INFORMATION QUALITY FOR PROCESS IMPROVEMENT

Oriol Camps-Lorente and Xavier Tort-Martorell
Department of Statistics and Operations Research, Universitat Politècnica de Catalunya (UPC), Barcelona, Spain
oriol.camps-lorente@upc.edu

As George E. P. Box once stated, every system executing a process is like a radio transmitter except that it transmits information instead of electromagnetic waves. Information flows from processes to management decisions, through systems involving people, technologies and operating procedures. However, quite frequently that informational chain gets disconnected and decisions are not properly based on evidence from processes. Errors in information capture, coding, representation, storage, recovery, analysis, communication or interpretation may cause such disconnections. A framework for analyzing the quality of information flows involved in managing particular processes is proposed and exemplified with a case study. The framework has interest for statisticians and its implications for statistical education are discussed, together with a case of pedagogical innovation.

INTRODUCTION

Information flows from processes to process management decisions in a continuous way. We can conceive that flow as a chain of activities that captures information carried by facts and transforms it into representations containing information suitable for decision-making. If such a chain works properly, representations (inside a report, a PowerPoint presentation or an e-mail) also carry information about process' facts, and therefore they permit to obtain knowledge. If flows do *not* work as they should, representations do not carry all the intended information and managers just obtain false or partial beliefs. Informed decision-making comes from correct information capture, codification, representation, storage, recovery, analysis, communication, or interpretation. Very roughly, a data set *D* carries the information that *x is P* (a fact, in general) if and only if *D* implies that *x is P*, its values have been properly captured and codified, and the fact that *x is P* actually happens in the process. A graphical representation carries the information that *x is P* if and only if it uses data that carry the information that *x is P* and has been correctly represented. And so on.

Statistical analysis assumes a data set, and statistical theory ensures that statistical results will carry information about the data set, but nothing but information flows ensures that results will carry information about *facts*. Therefore, we should adopt a *dynamic* point of view [information as a flow in a distributed system, for instance as treated in Barwise & Seligman (1989)], in contrast to a *static* one [information as a product, paradigmatically in Wang (1998)]. There is no information outside any information flow. Information has to do with *regularities* governing facts and processes, and they can be grasped and used to improve the adaptive and competitive capabilities of organizations and other agents. [A rigorous analysis of the information concept at the basis of our point of view can be found in Perry & Israel (1990), Barwise (1993), or Barwise & Seligman (1989), from the initial insights of Dretske (1982) and Barwise & Perry (1983)].

Information Quality Management will be understood, in this paper, as Information Flows Quality Management, and it have some implications for statistics and, particularly, statistical education. It consists, concisely, in study how informational activities (capture, codification…) can fail and how these failures affect information transmission from facts to decisions. In this paper we briefly present a methodology to do this and then we examine the implications for statistical education; we present a case study of both.

A FRAMEWORK FOR INFORMATION FLOWS QUALITY ANALYSIS

Let us assume that we are statisticians involved in quality management activities and, for the sake of simplicity, that we are concerned with only one particular process. Our goal is to improve the management of information emitted by the process and used in process management decisions. To judge this improvement we use the following criteria: (A) maximize availability of

In C. Reading (Ed.), *Data and context in statistics education: Towards an evidence-based society. Proceedings of the Eighth International Conference on Teaching Statistics (ICOTS8, July, 2010), Ljubljana, Slovenia*. Voorburg, The Netherlands: International Statistical Institute. www.stat.auckland.ac.nz/~iase/publications.php

non-redundant useful informative contents, (B) maximize efficiency of useful informative contents, (C) minimize useless informative contents, and (D) minimize non-informative contents.

'Useful' and 'useless informative contents' are process-dependent or even decision-dependent concepts. Our last assumption concerns the process *function* regarding organization's goals. We assume that we know, for a given process, its *adaptive* or *competitive* function, that is, why it is important for the adaption of the organization to its environment. After all, information matters because agents that can capture it from their environments and use it have more possibilities to succeed [as noted by Dretske (1982), Barwise & Perry (1983) or Perry & Israel (1990)].

In organizations, information flows through systems involving people, technology and operating procedures. Every process can generate more than a single information flow, and the system performing flows may not exactly correspond to people, technology and operating procedures of the process itself.

Analyst should be able to identify different *states* of the organizational system. Broadly speaking, states are defined by agents having information, or knowledge, or by technologies carrying or having information, representing or storing. Transitions among states roughly correspond to either information transmission between agents or information transformation. Even though every change in agents or technologies defines a state, these have to be defined by pragmatic criteria, maybe grouping some of them for the sake of analysis' efficiency.

At the end, the main objective is to understand and to assess the different *mechanisms* that perform state transitions. Such mechanisms implement the information flow in the concrete system under study. Particularly, they implement state transitions, or transmission and transformation operations. If mechanisms work, the entire flow works. Mechanisms should be assessed by taking account of its own *purpose* and adaptive function, that depend on the information flow's and the process' adaptive functions.

In short, the analytical framework includes the following steps:

- Identify and describe the process
- Identify which (and how many) information flows are generated by the process
- Identify the organizational system that performs flows' activities. It consists of agents, technology in a broad sense, and operating procedures.
- Identify for which *states* the system passes, and/or which are the main information transmission and information transformation operations.
- Understand which *mechanisms* perform the transitions among states, or transmission and transformation operations.
- Gather information to analyze the failure modes of the mechanisms.
- Link mechanisms' failures with mechanism's and process' purpose. Analyze how the main failures compromise the *adaptive function* of information flow.

CASE STUDY: COMPLAINT MANAGEMENT IN A MUNICIPAL ADMINISTRATION

A simple but interesting case of information flow, not primarily involving quantitative data and statistics, is that of complaints from customers or, as in the following example, citizens of a town. Without entering into the details of our actual intervention at Esplugues de Llobregat (Barcelona) municipal administration, we shall describe several basic features of the case to exemplify some points of the last section.

Esplugues del Llobregat is a small city of 46.500 inhabitants adjacent to Barcelona, the capital of Catalonia and Spain's second greatest city with 1.615.000 inhabitants. Esplugues' municipal government has enjoyed political stability and fairly managerial continuity since the early 80's. Initiatives to establish a management system have had a sustained political support and they have succeeded to a quite large extent. It adopts the EFQM European Model in 1996. In 2000 it wins the Premio Iberoamericano de la Calidad (Latin American Quality Award).

Council publishes a set of "Service Commitments", 132 currently, divided in several activity areas: security, environment, citizenship, and services to people. The process of Complaints, Suggestions and Observations (CSO) Management is central from the viewpoint of both Council's Managing Director and the Major, and it includes some Service Commitments. For instance, up to 48 hours from receiving a CSO, a courtesy contact has to be made indicating that the issue has been communicated to its responsible. Another commitment regarding CSO is that the 100% have to be *answered* into de boundaries of 10 days, even when a hypothetic solution requires much more time.

A simplified version of the process is roughly as follows. A citizen has many means to communicate a CSO: face-to-face at different contact points and personnel, by telephone or e-mail. CSO are centralized at Citizen Service Points (or CS-Points), where are registered in a database using a specific software. The Major receives via Blackberry *all* CSO at the moment of registration. She can stop the process' course if she wants to manage a CSO herself. The CS-Point managing the CSO assigns a technical responsible (architect, local police, maintenance chief, or anyone else), whom has to assume it. CSO assumption is an important part of the process because CS-Points can commit errors assigning a responsible. Usually, any waste of time at this task has an impact on Service Commitments. The town councilors receive, via Blackberry, all CSOs *of his area* at the moment of the technician's assignation.

The CSO responsible has to investigate it, solve it if possible, and *answer it*. He receives software alerts of every open CSO when commitment's deadline is near. When it is answered, he can close the CSO and CS-Point personnel can perform the process' final activities. They check the quality of the answer (which can be returned for revision), contact with the citizen usually by telephone, and if the contact succeed they finalize (or definitely close) the CSO at the database.

CSO Management process implies two main information flows:

- From the fact that motivates the CSO, which is collected by a citizen, to a response decision mostly made by a technical responsible, and finally to a closed register in the CSO database.
- From facts concerning CSO handling to decisions about the very process of CSO management.

The first flow is strictly parallel to the process, which in this case has a high informational component. In this example, both flows are executed by the same system:

- *Agents*. Citizen, CS-Point personnel, technical personnel (architects, local police, etc.), Major, town councilors, CSO Management Process Responsible.
- *Technology*. CSO Management Software, Blackberry, e-mail, telephone.
- *Procedures*. The CSO Management procedures are fully documented and include most of activities of the first flow. Activities for the second flow are not documented, and are not systematically approached.

Many states can be defined for this system. Usually it is possible to start by identifying an initial state, a decision state and a final state. In our case, we show some examples for the first flow:

- *Initial state*. The citizen has information about a fact that is cause of a CSO. Other agents remain ready to act.
- *Second state.* The citizen has communicated the CSO. CS-Point personnel have received it and have assigned a technician. CSO Database registers it.
- *Third state.* The technician has assumed the QSC. Major has received a notification via Blackberry. CSO remains opened at database.
- […]
- *Decision state.* The technician decides the answer that has to be done to the citizen. He writes it at the CSO software.
- […]
- *Final state.* CS-Point personnel definitely close the database register.

Any potential failure in the information flow will be able to be detected in a specific state transition. The concrete way to perform a transition is a *mechanism*, for instance, the mechanism of receiving CSO from citizens, the mechanism of assuming CSO by technicians, or the mechanism of definitely closing CSO registers. We are mainly concerned with the failure modes of such mechanisms, so it is important to identify them and list them.

At Esplugues, several mechanisms for the first flow presented actual or potential failures. For instance, the mechanism of deciding a response and writing it up at the database could be improved on. Answers did not fulfill quality criteria, mostly for (*i*) being too long and for (*ii*) using a too much technical language. Interviews with technicians revealed that some of them were not aware that their responses had to be read out by a non-technician employee to a non-technician citizen. In fact, they had no a broad perspective of the CSO management process, only of the particular processes the complaints could refer to. As a consequence, some citizens perceived that (1) the complaint was not answered, and (2) his interlocutor had no knowledge to solve it.

The mechanism of receiving CSO also generates some problems not easy to deal with. In each communication, the citizen usually transmits different contents corresponding to different complaints, suggestions or observations, perhaps referring to different services or matters. Of course, the citizen expects a single response to his communication. If the CSO is divided into different 8 registers, the citizen will receive 8 courtesy contacts up to 48 hours, and 8 final communications, one for each issue. If CSO is treated as one, to which technician have to be assigned? Actually, the current policy is that CS-Points divide each CSO by department. It is not optimal, because technicians tend to omit difficult issues' response when there are several. Another reason affects the second flow: dividing them by department, data about the number of CSO does not carry information about the number of CSO but about the number of citizen's communications.

The second flow is rather less structured than the first one. Periodical satisfaction survey from a sample of already managed CSO provides the main input to process monitoring. It consists of eight questions, uses 50-60 CSO of different departments and it is performed telephonically. About 3 of 7 questions have coincident answers systematically, which induces to think that the survey is redundant and not efficient. Process owner has access to aggregate data and counts from the database, but regularly only CSO out of deadline are listed. Weaknesses of satisfaction survey and lack of direct information from CSO caused that even if information from CSO is obsessively promoted, captured, transmitted and used, *information for process management* is quite disregarded.

However, both flows are important, because (*a*) CSO provide inexpensive information about a broad range of issues concerning the town, and (*b*) CSO increase citizen loyalty. To acquire the same amount of information using contracted personnel would be extremely expensive, and citizen loyalty have direct political rewards, so CSO information flows play a critical role for local government efficient knowledge of its own environment. The more volume of CSO received, the more adapted local administration becomes to its citizens and context. Therefore, the process has to be extremely efficient; otherwise an increasing CSO volume will become too costly. To achieve more efficiency it is necessary to improve the second information flow (information for process management) the initial step to obtain more control of the first flow and try to simplify or optimize it.

IMPLICATIONS FOR STATISTICAL EDUCATION

Information flows analysis is important for statistical education because it questions the *analysis-centered viewpoint*. Statistical theory provides methods for, *given* a data set, analyze it and obtain a result. This is the paradigm that directs statistical education: Statistics tasks starts with a given data set and ends with its analysis. This is true not only for descriptive statistics and statistical inference; even in Design of Experiments (before we get data) the main focus is to obtain an optimal design given a statistical model, that is, to optimize the further analysis.

Mallows (1998) introduces the question "how do data relate to the problem?" as a previous issue. We agree about the diagnosis: «textbook examples are typically concerned with training the student how to apply statistical techniques, and usually take the model for granted» [p. 4]. Applied Statistics, and therefore huge parts of Statistical education, have to think on how certain facts relate to the problem, and on how the data set carry information about these facts. That is an approach to Mallows' problem. In addition, they have to think on how statistical results are used to really solve the problem. This implies to understand how the results' information content is added to a set of premises for problem solving; that usually means different forms of representation, communication and, at the end, interpretation.

Statistical practice does not equal statistical theory, but statistical education should be, at least in many cases, focused on practice; otherwise troubles as the following would be disregarded:

- Even if objectives and problems are well-defined, sometimes it is not easy to discover *when* and *how* to obtain information.
- Statisticians sometimes have to *detect* and *filter* information, as in problems of *contents excess*.

- Sometimes it is not easy to choose which representation has to be used to adequately present relevant information. Or quite worst, people get used to specific representations and do not consider improving them.
- Without operative definitions of variables, most information requests become garbage.
- Even on the basis of correct analysis, wrong logical inferences can cause that the decision maker makes a mistake. Wrong conclusions can be induced by deficient representations.
- In model selection and hypothesis testing there are two aspects: information *selection* (choose a good question) and information *use* (infer conclusions from answers to the question). Experimental evidence [see Slowiaczec, Klayman, Sherman & Skov (1992)] suggests that people do best in information selection and worst in information use. *Cognitive biases* in assessing uncertainty make easier to choose discriminatory questions than correctly evaluate results' consequences for belief revision (maybe necessary for decision making).
- In decision making we may use representations or data sets which have (1) informative, true contents about a whole issue, (2) true but incomplete contents, informative but partial, (3) irrelevant contents, informative about facts of the situation but redundant or useless, and (4) uninformative, false, misleading contents. Sometimes there are not operative criteria to distinguish them.

CASE STUDY: A CLASS AT UPC'S STATISTICS DEGREE

As a part of a one semester introductory course in Quality Management at the Statistics Degree of the Universitat Politècnica de Catalunya, authors have included a two days lesson on *Information Quality*. The lesson's aim is double: (1) to show how information flows are a critical issue in organizations and particularly for Quality Management tasks, and (2) to relate students' statistical training to a broader view of information flows, from data capture to results' communication.

The class consists of a simulation game, a discussion on game's conclusions, and finally an exposition of a systematic summary on Information Flows Quality Management. We will present the game and its conclusions. The professor has a map of a urban situation on paper ("The Truth"), usually with streets, buildings and cars; looking at the map, it is obvious that a car crash will happen unless a couple of traffic lights and signs are placed.

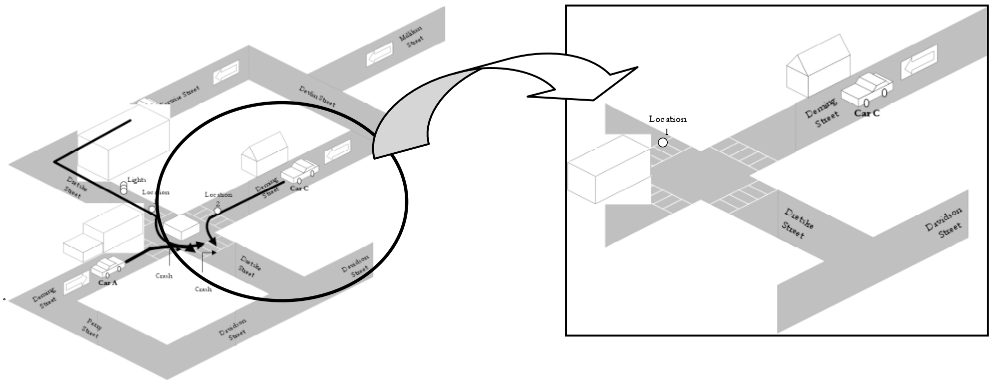

Figure 1. Examples of "The Truth" (left) and informer's notes or clues (right)

One student ("the coordinator") is alone in front of the blackboard and the rest ("informers") are separated, each one with a piece of paper with a partial version of the map and other written instructions (for instance, "car A is at left of car B"). The coordinator has up to 30 minutes to decide where to put a traffic light and a sign in concrete places, in time to avoid the car crash. The coordinator gets the information asking *concrete* questions to informers, whom respond *concrete* answers. Informers can only talk to the coordinator to respond him (except the last 2 minutes under professor criteria), and cannot speak between them during the activity.

The coordinator probably will begin drawing a map, and depending mostly on his ability the group will succeed or fail. Informers probably will get little confused because they heard of aspects that *may seem* incompatible with their version (but all their notes carry information about The Truth). Some of them may try to make inferences joining their version with previous interventions; if they err, the coordinator will be in serious trouble. The professor (and maybe different students with the role of “observers”) take notes and after the students draw their own conclusions. The following are some actual examples:

- Even in a simple simulation, it is not easy to discover *how* to obtain information. To whom ask the questions? In which order? Which questions? Is there an optimal strategy?
- Information sharing usually shortens decision making, but communication remains the most critical problem of organizations. Informers cannot speak between them as a rule, but people in actual organizations tend to do the same thing voluntarily.
- The last two minutes, if professor allow informers to talk to the coordinator, all of them will try to say his own portion of vital information, and the situation usually gets worst for the coordinator. It is a situation of *contents excess*.
- It is not trivial to choose a representation. Some coordinators begin with a list (“car is A at Barwise Street”, “Barwise Street parallel to Tukey Avenue”…), before trying to reproduce the map.
- Even in an aseptic simulation game operative definitions are relevant. “Is car A close to car B?” can be a quite confusing question.
- Wrong inferences can cause that the decision maker errs. As mentioned above, some informers will try to draw their own conclusions, even unconsciously.
- In the game, only *true* contents are allowed, though maybe not complete. Actual world is much more confusing, even setting aside intentional lies and manipulation.

CONCLUSION: BROADENING THE ANALYSIS-CENTERED STATISTICAL EDUCATION

The role of the statistician in his professional work mostly happens in multidisciplinary teams. His function usually consists in advising on methodological issues. An analysis-centered statistical education may cause a gap between the expert’s work and the statistician’s work. All statisticians know what happens when the expert assumes that data gathering needs no statistical approach. In business and industrial statistics this problem gets even worse because data is usually gathered and handled for many different purposes, so process information becomes impoverished for decision making.

Therefore, introducing the point of view of information flow in statistical education enables novice statisticians to intervening in a broadest set of issues. At least, students of Statistics should convince themselves of his role in defining projects, data codification, storage and recovery procedures, representation rules, communication standards, interpretation advices, and so on. Statisticians should want to be involved in decision making, not as mere input suppliers, but as experts on evidence-based decision making. They should know how to take evidences from facts and put them into the decision process, assuring the reliability of every intermediate step.

REFERENCES

Barwise, J. (1993) Constraints, Channels, and the Flow of Information. In Aczel, P., Israel, D., Katagiri, Y. & Peters, S. (eds.), *Situation Theory and Its Applications*, Stanford: CSLI Publications.

Barwise, J., & Perry, J. (1983). *Situations and Attitudes*. Cambridge, Mass.: MIT Press.

Barwise, J., & Seligman, J. (1997). *Information Flow. The Logic of Distributed Systems*. Cambridge Tracts in Theoretical Computer Science. Cambridge, UK: Cambridge University Press.

Dretske, F. (1981). *Knowledge and the Flow of Information*. Cambridge, Mass.: Bradford Books, MIT Press.

Mallows, C. (1998). The Zeroth Problem. *The American Statistician, 52*(1), 1-9.

Perry, J., & Israel, D. (1990). What is information? In Hanson, P. (Ed.), *Information, Language and Cognition*. Vancouver: University of British Columbia Press.

Slowiaczec, L. M., Klayman, J., Sherman, S. J., & Skov, R. B. (1992). Information Selection and Use in Hypothesis Testing: What is a Good Question and What is a Good Answer? *Memory and Cognition*, *20*(4), 392-405.

Wang, R. (1998). A Product Perspective to Total Data Quality Management. *Communications of the ACM*, *41*(2), 58-65.