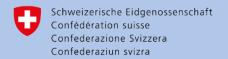
Bundesamt für Statistik BFS Office fédéral de la statistique OFS Ufficio federale di statistica UST Uffizi federal da statistica UST



Project "Metadata Access Automation"

Fabian Santi – IOR/IOS 2022-11-03

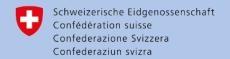




Content

- 1. Objectives
- 2. Expected added value
- 3. Problems encountered/challenges identified
- 4. Solutions provided
- 5. Lessons learned



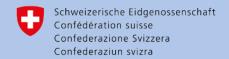




Current and desired situation: FSO project with external data users

- Metadata delivery
 - Often with Excel files
 - Data descriptions possibly in pdf files
 - Ex: municipalities list version xy
- Changes during the course of the project
 - Repetition of the delivery
 - Code adaptation and risk of error

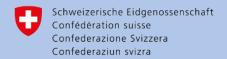
- Metadata delivery
 - URL + API
 - No separate delivery of metadata documents
- Changes during the course of the project
 - Always a uniform and up-to-date basis from the URL and API
 - Few or no code adaptations





Expected added value

- Integrated use of SMS 2.0 / I14Y (meta)data in all statistical projects, internal and external
- Better metadata "automatically" through multiple use
- Incentive to improve data harmonisation
- Incentive to describe other data sources via the I14Y platform
- Data access could be structured in a harmonised way in the long term.

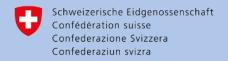




Problems encountered/challenges identified

- External access
- Ease of use
- Availability of necessary APIs



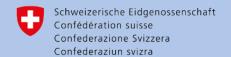




Problems encountered/challenges identified

- External access → Public Gitlab
- Ease of use → R/Python librairies
- Availability of necessary APIs
 → In development







Solutions provided: 2 PoC

fso.metadata (Public access)

R: https://gitlab.com/DSCC/fso-metadata

Python:

https://pypi.org/project/fso-metadata/

DSDValidatoR (Ease of use)

R:

https://bitbucket.bit.admin.ch/pr

ojects/BFS-FACH/repos/sis-

sms-

tools/browse/DSDValidatoR

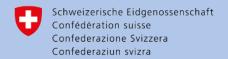


Demo

Demo 1: fso.metadata

Demo 2: **DSDValidatoR**







Lessons learned

- Development of a metadata system takes longer than expected
- Public APIs must be specified for several use cases and have a dedicated infrastructure, which slows down development
- The metadata system and public APIs are necessary conditions for the creation of user-friendly libraries
 - → We hope to have a stable version available in 2023