The role of communication in Statistical science and the strategies of communication for statistics users

ESAC Doc. 2018/28

The role of communication in Statistical Science and the strategies of communication for statistics users

David J. Hand[1] and Maurizio Vichi[2]
[1]Imperial College, London
[2]Sapienza University, Rome

1. Communication: Modes and Methods.

The word "communication", according to the Oxford dictionary, derives from the Latin "*communicare*" ("*to share*"), and has the meaning of imparting or exchanging of information by speaking, writing, or using some other medium. Communication can be "verbal" (or "non verbal"), meaning that it describes (or not) the processes of conveying information in the form of linguistic (or not-linguistic) representations. Interpersonal communication is in particular how people interact and exchange information and concepts. The general idea is that there is at least a "sender" and a "receiver" that exchange information when they communicate.

The modern age has given rise to an extraordinary velocity, volume and variety of communication. Moreover, modes and methods of communication have changed dramatically in recent decades. In the previous century, communication between people was mainly via face-to-face conversations based exclusively on verbal cues and interactions strengthened by gestures and the tone of voice (non-verbal cues) and telephone, and by mail (old-fashioned letters). In many situations this basic form of communication, where humans "transmit" information and "receive" an instant response, has become less frequent, while the rise of technology and the increasing use of the internet have developed a new form of Computer-Mediated-Communication (CMC) (Herring, 1996). In CMC, any human communication occurs through two or more smart devices having specific formats, such as instant messaging, email, chat rooms, online forums, social network services and text messaging. CMC represents an evolution of radio and telephone communication facilitated by the portability of the new devices and the ability to access the internet from everywhere. CMC can occur into two forms: synchronous and asynchronous. The first happens in real time with the communicators being simultaneously engaged, although not necessarily in the same location, such as in phone calls, video chats and video-audio calls. Asynchronous CMC take places when parties are not communicating at the same time and the sender does not receive an instant reply from the receiver, such as in text or voice messages and emails.

Compared to basic face-to-face communication, there are some advantages and disadvantages in using this modern mode of communication. Pros are that CMC is not limited by time and place, it can take place anytime and anywhere. Moreover, people can be available for communication at different times and in different places. CMC also allows people to reach out to a large number of other persons simultaneously. It also facilitates the storage of records of previous communications. Stored records can be used for a variety of diverse purposes, ranging as widely as tracing potential terrorists (e.g. Paganini, 2015) and studying communication strategies (e.g. Hand, 1986) Finally, CMC helps in breaking down the barriers of communication caused by reasons such as shyness, appearance or physical limitations.

However, compared to the basic face-to-face communication, there are some disadvantages: CMC limits the richness of communication since it is lacking in terms of emotional and non-verbal cues. It increases the risk of loss of confidentiality and trust. In fact, different people can successively use the same communication between

other persons with different purposes. Finally, CMC is limited by technology. Computers might be down; the connection might be poor, etc..

Having described the new modes and methods of communication, let us now explore if and how these are changing Statistical Science, especially in official statistics.

In official statistics, communication really begins with messages from the user to the producer, describing the information (indicators, summaries, etc) that they need. This involves an extensive exercise in user engagement. Moreover, it is not a one-off exercise, but rather one that continues and is refined as data collection and analysis exercises evolve.

Beyond the communication of needs, communication has a central and fundamental role in two parts of the discipline. First, it is the fundamental channel between people (interviewers and respondents), necessary for data collection. In fact, communication taking the form of one or more questions that interviewers submit to the respondents represents a classical form of data collection. Second, communication becomes the process of information transmission, between producers and users of statistics, which happens at the end of the statistics production in order to transfer the knowledge on the phenomenon under study, with qualitative and quantitative data that synthesise it. Thus, communication is essential at least at the beginning and end of the production process of statistics. In the next two sections we discuss the changes in communication and communication technologies that are modifying the discipline of Statistical Science, as well as how to take advantage from these new challenges.

2. New data collection processes due to the new methods and modes of communication

Face-to-face communication and CMC constitute two completely different statistical data collection methods. Classical data collection by interview, where an “interviewer” poses questions face-to-face to a “respondent”, is now integrated into CMC methods in situations where interviewers fill in questionnaires answered by respondents using a smart device (e.g. tablet or computer) with computer-assisted personal interviewing (CAPI) software and computer-assisted telephone interviewing (CATI) software. These methodologies have some disadvantages: they are quite expensive relative to the costs of the interviewers, and they may produce selection and exclusion biases (e.g., young vs older people or groups without telephone or affected by a digital divide).

Respondents can also self-complete questionnaires as in computer-assisted web interviewing (CAWI) software or in email surveys. This is certainly less expensive; however, although a high nonresponse rate to questionnaires is frequently observed. There has been substantial work, over many decades, in the related field of Computer Assisted Learning (Francis, 2016), and it is likely that the ideas and lessons from this area could usefully contribute to developing effective CAWI systems.

CAPI, CATI and CAWI have the merits that they directly store the data in a database ready for analysis and they allow data pre-processing for detecting errors directly during the data collection. The mixed mode data collection strategy experimented with by several National Statistical Institutes (e.g., Statistics Netherlands, Statistics Norway) suggests that the sequential use of different methods of data collection can produce better results in terms of lower costs, superior timeliness, larger coverage and enhanced response rate. The use of a mixture of CAPI, CATI and CAWI suggests the adoption of a new intelligent CAWI method with a virtual agent similar to the proposal of Conrad et al., (2015), to include in the self-complete data collection virtual personal features.

> *Recommendation* 1: Increase the use of CAWI data collection, by using advanced software. NSIs and Eurostat need to invest with European projects for “intelligent” CAWI software that includes a (voice) virtual assistant (similar to Apple SIRI, Google Assistant, Microsoft Cortana, etc) that would help

respondents, reproducing questions with a virtual voice. In particular, they should be able to explain the meaning of questions and how to answer them, and the respondent should be able to request information about the questions and receive explanations. The software should help to reduce errors linked to data collection, identifying logical inconsistencies between responses within a questionnaire, verifying eventual anomalous answers, and so on, directly during the data collection. This is also possible with CAPI, although this is more expensive ♦

Recommendation 2: There is a clear burden of data collection for the citizens that are contacted to answer questions, as well as to companies and other organisations, such as banks. This burden should be controlled by arranging that the minimum numbers of requests go to the same citizens or organisation. Centralised CAWI software with the characteristics described above, if not already proposed by NSIs, could verify how many times a person has been contacted for data collection. Some commercial bodies have developed tools for controlling respondent burden by limiting repeat requests to individual IP addresses (e.g. Qriously, which conducts surveys via mobile phone apps). ♦

Computer-assisted data collection methodologies do not take full advantage of the modern computer mediated communication available on the internet. In particular, data collected using CATI, CAPI and CAWI is “structured”, as is in the form of a set of questions and their corresponding answers, packaged into questionnaires - exactly as was classically the case in the pre-internet and pre-computer age. Such data collections are in some way already old fashioned.

Communication via the internet is developing a different form. The use of mobile phones, credit cards, customer-loyalty cards, web browsing, leave behind digital traces that represent information about people. In addition, social media platforms tend to broadcast continuously and show communication among people on all possible topics and phenomena relating to their lives. These data are "unstructured"; that is, phenomena to be observed are frequently mixed and they are discussed by people with data in the form of texts, videos, vocal messages and pictures. The information exchanged among people passes directly and continuously with data being recorded on web platforms. Thus, the new challenge for Statistical Science is to intercept these data, relate them to each other, and analyse them to usefully extract knowledge which effectively describes reality. This is a tough challenge, requiring sophisticated statistical tools, and probably even novel statistical concepts and methods - see, for example, the discussion of “smart statistics” in Vichi and Hand (2018). Intelligent web scraping software is needed: that is, automated www-robots navigating the web continuously, copying digital data exchanged in the communications and transforming these unstructured data into a structured form for subsequent statistical analysis and knowledge extraction.

In this scenario, in the near future, data collection will be less the result of purposive interviews and questionnaires and more the direct incidental observation of the data from many different sources flowing through the web.

Furthermore, the advent of the Internet of things (i.e., the network of “intelligent” physical devices, vehicles, home appliances and other items embedded with electronics, software, sensors) means that there will also be communication between smart devices, not merely among people. Smart devices will be able to automatically answer requests made to their owners without posing questions to those owners themselves. The data collected by different sensors used in a wide variety of locations (e.g., access badges, supermarket loyalty cards, smart security cameras, travel cards, etc.), will help to directly reconstruct the lives of people without having the need to interview them. These data are not produced for statistical purposes but they have huge potential for use in this mode. The reuse of administrative data for statistical purposes would complete the information on single individuals.

The opposite side of the considerable potential for good arising from the understanding of population behaviour presented by these new data technologies, is the potential for invasion of privacy. For this reason, the EU has introduced the General Data Protection Regulation (GDPR), controlling how personal data of all individuals within the European Union may be stored and used. The GDPR was adopted in 2016 and becomes enforceable on 25 May 2018. According to the GDPR personal data is any information relating to an individual, whether it relates to his or her private, professional or public life. It can be anything from a name, a home address, a photo, an email address, bank details, posts on social networking websites, medical information, or a computer's IP address.

The GDPR permits the reuse of personal data by National Statistical Institutes (NSIs). Article 89 gives appropriate safeguards and derogations to the GDPR for processing for archiving purposes in the public interest, for scientific or historical research purposes, or for statistical purposes, subject to appropriate safeguards.

Thus, National Statistical Institutes can become the "Official National Libraries of digital data" relating to individuals and things. NSIs would have to strictly guarantee the protection and privacy of data using appropriate cybersecurity tools. Individuals should be the unique owners of their data, insofar as "ownership" can be defined, but the NSIs should have the possibility to reuse these data for statistical and scientific purposes, adopting the general data protection and privacy regulation for their treatment.

As with current internet companies, this would represent an exchange, with citizens offering information and receiving services - new jobs, personalized medicine (network medicine), general opportunities, and so on - based on that information.

Recommendation 3: National Statistical Institutes should develop more projects in collaboration with Universities to allow and facilitate this new kind of data collection. Methodologically there is a need to intensify studies on machine (statistical) learning on text analysis and other unstructured and heterogeneous data. The process of transforming unstructured data into structured data ready for analysis with statistical tools has to be fully automated if the full promise is to be attained. ♦

Recommendation 4. National Statistical Institutes should promote at the European level a standard to make the data produced by smart devices linkable, by introducing standards. Universities and NSIs should develop projects to create methodologies and software for data linkage, data fusion, etc. ♦

3. Communication for transmitting knowledge at the end of the production process

We noted in section 1 that communication also has great importance at the end of the production of statistics, when the knowledge formed by qualitative and quantitative data synthesising the phenomenon under study has to be transferred by the producers of statistics to the users.

Communicating statistics involves three fundamental aspects (Maggino Trapani, 2009): (i) using correct and accurate contents, (ii) using appropriate tools, (iii) allowing the message transmitted by the producer to be easily received by a user. The use of correct and accurate content is an aspect of the quality of statistics, and we shall suppose that the quality is appropriate for the purpose with which the statistics are produced. Moreover, we take for granted that the appropriate tools are used in the statistical production, following the NSI's best statistical practices.

In order to allow the message to be easily understood by stakeholders, the process of information transmission has to be commensurate with the characteristics of the users. When the user is an expert in the field of the phenomenon under study and has a good statistical literacy, the communication can be more complex and will generally be different from communication offered to citizens who are not expert, who need simpler messages. In particular, laypeople may be uneasy with numbers and statistical figures.

This means that the first step is to establish a good classification of the users of statistics, ranging from inexpert citizens to expert statisticians, so that the most appropriate style of communication can be chosen for each segment of the population.

3.1 The classification of the European citizens as users of Statistics

Statistical Science is applied in many different fields and has an extraordinary ubiquity in the modern life of the citizens. They may not recognize that the debate around climate change and global warming is primarily based on statistical information, that the hedge fund industry is based on statistical models and analysis, that statistics plays a central role in modern drug development, that epidemiological issues such as severe acute respiratory syndrome, bovine spongiform encephalopathy, acquired immune deficiency syndrome and avian flu are fundamentally statistical, and so on (Hand, 2009). Any citizen is at least an involuntary user of data and statistics.

A classification of the users of statistics has been given in Vichi, Valente Rosa and Ruane, (2015). It divides users into two major categories: Institutional Users and Non-Institutional Users, with each set of users differentiated according to their interests. This distinction is a priori important for several reasons.

Institutional users generally need data for governmental and administrative decision making. Their data needs are considerable and usually have priority in terms of demands that are met by data producers. Furthermore, data for Institutional Users typically need to allow for comparison over time and space in order to verify the impact of decisions.

Non-Institutional users, for example journalists, have different needs, focused particularly on having statistics ready for communication and diffusion to wider audiences. For them it is important to be able to show new trends of interest and importance for developments in the everyday life of European citizens. It is clear that users may have several interests so they might actually belong to more than one class. However, here they are assigned to classes with the criterion of the most frequent and/or prevalent interest. In territorial terms, the classification has, for Institutional Users, a European level of aggregation of users. This is certainly the most aggregated classification, where users are represented at European level. However, the same classification can be produced at national and regional level. These are not reported here for reasons of space. Indeed, a territorial location of users is highly recommended also because needs frequently differ according to the national or regional interests of users.

The Institutional users include European organisations, such as the European Parliament and Council, European Economic and Social Committee, European Central Bank, Committee of the Regions, European Statistical System Committee, European Trade Union Confederation, Confederation of European Business, European Data Protection Supervisor, and the European Association of Craft, Small and medium-sized Enterprises.

Non-institutional users, individuals who may or may not belong to institutions, include the following groups, according to their interest in statistics:

- Users with a general interest (e.g., economic growth)
- Journalists and media
- Citizens

- Students (by level of education, or age) and Teachers (by level of teaching education)
- Users with a specific subject/domain interest (e.g., health)
- Other decision makers
- Policy analysts
- Marketing analysts
- Experts in a specific field
- Users with a research interest (e.g., innovation in enterprises)
- Scientific community – academics and researchers at universities and research institutions
- Consultants and researchers in Governmental Agencies and private sector

Another useful taxonomy, disaggregated at individual level, identifies users according to their frequency of statistical usage and proficiency:

a) Heavy users: researcher, specialist, politically or civically-engaged citizen, and others that use statistics on a daily basis. Typically this is the person who knows where to find data and how to interpret it. Within this category we can distinguish the

Very heavy users: researchers who would be routinely engaged in using disaggregated and micro data (AMFs and RMFs) in their research and who could contribute to the improvement of data quality by engaging with data producers

b) Light (occasional) users: users who from time to time check some figures. They would know the National Official Statistics and Eurostat websites but would find some difficulty in getting the data they needs and would not be looking for metadata.

c) Non-users who might be potential-users: all people who do not go looking for data, believing it is hard to understand and not being aware of the relevance and richness of data.

This last classification has some similarities with “Personas”, defined by the Digicom project organized by Eurostat with several NSIs (Digicon status, 2018 Q1), where 5 prototypes are defined : technically advanced users, analytically advanced users, data oriented users, visually oriented users, personally interested light users.

It is now clear that a good consultation and an efficient communication infrastructure must be commensurate to the typology of users. This allows us to focus on the different sets of users and different contexts. Each class of user is likely to have a different and specific set of competencies in terms of statistical literacy, interests and influence in society. These different competencies have to be represented in the design of the new communication infrastructure.

<table>
<tr><td rowspan="3">users</td><td colspan="3">institutional users</td></tr>
<tr><td rowspan="2">non-institutional users</td><td>users by interest</td><td>general interest, journalists, citizens, students, teachers, users with a specific interest, other decision makers, policy analysts, marketing analysts, Experts in a specific field, users with a research interest (e.g., innovation in enterprises), Scientific community – academics and researchers at universities and research institutions, consultants and researchers in Governmental Agencies and private sector</td></tr>
<tr><td>users by frequency of use</td><td>heavy users & very, heavy users
Light users</td></tr>
<tr><td>Non-users</td><td colspan="3">potential users</td></tr>
</table>

3.2 Strategies of communication for different groups of statistics stakeholders

Let us consider the last taxonomy, which identifies users according to their frequency of statistical usage and proficiency. Thus, we have:

a) Heavy users: researchers, specialists, politically or civically-engaged citizens, and others that use statistics on a daily basis. Typically, this is the person who knows where to find data and how to interpret it. Within this category we can distinguish the very heavy users: researchers who would be routinely engaged in using disaggregated and micro data (AMFs and RMFs) in their research and who could contribute to the improvement of data quality by engaging with data producers. These users can also become producers of data by using the new technologies of data collection discussed in section 2.

For these stakeholders, communication with producers of statistics should go in both directions from producers to users and vice-versa. These users need access to micro data of NSIs, and expert users can furnish methodologies, software, complex strategies of data analysis and data collected with innovative methods. What do we need to do to enable this?

A new communication tool could be provided by Eurostat in the form of a "research networking site", where researchers can share: data; papers on data, new practices of data production, new methodologies for data collection and analysis; ask and answer questions; and find collaborations between researchers of NSIs and Universities. Membership should be allowed to researchers of National and international Institutions of the Statistical Systems, to Universities, and to private experts with recognized experience. An accreditation system can be put in place (similar to the accreditation system allowing researchers to access data via the UK's Administrative Data Research Network). Each member should have a user profile where they can upload data, papers presentations and open software. Users could also be allowed to follow the activities of other users and propose discussions. The networking site could be inspired and have features similar to ResearchGate, Academia.edu, Google Scholar, Mendeley. However, the main difference with respect to these research networks should be the strict use and/or production and/or analysis of statistical data. Data should represent the common denominator amongst researchers (producers, methodologists and heavy users) of this network.

In fact, at the beginning of 2018, Eurostat launched a very nice initiative, the "European Statistics User Forum" dedicated to the discussion of statistical topics: Food Price Monitoring (experimental statistics), EU-SILC scientific use files, and Income, Consumption and Wealth (experimental statistics). Promotion efforts included internal promotion (intranet, e-mails, and mentions in presentations) and external promotion (tweet on EU_Eurostat, text on Eurostat and ESS website, e-mails to potentially interested groups). However, this *focus group* can only discuss topics defined centrally, and does not have the other features described above.

> *Recommendation 5:* Eurostat should extend the features of the website section of "Experimental Statistics" (ES) and the European Statistics User Forum and use a structure similar to the one described above. At present ES seems dedicated to NSIs, which can propose new data and statistics; while the proposed research network should have an individual membership level in order to better reach the segment of expert statistics users, understand their needs and encourage collaborations. Eurostat could use this network to ask questions requiring expert input and to propose projects. Questions could be used to identify missing information on existing datasets; identify new thematic needs; assist work program prioritisation, and so on. Users who are expert will be able to propose and give their opinions on the importance of new statistical products, for example, regarding: economic, financial and political crises, new forms of economic production with global value chains, global demographic trends, migrations, poverty and goals of sustainable development. Eurostat should financially support this initiative. ♦

For the segment of

b) Light (occasional) users: users who from time to time check some figures,

They would be expected to know the National Official Statistics and Eurostat websites, but would find some difficulty in getting the data they need and would not be looking for metadata. Thus, they would like to receive data and information when they request it.

> *Recommendation* 6: NSIs and Eurostat should support a registration system for light users of their websites. The registration should have a statistics search tool and a virtual assistant that helps the user to understand what statistics and data are relevant to his/her interest. The direction is the one taken by Eurostat with "Statistics Explained" (http://ec.europa.eu/eurostat/statistics-explained/index.php?title=Statistics_Explained) the new innovative electronic publishing platform and gateway for statistics information. Then, when the user seeks updates, they immediately receive the latest statistics and data for which they are registered. In this case, profiling of users is useful to give them what they need. Here the virtual assistant could give general information on the latest statistical developments of the NSIs and could collect requests for new information coming from this segment of users. As always, the obligation to register is the cost of obtaining the information the system can provide.♦

For the segment of

c) Non-users who might be potential-users: all people who do not go looking for data, believing it is something hard to understand and not being aware of the relevance and richness of data.

The lack of statistical knowledge for this segment may more frequently induce an inability to distinguish between official statistics and alternative figures. This can lead to misunderstandings and the acceptance of "fake news". Citizens are disoriented and lose their trust for any type of statistics. Thus, we endorse the policy approach of the European Statistical System Vision (2014), based on the need to create stronger engagement between statistics producers and potential users. The alliance has to be founded on the trust in official statistics and on the perception that they are valuable and useful. Thus, there is a need for greater awareness among citizens that official statistics can enhance their capacity to make good decisions for themselves in their daily life and provide them with a better understanding of the complex and fast evolving world.

However, this awareness can only be achieved in practice if citizens receive a basic training in data handling (The Data Manifesto, 2015), that is, they must have a basic education in the types of quantitative methods necessary to read and understand statistics (information-processing skills, PIAAC OECD 2013). We stress the need to create an alliance among producers, users and the scientific community (statistics stakeholders) to ensure that statistics is a key competence of a modern education, and becomes an important element in upskilling courses for *lifelong learning*. This alliance needs to be translated into a concrete action plan to be proposed to and further decided by the policy-makers at national, European and international level.

What can be proposed to promote the culture of statistics? ESAC (Vichi, Valente Rosa, Ruane, 2015) has already proposed: the European Statistics Day starting from 2016 with the specific objective of developing the awareness of European citizens of the importance and relevance of official statistics; and the Conference of European Statistics Stakeholders (Rome, 2014; Budapest, 2016; Bamberg, 2018) to foster the collaboration, improving discussion between Statistics Stakeholders (producers, users and methodologists).

Particular tools can also be developed further – see, for example, Babic and Herteliu (2016). To take a simple example, one strategy is to position individuals in rank order. A topical example would be pointing out that everyone but you in your peer group has walked 10,000 steps today.

We further propose:

Recommendation 7: Eurostat should synthesise and propagate the work on promoting the culture of statistics with materials prepared for the potential users of statistics based on such materials as:

Videos: development of short online training videos on statistical literacy, mainly for specialist disseminators (journalists and media) but also available to the wider public, based on deceptively simple questions (e.g.: what is the meaning of a percentage?; how to read a ratio?; official data is what?....;

Webinars: with interactive learning for large groups on line, such as students;

Short online thematic modules, to be used in schools, based on European statistics, to be integrated in curricula programs at school level (e.g.: some history facts about population trends, urbanisation, environment issues). Those modules – language and contents – should be prepared so that they can be used at different stages of the education system.

Strategies for communicating statistics using new and mobile media.♦

Communication between producers and potential users needs a clear and easy message, based on a storytelling approach with the use of graphical tools, and other instruments that simplify the presentation of statistical results. The approach should be similar to the best experiences given by data journalism, that is, with the approach of using data to tell stories on the phenomena statistically analysed. Digital storytelling is the activity of building a narrative, based on digital media such as video, audio, images, text, etc., around data sets to get the most relevant information and transfer this in a simplified manner to people.

Recommendation 8: Eurostat and NSIs should further develop, for each of the different themes described in official statistics (e.g., general and regional statistics, economy and finance, …, science and technology), a "digital storytelling", with which the user can interact, that presents the main information, trends and innovations on the theme, by using graphical tools and a simple language. In this way, people can be quickly informed on the main knowledge regarding themes being described. ♦

The publication should be similar to Facebook Stories. This way of communicating statistics should be similar to the communication tools of the young generation. It goes without saying that stories should be attractive so that people will be stimulated to stop and follow them.

A key challenge is that of making non-users more aware of the data and potential of the Eurostat website. This leads us to

Recommendation 9: Eurostat should promote its materials more widely to non-user communities which could take advantage of them.

4. Conclusion

Much modern communication involves transmission and receipt of data, either in the form of raw data or in the form of processed data. Raw data provide the input to statistical analyses, and processed data provide the results of statistical analyses. Considerations of statistical communication should look at both of these aspects:

how the data is initially captured, and how the results are then presented to potential users. For the former, much data is initially in unstructured form, and this needs to be translated into a form suitable for statistical analysis. For the latter, the style of presentation depends on the level of expertise of the intended recipient, which might range from the experienced and skilled professional statistician to the interested but statistically naive layperson - or even to the members of the public who have a suspicion of statistical results.

We present nine recommendations for facilitating these two complementary aspects of statistical communication. In summary, these are:

1. Increase the role of Computer-Assisted Web Interviewing;

2. Ensure that the minimum number of requests for data go to any individual;

3. NSIs should develop collaborations with universities to expedite new modes of data collection;

4. NSIs should promote, at the European level, a standard to make data produced by smart devices linkable.

5. Eurostat should extend the “Experimental Statistics” section of the website to include individual membership;

6. NSIs and Eurostat should encourage casual and occasional users to register by developing software tools to assist such users;

7. Eurostat should promote the culture of statistics through videos, webinars, and short online thematic modules for use in schools and other places;

8. NSIs and Eurostat should develop a “digital storytelling” facility, presenting statistics on official statistics themes in and accessible, interactive, and straightforward way.

9. Eurostat should promote its materials more widerly to non-users.

Finally, although we have not explicitly identified it as a recommendation, it is important to put in place a system for evaluating the effectiveness of communication.

Acknowledgements:

The authors are extremely grateful to Ineke Stoop and the other members of ESAC for valuable comments on an earlier draft of this paper. Any remaining errors or oversights are the responsibility of the authors.

References

Babic M. and Herteliu C. (2016). Picture, storytelling, and communicating statistics. *Journal for Communication Studies*, 9, 9-11.

Conrad F.G., Schober M.F., Jans M., Orlowski R.A., Nielsen D., and Levenstein R. (2015). Comprehension and engagement in survey interviews with virtual agents. *Frontiers in Psychology*, 20, https://doi.org/10.3389/fpsyg.2015.01578

ESS Vision 2020, (2014). Document approved by the European Statistical System Committee. http://ec.europa.eu/eurostat/documents/10186/756730/ESS-Vision-2020.pdf/8d97506b-b802-439e-9ea4-303e905f4255

General Data Protection Regulation; https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.L_.2016.119.01.0001.01.ENG

Hand D.J. (1986). Patterns in statistical strategy. In *Artificial Intelligence and Statistics*, ed. W.A.Gale, Addison-Wesley, Reading, Massachusetts.

Hand D.J. (2009). Modern statistics: the myth and the magic, Journal of the Royal Statistical Society, Series A.

Herring, S. (ed.) 1996. Computer mediated communication: Linguistic, social and cross-cultural perspectives. Amsterdam: John Benjamins.

Maggino F., Trapani M. (2009). Presenting and communicating statistics: principles, components, and their quality assessment. A proposal Paper, NTTS 2009, Bruxelles.

Paganini P. (2015) Skynet, the US top-secret program to track terrorists. https://securityaffairs.co/wordpress/36715/intelligence/skynet-us-top-secret-program.html

PIAAC (2013). The Survey of Adult Skills, http://skills.oecd.org/documents/SkillsOutlook_2013_Chapter1.pdf

RSS (2015), The Data Manifesto, https://www.rss.org.uk/Images/PDF/influencing-change/2016/RSS_Data%20Manifesto_2016_Online.pdf

Vichi M. and Hand D.J. (2018) Trusted smart statistics: the challenge of extracting usable aggregate information from new data sources. *DGINS Conference, Bucharest, 10-11 October.*

Vichi M., Valente Rosa M.J, Ruane F. (2015). The Users of Statistics and their role in the European Society, ESAC document approved by ESAC, http://ec.europa.eu/eurostat/web/european-statistical-advisory-committee-esac/other-documents