



# L'harmonisation des données à l'OFS



Fabrice Gouzi / Fabian Santi, SIS-DataStewards

Journées suisses de la statistique – 2021-09-06



# Inhalt

- **NaDB: Ein schweizweites Programm!**
- **Datenharmonisierung im BFS**
- Erste **Schlussfolgerungen** und **Erfahrungen**



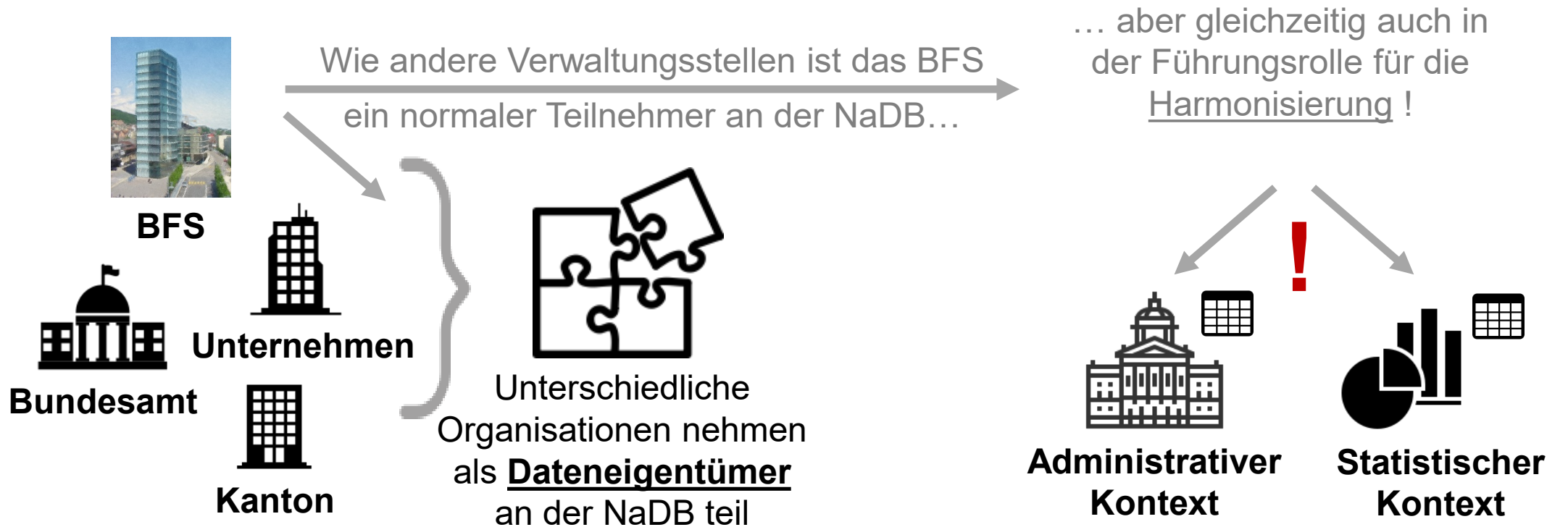


# NaDB und das «Once Only» - Prinzip

- **N**ationale **D**aten**b**ewirtschaftung (kurz: «NaDB»)
- Umsetzung des **Once Only Prinzips**
- Führung **BFS**
- Schweizweites Programm
- Beauftragt vom **Bundesrat (!)**
- **Datenharmonisierung** von Anfang an im Vordergrund
- Beginn der Arbeiten im BFS



# Harmonisierung: Unterschiedliche Ebenen



Harmonisierung muss bereits auf institutioneller Ebene beginnen!



# I14Y Interoperability platform

- **V1 online seit 30.6.2021**
- **Harmonisierte** Metadaten und Nomenklaturen
- Wo gibt es welche Daten ?
- **Daten verbleiben in den jeweiligen Verwaltungseinheiten**
- Abrufbar über **APIs**
- **Knowledge Graph** für Metadaten

The screenshot shows the I14Y Interoperability platform interface. The page title is 'Schweizer Berufsnomenklatur CH-ISCO-19'. The navigation menu includes 'Startseite', 'Nomenklaturen', 'Labs', 'About', and 'FAQ'. The main content area has tabs for 'Beschreibung', 'Inhalt', 'API', and 'Herunterladen'. The 'Beschreibung' tab is active, showing a detailed description of the CH-ISCO-19 classification and its relationship to the international ISCO standard. Below the description, there is a table with metadata:

Methodische Revisionen:	Anpassungen an die periodischen Revisionen der Internationalen Standardklassifikation der Berufe ISCO
Identifer	HCL_CH_ISCO_19_PROF@Federal_Statistical_Office
Publikationsdatum	18.10.2019
Änderungsdatum	27.03.2021

<https://www.i14y.admin.ch>



# SIS 2.0

- **S**tatistisches **I**nformations**s**ystem (kurz: «SIS»)
- Umsetzung der NaDB im BFS
- Harmonisierungs- und Standardisierungsprojekt (Daten und Applikationen)
- Web Portal als zentrales Tor zu **Prozessen** und **Daten**



# SIS 2.0 - Portal

SIS-Portal DE FR IT EN

Startseite Statistisches Portfolio Metadaten

Filter Mehrjahresprogramm

Mehrmjahresprogramm 2020-2023

- BFS: Information, Grundlagen und Infrastrukturen
- Register
- Wirtschaft und öffentliche Finanzen
  - Bautätigkeit
  - Tourismus
  - Geld, Banken und Versicherungen
  - Vorsorgeeinrichtungen (Pensionskassen)
  - Öffentliche Finanzen
  - Land- und Forstwirtschaft, Ernährung
  - Preise
  - Volkswirtschaft
  - Industrie und Dienstleistungen
  - Arbeit und Erwerb: Aus Sicht der Unternehmen
- Soziales
- Raum, Umwelt und nachhaltige Entwicklung
- Mobilität und Verkehr
- Gesundheit
- Gesellschaft
- Energie
- Bildung und Wissenschaft
- Bevölkerung
- Multithematische Analysen

Suche

Mehrmjahresprogramm 2020-2023

## Statistisches Mehrjahresprogramm des Bundes 2020 - 2023

Begin 2020 Ende 2023

Themenfelder	Informationsfelder
Multithematische Analysen	○
Bevölkerung	○
Bildung und Wissenschaft	○
Energie	○
Gesellschaft	○
Gesundheit	○
Mobilität und Verkehr	○
Raum, Umwelt und nachhaltige Entwicklung	○
Soziales	○
Wirtschaft und öffentliche Finanzen	○
Register	○
BFS: Information, Grundlagen und Infrastrukturen	○

**Web Portal ist nur  
der sichtbare Teil.  
SIS 2.0 ist viel mehr!**





# Datenharmonisierung

- **Die einzelnen Experten sind** verantwortlich für abgestimmte Merkmale
- **Definierte und verwendete Variablen**
- **Harmonisierung** aller Statistiken, abteilungsübergreifend
- Ist **vorbereitende Tätigkeit** zur Publikation der Metadaten auf I14Y!





# Definierte Variablen

Statistik der natürlichen Bevölkerungsbewegung (BEVNAT)

## DV\_SEXE

<https://sis-portal.bfs.admin.ch/stp/definedVariables/08d9407d-2412-bf1d-b8a8-4bc2b65456d6>

DE: Geschlecht      FR: Sexe      EN: Sex      IT: Sesso

Codes:

1 -> DE: Männer      FR: Hommes      EN: Male      IT: Uomini

2 -> DE: Frauen      FR: Femmes      EN: Female      IT: Donne

References:

ISO/IEC 5218:2004(en):

Information technology — Codes for the representation of human sexes

And more metadata!



# Definierte Variablen

Statistik der natürlichen Bevölkerungsbewegung (BEVNAT)

## DV\_SEXE

<https://sis-portal.bfs.admin.ch/stp/definedVariables/08d9407d-2412-bf1d-b8a8-4bc2b65456d6>

DE: Geschlecht      FR: Sexe      EN: Sex      IT: Sesso

Codes:

1 -> DE: Männer      FR: Hommes      EN: Male      IT: Uomini

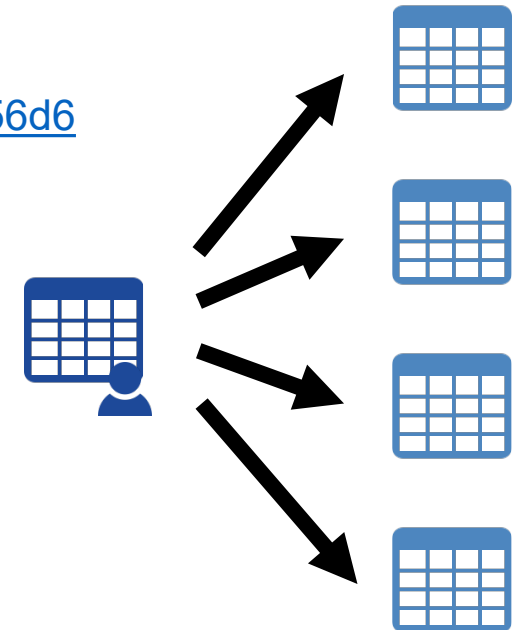
2 -> DE: Frauen      FR: Femmes      EN: Female      IT: Donne

References:

ISO/IEC 5218:2004(en):

Information technology — Codes for the representation of human sexes

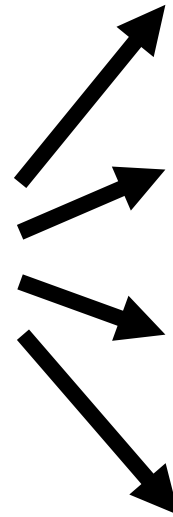
And more metadata!





# Definierte und Verwendete Variablen

Statistik der natürlichen  
Bevölkerungsbewegung (BEVNAT)



Landwirtschaftliche Strukturerhebung



Strukturerhebung (SE)



Reiseverhalten der Schweizer Wohnbevölkerung



Enquête sur le budget des ménages (EBM)



## Datenharmonisierung → Die Rolle des SIS-Datasteward

- **Iterativer Prozess:** Suche nach dem **Besitzer** einer **Variable / Codeliste**. **Unterstützt und begleitet** die Fachgespräche. Die endgültige **Entscheidung zur Harmonisierung** wird vom **Fach** getroffen.
- Fixiert die Werte für die **Qualitätscodes** und die **generischen Codelisten**.
- **Unterstützt** das Fach bei der **Erfassung** von Metadaten.
- Prüft, ob die **Anforderungen für die Registrierung eingehalten wurden** (Harmonisierung und Vollständigkeit der Metadaten).
- **Validiert definierte Variablen** für die Verwendung im System.
- **Validiert Datensätze** für die Verwendung im System.



# Die Dataset Structure Definition (DSD) Adoptions setzt sich hauptsächlich aus vorhandenen Variablen zusammen

- Basisdaten für die Statistik der [Adoptions](#).
- Datentabelle: 38 Spalten.  
→ 38 verwendeten Variablen.
- DSD bestehend aus 15 DV:  
→ 12 generische DVs  
→ 3 spezifische DVs
- [Demo DSD\\_ADOPTIONS](#)

Defined variable	Responsible	DV description	# UV
DV_ST 2019.1		Staaten und Gebiete	6
DV_JAHR		Jahr	6
DV_DATE		Null	4
DV_JA_NEIN		ja / nein	3
DV_ALTER_ERREICHT		Erreichtes Alter	3
DV_ALTER_ERFUELLT		Erfülltes Alter	3
DV_AHVN13		AHV-Versichertennummer (AH..	3
DV_GDE_KT 2020.3		Schweizer Kantone	2
DV_ZGB1_T		Adoption: Gemäss Art. 264	1
DV_SEXE		Geschlecht	1
DV_SCHAETZUNG		Codes, die eine Schätzung cha..	1
DV_LEIBLICHE_ELTERN_CD_N		Leibliche Eltern: Status	1
DV_LAUF_KS_N		Datensatz-Nr	1
DV_GDE_GDE 2020.3		Schweizer Gemeinden	1
DV_ETAT_CIVIL		Zivilstand	1
DV_AP_STATUS_N		Adoptiveltern(teil): Status	1



# Harmonisierungsbeispiele (Top 10)

Nom court	Name	Beschreibung	Zahl
<a href="#">DV_JA_NEIN</a>	ja / nein	Standardisierte definierte Variable der Codeliste ja/nein.	213
<a href="#">DV_ST_2019_1</a>	Staaten und Gebiete	Definierte Variable, welche die von der Schweiz anerkannten Staaten und Gebiete enthält.	100
<a href="#">DV_JAHR</a>	Jahr	Definierte Variable: Kalenderjahr (von 1. Januar bis 31. Dezember)	84
<a href="#">DV_DATE</a>	Datum	Datum Format ISO-8601 ( <a href="https://en.wikipedia.org/wiki/ISO_8601">https://en.wikipedia.org/wiki/ISO_8601</a> )	74
<a href="#">DV_SCHAETZUNG</a>	Codes, die eine Schätzung charakterisieren	Diese Codeliste gibt an, ob die Daten von einer Schätzung abgeleitet sind und welche Art von Schätzung angewandt wurde. Die Daten können mit den Original-/Quelldaten des Datenlieferanten übereinstimmen oder das Ergebnis einer Schätzung/Imputation eines bestimmten Typs sein.	61
<a href="#">DV_ALTER_ERFUELLT</a>	Erfülltes Alter	Das erfüllte Alter ist das Alter beim letzten Geburtstag, d.h. es entspricht der Anzahl der vollendeten Lebensjahre.	56
<a href="#">DV_AHVN13</a>	AHV-Versichertennummer (AHVN13)	Die Einführung der AHVN13 in den Personenregistern basiert hauptsächlich auf der Änderung des AHVG und dessen Verordnung, die am 1. Dezember 2007 in Kraft getreten ist.	52
<a href="#">DV_GDE_KT_2019.1</a>	Schweizer Kantone	Definierte Variable, die die Schweizer Kantone enthält.	48
<a href="#">DV_GDE_GDE_2019.1</a>	Schweizer Gemeinden	Definierte Variable, die die Schweizer Gemeinden enthält.	36
<a href="#">DV_ALTER_MT_ERFUELLT</a>	Alter in erfüllten Monaten	Das Alter in erfüllten Monaten entspricht der Anzahl der vollendeten Lebensmonate	32



# Erste Erfahrungen

- **Metadaten** und **Datenharmonisierung** sind eine grosse Herausforderung
- Aber: **alternativlos**
- Das BFS hat **früh damit begonnen** und ist bereit anderen beim selben Prozess zu unterstützen





# Fragen / Diskussion

