

Literacy in statistics for the public discourse

Walter J. Radermacher

Sapienza Università di Roma, Roma, Italia

E-mail: wjr@outlook.de

Abstract. If one assesses the quality of statistics according to whether they are fit for purpose, one must put the question of this very purpose at the beginning, not only for the production of statistics, but also for their use and the literacy required. In this contribution, public discourse, political communication and decision-making processes are placed at the beginning of the argument.

Official statistics work with a conceptual approach in which very much emphasis is placed on standardisation of products and processes, thus ensuring comparability of facts in regional and temporal terms. Only in this way can statistics be used as a common language to objectify conflicting issues. It is not about everyone being able to create his or her own statistical model of reality, quite the opposite. In this sense, public statistics are an infrastructure, comparable to rail transport. Moreover, statistical processes today are highly rationalised and industrialised, comparable to a factory. So, when approaching the question of what literacy is needed in this particular application area of statistics, the education and training of professional statisticians has accordingly specific requirements, which are in many ways comparable to what professions in other industries or in the operation of the railways should have as a basic qualification. For the citizen, the entrepreneur, the teacher, the student, etc., who wants to understand and apply the indicators of the public statistical sources, technical skills are of secondary importance. Rather, it is necessary to understand enough about the product and its properties to be able to judge its quality in the light of personal application goals and questions. This indeed already presupposes a lot of knowledge and experience in dealing with quantitative information. Such competencies do not necessarily belong to the field of mathematics but demand practice in interpreting indicators in their context, an assessment of the reliability of sources and processes, experience with graphical representations of statistics (including the flaws that may appear in them) and practice in assessing uncertainties, etc. One needs to know a certain amount about the data sources, the preparation processes, quality standards, etc., but not at the level that would be required if one were to carry out this work oneself.

Keywords: Data literacy, statistical literacy, official statistics, public discourse

1. Data and statistical literacy

Data literacy is one of the core competencies of our time! It doesn't take much to agree on such a statement. But what do we mean by this? A summary of the methodological work done to answer this question and the approach to define a corresponding general framing can be found in Katharina Schüller's report: "*Such a competence framework should map all stages of the knowledge... creation process from data; it should cover all dimensions of competence: (a) knowledge, (b) aptitudes, (c) skills, (d) motivation and attitude [1].*"

This framework is used in the following as a starting point to elaborate more specifically how literacy is to be understood in the context of public¹ statistics. This

is about creating the best possible conditions for good quality statistics to be designed, produced, communicated, interpreted and used, i.e. literacy at every stage and in every process. The goal is to achieve high quality statistics, for which education is an essential tool, an important prerequisite.

Before we continue, we will address (only very) briefly the question of whether it should be called 'data' literacy or 'statistical' literacy. This question quickly leads to the relationship between data science and statistics, about what is actually a subcategory of what, and so on. While it would probably be helpful and appropriate (also in terms of literacy) to place greater empha-

ducers and differ in their quality profile. Official statistics have a special role in this and are most closely committed to complying with codified quality standards.

¹Public statistics may be offered by different public sector pro-

sis on more sophisticated terminology,² the next sections proceed on the assumption that we are dealing with (more or less) the same subject area, regardless of whether we title it ‘data’ or ‘statistics’.

2. Public statistics – literacy for professionals

One could misdirect the answer to what constitutes solid preparation and training for a professional statistician with three very popular, nevertheless false hypotheses:

1. The most important part of the preparation for a professional statistician is a good knowledge of mathematical basics, data science and informatics.
2. The requirements for professionals are very different in the various branches, departments and sectors of a statistical institution (surveys, accounts, IT, etc.), so that there is neither a need nor would it be cost-effective to provide a common preliminary training.
3. In statistics, too, everyone should change fields at regular intervals, so specialisation makes no sense.

However, it is worth mentioning these hypotheses because all three are surprisingly common and are used by their respective proponents depending on their perspectives and interests. Our aim here is to get beyond these mutually exclusive positions in order to arrive at an overall view that, on the one hand, contains enough of what a statistician needs to know in order to understand his or her business (the ‘factory’) and to be able to situate his or her professional role within it. On the other hand, it has to be acknowledged that appropriate professional knowledge and skills sometimes require quite a long practical experience.

2.1. What a good mix of topics could look like?

The following is obviously not about the (highly) specialised part of professional education. For this part, a corresponding academic education is required plus further training ‘on the job’ as well as corresponding specialised courses. Rather, it is about what a good mix of topics, skills and knowledge could and should

look like, which the candidate statistician would have to compose as a basic prerequisite and preparation for employment in statistical organisations. One factor here is that the workflows, projects and tasks in a large organisation such as a national statistics institute are not (or only to a very limited extent) comparable with those in research, at universities or similar.

For this reason, the conceptual principles developed by W.E. Deming for management in industrial production are applicable to official statistics. In his later book, “*The New Economics*” [3], he developed his “*System of Profound Knowledge*” with four parts, all related to each other.³ These four parts are

- Appreciation for a system: understanding the overall processes involving suppliers, producers and customers (or recipients) of goods and services, understanding how interactions (i.e. feedback) between the elements of a system can result in internal restrictions that force the system to behave as a single organism that automatically seeks a steady state. It is this steady state that determines the output of the system rather than the individual elements. . . .
- Knowledge about variation: the range and causes of variation in quality and use of statistical sampling in measurements. In any business, there are always variations – between people, in output, in service and in product. . . .
- Theory of knowledge: the concepts explaining knowledge and the limits of what can be known. How do we know that what we think we know is really so? There is no true value of any characteristic, state or condition that is defined in terms of measurement or observation. . . .
- Knowledge of psychology: concepts of human nature. An organisation has a duty to create a system where people can take pride in what they do. . . .

The approach used in this paper follows Deming’s principles. It is all about profound knowledge for professionals working in official statistics. Obviously, there are some common basics that are vital to know and to teach as part of literacy. The most fundamental part is knowing how official statistics are organised and produced and what the brand identity is, i.e. the ‘DNA’ of official statistics [4]. The point is further to understand the possibilities and limits of measurement and knowledge creation, and what Deming calls ‘psychology’ in our case, accordingly, the sociology of the interactions between statistics and society.

²See the broader segmentation in [2] Radermacher WJ. Governing-by-the-numbers – Reflections on the future of official statistics in a digital and globalised society. Statistical Journal of the IAOS. 2019; 35: 519–37.

³See Deming’s System of Profound Knowledge in <https://blog.deming.org/2012/10/demingssystem-of-profound-knowledge/>.

2.2. Understanding the statistical factory and the individual statistician's role in it

In concrete terms, however, what knowledge and previous training is useful for understanding the statistical factory and its various interacting activities and organisational units? To begin with, the distinction introduced by A. Desrosières regarding the different stages of compiling statistics is followed, who separates three “aspects of statistics, (1) that of quantification properly speaking, the making of numbers, (2) that of the uses of numbers as variables, and finally, (3) the prospective inscription of variables in more complex constructions, models” [5]. Following Desrosières’ categorisation, a common denominator of statistics education should include components from all these aspects, in other words an overview of the operation and important methodologies of survey statistics as well as an overview of the functioning and structure of national accounts or other model calculations and the specific requirements for statistical indicators [6]. This kind of general overview is a prerequisite for the individual to be able to place his or her role and work in the larger context and understand that official statistics embodies a comprehensive system that is more than its individual parts, processes or products. Some essential elements from the methodology of survey statistics, the representativeness of samples or the interaction of surveys and external, already existing data, should also be considered in the basic training. The same applies to the basic methodologies and theories of national accounts. Furthermore, it should contain an understanding of the business architecture, for example, by teaching the logic and structure of the Generic Statistical Business Process Model⁴ or similar modern standards.⁵ It also should include basic training in topics of information quality, quality management and communication. A difficult but important subject is that of defining the boundaries and scope of official statistics. Especially those areas where modelling is used to increase the relevance of information, for example by estimating preliminary results or by standardisation, filtering (e.g. seasonal adjustment) or interpolation, should be elementary components of the basic training to become a professional statistician. In particular, the shifting of relevant boundaries in historical retrospect also reveals what is actually at the core of public sector statistics: the common elaboration,

understanding and agreement on conventions or even standards that fix methods and procedures for a certain period of time before they are modified in a process of revision.

2.3. The experience of EMOS in training official statisticians

With the aim of promoting such broad-based training and preparation for employment in official statistics, the ‘European Master in Official Statistics EMOS’ programme was launched in European Statistics. “EMOS is a label awarded by the European Statistical System Committee. . . . the EMOS network comprises 32 programmes in 18 countries and collaborating partners in statistical offices. The network builds on existing and nationally accredited Master programmes which, in line with the EMOS learning outcomes,⁶ familiarise the graduates with the system of official statistics, production models, statistical methods and dissemination. EMOS labelled Master programmes also collaborate actively with the National Statistical Institutes or other producers of official statistics for relevant master thesis topics and internships in the sphere of official statistics.”⁷

EMOS impressively demonstrates that comprehensive scientific preparation of professionals for official statistics is possible when academic and applied training are combined and well synchronised. However, the experience of the first years has also shown the difficulties: a relatively small number of interested students per participating university, capacity problems at the partner statistical institutes, overemphasis on survey statistics versus a lack of accounting and indicators, to name just a few of the issues. Nevertheless, it is important to learn from the experience and draw appropriate conclusions for the coming periods of the EMOS programme. In this context, it will be necessary to take into account newer methods of data science and the corresponding data sources, as well as the interaction with citizens in the sense of citizen sciences, which are both becoming increasingly important for official statistics.

However, it might prove more difficult in practice to adequately fill other gaps, such as those especially concerning accounts, indicators or the epistemology of statistics. Here it would be necessary to involve representatives of academic disciplines other than statistics,

⁴<https://statswiki.unece.org/display/GSBPM/Generic+Statistical+Business+Process+Model>.

⁵<https://statswiki.unece.org/display/GSBPM/GeoGSBPM>.

⁶https://ec.europa.eu/eurostat/cros/system/files/emos_learning_outcomes_2019_adapted.pdf.

⁷https://ec.europa.eu/eurostat/cros/content/emos-explained_en.

such as macroeconomics or sociology, and to actively involve them in the further development of the course structure. As far as the attractiveness of this academic training programme is concerned, it may also be opportune to offer individual courses from it for those students who are not interested in the whole package and its degree. Finally, what is at stake is the (personnel) capacities at the partnered statistical institutions; without a permanent and appropriate allocation of resources, this education programme will not be able to flourish.

3. Public statistics – users and literacy

The fact that initiatives to improve data literacy are gaining momentum, supported not only by business but also by politics and science, is extremely urgent and very welcome. Data literacy serves to promote maturity in a modern digitalised world and is important for all people – not only for specialists. As has already been pointed out, statistical education, like other education, is about several dimensions of competence: knowledge, skills and values [1].

However, such a broad and balanced approach is rarely found in practice. Rather, one gets the impression that a great deal of attention is paid to conveying technical skills, as the following examples illustrate:

- *“The need for data analysis skills grew by 86% from 2013 to 2018. Become data-driven with Google Cloud. Leverage data and gain real-time insights that improve your decision-making and accelerate innovation. Learn how to design and build data processing systems.”*⁸
- If one searches further in the Applied Digital Skills for Education for ‘Statistics’, you get one hit, namely *“Calculate Probability with Google Sheets”*.⁹

One is reminded of the Do-it-Yourself wave of the 1970s, which propagated screwing, repairing and constructing by anyone, sometimes certainly in cases where a good craftsman would have done the job better, faster and cheaper than an amateur.

For the citizen, the entrepreneur, the teacher, the student, etc., who wants to understand and apply the indicators of the public statistical sources, these skills are of secondary importance. If you want to buy fur-

niture for your flat, it wouldn’t make much sense to learn carpentry beforehand. Rather, it is important to understand enough about the product and its properties to be able to judge its quality in the light of personal application goals and questions. This indeed already requires a lot of knowledge and experience in dealing with quantitative information. Such competencies do not necessarily belong to the field of mathematics but demand practice in interpreting indicators in their context, an assessment of the reliability of sources and processes, experience with graphical representations of statistics (including the flaws that may appear in them) and practice in assessing uncertainties, etc.

3.1. Knowledge needed to assess the quality of the statistical product

The learning objective in this field of public statistics might become clear by comparison with the goal of healthy nutrition, where we would want people to eat (and live) better and promote appropriate education and culture for this. If I want to brew a good quality coffee, fairly and sustainably produced, I need a lot of knowledge about coffee as a raw material, the production process and value chain, the different brands and products (coffees and machines) on the market and finally the skills to handle my machine and the bought beans properly (which depends on my taste). Should I have a special interest in good coffee, I can learn a lot in so-called ‘Barista’ courses.¹⁰ What I get for myself in terms of quality and coffee enjoyment, depends on my skills and knowledge (and values, such as my willingness to pay for good quality). What I don’t need, however, is the skill of growing coffee beans, roasting them, etc.; there are trained specialists for that. In short, one could compare training for the consumers of public statistics with such a Barista course. For statistical information, too, one needs to know a certain amount about the data sources, the preparation processes, quality standards, etc., but not at the level that would be required if one were to carry out this work oneself. In a value chain, it is not necessary to have the skills and abilities of the upstream process steps in detail. It is sufficient to be able to assess (and use) the delivered product and its quality.

There has been ample opportunity in the pandemic crisis of recent months to observe the lack of success-

⁸Google Cloud Training “Smart analytics and data management” <https://cloud.google.com/training/data-ml>.

⁹<https://applieddigitalskills.withgoogle.com/c/en/curriculum.html>.

¹⁰See for example <https://www.baristatrainingacademy.net/barista-training-beginners-become-professional-barista/>.

ful communication between data experts, policy makers and citizens [7]. Another prominent case is already somewhat older, but similarly revealing: In 2002, the euro was introduced as the currency in many European countries. Subsequently, there was massive criticism and mistrust against the Consumer Price Index because it did not correspond to the inflation perceived by customers in their daily shopping, where they observed incorrect conversion of currency to prices of services and consumer goods [8]. For someone to understand what the consumer price index indicates (or does not), courses in data analytics or in-depth knowledge of probabilities do not provide much help. Rather, some understanding is required of descriptive statistics, what an arithmetic mean is with its weighting scheme, how the basket of goods is composed, how price changes of representative commodities are determined. It is of great advantage to have gained some experience with these empirical methods, in our example by changing the weighting scheme and observing the corresponding effects, by comparing the consumer price index with another type of price comparison (e.g. between the cost of living in different regions). In this sense, it could also be helpful to involve citizens more in the production of statistics. Such co-production could – besides possible provision of data sources – contribute to bridging the gap between producers and users of statistical indicators of this kind [9].

3.2. *Make-or-Buy*

The objective and tasks of data literacy initiatives are somewhat specific for managers and employees in public administrations, e.g. of municipalities. Under the impression of increased digitalisation requirements and corresponding expectations placed on them, employees may get the impression that a course in data analytics is nowadays part of their professional training programme if they want to be successful. Heads of administrative departments may feel called upon to set up their own data analytics units. Here, too, there is a danger of creating unrealistic incentives and misleading ambitions, organisationally as well as personally. If data analytics are propagated as part of the toolbox of modern administration, then the necessity of decisions between doing it oneself and cooperation with other parts of the same administration should also be pointed out, according to which a commissioning of a municipal statistics office could ensure the provision of adequate analysis and statistical monitoring for local politics. Wrong incentives will lead to the emergence

of parallel information worlds here as well, which are ineffective and inefficient and will tend to end-up in poor information quality.

4. FENStatS activities and the Data Literacy Charter

The Federation of European National Statistical Societies (FENStatS)¹¹ has started two activities in 2020, the first year of COVID-19, with different working groups related to statistics on the pandemic and to literacy. In the course of the work of the two groups, it became apparent that they have many overlaps in terms of content, so that it was decided to merge the two working groups. In the broadest sense, it is now about 'data informed decision making in the pandemic'.

In concrete terms, it has been achieved so far that the FENStatS website has been expanded by a category 'Guides and Resources',¹² where the working group provides material related to the pandemic, at national, European and international levels. In this way, the website facilitates access to relevant information to the public, organised in different packages (data, report, study, dashboard and a final one with tools for self-exploring the data available). What is currently being prepared is the creation of online learning units on the above-mentioned topic. The units are based on a data literacy framework, a jointly developed overall structure for the online course and a glossary. It is supported by the AI Campus in Germany.

Finally, FENStatS supports political initiatives to improve data literacy, such as the 'Data Literacy Charter'¹³ in Germany [10] or the 'Appeal for an urgent national data literacy campaign in Switzerland'¹⁴ or 'Advancing data literacy in the post-pandemic world'¹⁵ of Paris21, just to mention a few. FENStatS' Executive Committee has welcomed the Data Literacy Charter. It also encourages member societies to pay special attention to this Charter in order to customise it for their national purposes. Finally, it is considered beneficial and worthwhile to elaborate an international standard for literacy.

¹¹<https://fenstats.eu/>.

¹²<https://fenstats.eu/guides-resources>.

¹³<https://www.stifterverband.org/sites/default/files/data-literacy-charter.pdf>.

¹⁴<https://en.data-literacy.ch/>.

¹⁵https://paris21.org/sites/default/files/inline-files/DataLiteracy_Primer_0.pdf.

References

- [1] Schüller K. Future Skills: a Framework for Data Literacy. Working Paper No. 53. Berlin: Hochschulforum Digitalisierung; 2020 July 2020.
- [2] Radermacher WJ. Governing-by-the-numbers – reflections on the future of official statistics in a digital and globalised society. *Statistical Journal of the IAOS*. 2019; 35: 519–37.
- [3] Deming WE. *The New Economics: For Industry, Government, Education*. Second Edition ed. Cambridge MA: Massachusetts Institute of Technology, Center for advanced engineering study; 1994.
- [4] Radermacher WJ. *Official Statistics 4.0 – Verified Facts for People in the 21st Century*. Heidelberg: Springer Nature Switzerland AG; imprint Springer; 2020.
- [5] Desrosières A. A Politics of Knowledge-tools – The Case of Statistics. In: Sangolt L, ed. *Between Enlightenment and Disaster*. Brussels: P.I.E. Peter Lang; 2010.
- [6] Radermacher WJ. Guidelines on indicator methodology: a mission impossible? *Statistical Journal of the IAOS*. 2021; 37: 205–17.
- [7] Radermacher WJ. Governing-by-the numbers – Résumé after one and a half years. *Statistical Journal of the IAOS*. 2021; 37(2): 701–11.
- [8] Brachinger HW. Der euro als teuro? Die wahrgenommene inflation in deutschland. *Wirtschaft und Statistik*. 2005; 2005(9).
- [9] König Ariane. Can citizen science complement official data sources that serve as evidence-base for policies and practice to improve water quality? *Statistical Journal of the IAOS*. 2021; 37(1).
- [10] Schüller K, Koch H, Rampelt F. *Data Literacy Charter*. Berlin: Stifterverband; 2021. p. <https://www.stifterverband.org/sites/default/files/data-literacy-charter.pdf>.