Data Quality Requirements for Services of General Interest: The Case of Water Bills

Patricia Moura e Sa
NEAPP/UM - Núcleo de Estudos em Administração
e Políticas Públicas da Universidade do Minho
Faculty of Economics, University of Coimbra (Portugal)
email: pmourasa@fe.uc.pt (corresponding author)

Rita Martins
GEMF, Faculty of Economics, University of Coimbra (Portugal)
email: rvmartin@fe.uc.pt

Abstract

The Quality Framework for Services of General Interest in Europe (EU, 2011), alongside with Citizens Charters and Service Level Agreements emphasise the importance of giving quality information to consumers. Information needs to be provided in a way that is effective and that takes into account the needs of consumers with different levels of qualifications and expertise.

With some variations, Data Quality literature proposes a set of dimensions that include accessibility, clarity, relevance and accuracy. Invoices are a vehicle to convey information to all consumers, noticeably on prices and volumes.

This research takes the example of water bills, given the water double role of environmental and economic good and the fact that almost all water utilities are local monopolies. In this regard, identifying data quality requirements and ensuring the consumers' right to accurate, reliable and clear information is particularly important.

An initial list of 57 consumers' requirements was created based on the views of water consumers, collected by means of three focus groups. Using an affinity diagram, requirements were then organised into groups. Findings show that plain language efforts and strategies to enhance trust on the service provided by water utilities need to be further reinforced.

Consumers' requirements together with recommendations on water bills from the economic regulator also confirm the importance of the data quality dimensions identified in the literature. A few implications on what to consider when designing effective invoices can be derived from the current study.

1. Introduction

Several international frameworks, such as the Quality Framework for Services of General Interest in Europe (EU, 2011), alongside with Citizens Charters and Service Level Agreements call attention to the importance of giving quality information to consumers, as a way to ensure consumers‘ rights and promote adequate consumption behaviours. Transparency, in particular, is regarded as essential to enhance trust on service providers. According to the White Paper on Services of General Interest (Commission of the European Communities, 2004), consumer protection and transparency are among the several requirements (together with universality, continuity, quality, affordability) that must be observed.

Yet, as recent international reports emphasise (e.g. OCDE, 2013), ensuring consumers' right to information alone is not sufficient. It is essential to provide relevant information in a way consumers understand and feel useful and attractive. When identifying data quality requirements, considering consumers' needs is paramount. Moreover, because Services of General Interest (SGI) are provided to the population in general, when designing communication channels and documents, it is important to take into account that consumers are diverse and do not have the same levels of knowledge and expertise. The OECD's latest findings on adult skills (OECD, 2013) also stress that public services should be understandable and accessible to less literate people. Accordingly, the report highlights that public bodies ought to communicate in simple language. It is therefore pertinent to understand at what extent public services are intelligible when they communicate with customers, some of whom have low levels of literacy[1] and numeracy[2].

There are nowadays multiple channels to access information, namely the internet. Yet, one cannot ignore that a considerable group of people still do not have access to the web and/or do not feel at ease using ICT technologies. Making sure that information reaches the whole targeted population is even more critical in the case of SGI. Invoices are a channel that necessarily reaches all the customers on regular way. Therefore, it is crucial that the information disclosed through invoices is adequate and clearly understood.

With the aim of identifying data quality requirements from the costumers' perspective, this research takes the case of water invoices and uses a qualitative methodology to collect and analyse consumers' feedback. The use of focus groups makes it possible for consumers to express their views freely, in their own words, while benefiting from interaction with other customers with similar experiences. The meaning of their 'voices' is analysed with the assistance of an affinity diagram.

Studies on data quality requirements outside the context of information systems for organisations are scarce. By transposing data dimensions to the analysis of invoices (critical in communicating information on prices and volumes), the current research can represent a first step in building a quality framework to assess the quality of such documents.

The remainder of the paper is structured as follows. Section 2 makes a brief literature review on the concept, attributes and costs of data quality. Then, the Portuguese water industry (the focus of the current study) is briefly described in what concerns its market structure, the price regulation mechanisms and the importance given to water bills. Section 4 presents the methodological approach adopted in the study. Section 5 shows some key ideas

[1] Literacy is defined "as the ability to understand, evaluate, use and engage with written texts to participate in society, to achieve one's goals, and to develop one's knowledge and potential. Literacy encompasses a range of skills from the decoding of written words and sentences to the comprehension, interpretation, and evaluation of complex texts". (OCDE, 2013:59).

[2] "Numeracy is defined as the ability to access, use, interpret and communicate mathematical information and ideas in order to engage in and manage the mathematical demands of a range of situations in adult life. To this end, numeracy involves managing a situation or solving a problem in a real context, by responding to mathematical content/ information/ideas represented in multiple ways" (OECD, 2013:59).

that emerged from the focus groups related to the way people look at water bills, the problems they face in reading them and, finally, the resulting list of requirements. Section 6 discusses the way consumers' requirements relate to each other, pointing to some potential synergies and contradictions, and confronts the information collected with the regulatory body recommendations and with the data quality dimensions proposed in the literature. Finally, section 7 draws some conclusions and implications.

2. Data quality: concept, attributes and costs in a glance

The importance of 'good data quality' is obvious, since poor data quality can have substantial social and economic impacts. In a broad sense, poor data quality leads to "bad" decisions and jeopardises the implementation of many strategies and policies. The issues of poor data quality are extensively studied from the organisational point of view (see for e.g. Haug et al., 2011), but are rarely analysed from the consumers' or even from the public policies standpoints.

The distinction between 'data' and 'information' is well-established in the literature. Davenport and Prusak (1998, cited in Haug et al., 2011) stress that data is defined as "discrete, objective facts about events", whereas information relate to "data transformed by value-adding processes of contextualisation, categorization, calculation, correction and condensation". In practice, this distinction is not completely straightforward. The current research is focused on billing documents that communicate mostly data on prices and quantities to consumers, but while doing that some degree of treatment is inevitable and purely raw-data seldom exists. Yet, given that data on billing documents is relatively limited and treatment is not very sophisticated, we will use the expressions "data quality" and "data requirements".

Different authors propose different data dimensions. Usual data quality dimensions include: accessibility timeliness, precision, accuracy, reliability, currency, completeness and relevance (Wang and Strong, 1996). Wand and Wang (1996, cited in Haug et al, 2011) group data qualities into four categories: completeness, unambiguousness, meaningfulness and correctness. Later, Haug et al. (2009) propose three categories: intrinsic, accessibility and usefulness.

It becomes clear that the majority of the data quality attributes is rather dependent on data users and uses. Therefore, "intrinsic data quality" (that denotes that data have quality on their own), mainly related with accuracy and objectivity, needs to be complemented with "contextual data quality" (that emphasises the importance of looking at quality with reference to the context of the task at hand) (Wand and Strong, 1996).

In any case, data quality needs to be assessed taking into account the extent to which data meets customers' needs and expectations. Therefore, data clearly depends on the decisions to be made and on the people who use the data and/or will make such decisions. Thus, the idea of "fitness for use" should be adopted. As Ballou et al. (2004:10) stress, "If users of the data feel that its quality (…) is sufficient for their needs, then, from their perspective, at least, the quality of the information available to them is fine".

When data quality attributes are selected based on the idea that someone has of what is important to data consumers, the intuitive approach (Wang and Strong, 1996) is followed. Although this is often the case in many situations, the limitations of this approach are rather evident, especially the one that refers to the tendency to underestimate those who use the data. As Wang and Strong (1996) state, "they fail to capture the voice of the consumer". On the contrary, empirical approaches to data quality "analyses data collected from data consumers to determine the characteristics they use to assess whether data are fit for use in their tasks"

(Wang and Strong, 1996: 7). Based on the literature, 'fitness for use' in this regard implies (Wang and Strong, 1996; Haug et al., 2011):

- Accessibility: the consumer knows how to retrieve the data);
- Ability to be interpreted by those who use it;
- Relevance and timeliness (in accordance to the decision making process that is at stake);
- Accuracy: the consumer feels the data are correct, objective and come from reputable sources. In this understanding, accuracy is more than the simple difference between the correct value and the actual value used;
- Consistency

Such attributes clearly show that data quality as "fitness for use" goes well beyond traditional concerns with accuracy and consistency of data.

Representational data quality includes aspects related to the format of the data and the meaning of data, which cannot be discarded if data is to be interpretable and easy to understand by data consumers (Wang and Strong, 1996).

Tayi and Ballou (1998) call attention to the problem of semantics when multiple users are at stake. As the authors stress (1998:54), "the data gatherer and initial user may be fully aware of the nuances regarding the meaning of the various data items, but that will not be true for all of the other users. Thus, although the value may be correct, it can easily be misinterpreted".

Ensuring the quality of data for SGI is more difficult and more relevant than is the case with manufactured goods, given its intangibility nature and the difficulty of implementing effective quality control procedures. Yet, if that is to happen concerns start with good data quality planning.

Similarly to Wang and Strong study (1996), the current research looks at data as a product and listens to data users to understand their needs, identify how they can be met and evaluate the importance of each data attribute. 'Customer voices' are captured using focus groups, as explained later in section 4.1. In what concerns the methodological approach, the current study is completely different from that of Wang and Strong who have used surveys and asked users to directly assess data attributes. Our approach is more qualitative and gives users the opportunity to express their views on a more flexible context.

Data quality costs are the actual negative monetary effects that result from not reaching a desired data quality level (Eppler and Helfert, 2004). Such costs are extremely diverse, including among many others: higher maintenance costs, assessment costs, time costs of viewing irrelevant information, higher data administration costs, process failure costs, information scrap and rework costs (Eppler and Helfert, 2004). The same authors present a data quality costs categorization based on missing information quality attributes, which points mainly to the existence of: costs due to untimely arrival of information (e.g., missed opportunity); costs due to inaccurate information; costs due to inaccessible information (higher information gathering costs); costs due to inconsistent information, and costs due to unreliable information.

As it happens for all goods, investing in prevention and appraisal (by, for example, giving attention to low quality data detection, improving data quality format, or improving data processes) with the aim of minimising non-conformity situations is important, given the huge impact of low quality data on customer dissatisfaction costs and credibility lost costs. In the SGI context, the matter is even more critical.

3. SGI obligations: brief description of the Portuguese water industry case

Due to its characteristics of natural monopoly, water industry costs are minimized if services are provided by a single firm rather than if the output is provided by several

suppliers. When these market structures are in place, promoting competition is not efficient, which means that customers cannot choose their provider. Because water is essential for living (demand is price inelastic for residential essential needs), there are obligations of SGI that must be observed, such as universality, continuity, quality, affordability, equity, transparency and consumer protection. In what transparency and consumer protection is concerned, reasons for concern came from asymmetries of information between service providers and consumers. Even when considering developed economies, asymmetric information is not absent, being the consumption of bottled water for drinking purposes, instead of taped water, an indicator of the lack of consumers' confidence on the quality of taped water. Transparency and consumer protection is also relevant in what concerns to price schemes.

Monopoly prices tend to be higher than those resulting from market competition, making price regulation a key part of regulation models. Besides, price instruments are increasingly being used to simultaneously attain greater economic efficiency, equity, public health, environmental sustainability, financial stability, public and political acceptability and simplicity. As a consequence of this multi objective context, the prevailing pricing schemes are expected to be complex (Martins et al, 2013). However, if pricing policies are to be effective in driving adequate consumer behaviours, information on prices needs to be clearly perceived by consumers, which means that pricing mechanisms should be transparent and understandable.

The Portuguese water industry, as it happens almost elsewhere, is organized as multiple local monopolies, with natural monopoly characteristics. The industry is very fragmented, with almost one service provider per municipality (308 municipalities for less than 11,000 000 inhabitants) and with heterogeneous situations in terms of number of customers, water suppliers' management models and corresponding tariff schemes.

In 2009, the Portuguese Water and Waste Services Regulation Authority (ERSAR) started to produce recommendations on tariff schemes to be followed by all operators, with the aim of harmonizing tariff structures and ensuring the consumers' right to accurate, reliable and clear information.

The ERSAR intervention regarding the supervision and control of prices focuses on the accomplishment of guidelines and principles comprised in three main recommendations. Relevant recommendations include specific guidelines for the tariff structures (Recommendations 1/2009 - Tariff Guidelines and 2/2010 - Calculation Criteria) and the for the bills content (Recommendation 1/2010 - Bill Content).

According to the ERSAR recommendations, tariffs structures to be applied to end-users should always comprise two components: a fixed charge component (in order to allocate fixed costs for all consumers), and a volumetric component with four increasing blocks. In addition, it presents some criteria for tariffs differentiation, in particular taking into account the type of end-users (residential versus non-residential users).

The recommendation for implementing tariffs schemes with a high number of components (fixed and volumetric) as a response to multiple objectives contrasts with simplicity demands. The complexity of tariff structures translates into water bills which are difficult to be understood.

4. Methodological approach

This research took water bills as a case study to illustrate the need to understand the data quality requirements from the customers' perspective. Water invoices were selected because they are a fundamental communication channel between water utilities and consumers.

Invoices convey essential information mostly in terms of prices and quantities. Understanding tariff schemes and being able to monitor water consumption is very important to make water polices effective, especially those that concern the preservation of the resource.

The identification of the requirements of a "good" water invoice was based on the views of water consumers. Focus groups were used to collect the consumers' opinions on how invoices should look like. Details on how focus groups were used are given in section 4.1. Data thus collected was then analysed by means of an affinity diagram (KJ). The technique is briefly described in section 4.2. The output is thus a structured list of data quality requirements. Such list was confronted with the data quality dimensions proposed in the literature and also compared with recommendations on water bills from the economic regulator.

4.1. Using Focus Groups to collect data from domestic customers

Focus groups are widely used in qualitative research to collect data when listening to participants within the context of groups is important since the interaction among them can lead to interesting insights. The method is particularly suitable when the topic is poorly understood and the research stage is that of exploration and discovery. Identifying how the "ideal" water invoice should look like is not straightforward and the researchers considered that interesting ideas could emerge if groups of consumers were brought together to discuss the difficulties they had in understanding water bills and what they thought could be done to make their lives easier when contacting with such documents.

Domestic customers are the study target population. The first eligibility criterion to participate in the group was to be the person in charge of paying the water bill, since this person is typically the one who has more contact with water invoices. Due to the importance of considering different levels of expertise in analysing this type of documents, three focus groups were organised. One group (FG1) was made of people with a background in economics/management and who were supposed to feel at ease with tariff schemes. Moreover, given their background, FG1 participants were expected to understand better the water industry and its market mechanisms. On the other hand, FG 2 members had no academic degrees and typically have lower incomes than the participants of the other groups. They were expected to have a bit more difficulty in interpreting water bills, especially in terms of literacy and numeracy. Finally, FG3 is composed of people who deal with tasks of gathering and organising information on a daily basis in different fields and/or who have responsibilities in helping less knowledgeable persons in making sense of complex data. FG 3 participants were thus expected to be more aware of the problems people have in dealing with billing documents, while having ideas of their own on how to better present relevant data. Thus, not so much demographic characteristics but rather similarities on key experiences were taken into account. This segmentation allowed the researchers to listen to 21 individuals and ensured a certain degree of affinity among the members that facilitated interaction. Each focus group had 7 participants.

The researchers developed a flexible set of questions to guide conversations, but considerable freedom was given to the flow of ideas, as the technique requires. The sequencing of questions had in mind the importance of allowing the participants to become familiar with the topic, giving individuals the chance to recollect personal experiences and opinions and listen to others. The semi structured interview guide provided the broad topic and areas to be explored with each group (some common, some specific). The aim was to elicit unprompted responses from the customers on the characteristics that an "ideal" water invoice should have. To refresh the minds of the participants on how different water invoices might look like, the researchers brought to the meetings three invoices as possible examples.

With the participants' agreement, conversations were taped recorded. The duration of the focus groups varied a little, but each conversation took between 60 and 90 minutes. As

verbatim transcriptions of the focus groups were produced and those formed the basis of the subsequent analysis.

4.2. Analysing data using the affinity diagram

After data was collected and the transcriptions were done, it was necessary to make sense of what was said in the groups. A content analysis of the transcripts was undertaken to tease out categories, patterns and themes. Using different iterations, an adequate coding of data was found.

For the purposes of the current paper, data analysis starts with the identification of the water invoice requirements, as stated by the participants of the focus groups (see next section).

Affinity diagrams (also known as KJ, due to the initials of their creator, Jiro Kawakita) are used to structure qualitative data (Akao, 1997). Data are grouped according to their semantics. In order to foster divergent thinking and some creativity in the way data is sorted and grouped, the technique should be used in the context of teams.

The process of building a KJ diagram typically covers the following stages:

- Forming a team of 3 to 6 elements (as a guideline);
- Choosing the question/problem to be addressed and writing it in a visible spot;
- Writing a certain number of statements of fact in post-its, each of them corresponding to a view/requirement;
- Arranging facts in groups that reflect similar ideas or feelings and creating headers for groups;
- Arranging groups and showing relationships among them;
- Writing concluding statement and reflecting on the process.

5. Key findings

Before explicitly addressing how water customers wanted they water invoices to look like, some questions were raised to realize at what extent the participants analyse the information conveyed in water invoices and what kind of problems they have in understanding it.

5.1.Relevance given to water bills and invoice information

The vast majority of the participants comes in contact with water invoices every month, but its attention is mainly focused on the amount to be paid. A significant number of the focus groups members assumes that understanding the water bill is not a major concern, due to its relatively low weight on the family budget, especially if compared with other SGI, the electricity in particular.

Concerns with moderate consumption of a scarce resource were not much on the spotlight during the meetings. Only a minority, with the common feature of having children at school age, has explicitly mentioned the importance of the resource sustainability:

"I keep thinking… children, baths… I wonder if we are adopting a reasonable behaviour…"

The vast majority of the individuals only look in some detail to the invoices when an "out of the pattern" situation becomes apparent. Many participants have an amount in mind that works as a "benchmark" to identify such situations, whereas others have the tendency to compare the amount to be paid with the one from the previous month.

"If there is nothing abnormal, I do not look at the invoice".

"As far as there are no huge deviations from the previous month, everything is OK".

5.2.Main problems and issues to be addressed

When asked about the problems they have in understanding the information conveyed in water bills, the majority of the participants mentions the multiplicity of the items/services charged along with water consumption:

"It's the residuals, it's the sewage, it's goodness knows what… it's the environment".

Difficulties in understanding water bills seem to be associated with the invoices which are commonly regarded as complex. The following quotations exemplify the matter:

"The designations used are hard to understand".

"Service quote? What´s that? No one knows".

"One comes across "weird things".

To make things worse, some issues related to the way information is organised and presented are highlighted:

"Here the subtotals, there the total, this is poorly organised".

"The font size is too small".

Yet, the complexity of tariff schemes is also emphasised. To some participants, with higher qualification levels, the complexity of water bills comes from the tariff structure and not from the invoices on themselves, as illustrated bellow:

"The system [tariff structure] is too complex. It is impossible to explain it… A lot of items, some fixed, some dependent on consumption level, it absolutely impossible to have a clue of what's there".

5.3.Voices of the customers on the "ideal invoice"

During the focus group meetings several ideas on how the "ideal invoice" should look like have emerged, both as a way to overcome some of the problems discussed earlier and as a result of the group analysis of some water invoices brought to the meetings to refresh the participants' memory.

After analysing the transcriptions of the three group interviews (and following a process of eliminating duplications), fifty-seven customer voices (VOCs) were identified. As much as possible, the 'customer voices' preserved the participants' exactly words, but, in the few cases they sounded too vague or confused, the researchers re-wrote the requirements taking into consideration the context in which the idea had come up.

The final list follows:

- VOC1: Invoices should show the amount due;
- VOC2: Invoices should include the blocks;
- VOC3: Invoices should include "pictures"
- VOC4: Invoices should include the previous consumption - history;
- VOC5: The history of consumptions should be depicted in a graphic that can be quickly analysed;
- VOC6: Invoices should state the kind of meter reading (estimated/actual)
- VOC7: Invoices should be intuitive
- VOC8: Invoices should be visually attractive;
- VOC9: Invoices should be comprehensive;
- VOC10: Invoices should state how much is paid for each service (water, wastewater and solid waste);
- VOC11: Invoices should show if there are amounts due from previous periods;
- VOC12: Invoices should be informative;
- VOC13: Acronyms should not be used;
- VOC14: Invoices should not have small prints;
- VOC15: Invoices should indicate what is actually measured and what is indexed to other components;

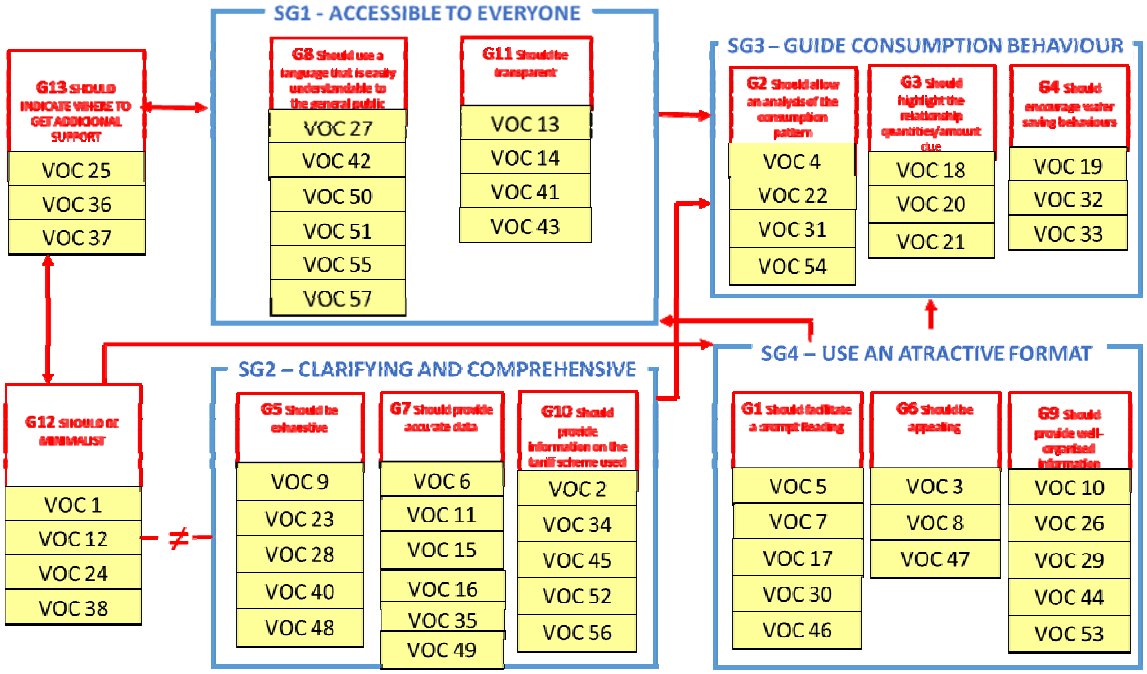
- VOC16: The graphics should indicate what is being represented/depicted;
- VOC17: The most relevant information should be on the first/front page;
- VOC18: Invoices should highlight what the consumer can somehow control (i.e. the quantities consumed);
- VOC19: Invoices should include a slogan saying “save water”;
- VOC20: Invoices should show the consequences of potential water savings;
- VOC21: Invoices should explain why the amount due is higher/lower than it was last month;
- VOC22: Invoices should convey information on the “standard” consumption of a household similar to mine;
- VOC23: Invoices should convey as much information as possible;
- VOC24: Invoices should state how much has been consumed;
- VOC25: Invoices should supply information on where/how to get the full tariff scheme;
- VOC26: Invoices should separately show variable charges, fixed charges and then the total amount due
- VOC27: The terminology used should be straightforward, with no double meanings;
- VOC28: Everything should be explained in detail;
- VOC29: Invoices should use tables;
- VOC30: Invoices should have a quick overview of what’s there;
- VOC 31: Comparisons should be possible;
- VOC32: Invoices should help to save water;
- VOC33: Invoices should state “turn off the tap”
- VOC34: Invoices should indicate the lower and upper limits of each block;
- VOC35: Invoices should explicitly state “the billing period is from day X up to day Y”;
- VOC36: Invoices should indicate how to make a complaint;
- VOC37: There should be a number to call in case we want to complain;
- VOC38: The invoice should be short (in length);
- VOC39: Invoices should communicate things in a way anyone understands;
- VOC40: Invoices should show the full tariff schedule whenever it changes;
- VOC41: Invoices should clarify what each component means;
- VOC42: Invoices should not use technical words;
- VOC43: Invoices should state in ‘plain Portuguese’ what we are paying for, in a discriminate way;
- VOC44: Invoices should show the weight of each component/service using a graphic;
- VOC45: The price per cubic metre should be clearly stated;
- VOC46: The amount due should be highlighted;
- VOC47: Invoices should be ‘clean’;
- VOC48: The values should be discriminated;
- VOC49: The units of measurement should be clearly stated;
- VOC50: Invoices should be simple;
- VOC51: Invoices should be clear;
- VOC52: Invoices should indicate whether each component is fixed or variable;
- VOC53: Information should be well-organised;
- VOC54: Invoices should help to monitor consumption over time;
- VOC55: Information should be easy to understand;
- VOC56: Invoices should emphasise the existence of increasing block tariffs;
- VOC57: Invoices should be “friendly”.

5.4.Underlying data quality requirements

Such voices were then structured by building a KJ diagram. The researchers and a small number of participants have looked at potential affinities among different VOC that thus shared some kind of meaning. The process is naturally somehow group-dependent and subjective, but highly rewarding and creative.

The grouping stage took long due to the large amount of data. At the end, as Figure 1 shows, thirteen first-order groups (red titles) and four second-order groups (blue titles) emerged (in the second level of aggregation two first-order groups remained as "lone wolves".

Figure 1: Affinity diagram



6. Discussion: data quality attributes of the 'the ideal' invoice

The affinity diagram provides an integrated and interesting view on data quality requirements on what water invoices is concerned.

The analysis of the diagram clearly draws attention to the difficulty of simultaneously meeting all requirements. In fact, contradictory demands exist: if, on one hand, consumers want to have everything explained in detail, without abbreviations and discriminating all items, on the other hand look for a short document, in certain cases even minimalist in its content. This tension cannot be totally overcome. Yet, one can argue that invoices possibly are not the most adequate channel to give complete information on everything that might be involved in service provision. What really matters is to state very clearly where and how additional information can be obtained. An example of the adequacy of such strategy is that of communicating the full tariff schedule. Although this information is highly relevant, providing the full schedule in every monthly bill would possibly be counterproductive. When no changes are made, information can be given on how to get it (website address, for instance).

An additional aspect that became immediately visible when conducting this study is the strategic importance of adopting an appropriate language. The issue was raised in various

circumstances and is also clearly depicted in the KJ diagram with six voices relating to consumers demands of a language that is easily understandable to the general public.

Looking at customer requirements it is also possible to state that water providers need to give attention to both content and format issues. Together they can make water bills accurate, quickly analysed and appealing.

It is legitimate to wonder if invoices could/should not be used to convey, on a simple and straightforward way, information on tariff schemes (that are intrinsically complex) so that parsimonious water consumptions would be effectively encouraged (regardless of the consumers' age, education or income level).

Unless concerns of clarity, accessibility, transparency and even 'attractiveness' are addressed, it is highly predictable that invoices will not entirely fulfil their role in driving consumption behaviours. The request for better showing the relationship between quantities consumed and amount due gives a clear indication in this regard.

It is interesting to notice that the various data quality dimensions proposed in the literature (see section 2) find expression in either the customer voices collected or in the guidelines given by regulatory bodies in recommendations regarding water bills (or in both). In this later case the Portuguese water bill recommendation (ERSAR's Recommendation 1/2010) was considered, since it explicitly establishes the contents that every operator should have in the water invoices it sends regularly to its customers. There is thus an obvious concern from the regulatory body to guarantee harmonised and transparent bills.

ERSAR Recommendation 1/2010 comprises some general principles, a few broad-spectrum provisions and then a set of more specific rules. In the preamble, the recommendation states "invoices should respect the principle of transparency and be easy understood by the final user, giving information on the operator and on the customer, specifying the services provided, the applicable charges, the forms of payment and other relevant information'. In the introduction, it also establishes that "invoices should adopt a simple format and use a straightforward language to facilitate reading and understanding".

Table 1 relates the data quality dimensions with the customer voices taking into account their meaning. Additionally, it uses an intensity scale to evaluate at what extent each data quality dimension is underlined on the various Recommendation guidelines. It is possible to observe in Table 1 that consumer requirements cover the various data quality dimensions, with the exception of consistency (since the idea of keeping uniform criteria over time or across users was not explicitly mentioned). In addition, the ERSAR Recommendation is in line with various data quality dimensions, some of them being explicitly addressed at various points.

Table 1. Data Quality Dimensions in the context of water invoices

Data quality dimensions	VOC number	Recommendation ERSAR 1/2010
Accuracy	6; 15; 35; 49	+++
Completeness	9; 10; 11; 21; 23; 28; 34; 45; 48	+++
Low level of detail	7; 30; 38; 47	n.f.
Relevance	1; 2; 12; 18; 19; 20; 24; 32; 33; 54; 56	++
Unambiguousness	16; 27; 41; 51; 52	n.f.
Understandability (by the general public)	13; 26; 39; 42; 43; 50; 55; 57	+
Consistency		+
Timeliness	40	+
Comparability	4; 22; 31	++
Adequate format	3; 5; 8; 14; 17; 29; 44; 46; 53	n.f.
Accessibility	25; 36; 37	++

Notes: +++ Explicitly and repeatedly addressed; ++ Explicitly addressed in more than one point; + Implicit or only occasionally addressed; n.f. not evidence of found in the text.

Completeness is a major concern. Several VOC highlight the importance of giving all information on water bills, detailing all items and fees. The regulatory body also has different guidelines that emphasise the operators' obligation of giving details on everything that is being charged to the consumers. As mentioned before, completeness is in potential conflict with the users' preference to a short and synthetic document. The challenge of combining exhaustiveness with synthesis is not easy. Giving emphasis to a quick and useful summary is recommended.

A broad dimension that necessarily means different things to different people is that of relevance. Consumers show the importance they give to the information conveyed in water bills in making decisions when they mention different content requirements, especially those that highlight the relationship between quantities and the amount due.

The regulator concerns with accuracy are clearly visible in the Recommendation, which calls for a high level of rigor and precision. On the other hand, the dimension is occasionally mentioned by the consumers.

Accessibility here is essentially related to giving details on contacts that can be used to obtain further information. There seems to be a certain consensus that invoices cannot include all possible data and that some information can (and should) be obtained through alternative channels.

On the other hand, understandability and unambiguousness are major priorities for the water consumers that participated in the study as the high number of voices on these matters indicate. For the regulator, some general principles reflect these concerns, but they are not translated into specific guidelines and rules, with the exception of one that refers the need to explain all abbreviations. For the users, such concerns are much broader, including the kind of language used. The regulator embraces some of such concerns not by directly giving indications on how to design invoices, but rather through a set of frequently asked questions (FAQs) in the ERSAR website about billing and reading of water and waste services.

Moreover, format issues, which have been up to now relatively neglected by the regulator, are regarded as important by water consumers. This reinforces the idea that both format and content need to be taken into consideration in designing effective invoices.

It shall also be noticed that some of these quality dimensions are similarly highlighted by regulatory authorities as well as by providers of water services around the world. Even without conducting an extensive and systematic review, it is possible to find some interesting examples. In Belgium, the lntercommunale Bruxelloise de Distribution d'Eau establishes that operators have the obligation to send clear and legible invoices. In the United States, "accuracy of invoices" is one of the indicators considered by American Water Works Association when analysing the performance of operators under the dimension "relationship with consumers". In Australia, invoices have to include the tariff schedules of all services that are charged and the operator is compelled to give free access to information on over-invoicing, payment forms and how to access to EWOV, a site that assists consumers in understanding bills. A service is also provided to consumers with listening or speaking disabilities to help them in reading and understanding their bills.

6. Conclusion

The current research looked at the data quality requirements that should be considered by water service providers when designing invoices. The water industry was chosen as an example of a SGI where special care has to be taken in communicating information to users, given their relatively dependency upon providers due to the lack of alternatives, the essentiality of the services at stake, and the existent information asymmetries.

Data quality issues, despite extensively discussed in the literature, have been almost exclusively analysed in the context of information systems for organisations. As it happens in other empirical studies, data quality dimensions were found to be associated with many of the requirements on water invoices identified by domestic consumers and with some of the recommendations of regulatory bodies on the matter. Moreover, the methodology used to uncover customer requirements is rather original in this context and was very successful in leading to a vast and diverse set of ideas on how the 'ideal' invoice should look like. The requirements relate to both content and format and give particular emphasis to language matters.

Regulatory authorities are aware of the importance of transparency and give recommendations on the bills content. In the Portuguese case, despite the importance of the guidelines produced, the concerns of domestic users go well beyond what is established.

Moreover, the research shows that concerns with users who have more difficulties in reading water bills need to be at the core when designing invoices. In fact, the appropriate communication of prices and volumes is special relevant due to the importance of protecting the rights of consumers. The consequences of poor data quality in the water services context are rather significant because consumers cannot switch between service providers, minimal levels of consumption are essential to living and the effectiveness of public policies is clearly jeopardised if consumers do not understand price signals.

Invoices are essential instruments in communicating information to service users. Findings show that plain language efforts and strategies to enhance trust on the service provided by water utilities need to be further reinforced.

This list of requirements produced can be regarded as the first stage in the development of a model to assess the quality of water invoices. The process can easily be replicated to other SGI.

A few principles, potentially applicable to different SGI, have emerged from the study carried out:

- Avoiding technical language (that is not accessible to all customers and that raises some suspicion);
- Making visible the consequences of different consumption levels on the amounts to be paid (as a way to protect the consumer and to improve resources sustainability);
- Developing multiple and transparent communication channels.

References

Akao, Y. (1995)**, "**QFD Toward Product Development Management", *Proceedings of International Symposium on Quality Function Deployment*. Tokyo: Union of Japanese Scientists e Engineers.

Ballou, D. P., Madnick, S. & Wang, R. (2004), "Assuring information quality", *Journal of Management Information Systems*, 20, 9-11.

Commission of the European Communities (2004)**,** *White Paper on Services of General Interest*, COM (2004) 374 final (12 May 2004). European Commission, Brussels.

Eppler, M. & Helfert, M. (2004), "A classification and analysis of data quality costs", *MIT International Conference on Information Quality*, November 5-6, Boston.

Haug, A.; Zacharissen, P. & van Liempd, D. (2011), "The cost of poor data quality", *Journal of Engineering and Management*, 4(2), 168-193.

Martins, R., Cruz, L., Barata, E. (2013), "Water Price Regulation: A Review of Portuguese Tariff Recommendations", *Public Organization Review*, 13(2), 197-205.

OECD (2013), *OECD Skills Outlook 2013: First Results from the Survey of Adult Skills*, OECD Publishing. *http://dx.doi.org/10.1787/9789264204256-en*

Recomendação ERSAR n.º 1/2010, *Conteúdos das faturas dos serviços públicos de abastecimento de água para consumo humano, de saneamento de águas residuais urbanas e de gestão de resíduos urbanos prestados aos utilizadores finais ("Conteúdos das faturas")*

Tayi, Y. & Ballou, D. (1998), "Examining data quality", *Communications of the ACM*, 41(2), 54-57.

Wang, R. & Strong, D. (1996), "Beyond Accuracy: What Quality Means to Data Consumers", *Journal of Management Information Systems*, 12(4), 5-33.