

Topic (iv): Editing in a generic process, standardisation and metadata driven processes

Discussion

Topic (iv): Points for discussion

1/4

- Questions for discussion taken from **GSDEM** presentation:
 - Is GSDEM well known in NSO? What is your experience?
 - If not enough, what action is needed to promote this standard model among responsables of surveys?
 - What feedback have you received from users?
 - Is there anything missing in this standard model?

- Questions for discussion taken from **Bosch *et al.***' s paper:
 - What is your experience about validation rules?
 - The process of validation rules was often considered as something different from SDE as a whole. This probably resulted in problems such as 'data ping pong'.
 - Some requirements for success:
 - Validation rules as a part of the whole SDE process. Consider that some of the validation rules are generic and already included in any SDE process.
 - For each rule identify the right place in the SDE process.
 - Don't force the use of a single environment or instrument to implement validation rules. Often there is resistance from responsables of a survey to change the instruments or software used for a process.

- Points for discussion taken from **Poland's** paper:
 - The paper presented the idea of a logical architecture of a reference architecture framework based on a process oriented model of statistical production. A lot of work has been done from a theoretical point of view, but much work still needs to be done in practice.
 - A good understanding of the concept of metadata is needed in the various contexts. Often misunderstandings or misinterpretations raises between experts on metadata and experts on survey process.
 - Another key issue: development of metadata exchange interfaces. The access must be automatic, not complicated, and user friendly. The system must also be feasible, understandable, and easy to manage. System maintenance must be performed timely and successfully.

- Points for discussion taken from **Germany's** paper:
 - The main goal: introducing an automated E&I process for a survey with millions of records. Three software/methods were considered: CANCEIS, Banff, and HoloClean (and also missForest).
 - Although machine learning method is one of the most innovative, it is not suitable for the case of study. This is mainly due to the target variables on which the system is based (categorical, alphanumeric, and discrete variables). CANCEIS, which is targeted also on numerical variables, revealed best performances.
 - Some considerations: innovative methods are not necessarily more suitable than 'conventional' ones; in this case, the former needs further developments to take into account quantitative variables and related errors, the latter also needs to be improved to overcome those limits they have for specific conditions, thus getting better performances.