledgement of the "AI and international regulations report"[28] is also to be classified here, for example. The measures included in this report call on Switzerland as a state to actively influence the design of the international regulatory framework on artificial intelligence at an international level and to thus contribute to moderate regulation. At the same time, however, individual administrative units are to play a leading role in international projects. In order to exploit the potential of international cooperation, coordination within the federal administration is essential.

6.2.4 Data science tools and platforms

Data science tools and platforms are needed to implement data science methods, techniques and practices. These cover, among other things, tools and platforms that can be used, for example, to visualise data, to check the plausibility of data, for collaborative development and for the use of data science algorithms, right through to the automated recognition of text, language and images. The choice of such tools and platforms is immense and growing continuously.[29] It is a challenge to keep track of and to understand which application is best suited to solving a particular data science problem. The administrative units should be enabled and supported to identify and use the best tool for solving the data science problem. In addition to the availability of the tools and platforms, conformity with the core principles and legal provisions (especially on data and information protection) must be ensured when using them.

6.2.5 Data stewardship and infrastructure

Data stewardship and infrastructure serve as the basis for optimal execution of the data science problem-solving process. Data science is possible with just a small amount of data (provided that the data quality is given). For such application cases, a large, superordinate data infrastructure is not always needed. However, to make use of the full potential of data science effectively, efficiently and sustainably, and to scale applications, a data infrastructure and architecture is needed that enables joint coordination of the acquisition, securing, cleaning, aggregation, provision and use of data. Furthermore, the regulatory and legal requirements regarding data and information protection and data security must also be taken into account. Such a data infrastructure comprises legal, technical and organisational building blocks that enable the creation, processing and sharing of data throughout their lifecycle.

The strategic core initiative SI-3 of the Federal Digitalisation Strategy 2020-2023 pursues the objective of "*developing service architectures for networked digital services through the implementation of the once-only principle*"[30]. This enabler ("cog") of the

---

[27] For example, see "Digital Foreign Policy Strategy 2021-2024"
[28] Artificial Intelligence and the International Regulatory Framework report
[29] For example, see "Machine Learning, Artificial Intelligence and Data (MAD) Landscape"
[30] See the corresponding measures of SI-3 in the "Federal Digitalisation Strategy 2020-2023".

data science ecosystem is thus covered by this strategic initiative. Nevertheless, further coordination and agreement are necessary.