Trust in the use of sensitive data is not nearly widespread. Even though there is no shortage in tools and methods to address the digitalisation of our society and industries, the acquisition and access to non-open data represent the greatest challenge. In the context of data monopolies, data regulation and general reluctance to share data, a paradigm shift is needed to regain control and sovereignty over our data. Sensitive – or restricted – data encompasses all the personal data that any Swiss resident shares in exchange for services. It also includes the process data generated by industries, health data, or data from the humanitarian sector. To unlock the data economy while reinforcing privacy and control, a new framework is needed to empower data users, to create a trusted standard in data management, to mitigate the regulations burden, and to help Switzerland maximise the benefits brought by the new digital age we live in.

What Is The Swiss Data Custodian?
The Swiss Data Custodian (SDC) refers to (a) a multi-sided platform which federates non-mutually trusted data providers and data consumers willing to enrich restricted data without sharing it; (b) a security-based ecosystem providing the tools to compute and store data without infringing on privacy; (c) a trust-based governance that monitors and maintains compliance with strict regulations. Above all, the SDC establishes trust and transparency in data usage. It is sector-independent and enables interoperability, portability and coetition by assigning the data usage consents to the rightful party. It is meant to protect the rights of individuals and organisations, and to generate new opportunities for the common good. The SDC follows the UN Guidance Note on Big Data 2030.

The Use of Data in Switzerland
On the one hand, the Open Data movement has already gained momentum in Switzerland. The objective is now to raise the threshold of open data accessible to all, and to value it. On the other hand, mobility data, health data or machine data from tech companies are often sensitive, valuable and/or personal. It cannot be opened, yet its potential for the whole society is huge. Today, most of not-opened data is held and duplicated in silos, and is barely reusable. Given the complexity of the matter, the SDC provides the tools to turn sensible data into an opportunity rather than an issue, while enforcing data protection regulations and more transparency over data usage.

Swiss Data Custodian Principles
- Sovereignty
  Clearly identifies who has the appropriate consent to use (derived) data at each step of a data processing workflow
- Authentication & authorisation
  SDC’s governance accompany data users to decide what data is kept, what is accessible, by whom, for what purpose, in what manner
- Security
  Guaranty respect of access control decisions and enforce privacy policies with no point of failure, human or machine
- Privacy
  Minimise information potentially inferred from data, protect from re-identification and disclosure of sensitive information
- Transparency
  Expose the data transformation lineage (who enriched the data, what for, what came out, how much do they know, etc) as well as structure, dependencies, and risks of the enforcing mechanisms/systems/methods

a privacy-based ecosystem to leverage the use of restricted data
A Collective Approach

At the request of politics, administrations, private companies, research institutes and other organisations, the Swiss Data Science Center (SDSC) has taken up the role of providing a Custodian ecosystem. Funded by the Swiss Confederation, SDSC has the mission to accompany the academic community and the industrial sector in their data science journey, facilitating the multidisciplinary exchange of data and knowledge. Therefore, SDSC plays a leading role in the design, development and adoption of the Swiss Data Custodian, while taking into account the expectations and requirements of each stakeholder.

Trust Is The Key

The SDC relies on trust. As building trust is not trivial, the Custodian enforces a solid privacy-based governance model. The Custodian Federation is also subject to audits, and a neutral entity will arbitrate and orchestrate the data usage consents allocation. External neutral challengers will also ensure the robustness of the security system. The Custodian aims at providing a cutting-edge solution to leverage the use of restricted data and empower users while ensuring best practices in terms of security and privacy.

Reference Architecture

The Reference Architecture is an open source solution developed by the ETH domain. Following the guidance of privacy-by-design, the Custodian is a network of decentralised Custodian-enabled infrastructure. These numerous instantiations across Switzerland are called Independent Administrative Domains (IAD), and allow any stakeholder to join the

for the data provider

- A Data Cockpit: a unique place to easily browse, read and edit data captured/accessed/stored/transformed by any third-party (real-time and time-machine)
- A Data Marketplace: possibility to securely trade the use of its own restricted data
- Regain trust/control over data-driven services and products
- Regain digital sovereignty: have the choice to manage personal data, or not
- Secondary benefits (longer-term): personalised data-driven services

for the data consumer

- A Data Store: propose data deals (matching and contracting) and leverage data
- A Data catalog: browse available data in the Custodian federation
- Possibility to embed the Cockpit/marketplace features in any app
- Member of the Custodian Federation:
  - enable collaboration between non-mutually trusted parties
  - outsource regulatory burden
  - bring a certain level of trustworthiness (through auditing body, challengers, etc)
Custodian Federation. The SDC is a collection of software services and utilities to be installed on premises. For facilitating the access to a Custodian-as-a-Service, renting a third-party Custodian-enabled cloud-service will also be an option. SDC software supports a governance layer and the ability to exchange information securely to other IADs. Once instantiated, the SDC allows business-as-usual and each IAD becomes a new node made of a Secure Data Vault and a Privacy-Preserving Computing engine. The Reference Architecture will be continuously improved and maintained up-to-date by the SDSC, in collaboration with academia and industries to provide the best solution possible.

Smallest Experiment - A Mobility Use Case
As a lever to promote the adoption of the Swiss Data Custodian, the Smallest Experiment focuses on travellers’ data. Mobility is a rapidly evolving sector, and the innovations in transportation systems and in ICTs generate more and more data with a high potential for the whole of society. The SDSC is collaborating with the Swiss Federal Railways (SBB AG) to bring to market an implementation of the Swiss Data Custodian. The experiment will demonstrate the feasibility of the technological solution and the strengths of the federation (i.e. combining data from non-mutually trusted parties). Through a mobile app, geolocation data will be captured and securely stored in the SBB’s Custodian IAD. The smallest experiment interface (i.e. a mobile/web app) will allow the data consumer to interact with the data providers, and to appreciate the adoption from both individuals and organisations. It will provide with a view on personal data usage and features to control the information and data travellers share. The Reference Architecture together with the Smallest Experiment form the Minimum Valuable Product (MVP).

Other Use Cases
Mobility is not the only data that will feed the Custodian Federation on a longer-term. Other use-cases are envisioned for the future developments. Below are two examples of
what the Custodian could enable in a near future.

(1) Telecommunication data presents a high potential when combined to other kind of data. Yet data combination can barely happen today and manipulating highly valuable, IP protected or personal data remains a challenge. The Custodian, acting as a trusted-intermediary, will allow this business-to-business or business-to-customer data combination while preserving data sovereignty for both parties. The usage over the enriched data, insights or any added-value will be fairly or contractually assigned to one or both party.

(2) Today, the domain of healthcare is highly silo’ed. Pharmaceutical companies develop treatments but strive to get feedbacks on how people react, given their medical antecedents and physical shape. Hospitals generate large data sets barely portable or reusable. And smart devices monitor very precisely individual patients. These three silos protect the privacy of the people, but also hinders the spread of personalised healthcare on a broad scale. The Custodian could reinforce these privacy protections while enabling the silos to work together.
The Multi-Sided Platform

The Secure Multi-Sided Platform (MSP) is based on the federation of multiple IADs. It characterises the exchange and trading of any resource between IADs and it mitigates frictions between non-trusted parties. Ethics is a central part of the design of the platform. It will set the scope of what is shareable and what is not. Concretely, MSP comes with (a) a set of features such as a matchmaker, a data catalog, advanced data management and control, or multi-party contracting; and (b) a stand-alone web/mobile app to let Consumers and Providers access their data environment.

The MSP can be seen as a market place allowing data consumers and data providers to make deals within the trusted Custodian Federation. The data usage consent is managed by the Custodian Technology and keeps the customer’s best interests in mind. The deals are not encouraged to be monetised, and are mostly expected as an exchange data - service, as it is today, but in a safe and trusted environment. The MSP also avoids duplication of data and considerably limits the information that companies can collect about individuals as well as minimising opacity around data usage.
Personal Digital Space and Individual Data Management

Our long-term perspective is that any sensitive or personal data captured in Switzerland will go through a Custodian node and remain in the federation. From the physical perspective, the data is stored in the IAD operated by the data collection app provider. From a logical perspective, data is segregated by consents and usage rights. This way, the consumers and providers keep track of the data they have a usage consent on. Therefore, anyone can log into a Personal Digital Space with their electronic identity, and access a unique knowledge graph of the data transformation lineage. Additional features can also be implemented, such as profile and settings edition, or time- and space-wise data sharing preferences. Theoretically, any individual can decide to host data and run computes on a personal hardware device, at home. This Custodian Pod would be yet another node of the federation.

App Store and Trustlets

Once the pieces of infrastructure previously described will be deployed, Switzerland will be ready to achieve the paradigm shift introduced before. Transparency and control will be enabled, and data will be securely decentralised yet visible and usable by the rightful party. In addition to the Marketplace and the Personal Digital Space that make the Custodian intelligible and valuable, an App Store will be developed. Instead of shipping the data to the service provider, the analytic apps – or Custodian Trustlets – will be downloaded locally on the IAD or Custodian Pod to bring value to sensitive data. User can therefore benefit from a third-party service without giving consent to use personal data. The Custodian Privacy Governance then decide what aggregated / anonymised information is shareable.

The SDC consists of a network of decentralised IADs that interact by means of standard API, services and utilities. Any stakeholder can install the software components from the reference architecture on premise and create an IAD. IADs can also be rented from Custodian-as-a-Service providers. Before joining the Custodian network, the IAD administrator must first register with one of the Custodian certificate authorities (conforming with RFC 5280) and obtain a x509 certificate from it.

Participating IADs interact in a flat hierarchy where mutual trust between different administrative domains is avoided. IADs are in full control of their respective platform services and resources, and manage them independently. However, an IAD has the option to delegate its decision to a controller IAD. This may be needed to ensure compliance with governance policies, or to adhere to federal regulations, or because the local context is insufficient to make an informed decision. Controller IADs are operated by trusted SDC institutions, possibly organised into a governance hierarchy, such as city, cantonal and federal institutions. They operate outside the data processing path, and mainly provide...
governance services. Under this scheme, data processing in the reference architecture is decentralised, but governance is distributed between decentralised and centralised instances.

Access to IAD data storage and/or computing resources are enforced by PrivBAC, an Adaptive Privacy Based Access Management Systems. Resources within the IAD are accessed through a Policy Enforcement Point (PEP) over TLS encrypted REST APIs. The PEP presents the access requests to its Policy Decision Point (PDP), with verifiable contextual attributes such as the requesting application signatures, application domain, user id, and user domain. The PDP combines this contextual information with pertinent information obtained from an RDF-based knowledge representation, called Policy Information Point (PIP), and evaluates the request against a set of boolean rules retrieved from a Policy Retrieval Point (PRP) in order to decide whether or not to grant the request. If needed the PDP can delegate the decision to a PDP placed elsewhere up the governance hierarchy if it is unable to reach a decision locally. The PEP remains responsible for enforcing the decision of the PDP. The PEP may leverage Privacy preserving capabilities (in the form of plugins), such as data encryption, possibly coordinated by third party
services, paving the way to Secure Multi Party Computation (MPC) foundations. The sequence of request-responses is appended to a verifiable token exchanged during the transaction, and the resulting events (such as an application has accessed a data set), are added to the PIP. The PIP contains a representation of the past and present states of the IAD. In addition to providing relevant information to the PDP, the PIP maintains a record of the IAD activities. For instance, through the PIP, data providers have full transparency of the lineage of their data, how it is being accessed, by whom, and for what purposes. In addition to the above components, the IAD includes a local Identity Management (IM) service linked to external IM services, along with service interfaces and utilities allowing users to manage their data, manage data processing workflows, manage access policies, and manage consent forms. Note that for the unexperienced user, managing access right is easily done incorrectly, with potentially disastrous consequences. Therefore, access is not managed directly, but is by default automatically inferred from Terms of Services (TOS) and consent forms.

Data consumer stakeholders develop and deploy application workflows that span one or multiple IAD. The application workflow can be co-authored in an MPC fashion in order to meet the privacy requirements, such as only accessing the minimum necessary information needed to accomplish its intended purpose. Note that the concretisation of a data analytic concept into a solution that features quantifiable privacy metric guarantees is probably the most complex and challenging part of the SDC application deployment process. Once the data processing workflow conforms to the privacy requirements, the executable components of the application are registered as certified Trustlets, and the PRP databases of the respective IAD hosting the application are updated in preparation to receive the application. The application is then deployed and a subscription to the desired population segment of data provider feeds is established. Data providers from the population segment are presented with a TOS provided that they have opted to receive such notifications. If they digitally sign the accompanying consent form, the access rights are automatically updated to authorise the access to the requested data.