

UNITED NATIONS ECONOMIC COMMISSION FOR EUROPE
CONFERENCE OF EUROPEAN STATISTICIANS

ModernStats World Workshop

21-22 October 2024, Geneva

A dataset catalogue as a tool for automated and metadata driven statistical production

Speaker: Henrik Andersson, Statistics Sweden

Author(s): Henrik Andersson, Statistics Sweden; Patrik Wahlgren, Statistics Sweden

Abstract

To have better knowledge and control over the data in the statistical production, Statistics Sweden is implementing a dataset catalogue for the statistical production. The catalogue describes the datasets in the different phases, steady states, of the statistical production, with references to the in-house metadata systems together with references to the actual datapoints. Thus, the dataset catalogue will be a reference point between all metadata along with references to where the actual datapoints are stored, which is stored will be in each producing system. In order to automate the statistical process, the dataset catalogue is equipped with a notifying system. The consumer system can therefore set up a subscription for datasets of interests and fetch the datapoints from the producer system whenever new datasets has been entered in the catalogue. By using a json schema to describe GSIM datastructure, the datapoints can be retrieved through a generic api. For large datasets we can will use other means of transport, but the datastructure schema remains the same.