Data quality requirements for water bills

Patrícia Moura e Sá and Rita Martins
Faculty of Economics, University of Coimbra, Coimbra, Portugal

Abstract
Purpose – The purpose of this paper is to uncover the customers' concerns with the information disclosed in water services invoices and to analyse them with reference to the data quality dimensions usually proposed in the literature. In the context of services of general interest invoices are particularly relevant as a vehicle to convey information to all consumers.

Design/methodology/approach – Based on the principles of quality planning, the research uses a qualitative approach to identify the data quality requirements of water invoices. Customer voices were collected by means of focus groups and their meanings analysed using an affinity diagram.

Findings – Findings show that plain language efforts and strategies to enhance trust on the service provided need to be further reinforced. Consumers' requirements together with the regulator recommendations also confirm the data quality dimensions identified in the literature.

Practical implications – This research highlights that avoiding technical language and making visible the consequences of different consumption levels on the amounts to be paid is essential when designing water invoices. Moreover, it emphasises that there is still room for improvement in the way the economic regulator performs its role in ensuring the provision of sound information.

Originality/value – This research addresses a literature gap by conducting a study on data quality requirements outside the context of information systems for organisations. The study is original because it looks at water invoices as a "product" that can be designed to meet the needs of their users.

Keywords Utilities, Qualitative methods, Customer requirements, Water industry

Paper type Research paper

1. Introduction
Data quality is not a recent topic of research, with data quality literature proposing a set of dimensions that typically include accessibility, clarity, relevance and accuracy. Yet, empirical studies outside the context of information systems for organisations are scarce and usually do not give information users a central role in defining data quality requirements. The current research is expected to address this gap.

The purpose of this research is twofold: to uncover the features of an "ideal" water invoice according to the perspective of domestic consumers; and to assess at what extent the data quality dimensions usually proposed in the literature are embedded on the customers' concerns with the information disclosed in water services invoices.

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 services of general interest (SGI). Invoices are a channel that necessarily reaches all the customers on a regular way, conveying information on quantities and prices. Therefore, it is crucial that the information disclosed through invoices is adequate and clearly understood.

Several international frameworks call attention to the importance of giving quality information to consumers as a way to ensure consumers' rights and promote adequate consumption behaviours. According to the white paper on SGI (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. The United Nations also acknowledges the importance of providing individuals with transparent and open information on water and sanitation services. In this regard "States must […] ensure that information relating to standards, as well as progress towards meeting those standards, is available and accessible, and that the mechanisms (including service delivery options) used to ensure that these standards are indeed met are available and accessible to all duties", Albuquerque (2014a, p. 30). Moreover, individuals must be aware of their rights and without the need for direct requests to get access to information.

Since 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 (2013) latest findings on adult skills emphasises that 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. This means 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 to overcome some of the problems raised by the low levels of literacy[1] and numeracy[2] of a significant part of the population even among developed countries.

This scenario calls attention to the importance of listening to consumers when designing several documents – in this case water invoices – so that their needs are met and communication problems are minimised. Quality planning methodologies and tools can be particularly useful in this regard.

Quality planning processes should be customer-driven to assure that the product/ service meets customer needs and expectations (Hauser and Clausing, 1988). Consequently, in quality planning methodologies, such is the case of the quality function deployment (QFD), everything starts with the voice of the customer (VOC) and product/service specifications derive from the customer needs rather than from the feelings of the firm technicians or marketers. As Akao and Mazur (2003, p. 20) state, the methodology assures that "customer needs are properly deployed throughout the design, build and delivery of a new product". The process, thus, converts user demands (or client requirements, Carnevalli and Miguel, 2008) into quality characteristics.

The current study is original given that it looks at water invoices as a "product" that can be designed to meet the needs of their users. Based on the principles of quality planning, customers' voices were collected by means of three focus group interviews. The use of focus groups makes it possible for domestic consumers to express their views freely, in their own words, while benefiting from interaction with other customers with similar experiences. Customers' voices were examined and translated into 56 customer requirements, which were analysed using an affinity diagram (KJ) (Akao, 1995). Those requirements were then compared with the water regulator recommendations on the matter. The process and the resulting outcome is thus very similar to the one traditionally used to identify the "WHATs" in the house of quality (HOQ) (Hauser and Clausing, 1988; Prasad, 1998). As Prasad highlights, WHATs are a list of customer wants or customer requirements that result from listening to customers at the first stage of the QFD methodology. The quality characteristics correspond to the

HOWs, which determine the set of alternate quality features to satisfy the customer wants (WHATs).

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, in Section 3 the Portuguese water industry (the focus of the current study) is briefly described in what concerns its market structure and price regulation mechanisms. Section 4 presents the methodological approach adopted in the study. Section 5 shows the key ideas that emerged from the focus groups related to the way people look at water bills, the problems they face in reading them and 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 authority recommendations on the same subject 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 from the public policies standpoints.

The distinction between "data" and "information" is well-established in the literature. Davenport and Prusak (1998) stress that data are defined as "discrete, objective facts about events", whereas information relate to "data transformed by value-adding processes of contextualisation, categorisation, calculation, correction and condensation". In practice, this distinction is not completely straightforward. The current research is focussed on invoices that are used to 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) group data qualities into four categories: completeness, unambiguousness, meaningfulness and correctness. Later, Haug *et al.* (2009) propose three categories: intrinsic, accessibility and usefulness. Similar attributes were identified by Leung (2001) when measuring software data quality. Based on the literature and taken into account the most cited data attributes and their description, Table I was built.

It becomes clear that the majority of the data quality attributes is rather dependent on data users and uses. End-users importance in assessing data quality is also emphasised by Leung (2001). 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), Wang and Strong (1996, p. 20).

In any case, data quality needs to be assessed taking into account the extent to which data meets customers' needs and expectations. Therefore, data quality 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", raised by Juran (1992), should

Main data attributes	Description
Relevancy	Data are applicable, interesting and usable
Accuracy	Data are error-free, correct, precise, with integrity
Ease of understanding	Data are clear, readable
Accessibility	Data are retrievable, available
Completeness	Breadth, depth, and scope of information contained in the data
Timeliness	Age of data
Objectivity	Data are unbiased
Reputation	Data comes from a reputed source
Representational consistency	Data are continuously presented in the same format, consistently formatted, data are compatible with previous data
Flexible	Data are adaptable, extendable
Cost-effectiveness	Cost of data accuracy, cost of data collection
Ease of operation	Data are easily updated, easily downloaded/uploaded, data can be used for multiple purposes, easily customised

Source: Adapted from Wang and Strong (1996)

Table I.
Summary of data
attributes identified
in the literature

be adopted. As Ballou *et al.* (2004, p. 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 "analyse 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, p. 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; and

- 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 are 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, p. 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".

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 categorisation 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, e.g. 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.

Ensuring the quality of data for SGI is more difficult and more relevant than is the case with manufactured goods, given their intangibility 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 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.

3. Brief description of the Portuguese water industry case
Portuguese water industry, as usual elsewhere, has natural monopoly characteristics, which means that costs are minimised if services are provided by a single firm rather than if the output(s) is (are) provided by several operators. When these market structures are in place, promoting competition is not efficient, thus customers cannot choose their provider.

Moreover, 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 regarded, reasons for concern came not only from monopolistic market structures, but also from asymmetries of information between service providers and consumers. Transparency and consumer protection is particularly relevant in what concerns to quality and price schemes.

In Portugal, the water industry is very fragmented, with almost one service provider per municipality (308 Portuguese municipalities for less than 11,000,000 inhabitants). Operators of water supply services are generally small but there are heterogeneous situations in terms of governance models, number of customers and tariffs charged.

There is an independent economic regulator, the Water and Waste Services Regulation Authority (ERSAR), whose scope is the regulation of water supply, wastewater and solid waste services. This authority also has competencies as the national authority for drinking water quality. Among its roles "ERSAR is engaged with

ensuring adequate protection for consumers and users of water supply and waste services by promoting the quality of service rendered by the operators and guaranteeing socially acceptable pricing, materialized in the fulfilment of the following principles: essentiality, indispensability, universality, equity, reliability and cost efficiency associated with the quality of the service" (ERSAR, 2013, p. 17).

In 2009, ERSAR started to produce recommendations on tariff schemes to be followed by all operators, with the aim of harmonising tariff structures and ensuring consumers' right to accurate, reliable and clear information. The ERSAR intervention regarding the supervision and control of prices focusses on the accomplishment of guidelines and principles comprised in three main recommendations. Relevant recommendations include specific guidelines for the tariff structures (IRAR, 2009; ERSAR, 2010b) and the for the bills content (ERSAR, 2010a).

According to the ERSAR recommendations, tariffs structures to be applied to end-users should always include two main components: a fixed charge component (in order to allocate fixed costs for all consumers), and a volumetric component with four increasing blocks. Special tariffs for low-income households and for larger households are recommended. The recommendation for implementing tariffs schemes with a high number of components as a response to multiple objectives (economic efficiency, environmental sustainability, financial stability, public and political acceptability) contrasts with simplicity demands. Even if compliance with tariffs content recommendation (ERSAR, 2010a) were 100 per cent, it would still be expected that invoices were complex due to the tariff schemes in place. Complexity also comes from the charge of wastewater services, and in almost all municipalities of solid waste charges, according to the volume of water consumed, as well as VAT and other taxes.

It is relevant to notice that each Portuguese household has a water metering device (the coverage rate is about 98 per cent) and the payment of the water service is partially dependent on the level of water of consumed. Moreover, each water provider has the obligation of issuing monthly water invoices.

4. Methodological approach

From the general purpose of analysing the data quality requirements for water invoices from the domestic consumers' point of view, the following research questions are addressed in this paper:

RQ1. What characteristics should an invoice have to meet the demands and concerns of domestic consumers?

RQ2. At what extent do the data quality dimensions identified in the literature cover the customer requirements that derive from such demands and concerns?

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 policies effective, especially those that concern the preservation of the resource.

The identification of the requirements of an "ideal" 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. Section 4.1 describes the essentials of QFD and

presents some tools associated with it and employed in this research, such are the cases
of focus groups and affinity diagrams. Then, in Section 4.2 details are given on how
focus groups were used within this project to collect customer voices. Such voices were
analysed using an affinity diagram, as explained in Section 4.3.

4.1 The QFD methodology

The QFD methodology was introduced in Japan in the 1970s to assist the development
of new products focussed on the customer needs. The main idea, as Mazur (1993) points
out, is that listening to the "voices of customers" throughout the development process is
essential to create real value to them. QFD then works as a planning methodology that
carries out the VOC through the different stages, while documenting the whole process
in a way that fosters teambuilding and consensus within the development team (which
may or may not include customer representatives). Using QFD the process of "seeking
out both spoken and unspoken needs, translating these into actionable (products and)
services, and communicating this throughout the organisation" (Mazur, 1993, p. 2)
becomes more effective and efficient.

For the purposes of the current project, the QFD methodology was used since it
"invites the participation of the customers who could affect or be affected by the
proposed design" (Eldin and Hikle, 2003, p. 314). The researchers considered that
collecting domestic customers' views would be essential to design an invoice that fulfils
their needs and can be easily understood by those who need to make decisions
regarding their water consumption behaviours.

The QFD methodology encompasses four major stages: product planning, part
deployment, process planning and production planning (Chan and Wu, 2002).
The first stage roughly corresponds to the fulfilment of the HOQ, which involves
collecting customer needs (called WHATs) and then transforming these needs
into technical requirements (called HOWs). Thus, product planning is of fundamental
and strategic importance in the QFD system. It is the basis of the whole development
process and is the focus of this research project too. Clearly, once defined the
features of the "ideal" invoice, efforts need to be put on creating a prototype and
implementing it.

The project has therefore started with the identification of the "water invoice
customers". Although different types of customers exist, in this situation ultimate
customers (domestic consumers) were considered the most important user of this kind
of document. Next, their needs were collected by means of focus groups interviews.
The process is described and justified in greater detail in the next section, but the key
point in collecting customer needs is to value customer voices, where needs are
expressed in their own words to avoid misinterpretations and/or the tendency to
"jumps" into conclusions regarding what customers want at early stages. As Chan and
Wu (2002) mention, there are several methods for collecting customer needs, including
surveys, focus groups, natural field contact, complaints, among others. Focus groups
were used in this research since they allow the collection of customer voices with few
constraints and in a way that benefits from the interaction among members. In fact,
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. Thus, as suggested in the QFD literature (Chan and Wu, 2002; Eldin and Hikle, 2003), the focus group technique allowed the researchers to identify the expectations, feelings and thinking patterns of the consumers regarding water invoices.

The VOC collection stage usually leads to a great amount of data that needs to be properly analysed. As Chan and Wu (2002, p. 28) stress, "customer words are usually too general and/or too detailed to be directly used as formal customer needs". Therefore, apart from extracting the VOC from the interviews transcripts, it is necessary to rephrase them as customer requirements and to understand their meaning. In this regard, some tools can be used to organise the words collected in a tree-like hierarchical structure, which calls attention to the various levels of customer needs (Chan and Wu, 2002). One of such tools is the affinity diagram.

Affinity diagrams (also known as KJ, due to the initials of their creator, Jiro Kawakita) are used to structure qualitative data (Akao, 1995). As Chan and Wu (2002, p. 28) point out, an affinity diagram is "a method of organizing qualitative data into natural and logical groups". Data are grouped according to their semantics. In order to foster divergent thinking and some creativity in the way data are sorted and grouped, the technique should be used in the context of teams. Within quality planning, affinity diagrams are particularly suitable since they foster research teams to group ideas/sentences with similar meaning and to evaluate possible associations/links among such groups of ideas. The use of affinity diagrams to structure requirements was applied within various QFD projects (see, e.g. Chan and Wu (2002), Eldin and Hikle (2003), Mazur (1993)). Mazur (1993, p. 5) argues that affinity diagrams can be used to surface "the deep structure in voice requirements". Groups of ideas with similar meanings (affinities) receive labels corresponding to the topic underlying them. These topics, in turn, are grouped into higher-level themes, and so on (Eldin and Hikle, 2003).

4.2 Using focus groups to collect data from domestic customers
In collecting customer voices using focus groups, attention has to be given to three main stages: selection of the group participants; development of questionnaires for the focus group sessions; and utilisation of qualified moderators (Eldin and Hikle, 2003).

Domestic customers are the study target population. The first eligibility criterion to participate in the focus groups 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, FG2 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 was 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. FG3 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 seven 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 within each group. The members were asked to explain how they use the information conveyed through invoices, to discuss the difficulties they might have in understanding their content and to state features they want an "ideal" invoice to have. The aim was to elicit unprompted responses from the customers on the requirements of an "ideal" water invoice. 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. As recommended in the literature (Eldin and Hikle, 2003), the researchers (who acted as session moderators) communicated the rules for the session, explaining the purpose, emphasising that there were no right or wrong answers, and encouraging the participants to share their points of view.

With the participants' agreement, conversations were tape-recorded. The duration of the focus groups varied a little, but each conversation took between 60 and 90 minutes. Verbatim transcriptions of the focus groups were produced and those formed the basis of the subsequent analysis.

4.3 Analysing data using the affinity diagram

After data were gathered and the transcriptions were done, it was necessary to make sense of what was said in the focus groups. For the purposes of the current paper, data analysis is focussed on the identification of the water invoice requirements, as stated by the participants of the focus groups (see next section). Since this research follows an inductive approach, the data analysis procedures applied aimed at extracting the key ideas expressed in the interviews and finding the meaning underlying such customers' voices.

Briefly stated, any content analysis aims to tease out categories, patterns and themes.

The process of building a KJ diagram typically covers the following stages (Akao, 1995):

- forming a team of three to six 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; and

- writing concluding statement and reflecting on the process.

In the current project, the group used the requirements collected in the focus group and wrote each of them on a separate card. Ideas were grouped according to their similarity and posted next to each other. As stressed in the literature (Mazur, 1993), the customer requirements hierarchy was built from the bottom up, and the relationships between

customer wants were based on the intuition of the team. The label for each cluster of customer requirements was found after an extensive team discussion. The process was interactive. After three rounds, where cards were moved around, consensus started to emerge and resulted into the tree-like structure presented in Figure 1.

5. Key findings

Before explicitly addressing how water customers wanted they water invoices to look like, some questions were raised to realise 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 focussed 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 […].

A very large majority of the individuals only look in some detail to 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 illustrate 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.
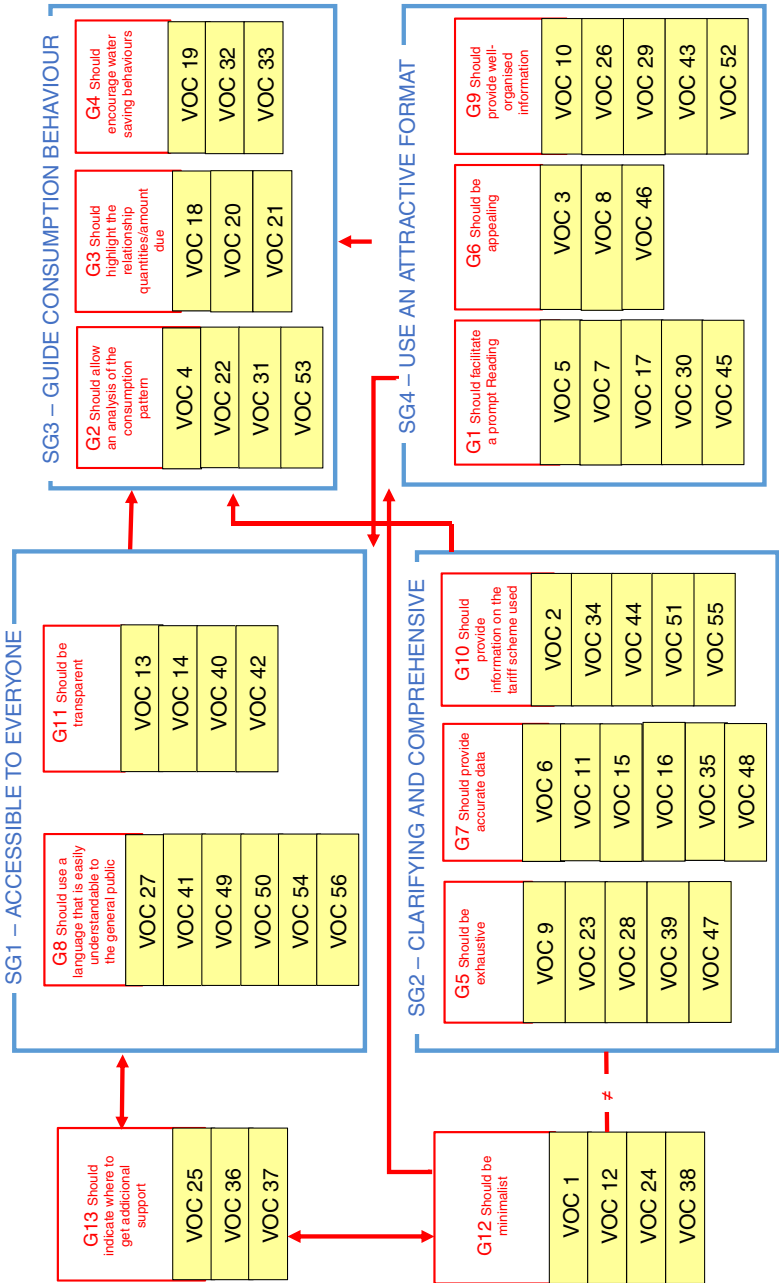
> The font size is too small.

Figure 1.
KJ for a water
invoice

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 is 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 focus group interviews (and subsequent to a process of eliminating duplications), 56 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 metre 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 cubic metres 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;
- VOC31: 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: invoices should be short (in length);
- VOC39: invoices should show the full tariff schedule whenever it changes;
- VOC40: invoices should clarify what each component means;
- VOC41: invoices should not use technical words;
- VOC42: invoices should state in "plain Portuguese" what we are paying for, in a discriminate way;
- VOC43: invoices should show the weight of each component/service using a graphic;
- VOC44: the price per cubic metre should be clearly stated;
- VOC45: the amount due should be highlighted;
- VOC46: invoices should be "clean";
- VOC47: the values should be discriminated;
- VOC48: the units of measurement should be clearly stated;
- VOC49: invoices should be simple;
- VOC50: invoices should be clear;
- VOC51: invoices should indicate whether each component is fixed or variable;

- VOC52: information should be well-organised;

- VOC53: invoices should help to monitor consumption over time;

- VOC54: information should be easy to understand;

- VOC55: invoices should emphasise the existence of increasing block tariffs; and

- VOC56: invoices should be "friendly".

5.4 Underlying data quality requirements
The 56 VOCs identified were structured by building a KJ diagram. The researchers and a small number of participants have together looked at potential affinities among different VOCs that shared some kind of meaning. The process is somehow individual-dependent and subjective, but highly rewarding and creative.

The grouping of VOCs stage took long due to the large number of customer voices. As Figure 1 shows, 13 first-order groups of voices – G1 to G13 (red titles) – and four second-order groups of voices – SG1 to SG4 (blue titles) – emerged (in the second level of aggregation two first-order groups of voices, G12 and G13, remained as "lone wolves"). The diagram also displays relationships among the second-order groups of voices.

One can observe a proliferation of arrows and links among groups, meaning that when addressing a particular issue attention needs to be given to its multiple impacts.

6. Discussion: data quality attributes of the "ideal" invoice
The affinity diagram presented in Section 5 guides the discussion on the attributes of an invoice capable of answering to most concerns revealed by domestic consumers.

Looking at the KJ diagram, four major issues that need to be taken into account when designing a proper invoice arise (see Figure 2): plain language; clear and comprehensive content; attractive format and guidance to parsimonious consumption.

Adopting "plain language" that everyone understands is of strategic importance. In this regard, an "ideal" invoice should not use technical jargon and should avoid as much as possible acronyms or other terms that jeopardise transparency goals. The issue was raised in various circumstances and is also clearly depicted in the KJ diagram with six VOCs relating to consumers demands of a language that is easily understandable to the general public.

In what a "clear and comprehensive content" is concerned, an "ideal" invoice is expected to provide accurate and complete information so that consumers can recalculate the total amount due. For that to happen, invoices should discriminate the various items (fixed and

Figure 2.
Demands placed upon an "ideal" invoice

volumetric charges; water supply, wastewater and solid waste services charged). Additionally, in some cases when "auxiliary" services (which are provided occasionally under request by customers or that result from breaches of contract attributable to the customer) are charged a link should be provided to the full justification of such value. In all circumstances it is important to ensure that units of measurement (e.g. m^3, €) are explicitly indicated by participants and that no inconsistencies are to be found.

An "attractive format" is a further concern since it seems to be essential to raise consumers' willingness to look at water invoices. Several VOCs referred to the preference for a short document that would include visual supports so that in a quick look customers could understand their bills. An aspect that can help to achieve an appealing format is a well-organised document.

Some participants stressed the relevance of disclosing information that would help consumers to assess the adequacy of their consumption levels (invoices should, for instance, provide graphics with previous consumption data, own consumption averages and comparable households consumption levels) and also called attention to the importance of including conservation messages.

Simultaneously, "Guidance to parsimonious consumption", an important role of an "ideal" invoice, implies conveying, on a simple and straightforward way, information on the tariff scheme applied (namely on the fact that increasing block tariffs are in charge to penalise higher consumption levels) and highlighting the relationship between quantities consumed and the correspondent amount due. The KJ diagram already revealed that unless concerns of clarity (SG2), accessibility and transparency (SG1) and even "attractiveness" (SG4) are addressed, it is highly predictable that invoices will not entirely fulfil their role in driving consumption behaviours (SG3).

Figure 2 also 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, discriminating all items and without abbreviations, on the other hand they 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. To achieve an appropriate balance, some information might not be directly given through invoices but should in any case a clear reference be provided on where and how such 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). The importance of the operators' websites in conveying information to end-users is also recognised by ERSAR, which in the Decree-Law194/2009 addresses the matter.

As it was expected, not all the focus groups gave the same importance to the four issues depicted in Figure 2. FG2 members, with lower qualification/income levels, revealed stronger feelings that invoices are complex and, because there are no switching options, operators can easily abuse their dominant position (market power). Besides, compared to FG1 and FG3, FG2 apparently had higher expectations concerning the invoices potential in promoting parsimonious consumption and its impact on their water bills. For FG2 individuals having everything fully explained was particularly important since this was associated with a "good faith" attitude from the water service providers'.

On the opposite, individuals who participated in FG1 were especially keen on a short document. FG3 members seemed to be more familiar with invoices, even if they also shared concerns with the use of easily understandable language, accessible to all. Some participants have explicitly mentioned that invoice should ensure the consumers' right to information, which requires communicating in plain language. It was also among FG3 members that visual aids and other format aspects were particularly addressed.

More qualified participants with children at school age revealed higher ecological awareness levels (sustainability concerns) and stressed the importance of invoices providing a way to evaluate whether they are consuming too much water.

With the aim of analysing the main data quality dimensions in the context of water invoices, Table II was designed. It combines the VOCs with the regulator's recommendations on the matter. The number of VOCs under each dimension gives an indication of its importance to the consumers. An intensity scale was used to evaluate at what extent each data quality dimension is underlined on the various recommendations guidelines. For this purpose, ERSAR, (2010a) was considered, since it explicitly establishes the contents of water invoices. Recomendação ERSAR No. 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 that "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".

The data quality dimensions in Table II correspond to those commonly mentioned in the literature (see Section 2). In some instances two data quality features were grouped into one (e.g. accuracy and precision) given their close meaning in this context. In other instances, it was decided to retain certain data quality dimensions less emphasised in the literature (such as comparability and understandability) due to their relevance in this specific study (e.g. comparability is associated with the identification of consumption patterns that are essential to promote parsimonious use of water resources).

Data quality dimensions (DQD)	VOC number	Recomendação ERSAR No. 1/2010
Main DQD from literature review		
Accessibility	25; 36; 37	++
Timeliness	39	+
Accuracy/precision	6; 15; 35; 48	+++
Reliability/consistency	nf	++
Completeness	9; 10; 11; 21; 23; 28; 34; 44; 47	+++
Relevance	1; 2; 12; 18; 19; 20; 24; 32; 33; 53; 55	++
Unambiguousness	16; 27; 40; 50; 51	nf
Others		
Level of detail (low)	7; 30; 38; 46	nf
Understandability (by the general public)	13; 26; 41; 42; 49; 54; 56	+
Comparability	4; 22; 31	++
Format (adequate)	3; 5; 8; 14; 17; 29; 43; 45; 52	+

Notes: nf, no evidence was found. +implicit or only occasionally addressed; ++explicitly addressed in more than one point; +++explicitly and repeatedly addressed

Table II.
Data quality dimensions in the context of water invoices

It is interesting to notice that the various data quality dimensions proposed in the literature (see Section 2) find expression in either the VOCs collected or in the guidelines given by ERSAR in recommendations regarding water bills (or in both).

It is possible to observe in Table II that consumer requirements cover the various data quality dimensions, with the exception of consistency (this did not come as a surprise since the idea of keeping uniform criteria over time or across users is somehow difficult to come to the consumers' mind when thinking about their own invoices at a single moment in time). In addition, the ERSAR Recommendation is in line with various data quality dimensions, some of them being explicitly addressed at various points.

Completeness is a major concern. Several VOCs 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 (low level of detail). 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 need to be promptly informed when tariffs schedules change was mentioned by the customers. The ERSAR's recommendation explicitly refers to the requirement of giving information to the customers on new tariff schemes prior to put them in place. Moreover, timeliness is also embedded in the need of communicating the approval of the new tariff scheme in the invoice that follows the date of its approval (Andrade *et al.*, 2012, p. 101).

The regulator concerns with accuracy are clearly visible in the ERSAR, (2010a), which calls for a high level of accuracy and precision. This dimension is also mentioned by the consumers even if with a lower intensity.

On the other hand, understandability and unambiguousness are major priorities for the water consumers that participated in the study, as the high number of VOCs 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 the abbreviations used. For the customers, 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 the development of a specific area for frequently asked questions (FAQs) about billing and reading. The FAQs are available on the ERSAR's website and in informative leaflets for end-users about "Rights and duties of water and waste services consumers".

Accessibility here is essentially related to giving details on contacts that can be used to obtain further information. The regulator acknowledges that invoices cannot include all possible data and that some information can (and should) be obtained through alternative channels, namely websites. In this regard, the Decree-Law 194/2009 defines that water and waste services operators' websites should provide information concerning: operator identification, competences and geographical scope; statutes and legal title for the management of the systems; annual reports and accounts; service regulations; tariffs; contract conditions; drinking water quality results, and other quality of service indicators; information on service suspension; contacts and opening

hours. In what the consumers is concerned, the need of keeping all details in invoices has raised divergent opinions.

Consumers have stressed the importance of invoices giving information in a way that enhances comparability so that they easily become aware of their consumption pattern and the reasonability of their water bill. By establishing that operators should provide users with a record of previous consumptions (in both euros and cubic metres) the regulator reveals a concern with comparability.

Moreover, format issues, which have been up to now relatively disregarded 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 when designing effective invoices.

7. Conclusion

Data quality attributes, despite extensively discussed in the literature, have been almost exclusively analysed in the context of information systems for organisations. Moreover, most studies on data quality literature follow an intuitive approach assuming that those who prepare the data know what its users need and value.

Water invoices are the basic (and key) channel to give information on quantities, prices and total amount due, as well as to convey other messages regarded as relevant at particular moments in time (such as interruptions, modifications in contractual conditions, etc.). In fact, the appropriate communication of such information is especially relevant due to the importance of protecting the rights of consumers when no "exit" behaviours are possible. Given the worldwide purposes of sustainability and the criticality of guiding adequate consumption behaviours, the importance of transmitting relevant and understandable information to consumers/citizens is a major concern but corresponds to an under-researched topic.

The purpose of this research was to explore the customers' concerns with the information disclosed in water services invoices identifying the customers' requirements of an "ideal" invoice and assessing at what extent the data quality dimensions proposed in the literature cover such requirements. To fulfil this purpose a set of consumers was listened to and their voices analysed.

By listening to the views of domestic consumers, collected by means of focus group interviews, this research gives important insights on what information needs to be provided in water invoices and how. Putting the consumers at the core of the process overcomes a drawback found in most studies that underestimate the users' perspective. This is an additional contribution of this paper.

The requirements identified in this study relate to both (comprehensive and clear) content and (attractive) format and give particular emphasis to (plain) language matters.

Customers' voices together with the recommendations of the Portuguese economic regulator cover most of the data quality dimensions proposed in the literature. Understandability and unambiguousness were particularly stressed by consumers and are also visible in recommendations issued by the regulator. Yet the concerns of domestic users go well beyond what is established specifically on the "recommendation on bill content".

Based on these findings, it is possible to derive some recommendations that should be considered by water service providers when designing water invoices:

- 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 (by including, for instance, the prices of subsequent blocks of consumption); and

- developing multiple and transparent communication channels (information technologies are essential but should not enhance any kind of exclusion, which might require in certain cases providing support to guarantee that all users have access to relevant data).

Concerns with language, which were transversal to the three focus groups, visibly show that, when designing invoices, attention needs to be given to users who have more difficulties in understanding water bills.

Worrying signs have emerged that the lack of understanding is jeopardising the users' trust on the service providers. It became clear that ensuring the citizens' right to information cannot be fully accomplished by taking care of information provision alone. It must be complemented with other data quality requirements, especially those related to clarity and friendliness.

This research highlights that avoiding technical language and making visible the consequences of different consumption levels on the amounts to be paid is essential when designing water invoices. Moreover, it emphasises that there is still room for improvement in the way the economic regulator performs its role in ensuring the provision of sound information.

The requirements identified in this study might be used in future research to evaluate the customers' perceptions on water bills using questionnaire surveys that measure satisfaction based on attributes indeed relevant to the customers.

The results of the current study also provide useful insights to conduct further research, especially on the potential of invoices to avoid illiteracy issues related to such documents and to the tariff schemes in charge. Besides, literature on the estimation of residential water demand functions can benefit from the results of this research by, for instance, testing the level of knowledge of domestic users on tariff schemes as an explanatory variable of water consumption.

Notes

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" (OECD, 2013, p. 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, p. 59).

References

Akao, Y. (1995), "QFD toward product development management", *Proceedings of International Symposium on Quality Function Deployment, Union of Japanese Scientists e Engineers, Tokyo*, pp. 1-10.

Akao, Y. and Mazur, G.H. (2003), "The leading edge in QFD: past, present and future", *International Journal of Quality & Reliability Management*, Vol. 20 No. 1, pp. 20-35.

Albuquerque, C. de (2014a), *Realising the Human Rights to Water and Sanitation: A Handbook by the UN special rapporteur Catarina de Albuquerque – Booklet 1. Introduction*, UN-OHCHR, Lisboa.

Andrade, I., Mesquita, A., Santos, C., Aleixo, C., Rosário, L., Igreja, M. and Ribeiro, P. (2012), Relação das entidades gestoras com os utilizadores dos serviços de águas e resíduos, Guia Técnico No. 20 (Technical Guide No. 20), ERSAR, Lisboa.

Ballou, D.P., Madnick, S. and Wang, R. (2004), "Assuring information quality", *Journal of Management Information Systems*, Vol. 20 No. 3, pp. 9-11.

Carnevalli, J. and Miguel, P. (2008), "Review, analysis and classification of the literature on QFD – types of research, difficulties and benefits", *International Journal of Production Economics*, Vol. 114 No. 2, pp. 737-754.

Chan, L. and Wu, M. (2002), "Quality function deployment: a comprehensive review of its concepts and methods", *Quality Engineering*, Vol. 15 No. 1, pp. 23-35.

Commission of the European Communities (2004, *White Paper on Services of General Interest*, COM (2004) No. 374 final, European Commission, Brussels, 12 May.

Davenport, T.H. and Prusak, L. (1998), *Working Knowledge: How Organizations Manage What They Know*, Harvard Business School Press, Cambridge, MA.

Eldin, N. and Hikle, V. (2003), "Pilot study of quality function deployment in construction projects", *Journal of Construction Engineering and Management*, Vol. 129 No. 3, pp. 314-329.

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

ERSAR (2010a), *Recommendation ERSAR No. 1/2010*, ERSAR, Lisbon.

ERSAR (2010b), *Recommendation ERSAR No. 2/2010*, ERSAR, Lisbon.

ERSAR (2013), "Annual report on water and waste services in portugal 2012 – executive summary", ERSAR, Lisbon.

Haug, A., Arlbjørn, J.S. and Pedersen, A. (2009), "A classification model of ERP system data quality", *Industrial Management & Data Systems*, Vol. 109 No. 8, pp. 1053-1068.

Haug, A., Zacharissen, P. and van Liempd, D. (2011), "The cost of poor data quality", *Journal of Engineering and Management*, Vol. 4 No. 2, pp. 168-193.

Hauser, J.R. and Clausing, D. (1988), "The house of quality", *Harvard Business Review*, Vol. 66 No. 3, pp. 63-73.

IRAR (2009), *Recommendation IRAR No. 1/2009*, IRAR, Lisbon, 28 August.

Juran, J. (1992), *Juran on Quality by Design*, Free Press, New York, NY.

Leung, H. (2001), "Quality metrics for intranet applications", *Information and Management*, Vol. 38 No. 3, pp. 137-152.

Mazur, G. (1993), "QFD for services industries: from voice of customer to task deployment", *The Fifth Symposium on Quality Function Deployment, Novi, MI, June*.

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

Prasad, B. (1998), "Review of QFD and related deployment techniques", *Journal of Manufacturing Systems*, Vol. 17 No. 3, pp. 221-234.

Tayi, Y. and Ballou, D. (1998), "Examining data quality", *Communications of the ACM*, Vol. 41 No. 2, pp. 54-57.

Wand, Y. and Wang, R.Y. (1996), "Anchoring data quality dimensions in ontological foundations", *Communications of the ACM*, Vol. 39 No. 11, pp. 86-95.

Wang, R. and Strong, D. (1996), "Beyond accuracy: what quality means to data consumers", *Journal of Management Information Systems*, Vol. 12 No. 4, pp. 5-33.

Further reading

Albuquerque, C. de (2014b), *Realising the Human Rights to Water and Sanitation: A Handbook by the UN special rapporteur Catarina de Albuquerque – Booklet 7. Principles*, UN-OHCHR, Lisboa.

About the authors

Patrícia Moura e Sá is an Assistant Professor at the Faculty of Economics, the University of Coimbra and a Researcher at the CICP/UM. She helds a PhD in Business from the Sheffield Hallam University (UK) and is the Author of several papers published in indexed journals, such as *Total Quality Management & Business Excellence*, *Technovation*, *Policy Studies* and *Electronic Journal of E-Government*. Her research topics include total quality management and organizational performance measurement in public administration and services of general interest. Patrícia Moura e Sá is the corresponding author and can be contacted at: pmourasa@fe.uc.pt

Rita Martins is an Assistant Professor at the Faculty of Economics, the University of Coimbra (FEUC). She is also a Researcher in GEMF, a research centre at the FEUC. She has published papers in international journals (such as *Journal of Productivity Analysis*, *Policy Studies*, *Water Policy*, *Public Organisation Review*). Her main research area of interest is regulatory economics especially in the context of water services.