

**UNITED NATIONS
ECONOMIC COMMISSION FOR EUROPE**

CONFERENCE OF EUROPEAN STATISTICIANS

Work Session on Statistical Data Editing
(Budapest, Hungary, 14-16 September 2015)

Topic (iii): Software tools and international collaboration

**The ValiDat Foundation Project:
Survey on the Different Approaches to Validation Applied Within the ESS**

Prepared by Sarah Giessing and Katrin Walsdorfer, Federal Statistical Office, Germany

I. Introduction

1. In late 2014, the one year ESSnet project ValiDat Foundation with participants from four Member States (Germany, Italy, Netherlands and Lithuania) was launched to accompany the ESS.VIP Validation aiming to harmonize data validation approaches in the ESS. Such an attempt can only be effective and contribute to the improvement of the ESS, if the interests of the Member States are adequately incorporated into the wider frame of a European approach. Therefore one of the goals of the ESSNET ValiDat foundation was to identify a common baseline for validation from a methodological and technological point of view. Finding a common language, using the same terms and understanding them in the same way, when talking about validation, is fundamental to the improvement of the validation process within and between statistical institutes.

2. To find out how the typical validation process is practically implemented in the Member States of the ESS a questionnaire for a survey has been sent to the National Statistical Institutes. The aim was to compile a comprehensive overview of validation methodologies and practical aspects of their application. The survey might be considered as a first attempt to systemize the approaches in the ESS. The questionnaire is subdivided in general and domain specific parts and tries to explore validation from a technical and methodological point of view. The paper will discuss and illustrate the questionnaire and its design in section II. The main focus is on presenting and commenting the results of the survey, see section III. We conclude with a summary section.

II. Planning and Design of the Survey

3. When planning the survey, the first issue to be resolved was regarding the scope of the topic: What is "Data Validation" anyway? Actually, validation is aimed at verifying whether data can be considered correct, that is to be considered acceptable. Any actions concerning further investigation into records identified as erroneous (error localization) or concerning the correction of errors are out of the scope of validation and hence out of the scope of the survey. Another issue was the question of participation. Although it might have been interesting to collect information not only from the National Statistical Institutes (NSIs), but also from other National Authorities with special responsibilities for collecting or compiling data for some particular statistics in the ESS, we decided to send the questionnaire to the NSIs only. This had certain organizational advantages, like e.g. availability of National contact points etc. .

4. The questionnaire was sent out to the NSIs by first of April 2015, with April 27 as deadline for the response. These were almost four weeks, but with the Easter break falling into the period. By end of

April, the response rate was about 50%. However, several questionnaires came in just a few days later. In the end, we received completed questionnaires from 25 of the 30 NSIs in the ESS.

A. Design of the general parts of the questionnaire

5. Given the tight schedule of the survey and assuming that most respondents would be comfortable with it, we decided to use an Excel workbook as format of the questionnaire. Considering that the intention of the survey was to catch a variety of information not usually available in one person, it seemed appropriate to split the questionnaire across several worksheets. Some of the worksheets were more general, collecting process related and technological information on the implementation of validation rules. Others were dedicated to individual domains to collect practical, specific information on validation rules. The intention was that each workbook should be filled in by another person in the institute.

6. This way, we ended up with three different worksheet templates. One for general information on the institute and its organization (sheet 1) and one concerned with general issues of technical implementation (sheet 2). The third template (sheet 3) was actually duplicated several times, meant to be filled in by experts for specific surveys or domains. Content and organization of those three templates were as follows:

- (a) Sheet 1 “Your Institute” to capture contact information and general background on organizational matters of official statistics in a country. E.g.
 - whether or not regional offices exist, possibly organized in a federal structure, or
 - whether or not there are other National authorities in the country providing official statistics and
 - whether or not the NSI co-operates with those authorities
- (b) Sheet 2 “Cross-domain” refers to the general usage of validation methodology, services and technology in an institute across statistical domains. Questions address
 - Validation methodology
 - Validation services
 - Validation technology

7. Concerning the *methodology*, issues considered in the questions relate to organizational aspects, e.g. the extent to which validation methodology is centralized in an institute, if there is a regular exchange of know-how, and if there are on-going efforts for standardization. The purpose of the questions about the validation *services* was to find out to what extent validation is centralized. In this context the term “validation service” refers to an IT service that is invoked or to which data are transferred, or it may refer to an organisational unit to which the task of validating data has been exclusively delegated. The respective questions were asked separately for each of the four GSBPM (= Generic Statistical Business Process Model) business processes that may involve validation, e.g. “Collect”, “Process”, “Analyse” and “Disseminate”.

8. The questions on validation *technology* concern the tools (application, frameworks, etc.) used. Respondents were asked to indicate the overall importance of tools used for validation by assigning a rank to each out of twelve choices offered. They should also indicate, which tools are used in which of the four GSBPM processes. Among the fifteen tools were well known standard packages for editing and imputation like BLAISE, BANFF, EDIT, CANCEIS. In addition to those, the list contained standard general statistical packages like SAS, SPSS and R, data base systems (Oracle, SQL, ACCESS), EXCEL and generic tools (GENEDI/SDMX converter). Respondents were also invited to add other tools used for validation to the list (and to briefly explain them). Finally, respondents were asked, in which of the four previously mentioned GSBPM processes the institute is standardizing on the tools, or specification language (i.e. to express validation rules), or data formats, or even on the rule sets.

B. Design of the statistic specific part (“Sheet 3”) of the questionnaire

9. Sheet 3 was implemented as a statistic specific sheet, duplicated several times. It was designed to collect survey-related information, dedicated to specific domains and to be filled in by the domain experts. Covering all domains was out of scope, of course. We selected a few statistics, trying to cover a broad field of applied validation methodology. In particular, we wanted to include a typical statistic where Member States deliver micro-data to Eurostat (as is the case for the Labour Force Survey), and also a typical statistic where only aggregated data are delivered (like the Structural Business Survey data). A good example for a statistics involving administrative data seemed to be the Census. From a European management perspective, it also makes sense to include a statistic subject of a running European task force, like Agricultural or Price Statistics.

10. Finally, we ended up with five statistic specific templates, e.g. for

- Census 2011,
- Labour Force Survey (LFS),
- Structural Business Survey (SBS),
- Agriculture,
- Prices,

asking respondents to answer the questions only with respect to her own particular field of work.

11. In the first part of the statistic specific part of the questionnaire, respondents should describe the validation process. Respective issues were

- documentation and methodological foundation of validation rule definition,
- relation to the data editing process,
- evaluation and update of the validation process / validation rules
- data sources of the domain
- co-operation with other domains / offices / national authorities

12. The second part concerned the validation rules. The intention was to collect information on “typical” validation rules and categories of validation rules. For that purpose the questionnaire used a technical/logical oriented typology of validation rules. It has been assumed that there are basically two sets (in the following: set A and set B) of rules:

13. Set A: Rules to ensure technical integrity of a datafile, like

- formal validity of entries (valid datatype, field length, characters, numerical range)
- presence of an entry
- no duplicate units
- all the values in a field of one dataset are contained in a field of another dataset (for instance contained in a codelist)
- each record has a valid number of related records (in a hierarchical file structure)

Set B: Rules for logical validation and consistency

14. For set B, for the survey a 2-way typology was assumed, with the following rule type categories in those two “dimensions”:

Typology dimension	Types of checks	
1	Identity checks	Range checks <ul style="list-style-type: none"> • bounds fixed • bounds depending on entries in other fields
2	Simple checks, based directly on the entry of a target field	More “complex” checks, combining more than one field by functions (like sums, differences, ratios)

Table 1: Categories of a 2-way typology for validation rules for logical validation and consistency

15. Also, rules are often implemented as conditional checks, i.e. they are only checked, if a certain condition holds. This can be regarded as another property of a rule and might be considered as additional “dimension” of the rule typologies (for both rule sets, A. and B.). Of course there might be several conditions combined by logical AND or OR statements. However, this issue has not been particularly addressed in the survey.

16. Regarding the technical integrity rules (set A), respondents were asked to provide an own example for each rule category and to explain, which tools are used to implements them. Regarding the logical consistency rules (set B), respondents have been asked to provide own examples for each category of the 2-way typology of rules implied by table 1. They should also provide some information on the frequency of the respective types of checks and if those types of rules were implemented rather as hard checks, identifying records or fields as "erroneous", or as soft checks, identifying records or fields as "suspicious" or "outliers". Finally, respondents should indicate in which part of the production process (e.g. in which of the four relevant GSBPM business processes) rules of a certain type are used “frequently”.

17. The final question of the statistic specific part of the questionnaire was concerning the workload. Respondents were asked to provide workload percentages for data validation on one hand, and editing and correction on the other hand. Eventually, at the end of the sheet, there was also some space for feedback on the survey itself and for suggestions for its improvement.

III. Results of the Survey

18. The final response rate of the survey was 83%. Most of the NSIs filled out the complete questionnaire. Some respondents did not complete all the statistic specific templates. So the response rates for the statistic specific templates ranged from 67% for prices to 80% for Census 2011 and Agriculture. Analysing the results of the survey, we have produced a wide range of figures, most of which will be presented in the present section, or in the appendix. The figures present simple response scores. Be aware that bars in figures may not always be directly comparable, as for example response rates may differ between questions, etc. The figure captures correspond to the exact wording of the respective question in the questionnaire.

A. Results of the general parts of the questionnaire

19. (a) Sheet 1 “Your institute”

In this first worksheet template we asked for general information on the institute and its organization to find out about the institutional differences between the NSIs. The first question deals with the organization of the institutes. 57% of the NSIs are organized as one central statistical office. 35% have in addition regional offices whereas 9 % of the institutes have a decentralized system. Regarding co-operation with other National Authorities, in all responding countries such authorities exist (e.g. in 22 countries), with the exception of one country. And almost all responding organizations (21) co-operate with such partners. Only in one country no such other authority exists.

20. (b) Sheet 2 “Cross-domain”

refers to the general usage of validation methodology, services and technology. On those issues, we got the following results:

21. Concerning the methodology, on the extent to which validation methodology is centralized, we found that half of the responding institutes (11) do not take a centralized approach. A regular knowledge-sharing process seems to be rather exceptional (4 institutes). In the majority of institutes there is no co-ordinated knowledge sharing (13), or only on an ad-hoc basis (2). Nevertheless, 15 institutes report some in-house efforts on standardisation/research of validation methodology. 16 institutes participate in international efforts, inside the ESS in most cases (11). On the topic of validation, only 4 institutes co-operate with the academia.

22. The second part of the Cross-domain sheet has addressed the issue of using central validation services during the different phases of the processing the data. During data collection, for data collected from web forms, the majority of the institutes (15) use central validation services. Not quite so many report using central validation services when collecting data through file upload (10) or through other channels (10).

23. In the three other GSBPM processes, e.g. Process, Analyse and Disseminate only one institute has reported that they use fully centralized validation services, e.g. in the “Process” phase (Sweden). In those three process phases in the majority of countries each domain runs its own validation process, but there is also a number of countries with centralized services for some specific domains, c.f. figure 2.

24. The final block of the cross-domain sheet is concerned with the tools used in the various business process steps. The two tools with the highest scores on overall importance for data validation are SAS and Excel, c.f. figure 1. This general impression is of course also reflected in figure 3 looking at the use of the tools for the particular process steps.

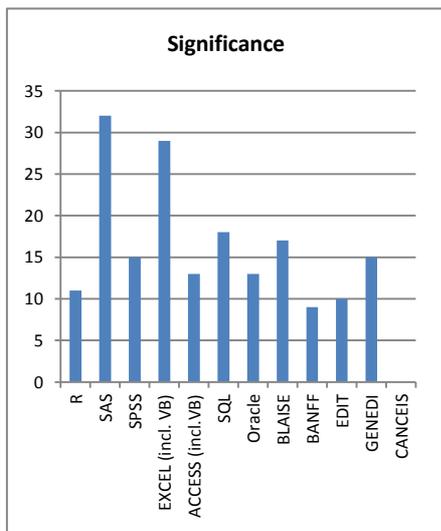


Figure 1: Which tools (application, frameworks, etc.) do you use in the respective business process? [Check 0-4 boxes/row]

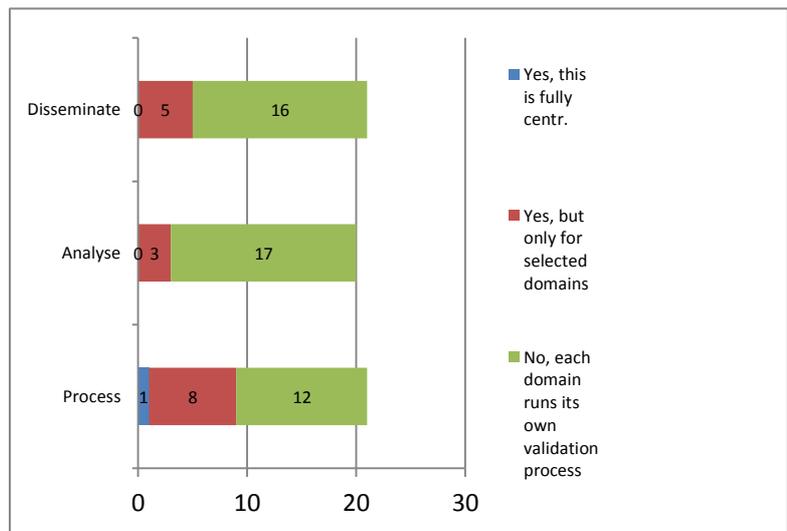


Figure 2: Do you use central validation services in the GSBPM business process...

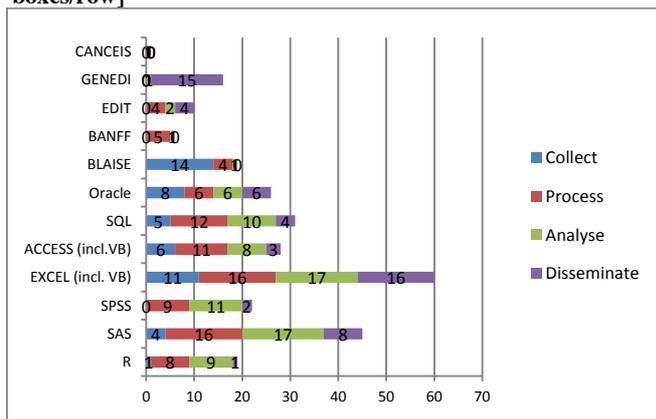


Figure 3: Which tools (application, frameworks, etc.) do you use in the respective business process? [Check 0-4 boxes/row]

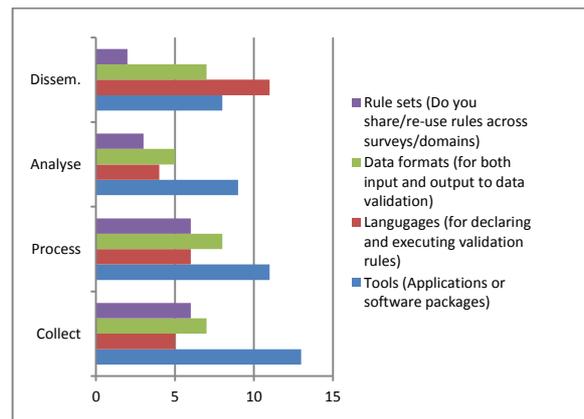


Figure 4: Please indicate for the respective process if you are standardising on validation tools and languages. [Check 0-4 boxes/row]

25. However, there are differences. The most important tool for the Collect phase is for example BLAISE, while Genedi / SDMX converter tools are relevant for the dissemination phase (and only for this phase).

26. Finally, respondents were asked about standardizing efforts on validation tools, languages, data formats and rule sets in the four process steps, c.f. figure 4. Obviously, the highest level of standardization is reached in the sharing of tools, especially during data

collection. Sharing rule-specification languages is most common during the dissemination phase. Sharing of rule sets across surveys and domains is not uncommon, but less frequent.

B. Results of the “general framework” part of the statistic specific part of the questionnaire

27. In this section we report the results of the first block of questions from the statistic specific part, supposed to establish the general framework for validation in the domains: Regarding documentation, the majority of respondents (100, across all areas) told us that the validation process is “documented in a programming language”. Only 67 reported documentation in a management system. Compared to the other three statistics, the discrepancy between these two options is much larger for price-statistics (20 vs. 11) and agriculture (25 vs. 15). For Census, on the other hand, the share is almost equal (17 vs. 14). Regarding methodological foundation of the validation rules, across all domains 81 respondents say they use methodology for defining validation rules (about 80% of the respondents for agriculture and price-statistics, vs. about only 50% on LFS). For graphic representation see figure A1 in the appendix.

28. The issue addressed by the next question concerned separation between implementation of the data validation process on one hand and the editing and imputation processes on the other: In three domains (e.g. Census, prices, SBS) the majority of respondents claim that these are separate processes. For agriculture and LFS on the other hand, a full separation is the case only for a minority of countries (38% in agriculture and 30% in LFS). See figure A2 in the appendix.

29. On the topic of an evaluation process for validation (rules) raised in another question, overall a slight majority answered “yes” (about 56%) (SBS: 67 %; Agriculture: 65 %; Census: 62 %; LFS: 48 %; Price- Statistics 32 %). A somehow related question was about whether validation process and/or rules are regularly updated. The responses gave a similar picture (as probably to be expected), but on a higher level, e.g. question answered with “yes” by about 81% of the respondents (SBS 95 %; Agriculture 93 %; LFS: 83 %; Census: 70 %; Prices 64 %). C.f. figures A3 and A4 in the appendix.

30. Another question concerned the role particular data sources play in an area:

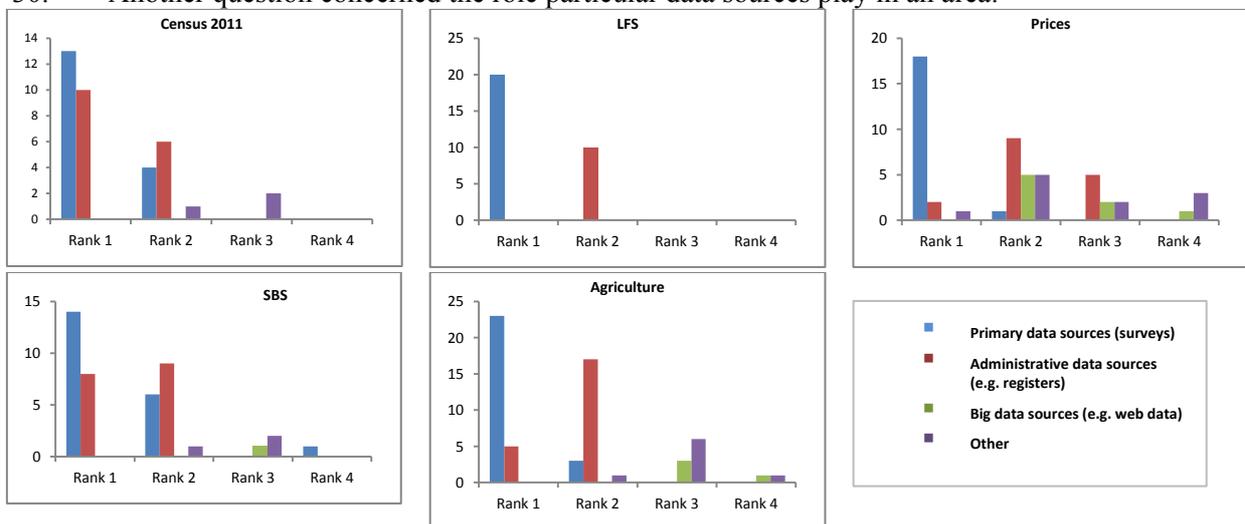


Figure 5: Please rank, if applicable, your domains's data sources according to their share in the data collected, specifying an integer starting from 1, where 1 is the highest rank. Ranks can be shared.

As expected, for all five statistics, the majority of the respondents assigned the highest rank to “primary data sources” (all respondents, in case of LFS, while for SBS eight respondents assigned this rank to the “administrative sources”). No surprise also (more or less just the flip side of the coin) that through all statistics “administrative data” got the largest score for rank 2. However, for price-statistics 5 countries gave rank 2 to “big/web data” while 5 other countries gave rank 2 to “other sources”. Apart from this high score for rank 2, the “other” sources received the highest score (i.e. 6 countries) as score for rank 3 (in agricultural statistics). The “big/web” data sources play a fairly important role mainly in price-statistics with 5 + 2 + 1 nominations for ranks 2, 3, and 4, resp., and are not totally unimportant in agricultural statistics (3 + 1 nominations for ranks 3 and 4).

31. At least somewhat related to that question was the final question in this block, asking about cooperation with other partners (other domain, national authorities, or statistical office), c.f. figure A5 in the appendix. About 55% of the respondents said they do, with the highest score (about 81 %) in agriculture, and the lowest (38 %) in LFS. The latter result fits into the picture since – when answering the “role of data sources” questions - none of the LFS respondents assigned a rank to “big/web” or “other” data sources. For all LFS respondents “primary sources” are most important (10 of them still putting “administrative sources” on rank 2).

C. Examples of typical rules and conditions provided by the respondents

32. Table 2 below presents one example¹ for each rule type in the set A of rules to ensure technical integrity (c.f. II.B.). For the rule types of set B (Rules for logical validation and consistency), table 3 provides examples¹.

Formal validity of... - datatype - field length - characters - numerical range	<i>Telephone number</i> : Numerical data. <i>Date</i> : If Date is given as text it should be 8 characters long <i>Date</i> : If Date is given as text it should contain only numbers. <i>Month</i> : Month of arrival in the country must be in {1,...,12}
Presence of an entry	<i>Persons in households</i> : It is checked whether all have responded
No duplicate units	<i>Holding ID</i> : Each holding has a unique ID number, duplicate ID numbers are not allowed within the dataset
All the values in a field of one dataset are contained in a field of another dataset (for instance contained in a codelist)	<i>Occupation</i> : Field “Occupation” must contain only entries from a list of valid ISCO-08(COM) codes at 3 digits level
Each record has a valid number of related records (in a hierarchical file structure)	<i>Number of members of a family</i> : the aggregated number of persons in each family must be equal to the number of individual rows in the dataset corresponding to the members of that family

Table 2: Examples of rules to ensure technical integrity of a datafile

	“Simple”	“Complex” (checks involving functions on field entries)
Identity check	<i>In a sheep survey</i> : “Milk production” must be equal to “milk disposal”	<i>Employment</i> : “Number of persons engaged” must be equal to the sum of “employees” and “self-employed persons”
Range checks - bounds fixed - bounds depending on entries in other fields	<i>Working hours (monthly)</i> : “Hours worked” must be between 0 and 168	<i>Average poultry weight</i> : “weight of poultry” divided by “number of poultry” must be between 0.03 and 40
	<i>Cereal production</i> : “Organic cereals” must be less or equal to “cereals”	<i>External services</i> : “Expenses on external services” must be greater or equal to “payment for agency workers” plus “telecommunications” plus “business trips of company personnel”

Table 3: Examples of rules for logical validation and consistency

¹ Examples selected from the examples provided by the survey respondents.

33. Typical conditions of a conditional check mentioned by the survey respondents are f.i.
- if “age under 15” (then marital status must be not married), or
 - if “legal form: Self-Employed” (then number of self-employed" must exceed 0), or
 - if “status in employment = compulsory military service” (then sex must be male), or
 - if “no. of employees not zero” (then wages and salaries must be greater than zero), or
 - if “enterprise reports production of goods” (then it should also report costs for raw material), etc.

D. Results of the “validation rules” part of the statistic specific part of the questionnaire

34. The first question of that part was about the tools for implementing the technical integrity rules (set A.). According to the responses it could be said that for all domains this type of rules is mainly implemented in a data base (Oracle, SQL) or “other” (in-house, etc) tools². Also SAS plays an important role. But there are some notable exceptions, like: for the LFS Blaise and SPSS are also very common tools, in agriculture ACCESS and EXCEL are widely used, and ACCESS is also important for SBS technical-rules implementations. Figure 6 presents the details.

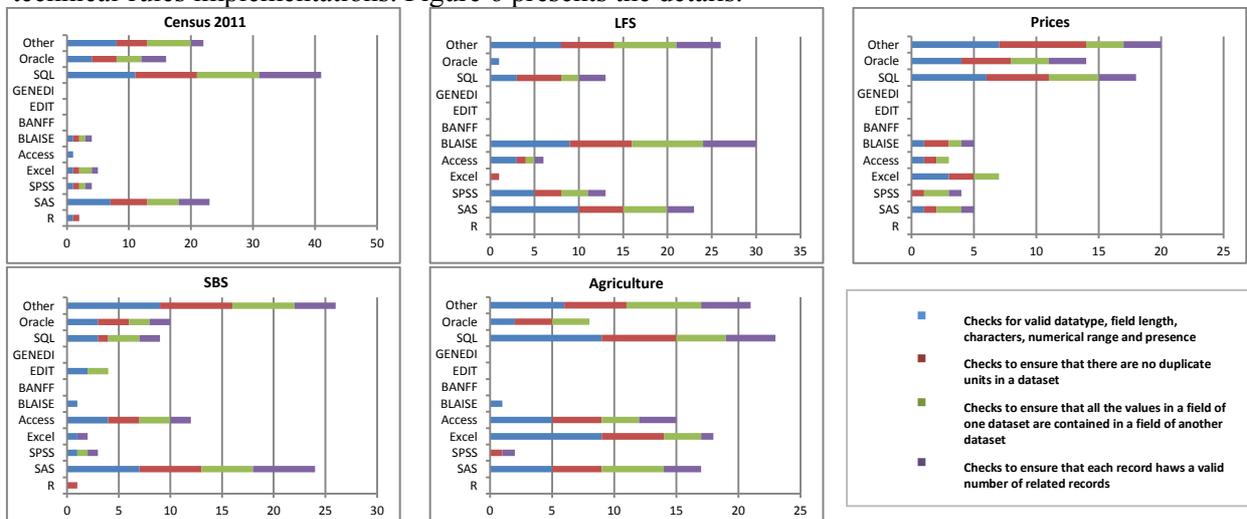


Figure 6: Which tools do you usually use for these types of validation rules?

35. The remaining questions were directed towards the rules for logical validation and consistency (i.e. from set B.) and how frequently certain types of rules are used.

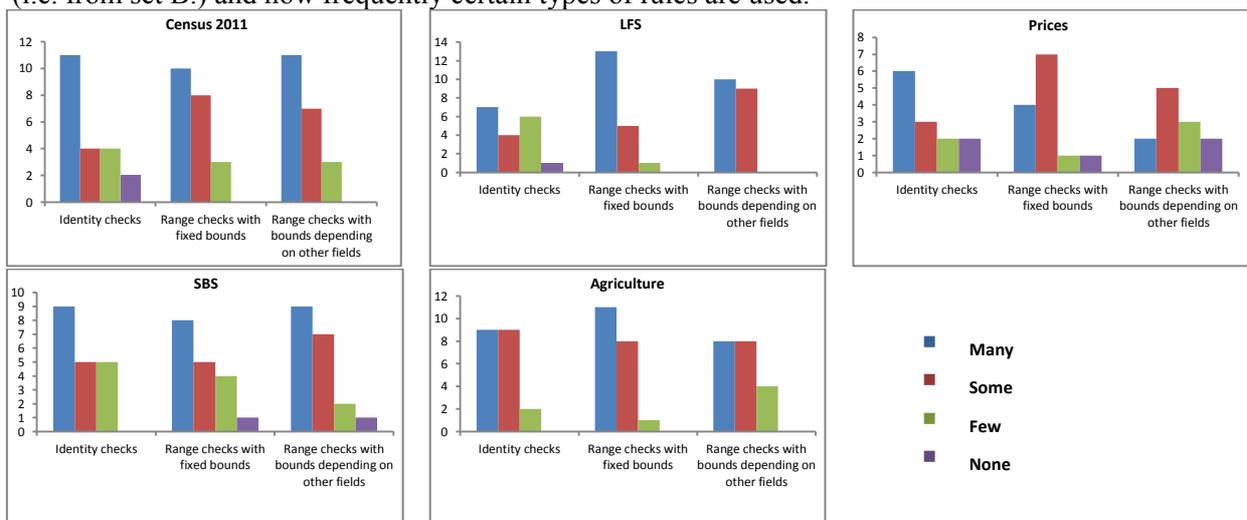


Figure 7: How many validation rules for logical checks do you use in your domain? [Check 1 box/row]

² A notable exception is LFS where Oracle is used by only one of the countries.

The general impression is that all domains claim they have many rules of both categories, i.e. identity checks as well as range checks (both with bounds fixed, and with bounds depending on entries in other fields). Comparing the domains, the impression is that in LFS and agriculture range checks with fixed bounds are slightly more important, while for price-statistics, identity checks play a bigger role. Especially for the Census and for SBS, this general picture does not change much, when considering (for the next question) how frequently those types of rules are implemented as conditional rules. For LFS, the impression is that particularly the range checks are implemented typically as conditional checks, while for agriculture it seems to be rather the other way round. For price-statistics, conditional checks seem to be comparatively less frequent. Many respondents claim that those checks are not used at all in combination with range checks, especially not with range-checks where the bounds depend on entries in other fields. See figure A6 (appendix).

36. Taking into account also the second dimension of our rule-typology, i.e. whether or not checks are complex in the sense that they involve sums or ratios of fields, seems to have been relatively difficult to answer for the respondents. Especially for price-statistics, the differences between the scores for the various response categories are low and some general caution is due when interpreting the result because respondents might have been tempted to rate range-checks, particularly those with bounds depending on other fields, as complex anyway, independent of the additional complexity due to the involvement of sums or ratios. What might still be said is that for SBS range-checks are quite often implemented as complex checks, while in price statistics the checks generally tend to be simple. With due caution, the general notion is that range-checks with bounds depending on other fields tend to be implemented more frequently involving sums or ratios, especially as opposed to identity checks. C.f. figure A7 (Appendix).

37. On the issue of hard and soft checks, we got a fairly clear picture: in the Census and in LFS the hard checks prevail, while for price-statistics it is rather the other way round. For SBS and agriculture, both types are common. Generally, identity checks tend to be implemented as hard-checks, whereas range checks are more frequently used for soft checks. Figure 8 presents the details.

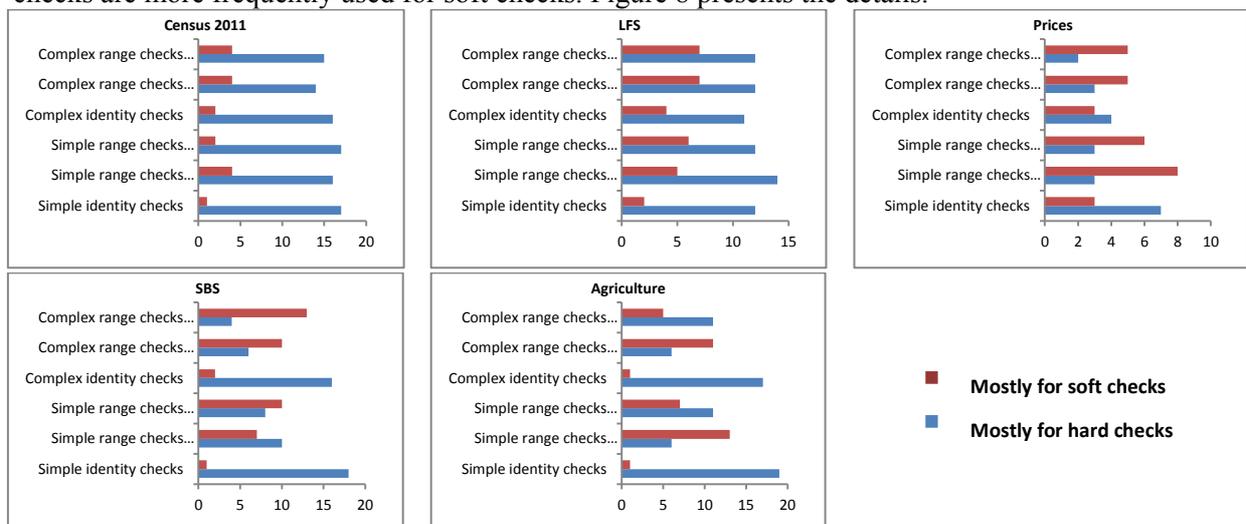


Figure 8: Checks can be implemented as hard checks, identifying records or fields as "erroneous". Or they may be implemented as soft checks, identifying records or fields as "suspicious" or "outliers".

Also, the results on the distributions of the checks by phases of the production process tend to be quite obvious: Generally, the share of checks in the final (disseminate) phase is low. For price-statistics, the share of checks decreases from phase to phase, starting from the Collect-phase. For LFS and agriculture, the majority of checks are in the Collect- and Process-phase. For the Census, the checks are more concentrated on the phases Process and Analyse, while for SBS the largest amount of checking seems to be during the Process-phase. Regarding the different types of checks, except for LFS there is an overall tendency that simple checks not involving sums or ratios are more frequently used in the Collect phase, especially in price-statistics. See figure 9.

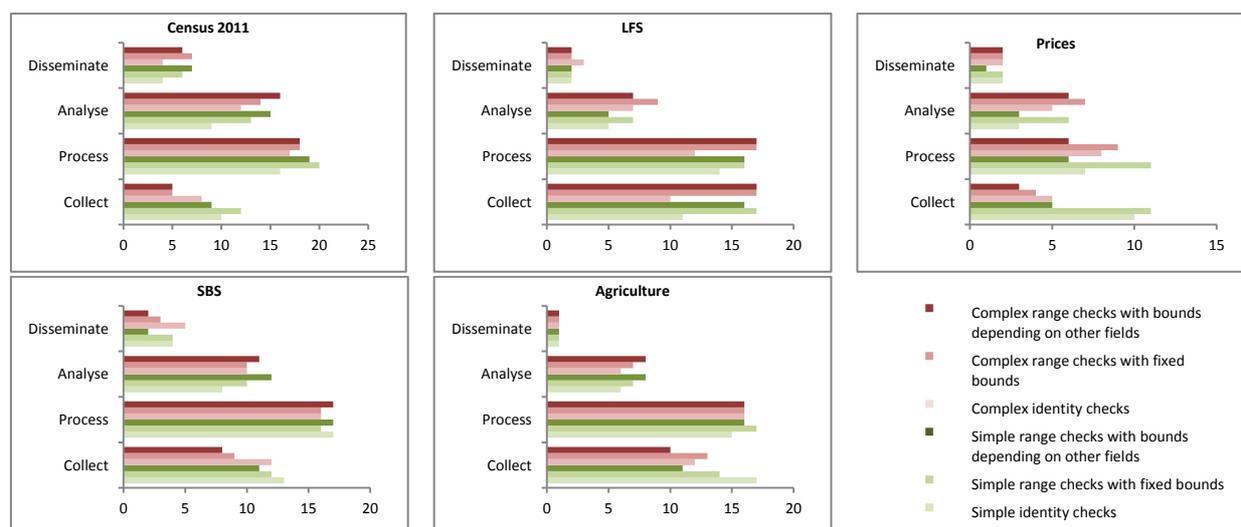


Figure 9: In which phase of the production process do you frequently use logical validation rules ? [Check 0-4 boxes/row]

38. The final question of that block was about workload. Using the 25% and the 75% quartiles of the percentages indicated by the respondents, gives the idea that the workload for the validation process ranges mostly between 10% and 35% of the total workload. This is quite remarkable, taking into account that we asked the respondents to consider only in-house workload, and only regular, recurring work. Specifically, they should not include one-time, preparatory work nor the workload of respondents or machine workload. This range is almost similar to the quartile ranges computed for the respective workload share of the editing and correction process (e.g. 11% to 40%). Naturally, there are some differences between domains. For example for Census, Agriculture and Prices, the lower quartile for the share of the validation process is even 15% (e.g. higher). For LFS, the upper quartile is only 20%, while for SBS it is even 40% (i.e. for SBS we have a large response variability in that question). Anyway, even taken with some caution, it is evident that the validation process takes a considerable amount of the workload resources. However, those workload shares are probably mostly rather rough estimates, considering especially that for many countries and domains the separation between the validation and the editing process is not very clear (c.f. result reported in III.B, or figure A2 in the appendix).

III. Summary and Final Remarks

39. This paper has reported on a survey carried out as one of the actions of the ESSNET ValiDat Foundation. It has explained in detail the structure and the design of the questionnaire which was composed mainly of a cross-domain section, and a domain specific template for the domains Census 2011, Labour Force Survey (LFS), Structural Business Survey (SBS), Agriculture, and Prices. The questionnaire was sent to the National Statistical Institutes in the European Statistical System. Questions of the cross-domain section were dealing with organisational issues and the state of standardization concerning validation methodology, services and technology, looking particularly at similarities and differences between the different stages of the statistical data production process. The answers provided by the institutes have been analysed in detail, and the results have been presented in this paper. A major focus of the individual, statistic specific questions was the design and structure of validation rules. The paper has explained the typology of rules used to collect information on typical rules, especially on how frequently certain rule types are implemented. The respondents were also asked to provide typical examples for the different rule types. A selection of those examples has also been presented in the paper.

40. Considering the background and perspective of our project, perhaps the main results could be summarized in this way: Firstly, the validation process takes a considerable share of the workload of a production process. Secondly, statistical institutes trying to standardize in the validation process seem to standardize on tools in the first place. However, in most NSIs each domain runs its own process. Looking at the different domains, for all the topics addressed in the survey, we found some similarities, but also a lot of dissimilarities. Overall, the results of the survey are certainly a step forward regarding the general objective of exploring validation practices.

Acknowledgements

41. This work has been supported by the ESSNET project “ValiDat Foundation” Nr. 11103.2014.002-2014.751. This support is appreciated. In the first place however, we owe gratitude to the many respondents of this survey. Without their remarkable input, this work would not have been possible.

Appendix

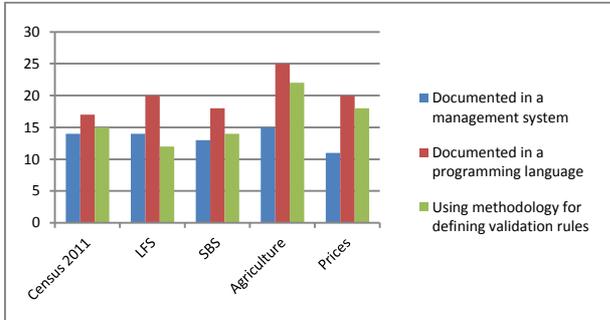


Figure A1: How would you characterize this process? [Check 0-3 boxes]

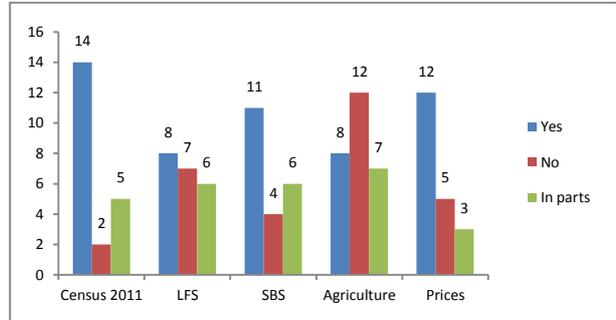


Figure A2: Are data validation and data editing implemented as separate processes in your domain?

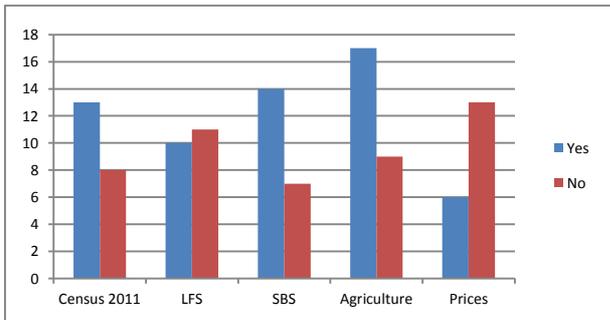


Figure A3: Is there an evaluation process to assess the validation process/the validation rules (coverage, quality)?

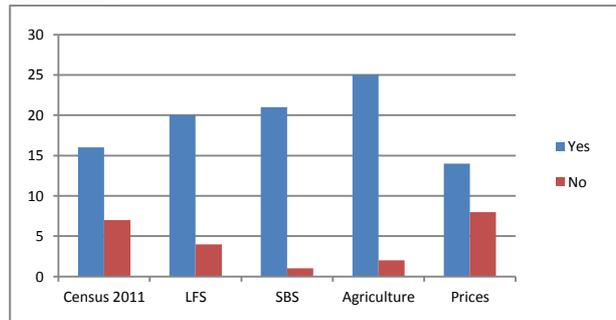


Figure A4: Do you update the validation process/the validation rules ?

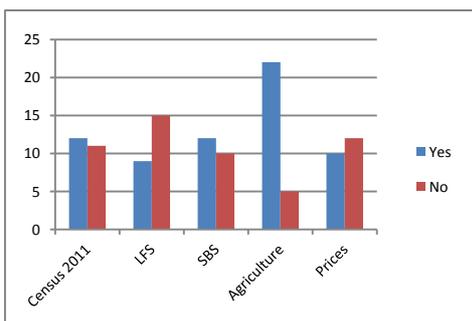


Figure A5: Do you cooperate with other statistical domains, statistical offices or national authorities in your domain?

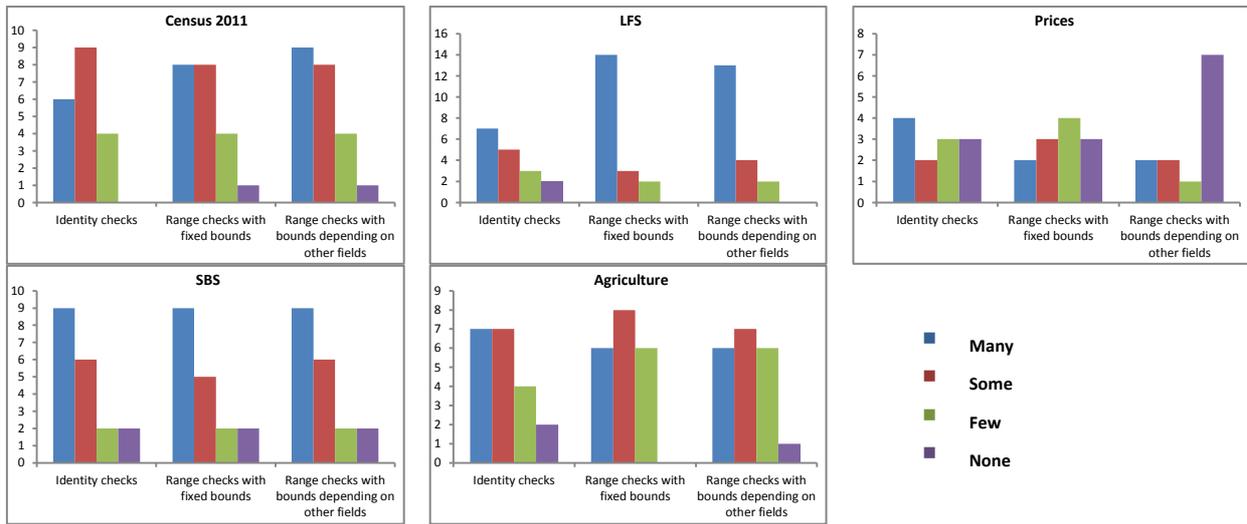


Figure A6: How many of those logical checks are checked in combination with a condition ? [Check 1 box/row]

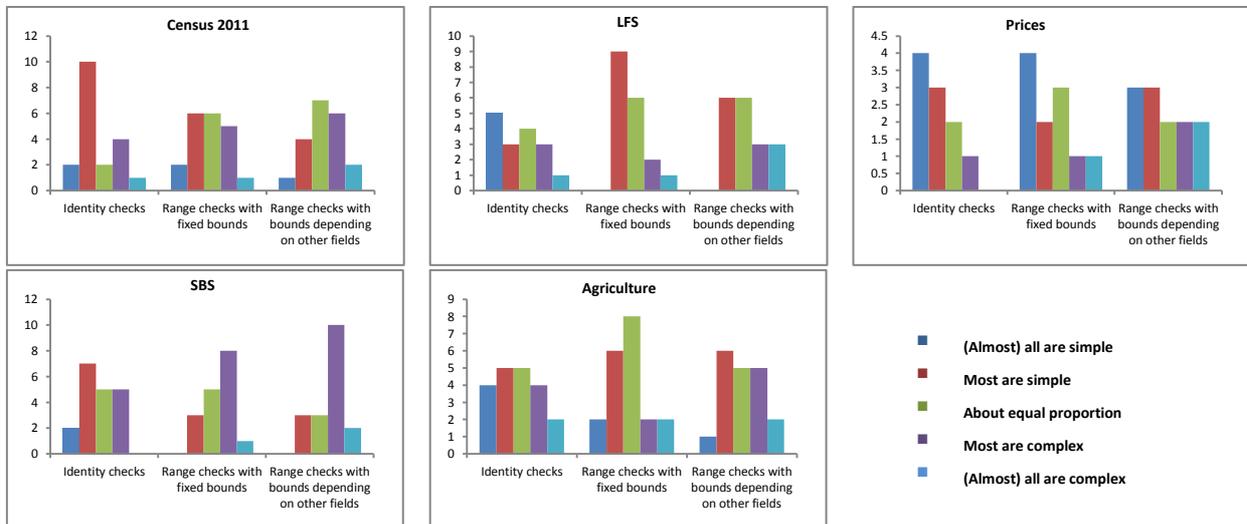


Figure A7: How big is the share of complex rules, combining more than one field (i.e. sums, differences, ratios,...)? [Check 1 box/row]