

Focus groups, meaning making and data quality

Abstract. Data quality represents a major strategic asset for organisations and privacy is one of the main influencing factors impacting on organisations' practices related to customers' data collection. Other elements which affects this relationship are individuals handling data within the organisation and their different needs, however there is limited research focussing on these aspects. This work presents early findings of an empirical qualitative research using focus groups within a social constructivist approach to delve into managers' perception of data quality and to understand the impact of privacy on organisational practices. Focus groups have been widely adopted in academic research as an exploratory method to collect people's opinion. This study shows how some features, like involving pre-existing groups, the adoption of questionnaires and the workshops' setting can positively influence focus groups' outcome, encouraging participants' co-operative interaction, helping them to clarify concepts and raising awareness about privacy and data quality.

Keywords: Focus groups, data quality, Accuracy, social constructivism.

1 A social constructivist perspective

The present research is part of the EnCoRe project, a 45 months research project studying how to improve the ease with which individuals can grant and revoke their consent to the use, storage and sharing of their personal data by others. This paper is part of PhD research looking at data quality and privacy related practices from a social constructivist perspective. The aim is to reveal how notions associated with data quality and privacy are constructed within organisations and how different needs and understanding of these concepts can affect organisational practices.

In a social constructionist view, “all knowledge, and therefore all meaningful reality as such, is contingent upon human practices, being constructed in and out of interaction between human beings and their world, and developed and transmitted within an essentially social context” (Crotty 1998:42). In the social constructivist paradigm, reality is divided in two realms: social or institutional facts and natural or brute facts. Brute facts are “phenomena having a being independent of our own volition” (Berger & Luckmann 1967:1): they exist independently of any human institution. Institutional facts require human institutions to exist, they depend on the presence of other human beings since they “are typically objective facts, but [...] they are only facts by human agreement or acceptance” (Searle 2010:11). Facts are constructed by a collective intentionality that imposes special status on a phenomenon since the phenomenon can not perform its functions based solely on its physical characteristics (Searle 1995): institutional facts therefore can exist only because of collective atti-

tudes and recognition. Intentionality is pivotal as it enhances collective agreement which leads to meaning making processes and to the construction of institutional facts. A constructivist approach focuses on what is done with words in terms of what is said because “constructivists are especially interested (...) in actively constructing and co-creating knowledge by working together with actors involved” (Charreire Petit & Huault 2008:75).

This research uses focus groups to elicit interactions and meaning making processes to reveal the organisational dynamics and unfold privacy and data quality related definitions: data collected result from a process where individuals are agents and co-producers of their reality (Charreire Petit & Huault 2008). Meaning making requires participants to share their opinion in a co-operative interactive process to come up with a new understanding (Albrecht et al. 1993, Jarowski & Coupland 1999, Mercer 2000). The new shared definitions result from conflicts which arises from participants' cognitive diversity and by the context (Arazy et al. 2011): since co-workers often assume concepts without discussing them with their colleagues, this can lead to misunderstanding which can be unfolded by the focus groups.

This paper analyses data collected in 4 workshops to reveal how social constructivist behaviours can be enhanced by focus groups and how focus groups can be effective in unfolding data quality related meanings.

2. Understanding focus groups

Focus groups are social events consisting of group discussions organised to explore some specific topics and to collect participants' opinions in a permissive and non threatening environment (Albrecht et al. 1993, Kitzinger 1994, Gibbs 1997). They reveal not only what participants think, but also the reasons behind attitudes and how meanings are constructed (Morgan 1997, Wibeck et al. 2007). During their interaction, participants stimulates' each others and have a certain degree of freedom over the topic, so that each contribution triggers more spontaneous interplays (Grabowski et al. 1992). Moreover, participants' control over the process facilitates the disclosure of issues which are more salient for the group (Gibbs 1997, Myers 1998, Warr 2005, Munday 2006). The researcher becomes more like an observer who keeps participants focussed on the topic (Gibbs 1997, Myers 1998) reducing the distance between the researcher and the social context. The less authoritative role of the researcher and participants' level of freedom results in an unstructured environment which generates 'unruly' data, as participants produce less coherently sequenced accounts of individual and collective experience (Warr 2005).

Although focus groups have become a common method in academic research, sometimes they are inappropriately referred as 'group interviews': Ho (2006:2) describes “focus group as a group interview without the alternate question-answer sequence found in typical interview sessions” but Gibbs (1997) clarifies that “group interviewing involves interviewing a number of people at the same time, the emphasis being on questions and responses between the researcher and participants”. On the other hand, focus groups are not used to generate multiple accounts of individual perspectives, but

they explicitly rely on the group interaction to produce and collect interaction-based data (Grabowski et al. 1992, Kitzinger 1994, Warr 2005, Wibeck et al. 2007). Since focus groups' unit of analysis is the group and its dynamics, unlike group interviews, participants' interaction is an integral part of the whole research process. Focus group relies on group interaction and considers interactions as data, while group interviews rely on participants' interactions with the researcher to elicit data.

2.1 Adopting focus groups: the method and the organisations

Despite many concerns (Morgan 1996, 1997, Myers 1998), the research involved pre-existing groups of co-workers and workshops were held at organisations' offices: this resulted in a natural-like setting which revealed fragments of interactions close to the spontaneous occurring ones (Kitzinger 1994, 1995, Warr 2005, Munday 2006). Hollander (2004) warns that one of the risks of such groups is the impact of what is being said on future relationships; to collect individuals' view and to reduce this pressure, the data collection included a preliminary questionnaire and a follow up email. The use of a questionnaire is also meant to activate participants' thinking and to provide inputs for the discussion (Kitzinger 1995). The researcher asked non-selective questions and adopted a non-directive approach to create a casual setting and facilitate the discussion. By limiting her interventions and putting herself as a learner, participants were fooled into acting as co-researchers. The methodology was pre-tested with 4 colleagues who participated in a role playing game-like activity in a 2 hours' long simulation.

The recruitment of the organisation has been conducted using LinkedIn, the business social network with more than 5 millions subscribers from the UK¹. LinkedIn members can get in touch and discuss business related topics by joining or creating one of the available groups. To recruit the organisations, the researcher signed up to the data quality related groups and started a discussion in June 2010 which has resulted in 107 contributions². This way the researcher had access to the profiles of relevant professionals and could contact them. Once the selected professional accepted, s/he was asked to involve 2 or more colleagues working in data quality and privacy area of their organisation. From October to December 2010 the researcher ran 4 focus groups with 4 UK based organisation from different business areas, involving 15 professionals with an average of 16 years of experience³. Focus groups took place at the organisations' offices engaging 3 to 5 participants each. Workshops were audio-recorded and transcribed.

The following table shows some data about the groups' composition; 'utterances' are communicative units consisting of a speech "clearly delimited by the change of speaking subjects" (Bakhtin 1986).

¹ Data from LinkedIn official website, <http://press.linkedin.com/about/> [accessed on May 5th 2011].

² The discussion is available here: <http://tinyurl.com/linkedin-dataquality> [created on 20th May 2011]

³ Participants' professional experience ranged between 7 to 25 years.

Table 1. The following table shows some data about the focus groups:

Type of organisation	FG length	Transcript size (words)	Number of utterances	Researcher generated utterances	Participants' generated utterances
Geographic Data [GEO]	1h28'	14464	295	22.70%	G1. Group Data Manager: 30.1% G2. Data Analyst Manager: 25% G3. Data Procurement Manager: 22%
Telecommunication [TEL]	1h37'	16549	357	19.60%	T1. Regional Head of Security: 10.6% T2. CISO: 23.8% T3. Principal Manager ICT security: 21% T4. Global Privacy Manager: 24.9%
Finance [FIN]	1h33'	13886	484	18.80%	F1. Enterprise Data Manager: 37.3% F2. Head of Customer and Market insight: 29.9% F3. Customer Segment Manager: 13.8%
Educational Public Agency [PUB]	1h58'	18650	334	18.50%	P1. Head of Records and Rights: 16.4% P2. Senior Data Project Manager: 28.4% P3. Council Solicitor: 20.3% P4. Interim Head of Data Manager: 11.3% P5. Head of information authority ⁴ : 4.7%

2.2 Focus groups and the construction of Accuracy

Since data represent a strategic competitive advantage with a direct impact both on customer satisfaction and organisations' revenues, organisations tend to collect as much data as possible, even without a clear strategy (Duran 2009) with inevitable consequences for data protection practices. The relationship between data quality, data collection and data protection needs to be clear within organisations, so that customers' privacy is protected and the quality of data is appropriate and accurate. Data management practices are regulated by laws: both the Directive 95/46/EC and the UK Data Protection Act underline this relation between data processing and privacy and organisations are also well aware of it:

P2: *So the dimensions of data quality are, are integrally linked to the privacy aspect, because [...] there's almost like a causal relationship between the two in the sense that the minute you begin to degrade data quality you're beginning to impact, in one way or another, on the person's privacy...* [Q1]

As the DPA has been enacted to conform to the European Directive, which requires member states to regulate personal data processing, its influence on the DPA is evident: both documents describe data protection principles which requires personal information to be processed fairly and lawfully and for limited purposes. The quantity of data collected should be adequate and relevant; data should be accurate and up to date and kept for no longer than necessary. Data should be processed in line with the rights of individuals, be secure and not transferred to other countries without adequate protection. Most of these principles are part of data quality dimensions, where 'dimen-

⁴ Participant was on the phone and was disconnected several times.

sions' are parameters and attributes meant to represent an aspect pertaining to data quality (Wang & Strong 1996).

The questionnaire presented a list of 12 dimensions related to these principles; participants were asked to rate 3 dimensions: 12 out of 15 partakers selected Accuracy and 11 ranked it as the most important. As the workshop began, participants were asked to sort the dimensions again and explain their choices. Once they answered, the researcher challenged the group asking them to agree on 2 dimensions; the question was not meant to look for a consensus, which in many cases was impossible to reach (i.e. Q10, Q11), but to stimulate the constructivist process. The analysis of groups' interaction to define Accuracy, provides an example of how focus groups can engage participants in a collaborative social constructivist activity.

All focus groups agreed that data do not have to be 100% accurate, though the processes and arguments which lead to this statement were different in each workshop.

Some groups indicated that Accuracy depends on external factors, like requirements:

G1: *Yeah that might be an interesting discussion to have, coz in my mind the data doesn't have to be 100% perfect as long as it meets the customer requirements ...* [Q2]

T2: *Accuracy does not mean that it has to be 100% accurate. It's [how] the requirement is stated...* [Q3]

But most of participants underlined the importance of understanding the degree of Accuracy of their data and defining the level of confidence:

G2: *[...] you need to understand how accurate your data is really. So accuracy doesn't mean it has to be 100% but it does mean to me you need to understand the accuracy you're looking at.*[Q4]

P2: *[...] Many measures of data quality, particularly for statistical decisions don't require 100% accuracy [...] I guess the best example of that is, you know, the psychological polls that you see at the elections or whatever, where they come with their, you know, + or - 3% ...*

P5: *[...] I would agree with you pretty much 100% apart from the fact that [...], it all has to be 100% accurate to a level of confidence and [...] it's that level of confidence that needs to be defined...* [Q5]

The previous quote (Q5) shows how, through a critical yet constructive behaviour, participants build the concept of 'confidence': partakers explored their different positions in order to build their shared knowledge by questioning and testing their viewpoint in the so called 'exploratory talk' (Mercer 2000). The following quote is another example of exploratory talk: here participants discuss data quality dimensions critically, revealing their views and challenging their positions:

F1: *[...] And consistency is an important one from my point of view, it's purely because if you don't have the consistency then the understandability goes out the window...[...]*

F2: *Do you mean consistency of accuracy? Or propagation?*

F1: *Consistency of the fact that I've... [...]. So I have a consistency through the database of, you know, the same thing. It's closer I suppose in a way to conforming to a standard, but it is consistent...*

F3: *It's about being reliable, isn't it?*

F1: *Yes. And therefore understandable...* [Q6]

The next quote shows a 'cumulative talk' (Mercer 2000) where participants build on each other's contributions in a supportive and uncritical way, so that the concept of 'confidence' emerges as constructed on the previous utterance:

F1: *[...] So it doesn't have to be 100% correct, depending on the application.*

F2: *[...] I would add to that it doesn't have to be 100% [accurate], but you need to know where it isn't.*[Q7]

Sometimes, in their attempt to construct the meaning of Accuracy, participants engaged in a struggle for definitions. The following quote is a cumulative talk showing participants' collective thinking leading a partaker to change their mind:

F2: *My third one was concise representation. Because it's... to me it translates to consistency and accuracy.*

F3: *And I've got timeliness as my... I had consistency, now I've got timeliness, coz it's about being accurate and up to date.* [Q8]

Accuracy and fit for purpose were mentioned as strongly related in all workshops as explained in the following quote, where the speaker explains this relationship by examples to strengthen his position:

F2: *...you know, [data] doesn't need to be 100%, if we got a feeling for how much market share they've got, they're approximately 8-9%, well that's fine [...] but if we're looking at accuracy of the customer data on the data warehouse and we're concerned to see the levels of sales etc., we want to know the exact level of sale [...]. So it's fit for purpose.* [Q9]

3. Discussion

Participants adopted diverse strategies to discuss and construct meanings, based on their personal experiences and background and on their knowledge of shared history. Difficulties in reaching a consensus arise from partakers' diverse needs and approaches as the groups themselves recognized:

T2: *I think what we say is we struggle to agree on a second [dimension]. There's obviously pools in different people's areas here like [he] is our data quality champion internally, so conformance to a schema so that the data remains accurate is obviously very close to his heart...* [Q10]

P3: *Yeah, and I think going back to your question, you say, you know, can we pick and can we agree on three, I think the simple answer to that is no... [...] so I think, because of the different approaches we have, maybe these three could agree and I'll just sit on the side on this one...* [Q11]

Diversity in participants' attitudes can create misunderstandings and confusion which the focus group helped to bring to light:

G3: *Yeah that was interesting coz you thought that and I was thinking completeness in terms of what fields there were that they were filled, I suppose. I was thinking of attributes [...]*...

G1: *Yeah it's true. I come from a geographical background so I was thinking geographical completeness and you were thinking attribute completeness...* [Q12]

The previous quotation demonstrates the general tendency to assume that others involved in the same context also share assumptions about values and purposes, not considering that words have a variable meaning (Mercer 2000). Therefore, although participants have been working together, they never felt the need to clarify the terms because they didn't know the term needed a clarification (Whitley 1996). This kind of misunderstandings can have a major impact on organisational practice where the same data set is used by several people from different departments with distinct purposes. This reflects also the lack of a commonly agreed meaning of dimensions in the data quality field where Accuracy is defined in various ways, including the degree of data reflecting real world objects (Batini & Scannapieco 2006, Duran 2009), as a quality related to the processes of data collection (Kenett & Shmuvli 2011), or in terms of data free of errors (Vannan 2001). The process revealed also the connections between Accuracy and Consistency (Q6, Q8); Olson (2003) suggests that Consistency is part of Accuracy as Accuracy “refers to whether the data values stored for an object are the correct values. To be correct, a data value must be the right value and must be represented in a consistent and unambiguous form.” (:29). A second emerging pattern relates Accuracy to the purpose of data collection and the context of use as explained in Q9 (Davenport Huges 1954, Olson 2003). Discussions around Accuracy associated it to 6 main areas: the impact of applications (Q7), customer (Q2) and technical (Q3)

requirements, the importance of the purpose (Q9), the influence of context (Q9), confidence in data and errors' tolerance (Q4, Q5). Each organisation focussed on specific issues to construct their meanings and even if the final outcome is the same, each group put in place different strategies and arguments to get to their conclusions as shown in Figure 1:

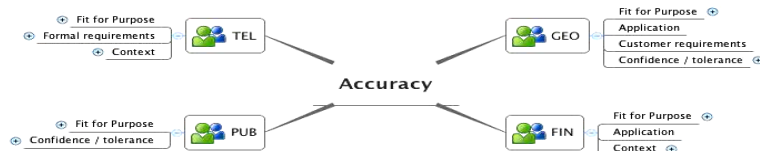


Figure 1. The figure shows which topics were considered in the discussion related to Accuracy.

4. Conclusions

If the constructivist researcher is expected “to show concretely how the actors' interaction can enhance thinking and the knowledge creation thinking” (Charreire Petit & Huault 2008:84), the examples prove the potential of focus groups in eliciting interaction and bringing out differences. Building a shared knowledge required participants to engage in a co-operative and creative endeavour (Mercer 2000); through collaborative behaviours and cumulative and exploratory talks, participants revealed their views and by building new meanings collectively they brought to light their differences and faced conflicts which contributed to the quality of the output (Arazy et al. 2011). As resulting from this analysis, focus groups proved to stimulate social constructivist behaviours and some features turned out to be particularly effective:

- Participants' control over the process: leaving partakers some freedom and autonomy resulted in spontaneous and natural-like interactions. Participants were free to contribute, while the researcher avoided making selective questions or forcing answers. Nevertheless, as shown in Table 1, participants' interaction was balanced and everyone engaged in the process. Moreover, partakers' control allowed them to raise topics spontaneously and this helped them to discover misunderstandings i.e. Q12.

- The context: running the focus group in the organisations' office created a natural-like setting where participants felt at ease. Reproducing a daily-like situation reduced pressure and elicited casual interactions. Such a spontaneity is also due to the groups' composition: participants were reassured by the familiar environment and felt comfortable to disclose stories and discuss even sensitive cases. The context was an incentive for confidential talks which would hardly emerge in a cross-organisational context, where participants would feel the pressures of the presence of competitors.

- The questionnaire: it induced participants to think about the topic beforehand. This helped the constructivist process as partakers reflected on the questions and made their own opinions individually. Moreover, questionnaires provide evidences of participants' pre-workshop opinions which can be compared with views emerged during the focus group to reveal the impact of the constructivist process, i.e. Q8.

These elements increased focus group efficacy by promoting participants' interaction and commitment and generated highly constructivist behaviours.

Participants' differences in approaches and needs (Q10, Q11) and the internal confusion (Q12) affected the process positively: from the study emerges that the more different participants views, the more they engaged in the meaning making process (Arazy et al. 2011). Furthermore, the workshops also increased participants' awareness and understanding of data quality concepts: by bringing together co-workers from different departments, they could have a better picture of their needs and expectations.

The role and importance of Accuracy as emerged in the preliminary research will require further analysis and a new round of focus groups is underway to improve understanding of Accuracy and allow for generalisations of results.

References

- Albrecht, T. L., Johnson, G. M. & Walther, J. B. (1993). Understanding Communication Processes in Focus Groups. In D. L. Morgan (Ed.), *Successful Focus Groups* (pp. 51-63). Newbury Park, CA: Sage.
- Arazy, O., Oded, N., Patterson, R., & Yeo, L. (2011). Information Quality in Wikipedia: The Effects of Group Composition and Task Conflict. *Journal of Management Information Systems* 27(4), 71 - 98
- Bakhtin, M. M., Emerson, C., & Holquist, M. (1986). *Speech genres, and other late essays*. Austin,: University of Texas Press.
- Batini, C., & Scannapieco, M. (2006). *Data quality: concepts, methodologies and techniques*. Berlin; New York: Springer.
- Berger, P. L., & Luckmann, T. (1967). *The social construction of reality: a treatise in the sociology of knowledge*. Harmondsworth: Penguin.
- Charreire, Petit, S., & Huault, I. (2008). From Practice-based Knowledge to the Practice of Research: Revisiting Constructivist Research Works on Knowledge. *Management Learning*, 39(1).
- Crotty, M. (1998). *The foundations of social research*. London: Sage.
- Davenport Hughes, C. (1954). Accuracy in Statistics. *Journal of the Royal Statistical Society*, 3(3), 146-149.
- Duran, E. (2009). *Customer data management: the impacts of poor customer data quality on customer relationship management and organizations*. Saarbrücken, Deutschland: VDM Verlag Dr. Müller.
- Gibbs, A. (1997). *Social Research Update 19: Focus Groups*. from <http://sru.soc.surrey.ac.uk/SRU19.html>
- Grabowski, M., Massey, A. P., & Wallace, W. A. (1992). Focus groups as a group knowledge acquisition technique. *Knowledge Acquisition*, 4(4), 407-425.
- Ho, D. (2006). The Focus Group Interview: Rising to the Challenge in Qualitative Research Methodology *Australian Review of Applied Linguistics*, 29(1), 5.1-5.19.
- Hollander, J. A. (2004). The Social Contexts of Focus Groups. *Journal of Contemporary Ethnography*, 33(5), 35.
- Jaworski, A., & Coupland, N. (1999). *The discourse reader*. London: Routledge.
- Kenett, R. S., & Shmueli, G. (2011). *On Information Quality*. R. H. Smith School Research Paper No. 06-100.
- Kitzinger, J. (1994). The methodology of focus groups: the importance of interactions between research participants. *Sociology of Health and Illness*, 16(1), 103-116.
- Kitzinger, J. (1995). *Introducing Focus Groups*. *BMJ: British Medical Journal*, 311(7000), 299-302.
- Mercer, N. (2000). *Words and minds: how we use language to think together*. London: Routledge.
- Morgan, D. L. (1996). *Focus Groups*. *Annual Review of Sociology*, 22, 129-152.
- Morgan, D. L. (1997). *Focus groups as qualitative research* (2nd ed.). Thousand Oaks, Calif.: Sage.
- Munday, J. (2006). Identity in focus: The use of focus groups to study the construction of collective identity. *Sociology*, 40(1), 16.
- Myers, G. (1998). Displaying opinions: Topics and disagreement in focus groups. *Language in Society*, 27(1), 85-111.

- Olson, J. E. (2003). *Data quality : the accuracy dimension*. San Francisco, Calif. ; London: Morgan Kaufmann.
- Searle, J. R. (1995). *The construction of social reality*. London: Allen Lane.
- Searle, J. R. (2010). *Making the social world*. Oxford: Oxford University Press.
- Vannan, E. (2001). Quality Data: An Improbable Dream? *Educause Quarterly*, 24(1), 56-58.
- Wang, R. J., & Strong, D. M. (1996). Beyond accuracy: what data quality means to data consumers. *Journal of Management Information Systems*, 12(4), 5 - 33.
- Warr, D. J. (2005). "It was fun... but we don't usually talk about these things": Analyzing sociable interaction in focus groups. *Qualitative Inquiry*, 11(2), 200-225.
- Whitley, E. A. (1996). Confusion, Social Knowledge and the Design of Intelligent Machines. *Journal of experimental and theoretical artificial intelligence*, 8(3/4), 365-381.
- Wibeck, V., Abrandt Dahlgren, M., & Öberg, G. (2007). Learning in focus groups: an analytical dimension for enhancing focus group research. *Qualitative Research*, 7(2), 18.

Acknowledgments

The present paper is an extended version of an earlier version presented at the Seventh International Summer School organised jointly by the IFIP Working Groups 9.2, 9.6/11.7, 11.4, 11.6 in Trento (Italy). A special thank to all participants who provided useful inputs and ideas to improve my work.