Subjects submitted by Insee to the 2018 European Conference on Quality in Official Statistics

Service statistique concerné	Participants	Sujet		Prise en charge DMCSI
		Thème de la conférence	Titre de l'intervention / article	
DMCSI – Unité Qualité	Marie-Hélène Kérouanton	Quality measurement, reporting and metadata	**The French statistical metadata repository, RMéS: managing metadata throughout the whole statistical process**	oui
DMCSI – Unité Qualité	Thomas Dubois	Innovation in data collection and dissemination	**Specification and immediate visualization of a questionnaire – a metadata-driven approach**	oui
DMCSI – Unité Qualité	Christine Fluxa	Coordination and cooperation in the national statistical systems, legal aspects **OU** Special Session: The revised European statistics Code of Practice – challenges and opportunities	**Quality guidelines as a tool for ensuring the coordination of quality in the French statistical system**	oui
DMCSI – Département des méthodes statistiques Division des méthodes et des référentiels géographiques	Vincent Loonis	Use of geospatial information to increase quality	**Improving the quality with spatial sampling**	oui
DSDS – Unité des études démographiques et sociales	Laurence Rioux Kévin Schmitt Mickaël Sicsic	Flash estimates	**Nowcasting of the poverty rate using microsimulation : estimations based on French data**	oui
Centre de formation de l'Insee de Libourne	Xavier Helfenstein	Innovation and quality culture in a statistical authority	**Sensibilization on quality at the training center of Insee**	non
SSM – Enseignement supérieur et de la recherche (Siès)	Thomas Balcone	Quality in innovation and research and development statistics	**Profiling : a new way to increase the quality of statistics on research and development**	non

The French statistical metadata repository, RMéS: managing metadata throughout the whole statistical process – *Marie-Hélène Kérouanton – Insee, Quality Unit*

Insee has had a long experience in metadata. Twenty years ago, Insee developed a software application for a system of structured documentation called DDS (Data structured documentation) . Nowadays, DDS is quite outdated. Furthermore, the patterns used to structure metadata have not been stringent enough and don't conform to any standard. And finally, all the documentary elements are actually collected at each phase of the process, via different files. They are entered at the end of the different steps into different instances of the DDS, generating redundancies and inconsistencies.

Therefore, a new statistical metadata repository, called RMéS, is being set up.

RMéS relies on two repositories.

The one dedicated to the questionnaires, variables and their codification is called a Colectica Repository. Information is stored in an international format, the DDI format (Data Documentation Initiative).

The other one hosts all the other metadata, described in more appropriate models. These models are compliant with W3C standards and with the Single Integrated Metadata Structure as required by Eurostat for the quality reports. In this repository we store not only concepts and classifications but also the description of all our statistical sources and their quality report in a RDF format.

These two repositories are linked together.

At the same time, we are developing management interfaces and services to enable other applications to use the metadata. We also created an application to design questionnaires, called Pogues.

With this project, we aim to gather and share in a single application all the metadata in accordance with international standards. What is more innovative is the idea of using metadata throughout the whole process, from the analysis of the needs to the statistical results and the assessment, in order to develop metadata driven processes.

Specification and immediate visualization of a questionnaire – a metadata-driven approach – *Thomas Dubois – Insee, Quality Unit*

Insee has been developing during the last years a metadata-driven questionnaire generator code-named Eno. This tool takes as input a formal description of the questionnaire complying with the DDI standard and executes a completely automated chain of transformations to produce the current survey questionnaire (web and paper questionnaires). The questionnaires can be personalized and, depending on the output format, a variable amount of flow logic can be implemented. Thanks to the automation of the process, Eno refering to DDI standard allows traceability of the changes in the questionnaires and reinforces the soundness of the data collection process.

Currently, nearly 15 survey questionnaires were produced by Eno over the last two years, and this figure is expected to grow to more than 40 survey questionnaires by the year 2020. Future work will allow these numbers to grow significantly with Eno providing interviewer questionnaire.

An additional module to this generator is a questionnaire design user interface, code-named Pogues, that connects with the generation process. More specifically, Pogues produces the DDI description of the questionnaire which is then submitted to an embedded instance of Eno. Using this tool, a survey manager or questionnaire designer can specify his web questionnaire in a friendly way and visualize the generated result in one click.

The current version supports the main functionalities needed for business surveys, and the roadmap foresees the development of more complex logic flows that can for example be found in household surveys. Other future developments will enhance the possibilities to specify questionnaire controls that could be used during the data collection process as well as during the data editing and imputation processes.

Pogues is open source and natively internationalized, and in consequence can be used directly by other statistical agencies. Other national statistical institutes have already shown their interest in this software.

Quality guidelines as a tool for ensuring the coordination of quality in the French statistical system – *Christine Fluxa – Insee, Quality Unit*

In compliance with the revised European regulation 223/2009, the head of the French NSI must have indicators to ensure the quality of the most important statistics produced in the public statistical system. In view of this, at the beginning of the year 2017, the French Quality Unit established quality guidelines in collaboration with the ministerial statistical services.

These guidelines formalize the framework for the co-ordination of the statistical system in terms of quality. They must be applied to European statistics, in accordance with regulation 223. But they also concern the national statistics that have been described as « structuring ». These « structuring » statistics are defined as both highly expected by the users and extremely prejudicial to the services if they were to be of poor quality. Their production and dissemination need to be insured against risks.

These guidelines are based on five orientations concerning governance, development of quality skills within the statistical services, the planning and realization of quality approaches to statistical processes; the fulfillment of European commitments and the systematic integration of users' needs and satisfaction in the designing of statistical products and services.

This article firstly describes the French quality guidelines and how they are initiating the management and the integration of quality in statistical processes. Then, it presents the procedure that has been followed in order to support each ministerial statistical services in the implementation of the guidelines. The article lastly shows how this implementation is monitored by the French NSI and how it will help to fulfill our commitments resulting from the 2014 peer review and to co-ordinate quality issues among the French statistical system.

Improving the quality with spatial sampling – *Vincent Loonis – Insee, Department of Statistical Methods*

Over the last years, many initiatives have been undertaken to help the National Statistical Institutes construct a fully geocoded information system. Such a point-based system is precisely the starting point of a handbook of spatial statistics being written by Insee and to be released by the end of June 2018. The latter, funded by Eurostat, will draw a list of statistical methods that rely on the availabilty of the (x,y) coordinates of the statistical units. Ranging from measures of spatial autocorrelation to spatial econometrics for panel data, these methods might be helpful to improve the production, the dissemination or the analysis of statistical results. The issue of spatial sampling fully falls within the scope of the handbook, and as such will be more precisely dealt. The presentation at the conference will focus on the issue of spatial sampling. It aims at proving that a geocoded sampling frame might help better carry out surveys. On the one hand, knowing the position of the statistical units help better organise the field work for face-to-face surveys. This can be done, for instance, thanks to Primary Units having very good spatial features. On the other hand, at the selection level, the sampling design might better spread the selected units over the territory. This strategy can be efficient to improve the precision estimation for high spatial correlated variables.

Nowcasting of the poverty rate using microsimulation : estimations based on French data - *Kevin Schmitt, Michaël Sicsic – Insee, Demographic and Social Studies Unit*

Every year in September N+2, INSEE publishes the poverty rate and the main indicators of inequalities in standard of living for year N. This delay is unsatisfactory for meeting the social requirements of users of these indicators. Of the 21 months between the end of the year under consideration and the publication of the poverty rate, about three-quarters of this time is taken up collecting tax and social data, and about one quarter with statistically matching Labour Force Survey (LFS) data, from which the Tax and Social Incomes Survey (ERFS) is produced. Nowcasting consists of producing an earlier indicator of the poverty rate for the target year N (in autumn N+1) based on the ERFS N-1. The method to be used here is microsimulation, which creates individuals' standard of living by imputing benefits and contributions on scales, and thus it is possible to take account of any legal changes made to these measures. The exercise is based on the INES model, which simulates the majority of French social security and tax legislation, based on any year of the ERFS. To implement nowcasting, one important step is ageing population by uprating incomes (using surveys about wages, aggregated tax data, inflation...) and calibration weighting (using margins from LFS and census). Reverse ageing is also used so that evaluations for year N and N-1 (and thus annual evolutions) are only based on the ERFS N-1 (that is minimising the sample bias). In this paper, we present the methodology and assess the quality of the early indicators thus produced. Indeed, we compare the results that would have been produced by microsimulation with those that were in fact disseminated from the ERFS. When applied to the target years 2010 to 2015, this method produced estimations similar to the actual figures published the following year.

Sensibilization on quality at the training center of Insee - *Xavier Helfenstein – Insee, training center in Libourne*

In the training center of Insee (Cefil), the culture of quality is disseminated to the futur staff of the NSI through a process of acculturation. Quality is omnipresent in all the courses and is presented to the trainees in various dimensions going from methodological to ethical dimension.

Different principles of quality are approached through study projects. For example, the project of concerted analysis of statistical tables demonstrates the necessity for the statistician to present the results of his works in a clear and understandable way. By organizing a forum, the skill transmission project emphasizes the importance of pedagogy. The statistical survey project places the trainees in situation to build a collection of data from scratch in order to answer to the request of a public actor. This exercise requires to use every skills of a collective, to meet the deadlines, to restore results in compliance with the statistical secret and to document the data. It teaches the trainees how to manage the impact of a data processing on the final result.

The Cefil also offers a classic sequence on quality to the trainees but the education of the training center wouldn't be efficient enough to the objective of professionalization if it dispensed only this sequence. Indeed, the support of quality is observed to be more induced by a daily behaviour than by a knowledge of an academic subject. Operating in project mode takes there all its interest. The approach of quality is thereby embedded in a holistic conception of the educational route of whom it is a component.

Profiling : a new way to increase the quality of statistics on research and development - *Thomas Balcone – Statistical Department of the Ministry of Higher Education and Research Service (Siès)*

Currently, statistics on Research and Development (R&D) carried out in the business sector are computed in France on the sole basis of legal units : firstly, a survey is addressed to them to collect the data and then, statistics on R&D are disseminated at legal unit level. Considering the increasing importance of the enterprise group in the French economy, it seems difficult today to go on using only the legal units to calculate business statistics. Indeed, assimilitating the legal unit to the enterprise is not relevant anymore for group's affiliates and subsidiaries. Taking into account the European definition of an enterprise will help to disseminate more consistent and relevant R&D statistics on the business sector.

The French business statistic register established by the French national statistical institute (INSEE), called SIRUS, contains notably all the legal units, all the enterprises and all the links between them. The main contribution of this register is to make possible the calculation and dissemination of statistics at an other level than the legal unit one : the enterprise level.

This article first describes why the data should go on being collected at legal unit level and not at enterprise one. Indeed, it seems that such a change in the data collection can be dangerous because it could result in a substantial increase of the response burden. Then, this article presents the process based on SIRUS that leads to the computation of key indicators on R&D at enterprise level. To conclude, it compares these key indicators with the ones calculated at the legal unit level to show the impact of moving to the enterprise level on French R&D statistics.