Contribution ID: **35**   Type: **Regular Presentation**

Documenting and validating administrative data with DDI

Wednesday, 29 November 2023 10:59 (30 minutes)

Official statistics increasingly rely on external sources, particularly administrative data, to produce statistics. This requires further industrialisation of the data integration before the downstream steps leading to dissemination.
In 2021, INSEE has launched a project named Resil with the objective of centralising administrative data ingestion for further processing of social statistics. Raw data received must be delivered to statisticians with documentation, in particular structural metadata.

DDI is used for describing the structure of the data stored. The data acquisition tool takes this formal description as input and generates the model of ingested data. So, DDI is used in an active way, since the DDI documentation is required to process the data received.

In order to assess the compliance between data and associated structural metadata (e.g., an integer value must be between limits, a code value must be present in a code list...), INSEE has developed a tool which generates VTL scripts (Validation and Transformation Language is a standard, part of SDMX) from structural metadata in DDI. The validation is then performed using Trevas, an open source VTL engine for big data distributed environments.

The presentation will focus on DDI as the pivot model for documenting and validating administrative data in the context of Resil.

Primary authors: TAILHURAT, Romain (Insee); Mr DUBOIS, Thomas (Insee)

Presenter: TAILHURAT, Romain (Insee)

Session Classification: DDI - New Directions

Track Classification: Proposed Topics of the Conference: User Needs, Efficient Infrastructures and Improved Quality