Bus Inf Syst Eng 59(1):3–21 (2017)
DOI 10.1007/s12599-016-0459-8

RESEARCH PAPER

Towards a Conceptualization of Data and Information Quality in Social Information Systems

Roman Tilly · Oliver Posegga · Kai Fischbach · Detlef Schoder

Received: 17 March 2016 / Accepted: 16 October 2016 / Published online: 19 December 2016

Abstract Data and information quality (DIQ) have been defined traditionally in an organizational context and with respect to traditional information systems (IS). Numerous frameworks have been developed to operationalize traditional DIQ accordingly. However, over the last decade, social information systems (SocIS) such as social media have emerged that enable social interaction and open collaboration of voluntary prosumers, rather than supporting specific tasks as do traditional IS in organizations. Based on a systematic literature review, the paper identifies and categorizes prevalent DIQ conceptualizations. The authors differentiate the various understandings of DIQ in light of the unique characteristics of SocIS and conclude that they do not capture DIQ in SocIS well, nor how it is defined, maintained, and improved through social interaction. The paper proposes a new conceptualization of DIQ in SocIS that can explain the interplay of existing conceptualizations and provides the foundation for future research on DIQ in SocIS.

Accepted after three revisions by the editors of the special issue.

Dipl.-Wirt.-Inf. R. Tilly (✉) · Prof. Dr. D. Schoder
Department of Information Systems and Information Management, University of Cologne, Pohligstraße 1, 50969 Cologne, Germany
e-mail: tilly@wim.uni-koeln.de

Prof. Dr. D. Schoder
e-mail: schoder@wim.uni-koeln.de

Dipl.-Wirt.-Inf. O. Posegga · Prof. Dr. K. Fischbach
Department of Information Systems and Social Networks, University of Bamberg, An der Weberei 5, 96052 Bamberg, Germany
e-mail: oliver.posegga@uni-bamberg.de

Prof. Dr. K. Fischbach
e-mail: kai.fischbach@uni-bamberg.de

Keywords Social information systems · Social media · Data quality · Information quality · Socio-technical processes

1 Introduction

Social information systems (SocIS) are, in essence, "information systems based on social technologies and open collaboration" (Schlagwein et al. 2011). They include, for example, the various forms of social media. Many people use SocIS to obtain and share general information, advice, or gossip, as well as for communication, entertainment, socializing, or political mobilizing (Parameswaran and Whinston 2007a, b; Kaplan and Haenlein 2010; Kane et al. 2014). Questions of data and information quality (DIQ) potentially affect all these uses: Are users interested in and do they actually talk about the same phenomena? Does the social medium allow producers of data to express their perceptions so that consumers of information will understand what they meant? How can producers know what consumers are interested in so they can supply them with high-quality information? Who decides about DIQ? These and other issues cannot necessarily be resolved successfully in SocIS.

Given the past decades of research on DIQ in traditional information systems (IS) (for an overview, see Lee et al. 2002; Madnick et al. 2009; Sadiq et al. 2011; Xiao et al. 2014), one might assume that understanding DIQ in SocIS is merely a matter of transferring existing definitions, frameworks, and measures to a new domain. In fact, several approaches have aimed at applying traditional DIQ concepts to SocIS (for an overview, see, e.g., Chai et al. 2009). Also, with respect to IS success, DIQ has been included in studies that apply the DeLone and McLean model of IS success (DeLone and McLean 1992, 2003) to

SocIS to explain the success of, for example, corporate intranets (Barnes and Vidgen 2009), online communities (Zheng et al. 2009; Lili and Rong 2013), and social microblogging services (Ou et al. 2011).

We argue, however, that traditional definitions of DIQ are insufficient for capturing the characteristics of SocIS, presumably because they have been developed for traditional IS in an organizational context. Traditional definitions make assumptions about, for instance, users, user behavior, tasks, contexts, governance, and relation of data/information production to consumption that conflict with the characteristics of SocIS, which afford social interactions to an open, heterogeneous virtual community of users who are both producers and consumers of content, interact in different social subsets, and distribute IS governance amongst themselves. To further the understanding of DIQ in SocIS, we pose the following research question: *How is DIQ conceptualized in IS research and do prevalent DIQ conceptualizations accommodate the characteristics of SocIS?*

To answer this question, we identify and categorize prevalent DIQ conceptualizations by means of a comprehensive and systematic literature review. We build on a generic definition of quality to differentiate the assumptions behind the various DIQ conceptualizations. Further, we provide an overview of the most important characteristics of SocIS, which we use to analyze traditional DIQ conceptualizations from a novel perspective. Our analysis reveals that traditional DIQ conceptualizations do not account for the unique characteristics of SocIS. In our discussion, we describe novel aspects of DIQ that arise in SocIS and that are important to advance our understanding and develop a conceptualization of DIQ in SocIS. In doing so, we follow a conceptual research approach and differentiate well-established concepts of a prominent area of IS research while revising them from a novel perspective (MacInnis 2011).

The remainder of this article is organized as follows. In Sect. 2, we provide the conceptual foundations of our work. Section 3 presents an overview of the methodology underlying differentiation and revising. In Sect. 4, we present a categorization of the DIQ conceptualizations identified. In Sect. 5, we differentiate these conceptualizations and analyze them critically in light of SocIS characteristics. We discuss our findings in Sect. 6, and conclude in Sect. 7 by outlining contributions and limitations of our study as well as providing an outlook on future research.

2 Conceptual Foundations

2.1 Data, Information, and Information Systems

We share the long-held view of IS as socio-technical systems (e.g., Lee 2010; Boell and Cecez-Kecmanovic 2015) comprising both social (humans and groups) and technical (hardware and software) components that interact to generate, process, and store information and data. In IS research, the domain of an IS and what it is supposed to be used for has traditionally been seen as defined by the larger organizational system in which IS are embedded (Hirschheim and Klein 2012; Winter et al. 2014). Thus, IS have been primarily thought to support specific (groups of) organizational users in performing certain tasks and thus "aim to provide instrumental value to the user" (van der Heijden 2004) as well as to the organization as a whole. We refer to these IS that are designed for and used in organizations, and that account for most of the IS research to date, as "traditional IS." Traditional IS entail various classes of IS such as transaction processing systems, management information systems, and decision support systems. Though these systems serve different purposes in an organization, they share in common that they retrieve, store, and process data that can be presented to human users (employees or customers) as information about real-world phenomena related to the organization, its activities, and its problems (Mason and Mitroff 1973).

We define data as what is stored in a database and processed by an IS: signs that are used according to certain syntactic rules, are objective, and may represent facts about relevant phenomena external to the IS, that is, in the real/physical world (Wand and Wang 1996; English 1999; Price and Shanks 2005). Data become information when a human user in an IS receives, perceives, and interprets data, puts them into context, and thus gives them a (subjective) meaning (English 1999; Price and Shanks 2005; Glowalla and Sunyaev 2014). This delineation of data and information is in line with what has become a "General Definition of Information (GDI) in terms of data + meaning" (Floridi 2011) over the last decades in many disciplines concerned with information.

How people search, filter, acquire, interpret, use, or share information is summarized under the term of human information behavior. Wilson (2000) defines it as "the totality of human behavior in relation to sources and channels of information, including both active and passive information seeking, and information use." Research on human information behavior has traditionally been focused primarily on individuals and individual behavior towards information (Hansen and Järvelin 2000, 2004), treating behavior that includes more than one individual as "one-way process in which an individual consults another individual" (Talja and Hansen 2006). However, recent research emphasizes that collaborative information behavior is as important and common as individual information behavior (Hansen and Järvelin 2000, 2004; Talja 2002; McKenzie 2003). Collective information behavior means that information-related tasks such as search, filtering, evaluation are

purposefully distributed and integrated among multiple individuals, rather than multiple individuals acting independently (Talja and Hansen 2006). IS generally have the potential to (also) support collective information behavior, but it will make a difference whether they are designed under the premise of supporting individual or collective/collaborative information behavior. As can be seen from the discussion of the specific characteristics of SocIS (Sect. 2.2), SocIS are conceptually a type of IS that supports collective information behavior.

2.2 Social Information Systems

Since this article strives to review DIQ conceptualizations in light of the characteristics specific to SocIS, we must first specify what a social information system is and how it differs from traditional IS. A minimal definition is that SocIS are IS that are (1) based on social technologies (also termed "social software," e.g., wikis, blogs, online social networks) and (2) enable or promote open collaboration (Schlagwein et al. 2011). SocIS are covered by the broader definition of IS as socio-technical systems that acquire information, store and process data, and present information to humans. However, applications of SocIS extend beyond organizational contexts and use cases of traditional IS (Parameswaran and Whinston 2007a, b). They "shift[] computing to the edges of the network, and empower individual users … to manifest their creativity, engage in social interaction, contribute their expertise, share content, collectively build new tools, disseminate information and propaganda, and assimilate collective bargaining power" (Parameswaran and Whinston 2007a).

In the following, we briefly outline six important and constituent characteristics of SocIS: digital sociability, prosumer role, continuity, open virtual community, reach, and co-governance. SocIS are IS that afford various digital social interactions such as coordination, communication, and collaboration between humans through IT artifacts (social software/technologies) (Butler 2001; Bagozzi and Dholakia 2002; Schlagwein et al. 2011). In addition to these basic social functions, they facilitate the emergence of more complex social phenomena such as collective action and community formation (Ali-Hassan and Nevo 2009) and create new interaction dynamics (Agarwal et al. 2008) (we term this the *digital sociability* characteristics of SocIS). Affordances of social interactions are provided to users who – to execute social interactions – produce, modify, exchange, and consume digital, so-called user-generated content. Users thus become "prosumers" who can consume and produce the data and information captured by SocIS (Parameswaran and Whinston 2007a, b; Agarwal et al. 2008; Ali-Hassan and Nevo 2009; Kaplan and Haenlein 2010) (*prosumer role*). Further, SocIS afford continuous social interactions and content creation/modification (Bagozzi and Dholakia 2002), that is, not limited to a certain occasion, project, or otherwise predefined timeframe (*continuity*). Digital sociability afforded continuously to prosumers enables the emergence of a virtual community (Rheingold 1993; Bagozzi and Dholakia 2002) that is open in the sense that its members are not predetermined and coerced to use SocIS but rather are self-motivated and use SocIS voluntarily. Hence, the community emerging around a social information system is potentially large, heterogeneous, and changing (Gu et al. 2007; Ma and Agarwal 2007; Agarwal et al. 2008; Schlagwein et al. 2011; Xu et al. 2014) (*open virtual community*). The virtual community is, however, not an inseparable group; rather, its members typically interact in nuanced subsets of different social reach, such as one-on-one, different groups, or community-wide (*reach*). Finally, governance of these virtual communities in SocIS is typically described as decentralized, bottom-up, informal, and reliant on consensus and agreement (Ali-Hassan and Nevo 2009). The term *co-governance* has been introduced to describe this mode of governance executed by community members. It is often seen as democratic or meritocratic, that is, the assignment of more prominent roles is based on voting and/or members' reputation or achievement (Parameswaran and Whinston 2007a, b).

In summary, SocIS provide unique conditions and affordances for the production and consumption of information. We use the characteristics of SocIS outlined above to discuss prevalent DIQ conceptualizations from a novel perspective and identify the shortcomings of such conceptualizations in light of SocIS.

2.3 The Concept of Quality in Data and Information Quality Research

In this subsection, we briefly introduce and provide a generic conceptualization of "quality" that will later serve as a framework to juxtapose different conceptualizations for DIQ we obtain from the literature.

Juran and Godfrey (1999) define quality both as "those *features of products* which meet customer needs and thereby provide customer satisfaction" (Juran and Godfrey 1999, emphasis in original) and as "*freedom from deficiencies* – freedom from errors that require doing work over again (rework) or that result in field failures, customer dissatisfaction, customer claims, and so on" (Juran and Godfrey 1999, emphasis in original). They explicitly delineate this definition from earlier conceptualizations of quality as "conformance to specification." The ISO 9000 norm also lists several different definitions of "quality," for example, "degree of excellence," "fitness for use," "fitness for purpose," and "the totality of characteristics of

an entity that bear on its ability to satisfy stated or implied needs" (Hoyle 2006). It includes the conceptualizations mentioned by Juran and Godfrey (1999). The ISO 9000 standard then defines quality as "the degree to which a set of inherent characteristics fulfills a need or expectation that is stated, generally implied or obligatory" (Hoyle 2006).

These definitions have in common that quality is conceptualized as the relation between (a) a *target level* and (b) an *actual level* of (c) one or multiple specified *quality dimensions* (or quality criteria) of an entity. Quality increases as actual levels of dimensions approach the target levels, measured by some metric/s that operationalize/s the qualitative dimension. For example, in the first definition by Juran and Godfrey, quality of a product is defined by how the state (actual level) of certain product features (quality dimensions) compares to what a customer needs (target level). The ISO 9000 definition also views quality as a relational concept in which a set of inherent characteristics (actual level of quality dimensions) are compared to needs or expectations (target level).

This relational understanding of quality in general as actual levels compared to target levels of quality dimensions can be used as an analytical framework to compare different DIQ definitions conceptually by answering three questions for each definition:

1. How is the required *target level* of DIQ dimensions defined?
2. How is the *actual level* of DIQ dimensions determined?
3. How are the relevant DIQ *dimensions* specified?

We refer to this analytical framework for conceptual DIQ definitions as the Target-Actual-Dimension (TAD) framework. A conceptual definition of DIQ (or "general definition" as referred to by Illari 2014) is one that states the core theoretical idea of what is likely a more expansive DIQ definition. While the *conceptual level* provides a rather abstract definition, *dimensions* detail the (most) relevant facets of DIQ according to and shaped by the conceptual level, and *metrics* operationalize them. The intuition is, however, not top-down, but that all layers are equally important and mutually dependent in defining DIQ, for example, for a research study or during implementation of an IS. While the conceptual definition guides the selection and specification of dimensions and metrics, dimensions and metrics make a conceptual definition applicable only to a research question or a practical problem (Kahn et al. 2002; Illari 2014).

The conceptual background provided in this section serves as a foundation to systematically review DIQ conceptualizations found in the literature. In the following sections, we (1) identify and categorize existing DIQ conceptualizations by conducting a systematic literature review, which we (2) compare based on the three dimensions of the TAD framework and then (3) discuss in light of the characteristics of SocIS. In doing so, we answer the following research question: *How is DIQ conceptualized in IS research and do prevalent DIQ conceptualizations accommodate the characteristics of SocIS?*

3 Methodology

Our methodology comprises two steps that follow from the two parts of our research question, namely, (1) identify the DIQ conceptualizations that exist in IS research and (2) analyze their applicability to SocIS. Both steps are conceptual in nature, but with different goals.

In the first step, we build a taxonomy of existing DIQ conceptualizations. The conceptual contribution is what MacInnis (2011) calls a *differentiation*. It aims at adding clarity by distinguishing entities through, for instance, a taxonomy or typology. We build the taxonomy by means of a structured literature review in which we identify DIQ definitions in IS research studies and group them to distinct conceptualizations in an inductive way.

The resulting taxonomy of DIQ conceptualizations provides the input for the second step of our methodology in which we analyze critically the applicability of these conceptualizations to SocIS. We do so by revealing the conceptualizations' assumptions about how target level, actual level, and DIQ dimensions are determined, and then by comparing these assumptions to the characteristics of SocIS. The type of conceptual contribution in this step is *revising*, that is, "taking a novel perspective on something that has already been identified" (MacInnis 2011). DIQ is an established topic in IS research and there are already studies that investigate DIQ in SocIS. However, we analyze whether existing DIQ conceptualizations are applicable to the new class of SocIS.

3.1 Differentiation: Building a Taxonomy of Data and Information Conceptualizations

We conducted a structured literature search to identify DIQ definitions in the IS literature. The details of the search process are documented in Appendix A1. We obtained from the search process a set of articles with specific DIQ definitions which were then used to build a taxonomy of DIQ conceptualizations. We followed the guidelines for taxonomy development provided by Nickerson et al. (2012). In the following, we present the methodology used in this step.

Nickerson et al. (2012) propose an iterative method to develop taxonomies (for a brief summary see Appendix A2.1). A taxonomy is a set of (one or more) dimensions

each consisting of (two or more) characteristics that are required to be mutually exclusive and collectively exhaustive (Nickerson et al. 2012). This means that each object that ought to be categorized according to the taxonomy has to have exactly one characteristic for each dimension. The challenge is then to develop a taxonomy – that is, a set of dimensions and characteristics – that effectively or usefully discriminates between empirical objects (Nickerson et al. 2012). What is effective/useful is determined by the specific purpose for which the taxonomy should be used.

In our case, the taxonomy is supposed to facilitate a discussion of conceptually different definitions of DIQ with respect to SocIS. Hence, as a meta-characteristic, we chose whether definitions are conceptually different. As already mentioned, we define a DIQ conceptualization as the core theoretical idea of a DIQ definition that is probably larger. The conceptualization guides theoretically both the selection of DIQ dimensions and their operationalization through DIQ metrics. Further, the iterative method requires the specification of ending conditions. Objective ending conditions were taken from Nickerson et al. (2012) (i.e., all objects examined; no object merged/split in the last iteration; no characteristic added/merged/split in the last iteration; at least one object under each characteristic; do duplicate characteristics). As subjective ending conditions, we specified that the taxonomy should be both robust (characteristics allow for differentiation among DIQ definitions) and comprehensive (all definitions from the studies in our search result can be categorized).

We alternated between conceptual-to-empirical and empirical-to-conceptual loops in our iterations. The criteria to assign a study to a characteristic (DIQ conceptualization) were whether the study defined DIQ according the DIQ conceptualization and whether it cited at least one of the key publications. Studies that did provide a DIQ conceptualization but were different from the conceptualizations we had obtained to that point were deferred to subsequent iterations.

In each conceptual-to-empirical loop, we examined whether studies that had not yet been categorized could be categorized according to the current taxonomy. Further, we analyzed whether characteristics should be merged or split given the empirical DIQ definitions.

In each empirical-to-conceptual loop, studies were examined for common DIQ conceptualizations, identified through conceptual similarity and common references to DIQ definitions. If necessary, we adapted the taxonomy by adding new or merging/splitting existing characteristics. For example, some studies did not cite one the key publications, but conceptualized DIQ similar to existing conceptualizations. These implicit applications of conceptualizations were merged with the conceptualizations based on the respective key publications.

We began conceptual-to-empirical with a set of DIQ conceptualizations from our knowledge of the DIQ domain. All studies were examined and categorized by two of the authors. Disagreements were resolved through discussion. The iterative process was repeated until the ending conditions were met.

3.2 Revising: Critical Analysis of Data and Information Quality Conceptualizations for Social Information Systems

Revising can be achieved by "revealing and questioning the validity of hidden or explicit assumptions, foundational premises, or tenets in the extant view and indicating their limiting features" (MacInnis 2011). We questioned assumptions of existing DIQ conceptualizations in light of a new phenomenon – SocIS. The analytical device for this step is the TAD framework derived earlier from the relational nature of the quality concept (Juran and Godfrey 1999; Hoyle 2006). For each DIQ conceptualization, we asked: How is the required target level of DIQ dimensions defined according to the conceptualization? How is the actual level of DIQ dimensions determined? How are the relevant DIQ dimensions specified? Having thus identified the assumptions of each conceptualization with respect to these questions, we critically analyzed whether they are theoretically compatible with the specific characteristics of SocIS.

4 Existing Conceptualizations of Data and Information Quality

In this section, we present the categories of DIQ conceptualizations identified in our taxonomy. The conceptualizations of DIQ as correspondence, fitness for use, and semiotic provided our initial set of categories. The categories of conformance, perceived, organizational, user-generated content, hybrid, and only dimensions were identified during the process of taxonomy building. Quantitative results regarding the prevalence of conceptualizations in the literature are given in Appendix A2.2. Conformance, correspondence, and fitness for use are presented first because they are referred to by other conceptualizations presented later. Hybrid and only dimensions are presented last because they need not be further discussed with respect to SocIS, as we will argue. The remaining conceptualizations are presented in no particular order.

4.1 Conformance

First, we identified a conceptualization of DIQ as conformance of data to data-related constraints such as meta-data or integrity rules. Weber et al. further distinguish hard from soft constraints, providing as examples that "an attribute value for email address has to contain an @ in order to be valid" (hard) and that "an attribute description could be recommended to be longer than 30 characters" (soft) (Weber et al. 2013). Link and Memari (2013) define DIQ in terms of whether data meet referential integrity constraints. This conceptualization is arguably rather technical because DIQ is defined with respect to conformance to rules/constraints that are formally specified, for example, in a database management system and can be evaluated without human interference. Hence, we termed this conceptualization as *conformance* DIQ.

4.2 Correspondence

The second conceptualization is DIQ as "the measure of the agreement between the data views presented by an information system and that same data in the real world" (Orr 1998). This conceptualization is rooted in "the role of an information system … to provide a representation of an application domain (also termed the real-world system) as perceived by the user" (Wand and Wang 1996). We termed this the *correspondence* conceptualization because the basic idea is that data are of high quality if they correspond to the phenomena they ought to describe. This conceptualization is also sometimes termed as "intrinsic" DIQ because it is often seen as being use-independent and hence intrinsic to data (Wand and Wang 1996). However, the use of "intrinsic" is not consistent in the literature. For example, Wang and Strong (1996) use "intrinsic data quality" to label a set of quality dimensions within their larger "fitness for use" conceptualization of DIQ, but with a different, use-dependent meaning. Hence, we decided to apply a distinct label for the "correspondence" category to avoid misunderstandings. Some studies defined DIQ as correspondence, but without references to our initial key publications (i.e., Wand and Wang 1996; Orr 1998). These studies were merged into one correspondence characteristic during the process of taxonomy building.

4.3 Fitness for Use

Third, the *fitness for use* conceptualization defines DIQ as the degree to which data/information "are fit for use by data consumers" (Wang and Strong 1996). The authors explicitly derive this conceptualization from the marketing and product quality literature. The notion is that data/information are produced to be used by a consumer for a specific task. Hence, DIQ needs to be evaluated with respect to how well information can be perceived, interpreted, and applied to a task by the consumer of that information, based on data she receives (see, e.g., Wang and Strong 1996; Strong et al. 1997; Ballou et al. 2003; Madnick et al. 2009). The label for this conceptualization is directly derived from the key publications (i.e., Wang and Strong 1996; Strong et al. 1997; Wang 1998). Again, some studies did not cite any of the key publications but conceptualized DIQ similarly.

4.4 Semiotic

Fourth, the *semiotic* DIQ conceptualization – key publications being Price and Shanks (2005) and Shanks and Darke (1998) – actually integrates three conceptualizations based on semiotic theory (Peirce 1931; Morris 1938; Price and Shanks 2005). Briefly summarized, semiotic theory distinguishes the sign (e.g., a character, word, icon), its referent or (intended) meaning (what the sign is supposed to refer to), and its use or interpretation (how the sign is understood and used by the interpreter). Further, the relationships between these three components are termed syntactic (between multiple signs), semantic (between signs and their respective intended meanings), and pragmatic (between signs and their interpretation and use by humans). The semiotic DIQ conceptualization states that DIQ comprises all three relationships, that is, syntactic DIQ, i.e., "the degree to which stored data conform to stored metadata"; semantic DIQ, i.e., "the degree to which stored data correspond to …represented external phenomena"; and pragmatic DIQ, i.e., "the degree to which stored data are suitable and worthwhile for a given use" (all quotes from Price and Shanks (2005)). Thus, the semiotic DIQ conceptualization integrates theoretically conformance DIQ (syntactic), correspondence DIQ (semantic) and fitness for use DIQ (pragmatic). Studies that cited semiotic DIQ *and* correspondence DIQ/fitness for use DIQ were categorized as semiotic DIQ.

4.5 Perceived

Fifth, in some studies DIQ is conceptualized as a feature of information that is perceived and attributed by individuals who are probably – but not necessarily – users of that information. The most significant difference with fitness for use is that DIQ is not explicitly constituted in relation to task and context. However, it is also different from conformance DIQ because it requires evaluation by humans, and from correspondence DIQ because it explicitly allows for subjective assessment of DIQ. Hence, we termed this conceptualization *perceived* DIQ. It is typically applied in studies that investigate phenomena such as IS adoption and

IS success that involve individual-level beliefs, attitudes, and behavior towards information and IS. These studies assume nomothetic associations of constructs in variance models, which they try to identify by means of survey data and quantitative methods (e.g., structural equation modeling). Perceived DIQ is included in these models as a first- or second-order construct similar to perceived ease of use or perceived usefulness. The most prominent model in this respect is the DeLone and McLean IS success model, which argues that IS success has six (DeLone and McLean 1992) or – in the updated version – seven (DeLone and McLean 2003) distinct but interdependent success dimensions, among which DIQ is defined as "the quality of the information that the system produces" (DeLone and McLean 1992). Studies refer to the DeLone & McLean IS success model directly and/or to one of its extensions/revisions (e.g., Seddon 1997; Rai et al. 2002) and may combine the structural model with specific measurement instruments for the DIQ construct from the literature (e.g., Doll and Torkzadeh 1988).

4.6 Organizational

Sixth, one study (van der Pijl 1994) argues that DIQ in an organization should be conceptualized as the fit between what information is needed in the organization (teleological perspective) and what information is produced by the organization's IS (causal perspective). The teleological perspective is determined by goals and targets on different organizational levels, namely, individual users and providers of information, business processes, business units, and the organization as a whole, including its market position and strategy towards competitors. The causal perspective sees DIQ as "the result of the quality of the process in which it is produced" (van der Pijl 1994), including analysis, design, and implementation of IS and data processing. We termed this the *organizational* DIQ perspective. Van der Pijl (1994) writes explicitly of fitness for use as one important perspective on quality (citing Juran et al. 1974), and the organizational conceptualization is, in fact, akin to fitness for use but takes a broader view that goes beyond individual use to include organizational goals and uses of information.

4.7 User-generated Content

The seventh category is constituted by three studies, each of which takes up one aspect of user-generated content that needs to be accounted for when conceptualizing DIQ in SocIS. Although these studies do not refer to an established conceptualization, they are connected to each other through the specific phenomenon – namely, user-generated content – for which they try to define and investigate DIQ. Hence, we termed this the *user-generated content* conceptualization of DIQ. Valecha et al. (2013) study contributions and DIQ in a collaborative crisis response IS (named "Ushahidi") during the aftermath of the 2010 Haiti earthquake. The study's empirical evaluation of Ushahidi aid-requesting threads with respect to these dimensions is reminiscent of and cites the fitness for use conceptualization, but the authors highlight explicitly the essential role of users/victims and their respective contributions (i.e., user-generated content) without which crisis response through Ushahidi would not have worked, and thus go beyond fitness for use DIQ. Lukyanenko et al. 2014a propose the DIQ conceptualization of "crowd information quality" as "the extent to which stored information represents the phenomena of interest to data consumers (and project sponsors), as perceived by information contributors". The "crowd information quality" conceptualization highlights the importance of user contributions and the need for IS to provide ways to capture information that are suitable for the contributors, while acknowledging that this may come at the cost of fitness for use. An empirical study by Kane and Ransbotham (2012) investigates DIQ of articles in Wikipedia's Medicine project and uses as a measure of DIQ the quality rating assigned to each article by the Wikipedia community. The study demonstrates a way in which prosumers of SocIS can explicate and argue their assessments of DIQ in user-generated content and then vote to agree upon the current state of DIQ, but also to improve the quality and negotiate and defined normative DIQ standards in social interaction. These three studies of DIQ in SocIS emphasize that content contributions and producers are vital for SocIS because they decide on the data they actually (want to) contribute. Hence, moreover, SocIS should be able to accept content in ways as flexible and adaptable to the producers as possible, while expecting a variety of content. They must also provide means by which consumers can find and receive data/information they need. Last, to improve the match between what is produced and what is/would be consumed, SocIS should provide means by which their prosumers can negotiate what quality means to them, thus constituting a normative understanding of DIQ through a socio-technical process.

4.8 Hybrid and Only Dimensions

Last, some other studies explicitly combine (at least) two of the above conceptualizations, but without adding further conceptualization (like the semiotic DIQ framework does).[1] For example, some studies cite the product and

[1] The combination of multiple DIQ conceptualizations should not be confused with the combination of multiple DIQ dimensions/metrics; the latter is common to most of the DIQ definitions (see Sect. 2.3 on the levels of DIQ definitions) and also occurs without a DIQ conceptualization (category "only dimensions").

service performance model for information quality (Kahn and Strong 1998; Kahn et al. 2002), which defines DIQ as "conforming to specifications and meeting or exceeding consumer expectations" (Kahn et al. 2002), thus combining conformance DIQ and fitness for use DIQ. We termed these *hybrid* conceptualizations. Further, some studies do not state a conceptual-level definition of DIQ at all, but merely combine (usually multiple) quality dimensions and metrics from the literature and existing frameworks of DIQ dimensions to define and operationalize DIQ. We categorized these as *only dimensions*. Both groups were excluded from further discussions of DIQ because they do not add DIQ conceptualizations to the discussion, either because they have no conceptualization (only dimensions) or at least no new one (hybrid).

5 Relating Data and Information Quality Conceptualizations to Social Information Systems

In this section, we analyze the DIQ conceptualizations by mapping them to the TAD framework and comparing them to the characteristics of SocIS, namely: enabling various forms of digital social interaction and collaboration (digital sociability); offering affordances for content production and consumption to users as a means for interaction (prosumer role); doing so without restriction to occasions or time frames (continuity); thus allowing for the emergence of a virtual community that is open to diverse prosumers (open virtual community); offering the potential for prosumers to interact nuanced subsets (reach); and being governed by the community members themselves (co-governance).

5.1 Conformance

The conformance conceptualization states that DIQ is the degree to which data conform to formally specified rules/constraints. While conformance certainly is, *ceteris paribus*, also desirable in SocIS, other facets of DIQ – ones not captured by conformance DIQ, such as understandability and honesty – will probably be more important for the purpose of social interaction (cf. digital sociability). Voluntary, non-professional prosumers of unstructured or semi-structured user-generated content outside a formal work context and task-description (cf. prosumer role) might be more willing and able to focus on those other DIQ facets, thus, sacrificing conformance DIQ to some degree; they can hardly be forced to take care of conformance (cf. open virtual community, prosumer role). Hence, prioritizing only this conceptualization of DIQ could come at the cost of other aspects of quality that are important for prosumers, or might even discourage production.

Further, the conformance conceptualization does not take into account individual, context-, and task-related perspectives on DIQ (cf. prosumer role, continuity). In general, the community should be able to decide which DIQ dimensions are relevant and which are valid states, as well as communicate and continuously adapt this definition to a changing and heterogeneous group of prosumers (cf. open virtual community, co-governance).

5.2 Correspondence

If DIQ is conceptualized as correspondence of data to external phenomena, the actual level of correspondence dimensions can be assessed objectively by comparing data to the respective values of external phenomena "seen" through the lens of a data model, either by technical means or by humans. The target level is defined by thresholds for desired degrees of correspondence between data and external phenomena. DIQ dimensions are specified by explicit definitions of different facets of correspondence (e.g., timeliness, accuracy, completeness). For example, quality of data in an inventory management system may be assessed with respect to accuracy (DIQ dimension) – operationalized as "numerical difference between stored and real-world counts of items" (DIQ metric) – by measuring the difference between data about the numbers of specific goods that should be available (according to the inventory management system) to those numbers in the real-world inventory (actual level) and comparing the results to reference values (target level).

However, we argue that the SocIS characteristics are not appropriately reflected in the correspondence conceptualization. Prosumers are not explicitly involved in defining the correspondence thresholds (cf. prosumer role) and the conceptualization does not include possible conflicts and necessary arbitration between different thresholds within in the community or different subsets of prosumers (cf. open virtual community, reach). Likewise, prosumers in SocIS will have individual perceptions of relevant DIQ dimensions and different perceptions will require arbitration between prosumers as well, while the correspondence conceptualization assumes dimensions to be explicit and agreed upon (cf. open virtual community, reach). The same applies to assessments of the actual level of correspondence assumed to be objective, although prosumers will have different subjective perceptions of the reference external phenomena and subjective interpretation of data (cf. prosumer role). Virtual communities are seen to be self-organizing (to some degree) and hence the community will decide these DIQ-related questions (cf. co-governance).

Similar to conformance, correspondence DIQ might be desirable in principle, but other aspects of DIQ will

probably be more important for social interaction (cf. digital sociability) and enforcing correspondence DIQ on voluntary, untrained prosumers seems difficult (cf. open virtual community, prosumer role). Even more problematic is that prioritizing correspondence might very well have negative effects. For example, in the context of content production by users/customers (e.g., citizen science, open innovation, social media), Lukyanenko et al. (2014b) argue that the conventional definition of the DIQ dimension "completeness" as "the ability of an information system to represent every meaningful state of the represented real world system" (Wand and Wang 1996; cited in Lukyanenko et al. 2014b) underrepresents the importance of the prosumers' role as content producers. However, voluntary, heterogeneous content producers may be unwilling or unable to provide data that are complete. Nevertheless, consumers may still be interested in what producers can provide. Thus, there is a tradeoff between completeness (complete representation of external phenomena) and, for example, accuracy (e.g., producers may provide dummy data only to complete their input), or even having any content at all (producers may be discouraged when faced with the required complete input).

5.3 Fitness for Use

Fitness for use DIQ is conceptualized as the extent to which information can be easily perceived, interpreted, and applied to a task by the consumer of that information, based on data she receives (Wang and Strong 1996). In this context, the information consumer largely determines the target level, actual level, and relevant quality dimensions by defining the "use." These elements determine the subjective assessment of the actual fit of some data/information (actual level), the implicit or explicit definition of desired fit (target level), and the implicit or explicit definition of relevant dimensions of fit (DIQ dimensions). Because information is data interpreted by humans, the consumer is involved in the manifestation of information as well as its quality assessment.

While the fitness for use conceptualization is typically applied to traditional IS, several studies also apply it to DIQ in SocIS. For instance, Arazy et al. (2011), studying antecedents of DIQ in Wikipedia articles, explicitly adapt the fitness for use DIQ and employ the dimensions of accuracy, objectivity, completeness, and representation from Lee et al. (2002) to conceptualize it further. Scholz and Dorner (2013), investigating antecedents of product reviews' helpfulness, motivate and structure textual features and meta-information of reviews along the consumer-centric DIQ framework established by Wang and Strong (1996).

We argue, however, that there are several problems when applying the fitness for use conceptualization of DIQ to SocIS. To begin with, prosumers not only use data/information but also produce them (cf. prosumer role). Prioritization of consumption is hence inappropriate because production and consumption are mutually dependent in SocIS with social interaction and open collaboration or prosumers (cf. prosumer role, digital sociability). Further, what "use" means is usually unknown *ex ante* (i.e., before the system is actually in use), heterogeneous, and changing because the prosumer groups of SocIS are usually open, possibly large, heterogeneous, and changing (cf. open virtual community) and because contexts and devices of prosumers change (cf. continuity). The same applies to production. Similar to problems with the correspondence conceptualization, solutions in SocIS will be rooted in the self-organizing capabilities of SocIS (cf. co-governance) and possibilities to bring together prosumers with complementary understandings of DIQ (cf. reach).

Technology and design in SocIS must accommodate data consumption by unknown/heterogeneous data consumers and hence provide more flexible or adaptable mechanisms to select and present data. They must further accommodate convenient, adaptable data production that relies on voluntary, self-motivated, non-professional producers. Hence, focusing only on fitness for use during consumption ignores the important role technology plays in SocIS in capturing data and bringing together prosumers who wish to collaborate (cf. prosumer role, digital sociability).

5.4 Semiotic

The semiotic DIQ framework integrates conformance, correspondence, and fitness for use DIQ. Hence, most of what can be criticized with respect to these three levels of semiotic DIQ and SocIS characteristics applies, and hence does not need to be repeated here. Nevertheless, since the semiotic DIQ conceptualization is explicitly theory-based, it would be interesting to investigate how DIQ could be extended to SocIS based on semiotics. In fact, Shanks and Corbitt (1999) proposed to add a social level of DIQ "on top" of the other three levels (syntactic, semantic, pragmatic), building upon the semiotic DIQ definition of Shanks and Darke (1998) and an extended taxonomy of semiotic levels by Stamper (1992). Shanks and Corbitt define (semiotic) social DIQ as "the *shared understanding* of the meaning of symbols. The goals for social DIQ are an *understanding* of different stakeholder viewpoints and an *awareness* of any biases and other cultural and political issues involved" (Shanks and Corbitt 1999; emphasis in original).

Shared understanding of the meaning of symbols (i.e., user-generated content) is an important aspect of DIQ in SocIS as well. However, the definition does not capture that, in SocIS, the prosumers, the content, and how the SocIS will be used by its community are not defined *ex ante*. These aspects are rather constituted in use and are hence dynamic (cf. open virtual community, prosumer role, continuity). Further, assuming that one social information system might potentially have a very large number of prosumers, "shared understanding" does not mean that *all* prosumers must share the same understanding. Rather, people with shared understanding should be able to find each other in the population of prosumers (cf. reach). In other words, a definition of DIQ in SocIS should incorporate the ideas of a partially shared understanding among prosumers with respect to *which content* is or should be in the SocIS, and what the *content means*.

5.5 Perceived

Since humans evaluate perceived DIQ, target levels are determined by their individual normative perceptions of how information should be in terms of quality, and actual levels are determined by their perception of the actual state of information. In principle, the specification of relevant DIQ dimensions would also be on the part of the individuals. However, because this conceptualization is often applied in quantitative survey studies of multiple constructs (including, among others, perceived DIQ), researchers often determine DIQ dimensions and related metrics as part of their selection/design of the measurement instrument for DIQ.

Much of what has been criticized with respect to the fitness for use conceptualization also applies to perceived DIQ: DIQ is only considered during consumption of data/information, not during production (cf. prosumer role) which does include that individuals can define DIQ in SocIS through continuous social interaction (cf. continuity); let alone the priming effect of measurement instruments in many survey studies with respect to DIQ dimensions which conflicts with the definition of rules of the virtual community by itself (cf. co-governance). Further and also related to the prevalent survey study type in which perceived DIQ is applied, DIQ perceptions are assumed to be homogeneous and can hence be measured using the same instruments across multiple individuals. This conflicts with the characteristic of SocIS to allow for heterogeneous notions of DIQ to co-exist (cf. reach).

5.6 Organizational

Organizational DIQ is the fit between an organization's goals and targets (on different levels) for which information are needed and organizational IS that produce information (for use on different levels) (van der Pijl 1994). Hence, target levels and relevant dimensions of DIQ, that is, which information are required and how should they be mannered to achieve goals and targets are determined by the teleological perspective. The causal perspective of organizational DIQ explains how actual information and its quality is constituted by current IS (or could be by others).

However, such a DIQ conceptualization is also not well suited for application to SocIS, partly because of what has already been said about fitness for use with respect to the role of prosumers in mutual production and consumption of content for social interaction (cf. prosumer role, digital sociability) and partly because in SocIS there is no hierarchy of organization and its goals, nor are there business processes and respective targets that could be fit to IS design and data processing (cf. open virtual community, co-governance).

5.7 User-generated Content

Studies in this category emphasize aspects of user-generated content that are also relevant for conceptualizing DIQ in SocIS. First, the voluntary user contributions ensure that there is any data/information at all; second to promote these contributions, entering content should be suitable to the contributors, probably at the cost of some DIQ on the part of the consumers; third, DIQ in user-generated content is not static but can be negotiated and defined in social interaction. However, these studies do not take the insights further to develop a conceptualization of DIQ in SocIS (which they also do not claim to do). Although the potential for interaction between producers and consumers is acknowledged in principle, the target and actual levels of DIQ and relevant dimensions are still primarily defined by the content consumers and not in interaction with content producers (cf. digital sociability, co-governance), while the role of consumers and producers are not fixed but interchangeable (cf. prosumer role).

For example, Lukyanenko et al.'s (2014a) "crowd information quality" is still specific to the crowd-sourcing context in which it has been proposed because it does not treat producers and consumers of data equally, since only consumers (and project sponsors) define the "phenomena of interest" and roles of producer and consumer are not considered interchangeable (cf. prosumer role). Hence, the definition does not allow for social interaction of prosumers within a virtual community (cf. digital sociability, open virtual community). Kane and Ransbotham (2012) do mention the potential of Wikipedia (and social media platforms in general) to enable collaborative knowledge management. However, they treat DIQ and related quantitative measures in Wikipedia as output variable

influenced by the contributor-article-network rather than conceptualizing it as a subject of interaction itself. As a result, the user-generated content conceptualization cannot serve as a conceptualization of DIQ in SocIS.

Table 2 (Appendix A2.2) summarizes the comparison and discussion of DIQ conceptualizations.

6 Discussion

Our review of existing DIQ conceptualizations has revealed several shortcomings when they are applied to SocIS. It has thus demonstrated the need for research on DIQ in SocIS that accounts for the specific characteristics that make SocIS different from traditional IS. In this section, we propose a new conceptualization of DIQ in SocIS. We begin by briefly summarizing the general problems of existing DIQ conceptualizations with respect to SocIS. From these, we derive three fundamental conditions of DIQ in SocIS and propose a new conceptualization that takes them into account. Further, we illustrate the conceptualization by means of the TAD framework, similar to the analysis of existing conceptualizations.

The general problems of existing conceptualizations can be summarized under three major themes, all rooted in specific assumptions about (traditional) IS. First, when human IS users are considered in conceptualizing DIQ, their role as information consumers is prioritized over their role as information producers (e.g., fitness for use, perceived). Not to mention conformance DIQ, in which neither consumers nor producers are included conceptually. This conflicts with the prosumer role in SocIS, that is, that both user roles – as producers and consumers of content – are equally important and mutually dependent for (digital) social interaction and collaboration. We conclude that DIQ in SocIS needs to be conceptualized as *reciprocal* between prosumers because DIQ in SocIS is inherently an interplay of different individual DIQ perceptions. Second, existing conceptualizations often assume that data/information and DIQ perceptions are homogeneous and static (e.g., perceived, semiotic). Such is not the case for SocIS, in which perceptions of DIQ and, hence, contribution and consumption of data/information may vary across heterogeneous (groups of) prosumers, time, and contexts. DIQ in SocIS is, hence, inherently *dynamic*. Third, specific aspects of IS use are assumed to be explicit so that DIQ management can be purposefully designed and evaluated (e.g., fitness for use, organizational, correspondence). For example, context, task, and real-world reference systems of IS use are derived from functional roles, business processes, and organizational goals. Such is not the case for SocIS, in which many aspects of IS use by often unknown, heterogeneous, and changing users are instead *implicit*, but nevertheless shape human information behavior and DIQ perceptions.

Following from this, we propose to conceptualize DIQ in SocIS as a *reciprocal, dynamic, and implicit socio-technical process that enables the matching of individual information supply by some prosumers and information demand by others*. The perspective of individual prosumers is important because whether and how they participate in SocIS and contribute or consume content is driven by their (information) behavior. However, when conceptualizing DIQ in SocIS, the individual level is not sufficient because social interaction and collaboration include multiple individuals. Hence, we propose to conceptualize DIQ in SocIS as a process of matching information supply and demand between multiple prosumers. This matching is reciprocal because DIQ perceptions of one prosumer that shape data/information during production are evaluated by other prosumers and their respective DIQ perceptions during consumption. It is dynamic because DIQ perceptions change across users, contexts, and time. It is also implicit because which DIQ perceptions and evaluations become effective during individual production and consumption can usually not be directly observed for other prosumers. Further, we conceptualize DIQ in SocIS as a process and speak of "information supply and demand" rather than "contributed and consumed content" because DIQ in SocIS is not restricted to data/information that have already been contributed and consumed at a given time. Rather, in interactive and collaborative SocIS, DIQ also includes the potential for future contribution and consumption given the prosumers of the SocIS, their perceptions of certain phenomena, their perceptions of DIQ, and their motivation and interest to participate in the SocIS. In other words, if one prosumer cannot find certain information in existing user-generated content or finds it to be lacking certain dimensions of quality, she can interact instantaneously with other prosumers and ask for contribution or improvement of that piece of information. The observable and measureable state of DIQ of some user-generated content in some SocIS as evaluated by some prosumers at a specific time can at best be indicative of the larger DIQ process of matching information supply and demand. Last, the process is socio-technical as it involves human prosumers engaging with technical features of an IT artifact.

Further, we propose to view the larger socio-technical process of matching information supply and demand as being composed of different socio-technical mechanisms that are actualized repeatedly by the prosumers, whether consciously or subconsciously. For instance, a prosumer may be brought into contact with other prosumers who match or produce content that matches her individual DIQ definition (*allocation*). A group of prosumers within the larger community may compare and discuss individual

DIQ definitions and negotiate some compromise (*negotiation*), resulting in a locally accepted definition of target levels, actual levels, and quality dimensions (*consensus*). New prosumers may learn accepted DIQ definitions from veterans in the community and explicitly formulated norms (*socialization*). Taking part in these activities and using such socio-technical mechanisms to mediate and arbitrate data/information and DIQ is part of collective information behavior in SocIS and hence should be considered when conceptualizing DIQ in SocIS.

With respect to existing DIQ conceptualizations, we consider the new conceptualization of DIQ in SocIS to augment existing conceptualizations and to provide a theoretical explanation of their interplay in SocIS. For example, fitness for use as a perspective on DIQ can be applied very well to a situation in which a prosumer wants to buy a product and hence reads product reviews to learn about it (see, for example, Scholz and Dorner 2013). In this scenario, she (in her role as a consumer of information) will evaluate information from reviews with reference to the task of product assessment in a specific context. Hence, the scenario described closely resembles the conceptualization of fitness for use DIQ. It is, however, incomplete in the context of SocIS because it does not include DIQ conceptualizations of other prosumers and possible DIQ-related interactions among prosumers. The reader could, for example, ask others for more/better information if she feels something is missing. Others might contribute additional information or refuse to do so, or even try to convince her that she is asking the wrong questions about the product and should modify her information demand.

How can our conceptualization explain the interplay of different existing DIQ conceptualizations in SocIS? For example, a prosumer might define correspondence to be the general definition of quality when maintaining her user profile in an online social network. There might also be other prosumers who think similarly that correspondence is important for profile information. However, their individual understanding of correspondence can be very different. Some may emphasize currency of profile information, that is, that when real-world information covered by the online profile changes (e.g., phone number, relationship status), one should update the online profile as soon as possible. Others may place value on veracity of information, that is, that all information presented in an online profile should be true. Hence, in this example, information contribution is affected by different definitions of correspondence that require arbitration through socio-technical mechanisms such as those mentioned above. Further, even if prosumers agree on a specific DIQ definition based on the correspondence conceptualization, prosumers who consume profile information may still not find the resulting profile information fit for their use, meaning that correspondence DIQ during production does not necessarily match fitness for use DIQ during consumption. Hence, further arbitration is needed to match different DIQ perspectives through a socio-technical process, as our conceptualization of DIQ in SocIS proposed.

Our taxonomy of DIQ conceptualizations and the newly proposed conceptualization of matching DIQ also provide a framework to discuss existing research on DIQ in a specific SocIS. For example, several authors studied DIQ in Wikipedia, but conceptualized it differently. Giles (2005) investigated the factual accuracy of articles edited collaboratively on Wikipedia and articles on the same topics in Encyclopædia Britannica Online. Accuracy was assessed by experts in the relevant fields. This approach follows correspondence DIQ as it conceptualizes DIQ as the degree to which information in the IS (i.e., Wikipedia and Encyclopædia Britannica Online, respectively) corresponds to the same information in the real-world reference system (i.e., scholarly knowledge). Arazy et al. (2011) analyzed how group composition and task conflict in groups of editors on Wikipedia that work collaboratively on one article affect article quality explicitly conceptualized as fitness for use. Senior librarians conducted the quantitative empirical assessment of fitness for use DIQ of the sampled Wikipedia articles in the study. As mentioned earlier, Kane and Ransbotham (2012) studied the contributor-article network of a sample of Wikipedia articles in the Medicine Wikiproject[2] to identify features of the network (e.g., number of contributors to the article, number of articles to which contributors also contribute) that positively affect an article's quality. As a measure of article quality, they used ratings that had been assigned in a collaborative process by the contributors according to the Wikiproject's article quality-grading schema. We categorized this study as one of those that explicitly conceptualized DIQ with respect to user-generated content based on quality assessments, which are the result of the same collaborative process that produces the evaluated information itself. Our conceptualization of DIQ as a socio-technical process in which different DIQ perceptions can co-exist and be arbitrated provides a foundation for understanding how studies that apply different DIQ conceptualizations to the same SocIS – like those illustrated above – can come to different conclusions. Further, it raises interesting questions regarding the co-existence and co-evolution of different DIQ conceptualizations among the users of a SocIS. How is a shared understanding of DIQ in SocIS related to the various DIQ conceptualizations described in our taxonomy and how does it emerge from a heterogeneous, open

[2] "A Wikiproject refers to a group of contributors who are dedicated to developing, maintaining, and organizing articles related to a particular topic" (Kane and Ransbotham 2012).

community of a SocIS? Do SocIS users change their understanding of DIQ over time and/or in specific contexts and, if so, how does this affect their information behavior?

Similar to the existing DIQ conceptualization, we can analyze the matching conceptualization according to the TAD framework. Target levels regarding DIQ dimensions are defined in principle by individual prosumers and take effect during production and consumption. However, prosumers may explicate and arbitrate target levels. Likewise, individual assessments of actual levels of data/information on DIQ dimensions during production and consumption are conducted by individuals but can be explicated, compared to each other between multiple prosumers, and probably revised. The same applies to the question of which DIQ dimensions are taken into account. In summary, DIQ is evaluated by individual prosumers, but individual evaluations are not isolated; rather, they are mediated and arbitrated.

7 Conclusion

In this section, we summarize our key findings and contributions, and also discuss limitations of our work and possibilities for future research.

First, we provide a comprehensive overview of existing conceptualizations of DIQ in IS literature and a comparative analysis of these conceptualizations. Our research is thus different from many other comparisons of DIQ definitions that focus on the comparison of DIQ dimensions used in different definitions/studies (e.g., Lee et al. 2002; Jayawardene et al. 2013). These comparisons are limited because similarly labeled but qualitatively described DIQ dimensions such as "accuracy" and "completeness" bear different meanings across different studies, according to how DIQ is defined at a conceptual level (Illari 2014). Hence, our review can provide structure and orientation in the field of DIQ research. However, we limited our review to the Senior Scholars' Basket (Association for Information Systems 2011) and the AIS Electronic Library to capture the state of the start of high-quality research. Future research should extend beyond this and also include literature from specific journals and conferences in the information or communication disciplines.

Second, our analysis of DIQ conceptualizations in light of characteristics of novel SocIS has revealed that existing conceptualizations have several shortcomings and do not capture specifics of DIQ in SocIS. They are thus limited with respect to describing, explaining, and influencing DIQ in SocIS. However, as our work is conceptual in nature, we do not provide an empirical assessment of existing DIQ conceptualizations in SocIS. Research on DIQ in SocIS seems to be at a very early stage, and empirical studies investigating DIQ conceptualizations in SocIS might be an interesting avenue for future research. Hence, our study should also be understood as a substantiated call for research into DIQ in SocIS.

Third, based on our review and the characteristics of SocIS, we provide a new conceptualization of DIQ in SocIS as the reciprocal, dynamic, implicit matching of individual information supply and demand of prosumers. We show how existing conceptualizations can be integrated into the larger matching conceptualization and that socio-technical mechanisms are important to achieve matching DIQ. We thus establish specific research themes for DIQ in SocIS that can serve as a foundation for conceptualizing DIQ in SocIS, but also as an agenda for more empirical research in this field. Future research should try in particular to determine how individual-level definitions of DIQ are constituted in SocIS; which existing DIQ conceptualizations can be applied; how DIQ conceptualizations and definitions co-exist and are mediated among multiple prosumers through socio-technical mechanisms; how these mechanisms shape individual and collective information behavior, including information supply and demand on an individual level and their matching on a system-level; and which types of socio-technical mechanisms best support matching DIQ.

Acknowledgements The authors thank three anonymous reviewers, the guest editors of the special issue on human information behavior, as well as the discussants at the International Conference on Information Systems 2015 for their input and valuable comments on earlier drafts of this article.

Appendix

A1 Structured Literature Search

To develop a taxonomy of existing DIQ conceptualizations, we first conducted a structured literature search of the DIQ domain in general. We identified relevant DIQ conceptualizations and definitions we used to develop a taxonomy. Our literature review followed the best-practice approaches of the IS discipline (Webster and Watson 2002; Kitchenham and Charters 2007).

A1.1 Search Process

We identified relevant articles by searching systematically the titles, keywords, and abstracts of all articles published

Table 1 Results of the search process

Initial papers identified	730
Removed papers: exclusion criteria	388
Removed papers: inclusion criteria	93
Final relevant papers	249

Table 2 Taxonomy and discussion of existing DIQ conceptualizations

Label	Key publications or examples	Conceptualization of DIQ	Target DIQ level is defined by …	Actual DIQ level is assessed by …	DIQ dimensions are specified by …	Problems with SocIS characteristics	Hits
Conformance	Examples: Link and Memari (2013), Weber et al. (2013)	DIQ is the degree to which data conform with formally specified data-related constraints such as meta-data or integrity rules	Valid states of data-related rules/constraints, defined in a database management system	Automated (algorithmic) comparison of data to other (meta-)data and rules/constraints	Formal data-related rules/constraints, specified in a database management system	Very restrictive for voluntary, non-professional production of (often unstructured or semi-structured) data (cf. prosumer role) Relevant DIQ dimensions for prosumers in social interaction go beyond integrity (cf. digital sociability) No individual, context-, and task-related perspectives on DIQ (cf. prosumer role, continuity) Community should decide about dimensions and target level and adapt them (cf. open virtual community, co-governance)	2
Correspondence	Key: Wand and Wang (1996), Orr (1998)	DIQ is the degree to which data in an IS correspond to the same data of relevant phenomena in the real world	Objective correspondence thresholds	Technical means or humans through objective comparison of data to external phenomena (as "seen" through a data model)	Explicitly defined dimensions of correspondence	Prosumers are not (explicitly) involved in defining correspondence thresholds and (individual) dimensions (cf. prosumer role) Individual and community-level correspondence thresholds/ dimensions exist and require arbitration (cf. open virtual community, reach) Prosumers have subjective perceptions of external phenomena and interpretations of data (cf. prosumer role) Relevant DIQ dimensions for prosumers go beyond correspondence (cf. digital sociability) and voluntary, untrained prosumers will have other strengths and needs (cf. open virtual community, prosumer role)	16
(Of which were implicit)	Example: Bardaki et al. (2013)						(9)

Table 2 continued

Label	Key publications or examples	Conceptualization of DIQ	Target DIQ level is defined by …	Actual DIQ level is assessed by …	DIQ dimensions are specified by …	Problems with SocIS characteristics	Hits
Fitness for use	Key: Wang and Strong (1996), Strong et al. (1997), Wang (1998)	DIQ is the degree to which data/information are fit for use by data consumers for a specific task	Data consumers according to their "use" of data (either implicitly or explicitly)	Data consumers by (subjective) comparison of data to their intended "use"	Data consumers according to their "use" of data (either implicitly or explicitly)	Use and production of data/information are equally important and mutually dependent in social interaction (cf. prosumer role, digital sociability) Prosumers, use, and production are unknown (at least *ex ante*), heterogeneous, and changing (cf. open virtual community) Contexts and devices of prosumers change (cf. continuity) Self-organizing capabilities and possibilities for complementary matchings need to be considered (cf. co-governance, reach) SocIS technology must accommodate convenient modes for production and consumption of content in social interaction (cf. prosumer role, digital sociability)	82
(Of which were implicit)	Examples: Prestipino et al. (2006), Cheong and Chang (2007)						(10)
Semiotic	Key: Shanks and Darke (1998), Price and Shanks (2005)	DIQ comprises all three semiotic relationships, namely, syntactic DIQ (conformance of data to metadata), semantic DIQ (correspondence of data to real-world phenomena), and pragmatic DIQ (suitability and value of data for a specific use)	Respective semiotic level, that is, syntactic, semantic, or pragmatic, which map to "conformance," "correspondence," and "fitness for use," respectively	Respective semiotic level, that is, syntactic, semantic, or pragmatic, which map to "conformance," "correspondence," and "fitness for use," respectively	Respective semiotic level, that is, syntactic, semantic, or pragmatic, which map to "conformance," "correspondence," and "fitness for use," respectively	Syntactic level: see conformance Semantic level: see correspondence Pragmatic level: see fitness for use (Semiotic) social DIQ level: DIQ should be conceptualized to be dynamic because the prosumers and content and how the social information system will be used by its community are constituted in use (cf. open virtual community, prosumer role, continuity) "Shared understanding" not necessarily community-wide, but in subgroups (cf. reach)	10

Table 2 continued

Label	Key publications or examples	Conceptualization of DIQ	Target DIQ level is defined by …	Actual DIQ level is assessed by …	DIQ dimensions are specified by …	Problems with SocIS characteristics	Hits
Perceived	Key: Doll and Torkzadeh (1988), Delone and McLean (1992, 2003), Seddon (1997), Rai et al. (2002), Wixom and Todd (2005)	DIQ is a feature of information that is perceived and attributed by individuals who are probably – but not necessarily – users of that information	Normative perceptions of individuals (usually survey participants) who may or may not be regular consumers of that information	Individuals (usually survey participants; may be regular information consumers) and their perceptions	Individuals, in principle, but actually often by researchers through survey instruments	DIQ considered only in consumption, not in production (cf. prosumer role) Negotiation of DIQ through continuous interaction not considered (cf. continuity) DIQ dimensions often primed by survey measurement instrument, not defined by the virtual community (cf. co-governance) Homogeneous DIQ perception assumed rather than co-existence of heterogeneity (cf. reach)	60
Organizational	Key: van der Pijl (1994)	DIQ in an organization should be conceptualized as the fit between which information are needed on different levels of the organization (teleological perspective) and which information are produced by the organization's IS (causal perspective)	How information must be to support and organization in best achieving its goals and targets (teleological perspective)	How information is (or could be) produced within the organization by its IS (causal perspective)	How information must be to support an organization in best achieving its goals and targets (teleological perspective)	Not compatible with the role of prosumers in mutual content production and consumption for social interaction (cf. prosumer role, digital sociability) Neither organizational hierarchy nor goals nor business processes and respective targets in SocIS (cf. open virtual community, co-governance)	1
User-generated content	Examples: Valecha et al. (2013), Lukyanenko et al. (2014a), Kane and Ransbotham (2012)	Studies connected by a shared phenomenon (user-generated content); DIQ of user-generated content to be evaluated by users/consumers	Consumers of information in user-generated content	Consumers of information in user-generated content	Consumers of information in user-generated content	DIQ evaluation by user-generated content consumers, not in interaction with producers (cf. prosumer role, digital sociability) DIQ as a result of social interaction, not as a subject (cf. co-governance)	3
Hybrid	Examples: studies citing the product and service performance model for information quality (Kahn and Strong 1998; Kahn et al. 2002)	Studies explicitly combining (at least) two of the above conceptualizations, but without adding further conceptualization	n/a	n/a	n/a	n/a	13

Table 2 continued

Label	Key publications or examples	Conceptualization of DIQ	Target DIQ level is defined by …	Actual DIQ level is assessed by …	DIQ dimensions are specified by …	Problems with SocIS characteristics	Hits
Only dimensions	Examples: Bansal et al. (1993), Fung and Lee (1999), Rotchanakitumnuai (2006), Zach (2011)	Studies do not state a conceptual-level definition of DIQ at all, but merely combine (usually multiple) quality dimensions and metrics from the literature and existing frameworks of DIQ dimensions to define and operationalize DIQ	n/a	n/a	n/a	n/a	62
Total							249

in the Senior Scholars' Basket (Association for Information Systems 2011) that is, *European Journal of Information Systems*, *Information Systems Journal*, *Information Systems Research*, *Journal of AIS*, *Journal of Information Technology*, *Journal of MIS*, *Journal of Strategic Information Systems*, and *MIS Quarterly*.

We conducted a keyword-based search (Kitchenham 2004; Kitchenham and Charters 2007) using two combinations of keywords: "information AND quality" and "data AND quality." We also searched the titles, keywords, and abstracts of all articles archived in the AIS Electronic Library (AISeL) for the keywords "information quality" and "data quality." We collected all papers published before 20 April 2015 that matched these keywords.

We screened the results manually, removed duplicates, and excluded articles that did not cover at least one of the concepts 'data quality', 'information quality', and DIQ. We included only articles that stated explicitly or referred to a definition of data quality and/or information quality.

A1.2 Results

Our search process resulted in a set of 730 articles. After removing duplicates and all articles that did match our exclusion criteria (see above), we identified 342 potentially relevant papers. We then screened each of these articles for their respective definitions of DIQ and decided to submit 249 articles to further analysis (see Table 1).

A2 Taxonomy

A2.1 Process of Taxonomy Development after Nickerson et al. (2012)

Nickerson et al. 2012 propose an iterative method to develop taxonomies. Briefly summarized, the method proceeds as follows: (1) based on the purpose of the taxonomy, determine a meta-characteristic that informs the selection of characteristics in later stages; (2) determine objective and subjective ending conditions for the iterative cycle to stop; and (3) choose whether to proceed "empirical-to-conceptual" or "conceptual-to-empirical." Then, in "empirical-to-conceptual": (4e) identify objects; (5e) identify their common characteristics; and (6e) group characteristics into dimensions and create/revise the taxonomy. In "conceptual-to-empirical," the process is: (4c) deduce characteristics and dimensions from prior knowledge, experience, or theory; (5c) examine whether objects for characteristics and dimensions; and (6c) create/revise the taxonomy. The process then continues as follows: (7) evaluate objective and subjective ending conditions and

either enter into the next iteration (step 3) or terminate, if all conditions are met. Note that in this method, it is possible to alternate between "empirical-to-conceptual" and "conceptual-to-empirical" iterations.

A2.2 Resulting Taxonomy

Table 2 presents an overview of the existing DIQ conceptualizations that were identified through the process of taxonomy building including short descriptions each conceptualization, key publications or examples for this conceptualization, how it maps to the TAD framework, its main points of conflict with SocIS characteristics, and how many studies were assigned to each DIQ conceptualization.

References

Agarwal R, Gupta AK, Kraut R (2008) The interplay between digital and social networks. Inf Syst Res 19:243–252

Ali-Hassan H, Nevo D (2009) Identifying social computing dimensions: a multidimensional scaling study. In: Proceedings of the international conference on information systems (ICIS) 2009

Arazy O, Nov O, Patterson R, Yeo L (2011) Information quality in Wikipedia: the effects of group composition and task conflict. J Manag Inf Syst 27:71–98

Association for Information Systems (2011) Senior scholars' basket of journals. In: Res. – Assoc. Inf. Syst. http://aisnet.org/?SeniorScholarBasket. Accessed 28 Apr 2015

Bagozzi RP, Dholakia UM (2002) Intentional social action in virtual communities. J Interact Mark 16:2–21

Ballou D, Madnick S, Wang R (2003) Special section: assuring information quality. J Manag Inf Syst 20:9–11

Bansal A, Kauffman RJ, Weitz RR (1993) Comparing the modeling performance of regression and neural networks as data quality varies: a business value approach. J Manag Inf Syst 10:11–32

Bardaki C, Kourouthanassis P, Pramatari K, Doukidis G (2013) An information quality evaluation framework of object tracking systems. In: Proceedings of the international conference on information systems (ICIS) 2013

Barnes S, Vidgen R (2009) An evaluation of user acceptance of a corporate intranet. In: Proceedings of the european conference on information systems (ECIS) 2009

Boell SK, Cecez-Kecmanovic D (2015) What is an information system? In: Proceedings of the annual hawaii international conference on system sciences (HICSS) 2015

Butler BS (2001) Membership size, communication activity, and sustainability: a resource-based model of online social structures. Inf Syst Res 12:346–362

Chai K, Potdar V, Dillon T (2009) Content quality assessment related frameworks for social media. In: Proc Int Conf Comput Sci its Appl ICCSA 09, pp. 791–805

Cheong LK, Chang V (2007) The need for data governance: a case study. In: Proceedings of the australasian conference on information systems (ACIS) 2007

DeLone WH, McLean ER (1992) Information systems success: the quest for the dependent variable. Inf Syst Res 3:60–95

DeLone WH, McLean ER (2003) The DeLone and McLean model of information systems success: a ten-year update. J Manag Inf Syst 19:9–30

Doll WJ, Torkzadeh G (1988) The measurement of end-user computing satisfaction. MIS Q 12:259–274

English LP (1999) Improving data warehouse and business information quality: methods for reducing costs and increasing profits. Wiley, New York

Floridi L (2011) The philosophy of information. Oxford University Press, Oxford

Fung R, Lee M (1999) EC-Trust (trust in electronic commerce): exploring the antecedent factors. In: Proceedings of the Americas conference on information systems (AMCIS) 1999

Giles J (2005) Internet encyclopaedias go head to head. Nature 438:900–901

Glowalla P, Sunyaev A (2014) Process-driven data quality management: a critical review on the application of process modeling languages. ACM J Data Inf Qual 5:1–7

Gu B, Konana P, Rajagopalan B, Chen H-WM (2007) Competition among virtual communities and user valuation: the case of investing-related communities. Inf Syst Res 18:68–85

Hansen P, Järvelin K (2000) The information seeking and retrieval process at the Swedish patent- and registration office. Moving from lab-based to real life work-task environment. In: Proceedings of the SIGIR 2000 workshop on patent retrieval, pp 43–53

Hansen P, Järvelin K (2004) Collaborative information searching in an information-intensive work domain: preliminary results. J Digit Inf Manag 2:26–30

Hirschheim R, Klein HK (2012) A glorious and not-so-short history of the information systems field. J Assoc Inf Syst 13:188–235

Hoyle D (2006) ISO 9000 Quality systems handbook, 5th edn. Butterworth-Heinemann, Oxford

Illari P (2014) IQ: Purpose and DIMENSIONS. In: Floridi L, Illari P (eds) The philosophy of information quality. Springer, Cham, pp 281–301

Jayawardene V, Sadiq S, Indulska M (2013) An analysis of data quality dimensions. IEEE Tech Rep No. 2013–01

Juran JM, Godfrey AB (eds) (1999) Juran's quality handbook, 5th edn. McGraw-Hill, New York

Juran JM, Gryna FM, Bingham RS (eds) (1974) Quality control handbook, 3rd edn. McGraw-Hill, New York

Kahn BK, Strong DM (1998) Product and service performance model for information quality: an update. In: Proceedings of the 1998 conference on information quality

Kahn BK, Strong DM, Wang RY (2002) Information quality benchmarks: product and service performance. Commun ACM 45:184–192

Kane GC, Ransbotham S (2012) Codification and collaboration: information quality in social media. In: Proceedings of the international conference on information systems (ICIS) 2012

Kane GC, Alavi M, Labianca GJ, Borgatti SP (2014) What's different about social media networks? A framework and research agenda. MIS Q 38:275–304

Kaplan AM, Haenlein M (2010) Users of the world, unite! The challenges and opportunities of social media. Bus Horiz 53:59–68

Kitchenham B (2004) Procedures for performing systematic reviews. Department of Computer Science, Keele University and National ICT Australia Ltd, Keele

Kitchenham B, Charters S (2007) Guidelines for performing systematic literature reviews in software engineering. Keele University and Durham University Joint Report, Keele and Durham

Lee AS (2010) Retrospect and prospect: information systems research in the last and next 25 years. J Inf Technol 25:336–348

Lee YW, Strong DM, Kahn BK, Wang RY (2002) AIMQ: a methodology for information quality assessment. Inf Manag 40:133–146

Lili L, Rong D (2013) Roles of community commitment and community atmosphere: an empirical study of online community success. In: Proceedings of the Wuhan international conference on E-business (WHICEB) 2013

Link S, Memari M (2013) Static analysis of partial referential integrity for better quality SQL data. In: Proceedings of the americas conference on information systems (AMCIS) 2013

Lukyanenko R, Parsons J, Wiersma YF (2014a) The IQ of the crowd: understanding and improving information quality in structured user-generated content. Inf Syst Res 25:669–689

Lukyanenko R, Parsons J, Wiersma YF (2014b) The impact of conceptual modeling on dataset completeness: a field experiment. In: Proceedings of the international conference on information systems (ICIS) 2014

Ma M, Agarwal R (2007) Through a glass darkly: information technology design, identity verification, and knowledge contribution in online communities. Inf Syst Res 18:42–67

MacInnis DJ (2011) A framework for conceptual contributions in marketing. J Mark 75:136–154

Madnick SE, Wang RY, Lee YW, Zhu H (2009) Overview and framework for data and information quality research. ACM J Data Inf Qual 1:1–2

Mason RO, Mitroff II (1973) A program for research on management information systems. Manag Sci 19:475–487

McKenzie P (2003) A model of information practices in accounts of everyday-life information seeking. J Doc 59:19–40

Morris CW (1938) Foundations of the theory of signs. In: International encyclopedia of unified science. University of Chicago Press, London

Nickerson RC, Varshney U, Muntermann J (2012) A method for taxonomy development and its application in information systems. Eur J Inf Syst 22:336–359

Orr K (1998) Data quality and systems theory. Commun ACM 41:66–71

Ou CXJ, Davison RM, Cheng NCK (2011) Why are social networking applications successful: an empirical study of Twitter. In: Proceedings of the Pacific Asia conference on information system (PACIS) 2011

Parameswaran M, Whinston AB (2007a) Social computing: an overview. Commun Assoc Inf Syst 19:762–780

Parameswaran M, Whinston AB (2007b) Research issues in social computing. J Assoc Inf Syst 8:336–350

Peirce CS (1931) Collected papers. Harv Univ Press, Cambridge

Prestipino M, Aschoff F-R, Schwabe G (2006) What's the use of guidebooks in the age of collaborative media? Empirical evaluation of free and commercial travel information. In: Proceedings of the Bled eConference (BLED) 2006

Price R, Shanks G (2005) A semiotic information quality framework: development and comparative analysis. J Inf Technol 20:88–102

Rai A, Lang SS, Welker RB (2002) Assessing the validity of is success models: an empirical test and theoretical analysis. Inf Syst Res 13:50–69

Rheingold H (1993) The virtual community: homesteading on the electronic frontier. Addison-Wesley, Reading

Rotchanakitumnuai S (2006) Developing the electronic service acceptance model from internet securities trading system. In: Proceedings of the australian conference on information systems (ACIS) 2006

Sadiq S, Yeganeh NK, Indulska M (2011) 20 years of data quality research: themes, trends and synergies. In: Proceedings of the Australasian database conference (ADC) 2011, pp 153–162

Schlagwein D, Schoder D, Fischbach K (2011) Social information systems: review, framework, and research agenda. In: Research-in-progress, proceedings of the international conference on information systems (ICIS) 2011

Scholz M, Dorner V (2013) The recipe for the perfect review? An investigation into the determinants of review helpfulness. Bus Inf Syst Eng 5:141–151

Seddon PB (1997) A respecification and extension of the DeLone and McLean model of IS success. Inf Syst Res 8:240

Shanks G, Corbitt B (1999) Understanding data quality: SOCIAL and cultural aspects. In: Proceedings of the Australasian conference on information systems (ACIS) 1999. pp 785–797

Shanks G, Darke P (1998) Understanding data quality in data warehousing: a semiotic approach. In: Proceedings of the international conference on information quality (ICIQ) 1998

Stamper R (1992) Signs, organisations, norms and information systems. In: Proceedings of the Australian conference on information systems (ACIS) 1992

Strong DM, Lee YW, Wang RY (1997) Data quality in context. Commun ACM 40:103–110

Talja S (2002) Information sharing in academic communities: types and levels of collaboration in information seeking and use. New Rev Inf Behav Res 3:143–160

Talja S, Hansen P (2006) Information sharing. In: Cole C (ed) Spink A. Springer, Dordrecht, pp 113–134

Valecha R, Oh O, Rao R (2013) An exploration of collaboration over time in collective crisis response during the Haiti 2010 earthquake. In: Research-in-progress, proceedings of the international conference on information system (ICIS) 2013

van der Heijden H (2004) User acceptance of hedonic information systems. MIS Q 28:695–704

van der Pijl G (1994) Measuring the strategic dimensions of the quality of information. J Strateg Inf Syst 3:179–190

Wand Y, Wang RY (1996) Anchoring data quality dimensions in ontological foundations. Commun ACM 39:86–95

Wang RY (1998) A product perspective on total data quality management. Commun ACM 41:58–65

Wang RW, Strong DM (1996) Beyond accuracy: what data quality means to data consumers. J Manag Inf Syst 12:5–33

Weber D, Leone S, Norrie M (2013) Constraint-based data quality management framework for object databases. In: Proceedings of the European conference on information systems (ECIS) 2013

Webster J, Watson RT (2002) Analyzing the past to prepare for the future: writing a literature review. MIS Q 26:xiii–xxiii

Wilson TD (2000) Human information behavior. Inf Sci J 3:49–55

Winter S, Berente N, Howison J, Butler B (2014) Beyond the organizational "container": conceptualizing 21st century sociotechnical work. Inf Organ 24:250–269

Wixom BH, Todd PA (2005) A theoretical integration of user satisfaction and technology acceptance. Inf Syst Res 16:85–102

Xiao Y, Lu LYY, Liu JS, Zhou Z (2014) Knowledge diffusion path analysis of data quality literature: a main path analysis. J Informetr 8:594–605

Xu YC, Yang Y, Cheng Z, Lim J (2014) Retaining and attracting users in social networking services: an empirical investigation of cyber migration. J Strateg Inf Syst 23:239–253

Zach O (2011) Exploring ERP system outcomes in SMEs: a multiple case study. In: Proceedings of the european conference on information systems (ECIS) 2011

Zheng Y, Zhao K, Stylianou A (2009) Information quality and system quality in online communities: an empirical investigation. In: Proceedings of the special interest group on human computer interaction (SIGHCI) 2009