National Data Management:
A Modern Data Stewardship Approach
for the Swiss Administration Enabled by
the New Platform "I14Y Interoperability"

Manuela Lenk
Swiss Federal Statistical Office

Georges-Simon Ulrich
Swiss Federal Statistical Office

Abstract

To create the basis for the implementation of the Once-Only-Principle in Switzerland, the Swiss Federal Council launched a program called "National Data Management" in 2019. The Once-Only-Principle aims to reduce the administrative burden on citizens and companies. Instead of having to contact them repetitively for the same information, the responsible authorities should strive to exchange the required data. To enable this, we need to actively work on the data landscape of the Swiss administration. The concept of data stewardship is known to be an efficient tool to promote and ensure a consistent high level of data quality. Nowadays public administrations of many countries strive to embrace the idea of data stewardship. This paper describes the approach we have developed for the Swiss administration.

Keywords: National data management, data stewardship, once-only-principle, interoperability.

1. Introduction

Data stewardship can be an important tool to achieve and maintain a high degree of data quality (Plotkin 2014). This is particularly true for a distributed and diverse data landscape as it is so often the case in the public administration. There is no centralized authority in the classic stewardship approach as originally proposed in 1996 (Dawes 1996). The very nature of a modern state with its division of power and responsibilities across many institutions, seems to make its application especially well suited.

However, data stewardship strategies differ vastly among the various countries, which decided to implement them (van Donge, Bharosa, and Janssen 2022). As a contribution to the public discussion, we want to outline our approach for the Swiss administration as introduced by the federal program "National Data Management".

2. National data management

The program is launched by the Swiss Federal Council with the intention to create the basis

for the implementation of the "Once-Only-Principle" in Switzerland (Swiss Federal Statistical Office 2020): Instead of having to report the same information to different government agencies multiple times, administrative units should aim to exchange the required information directly in between them. This way, countries might be able to reduce the administrative burden on businesses and citizens significantly. Currently that approach is taken in a wide range of states and is even promoted on a supranational level (Krimmer R., Prentza A. and Mamrot S. 2022)(European Union 2017). A regular data exchange between different government agencies necessitates - beside the legal foundation - a couple of technical prerequisites. Among these a common view on the concepts present in the administrative data, seems to be the most important one to us. That is the reason why we made the effort to design a comprehensive collaboration model, which defines the necessary roles that all participating administrative units (on federal level) must implement. We see our model as crucial to enable the data harmonization process across all relevant governmental layers. In Switzerland, this includes three governmental levels: federal, cantonal as well as municipal. The idea of data stewardship is at the very core of our model.

3. Data stewardship

The Swiss Federal Council assigned the responsibility for the implementation of the program "National Data Management" to the Federal Statistical Office ("FSO"). At FSO, we have a long history of dealing with a wide range of diverse data while also ensuring a high degree of data quality. Like other national statistics institutes, we embraced early on the idea of data stewardship within our organization. Since we receive data from nearly all parts of the Swiss administration, we already had to establish an (internal) data harmonization process in the past. We developed an in-house solution called "Statistical Information System" ("SIS"). The kernel can be best described as a statistical data warehouse in which the various domain specific datasets are loaded. Those Data Owners, who have the responsibility for our statistical business process, administer the datasets in SIS. They direct their long-term development and restrict the access according to our legal framework. However, to place their datasets in the system, Data Owners need to get the approval from the responsible Data Steward. It is at this point, where data harmonization policies are enforced: Data Stewards examine incoming upload requests and verify if they comply with all applicable standards, e. g. if underlying concepts, such as "occupation" or "age group", are encoded as agreed upon at FSO.

Our in-house methods have not only provided us with essential knowledge for executing data harmonization on a large scale consistently and sustainably, but also highlighted the central importance of Data Stewards in everyday tasks. Based on our experience, the introduction of Data Stewards is a prerequisite for effective cross-departmental and - in particular - cross-organizational data management. However, Data Stewards alone cannot guarantee high data quality. For their ongoing work, they require several points of contact who have specific, well-defined tasks and responsibilities. These points of contact need to exist in all involved institutions and at all levels. It is this key insight, among others, that has motivated us to develop the approach described in the following sections.

This model is based on our yearlong experience and was extensively discussed with all stakeholders in the Swiss administration. While the overall data stewardship strategy is in its final form, the actual implementation is an ongoing process at the time of writing this article.

4. Institutional level

Our model distinguishes two levels: institutional and federal. The institutional level consists of all administrative units that participate in the program "National Data Management" by offering access to their domain specific core datasets. It is important to note that this includes not only federal government agencies such as FSO or the Federal Tax Administration but also cantonal institutions and city administrations. In other words: "Institutional" does not point

to a specific governmental level, but rather to the level where the data is administered.

At this level, our model defines five roles, two of them being mandatory for every participating institution: Data Owners and Local Data Stewards. They are the bare minimum set of roles, because both are needed to efficiently communicate not only with other institutions but also with other roles at the federal level. As already pointed out, data stewardship is the core concept of our approach. Therefore, Local Data Stewards are one of the main forces we rely on, to drive forward the process of data harmonization in Switzerland. They are usually domain experts and strive to fit all datasets into a consistent data landscape of their organizations. Data Owners are also of central importance: They control access to those datasets and work closely together with Local Data Stewards to guide the available development resources into a meaningful direction.
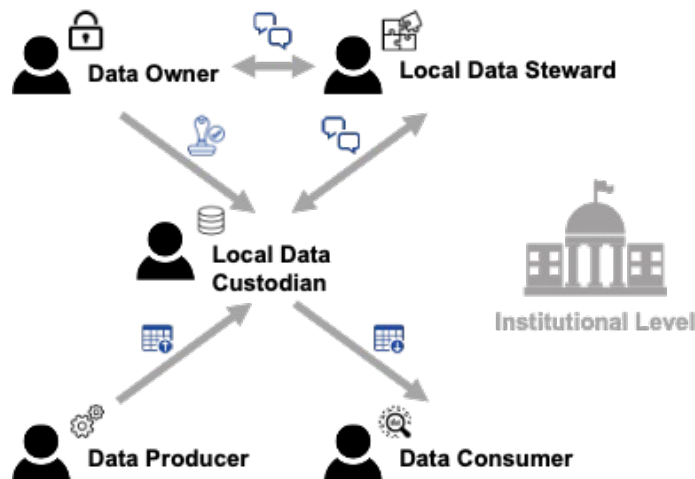

Figure 1: Institutional level

Local Data Custodians handle the access to the actual technical infrastructure. Although this role is optional, it is nearly everywhere present and usually filled out by the IT department. Local Data Custodians always act on behalf (and the explicit approval) of the responsible Data Owners.

Data Producers sit at the very base of where the actual data creation takes place. Their view is narrowed down due to the focus on the day-to-day task of generating the most accurate data possible. That involves constant monitoring and correcting incoming data. On the exact opposite side of the local data chain sit Data Consumers. They are typically – but not exclusively – statisticians, data scientists, analysts or even decision makers and rely heavily on a consistent data landscape. Any missed opportunities during the harmonization process puts at best an extra workload on them. In worst case, it leads to an incomplete analysis or even wrong conclusions.

During repeated consultations with our stakeholders in the Swiss administration, we found out that except for data harmonization tasks, all other responsibilities described here, are already carried out in nearly all organizations we examined.

Now that we established a common set of roles at the institutional level, we can proceed to lay out how the cross-organizational collaboration takes place.

5. Collaboration across organizational boundaries

Local Data Stewards represent together with Data Owners an interface to their organizations for other entities in the Swiss administration. There are two key collaborations: One is with the Swiss Data Steward during the participation of the Local Data Stewards in specific working groups and the other is with our central governing committee in the program "National

Data Management".

The concept of a nation-wide Data Steward is already explored and successfully implemented in other countries (UNECE/CES Task Force 2022). Similarly to the approach in New Zealand (New Zealand Government 2021), the role of Swiss Data Steward is fulfilled by the directorate of FSO. The Swiss Data Steward has the high-level overview, which enables him or her to support the data harmonization process of the overall data landscape. Here we mean with "overall" indeed the universal sense of the term: The Swiss Data Steward carries the responsibility to direct all Local Data Stewards to develop and subsequently use the same data standards for the same underlying concepts. A recent example is the "Swiss Standard Classification of Occupations" which is now a standard throughout the federal administration (and strongly recommended for all other governmental levels) for the representation of a person's profession. Such standards are published on the "I14Y Interoperability Platform" ("114Y IOP", see https://www.i14y.admin.ch) – our main technical tool in the program "National Data Management" – and made accessible by an open API so that cross-organizational processes can retrieve these declarations in an automated way.
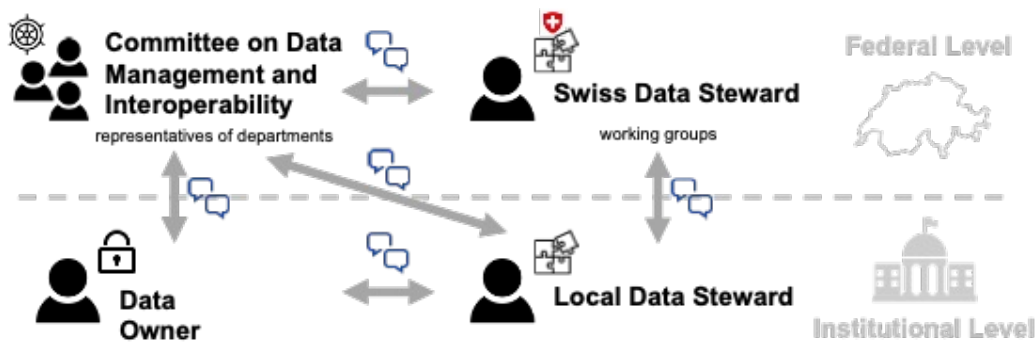

Figure 2: Interplay between institutional and federal level

The "Committee on Data Management and Interoperability" is comprised out of representatives of all Swiss departments. In this panel the most fundamental strategic decisions are made in collaboration with all Data Stewards (local as well as federal) and Data Owners. Besides providing the overall roadmap for the program "National Data Management", the committee also defines towards which part of the data landscape the efforts will be directed to in the coming months.

6. Cross sectional domains

"Cross Sectional Domains" are a concept, which we developed out of our experience with data harmonization. As it turns out the harmonization of a data space must be layered towards – usually – more than just one aspect: While a certain concept might be encoded in an "Open Government Dataset" in a specific way, the very same representation could be much less useful in a statistical context. It is exactly these "contexts" what we call "Cross Sectional Domains". We regard them as dimensions of the data space that must be harmonized by their own.

For instance, encoding the age as a categorical variable (e.g. "0-49 and 49+") in an open government dataset may not only be meaningful but even particularly desirable due to data protection considerations. Nevertheless, this format substantially compromises the dataset's utility for statistical analyses. Another example involves hierarchies, such as geographical subdivisions. For instance, summarizing the surrounding municipalities of a city with the city itself into an urban settlement area might make sense for a particular statistical context, while for other evaluations, administrative boundaries may be paramount.
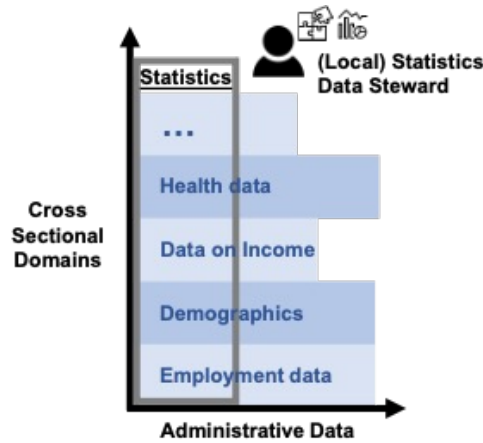
Figure 3: Cross sectional domains and data stewardship

Therefore, our collaboration model provides a separate branch of specialized Data Stewards such as "Statistics Data Steward" at the institutional as well as on federal level. They look out to direct and develop the datasets to maximize their usefulness within their very specific domain. Without this additional layer of data stewardship, the process to agree upon a common representation for a set of concepts, tends to be overly cumbersome and – in our experience – often fails to deliver the degree of consolidation one would wish for. We attribute this to the different requirements (and sometimes hidden characteristics) of the subparts of the data space which we meanwhile call "Cross Sectional Domains".

Especially the domain of official statistics benefits from this approach: It guarantees a degree of freedom in designing and harmonizing of that part of the data landscape, which might otherwise not exist.

Table 1: Roles within the program "National Data Management"

| Role	Description
Swiss Data Steward	Coordinates the standardisation and harmonisation of administrative data across all institutions. This role is currently fulfilled by FSO.
Statistics Data Steward	Coordinates the standardisation and harmonisation especially of statistical data across all institutions. This role is also currently fulfilled by FSO.
Data Owner	Directs the development of a particular dataset in an institution and controls access to it. He or she is ultimately the decision-maker when it comes to propagating changes within their data domain.
Local Data Steward	Drives the standardisation and harmonisation of administrative data within one specific institution.
Local Statistics Data Steward	Drives the standardisation and harmonisation of statistical data within one specific institution.
Local Data Custodian	Is technically responsible for hosting the data. Acts exclusively on behalf of the Data Owner.
Data Producer	Creates a particular dataset based on internal or external sources in an institution.
Data Consumer	Utilizes one or more datasets in an institution.

7. Complete data stewardship model
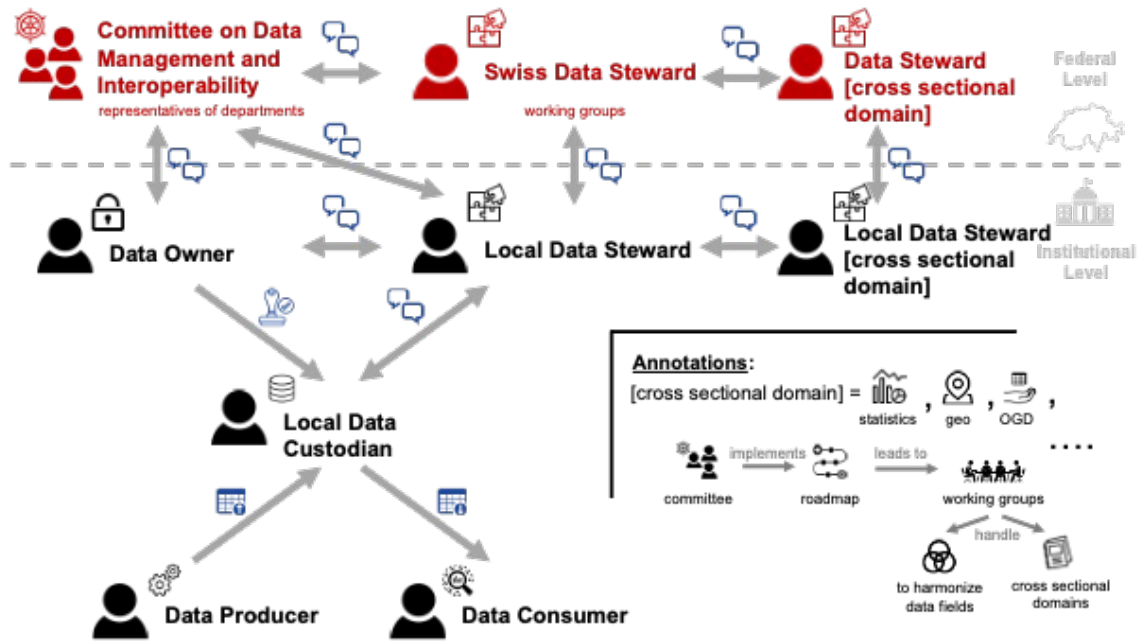
Figure 4: Complete data stewardship model of the program "National Data Management"

Now that we described both levels of our approach separately, it is much easier to walk through the complete model, which otherwise might have appeared to be unnecessarily complex. The figure above depicts both levels separated by a grey dotted line. In the lower part, we find all roles working in a data producing (or providing) institution. As explained before, they strive to align and harmonize all datasets to fit them into their common vision of an orderly (local) data landscape.

Well-maintained datasets at an organizational level do not automatically lead to an overall consistent data landscape. That's why Local Data Stewards consult regularly with their counter parts on federal level, such as the Swiss Data Steward or Statistics Data Steward. Since this collaboration is so crucial, we ensure its existence – as already mentioned – by making the roles Data Owner and Local Data Steward mandatory for federal institutions and strongly recommend them for all other participants.

Looking at the overall process from a data perspective, we can describe the workflow with a simplified example: While a Local Statistics Data Steward takes care that all spatial information is encoded the same way in datasets which are subject to statistical analysis in his or her own organization, the Statistics Data Steward makes sure to develop and establish a common standard for this purpose throughout the whole Swiss administration. – A job that can only be meaningfully performed by the help of a close collaboration with all Local Statistics Data Stewards: without them, the Statistics Data Steward would probably not be able to elaborate a complete standard covering all requirements and worse, he or she might even overlook the principle need for a certain data standard.

8. I14Y Interoperability platform

As already mentioned, I14Y IOP is the main technical tool of our federal program "National Data Management". The platform hosts all jointly developed data standards and enables interoperable processes by also being the central metadata catalogue for the relevant datasets of the Swiss administration: If you are a public institution in Switzerland and want to search for data you might need in one of your administrative processes, you can simply use the catalogue search to find out who in the administration might be able to deliver that data
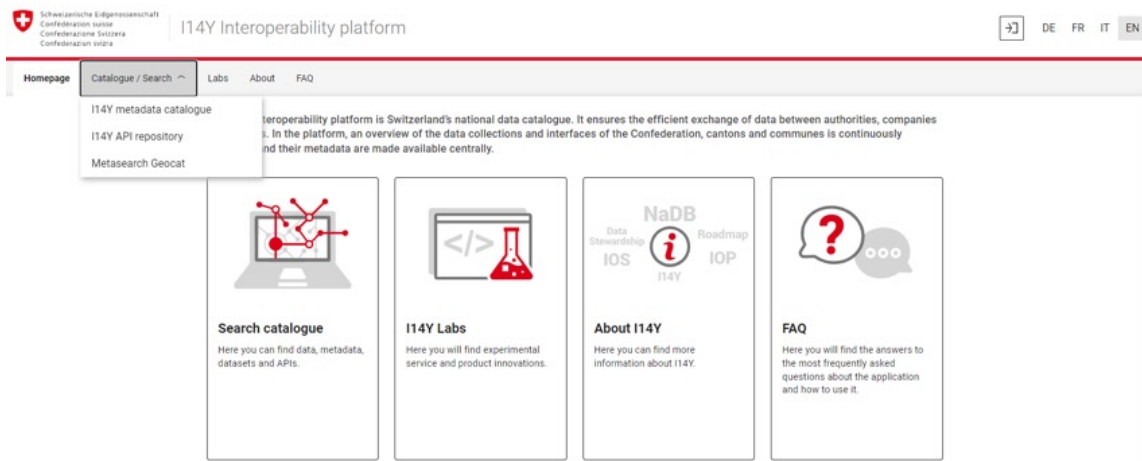
Figure 5: Home screen of the new developed "I14Y Interoperability platform"

(provided you are legally authorized to ask for it). On top of that, I14Y IOP also serves as an API repository, meaning that the metadata on it not only describes all datasets in detail (column types, encodings etc.) but also how to access them programmatically.

Furthermore, over the last year we extended the platform to serve also as a private metadata catalogue, so that other institutions can use I14Y IOP for internal purposes: They are now able to describe their internal data landscape without having to do their own (redundant) implementation of a local metadata catalogue, thus saving crucial development resources. OFS has the plan to open source the whole platform in the very near future. International institutions will then be able to contribute to the project or custom-fit the code base for their own needs.

9. Data ecosystem

For the time being, we have focused our program on the core data landscape of the public administration. This typically involves highly structured, relational data, which usually contain one or more well-defined primary keys. In many domains, we already have identifiers that are recognized and utilized across organizational boundaries. Our main role as a National Statistics Institute in this context is to encourage the broad and consistent use of these identifiers. A good example of this is the AHV number, which is a unique person identifier managed by a governmental authority called the "Central Compensation Office" ("CCO"). The CCO is not only responsible for issuing AHV numbers, it actively manages them to ensure consistency and uniqueness, for example, by regularly publishing lists of inactive and annulled identifiers. Following a law amendment effective from January 1, 2022, the AHV number may now be systematically used by all authorities to fulfill their statutory duties.

10. Legal framework

In this paper, we have intentionally not delved into the legal foundations. This has several reasons. Firstly, we anticipate that the situation in Switzerland - with its extensive network of cantonal and federal legislation - may be somewhat unique in an international comparison and thus may not relate to the legal situation of an international reader. Secondly, an accurate representation of the overall situation would quickly become entangled in numerous legal details. Many of the regulations relevant to our work stem from the combined interpretation of various legal sources. While the Federal Statistics Act is the most central of these for our institution, numerous complementary regulations come into play, particularly within the

framework of the program "National Data Management". Lastly, it should also be noted that this specific part of the legal landscape is still evolving due to the currency of this subject.

11. Conclusions

We described the data stewardship approach taken by the federal program "National Data Management". The model is in its final form and we are currently in the implementation process. It consists of two levels and defines a mandatory and optional set of roles for participating organizations. As far as our experience goes, a huge data landscape as the one, we are confronted with at the Swiss administration, can only be harmonized in a meaningful away, if those underlying dimensions, which we call "Cross Sectional Domains" are identified and separately developed by dedicated Data Stewards. As in other countries such as New Zealand, the Federal Statistical Office is serving as Swiss Data Steward (as well as Statistics Data Steward). This makes sense given the fact that national statistics institutes usually have a long history of maintaining, harmonizing and quality assuring datasets from a diverse set of sources.

References

Dawes S (1996). "Interagency Information Sharing: Expected Benefits, Manageable Risks." *Journal of Policy Analysis and Management*, **15**(3), 377–394. `doi:https://doi.org/10.1002/(SICI)1520-6688(199622)15:3<377::AID-PAM3>3.0.CO;2-F`.

European Union (2017). "Tallinn Declaration on eGovernment." URL `https://ec.europa.eu/newsroom/document.cfm?doc_id=47559`.

Krimmer R, Prentza A and Mamrot S (ed.) (2022). *The Once-Only Principle.* Springer. `doi:10.1007/978-3-030-79851-2`. ISBN 978-3-030-79851-2, URL `https://link.springer.com/book/10.1007/978-3-030-79851-2`.

New Zealand Government (2021). "A Data Stewardship Framework for NZ." URL `https://data.govt.nz/toolkit/data-stewardship/a-data-stewardship-framework-for-nz/`.

Plotkin D (2014). *Data Stewardship: An Actionable Guide to Effective Data Management and Data Governance.* Morgan Kaufmann. `doi:10.1016/C2012-0-07057-3`. ISBN 978-0-12-410389-4, URL `http://www.sciencedirect.com/science/book/9780124103894`.

Swiss Federal Statistical Office (2020). *NaDB - in brief.* 2014-2000. Bundesamt für Statistik (BFS), Neuchâtel. URL `https://dam-api.bfs.admin.ch/hub/api/dam/assets/12307690/master`.

UNECE/CES Task Force (2022). "Data Stewardship." URL `https://unece.org/sites/default/files/2022-06/Data_Stewardship_ver_1_020622%20-%20for%20consultation.pdf`.

van Donge W, Bharosa N, Janssen M (2022). "Data-driven Government: Cross-case Comparison of Data Stewardship in Data Ecosystems." *Government Information Quarterly: An International Journal of Information Technology Management, Policies, and Practices*, **39**(2). `doi:10.1016/j.giq.2021.101642`.

Affiliation:

Manuela Lenk
Swiss Federal Statistical Office
Neuchâtel, Switzerland
E-mail: manuela.lenk@bfs.admin.ch
URL: https://www.bfs.admin.ch/bfs/en/home/fso/swiss-federal-statistical-office/overview/registers.html

Georges-Simon Ulrich
Swiss Federal Statistical Office
Neuchâtel, Switzerland
E-mail: georges-simon.ulrich@bfs.admin.ch
URL: https://www.bfs.admin.ch/bfs/en/home/fso/swiss-federal-statistical-office/overview/directorate.html