What is Quality of Data and Information in Social Information Systems? Towards a Definition and Ontology

Abstract

Data and information quality (DIQ) have been defined traditionally in an organizational context and with respect to traditional information systems (IS) as, for example, “fitness for use” of data or “intrinsic” DIQ. Numerous frameworks have been developed that operationalize traditional DIQ accordingly. However, over the last decade, social IS such as social media have emerged that enable social interaction and open collaboration of voluntary prosumers, rather than simply supporting specific tasks as do traditional IS in organizations. After reviewing the current definitions of DIQ, we conclude that these definitions do not capture DIQ in social IS well, nor how it is defined, maintained, and improved through social interaction. Hence, we propose a new definition of DIQ in social IS based on the notion of “matching” between dynamic, voluntary, and heterogeneous supply and demand of data/information. We illustrate our definition with an ontological framework and discuss its implications.

Keywords: Social information systems; Social media; Data quality; Information quality; Ontology
Introduction

In 1992, The Wall Street Journal wrote, “Thanks to computers, huge databases brimming with information are at our fingertips, just waiting to be tapped. They can be mined to find sales prospects among existing customers; they can be analyzed to unearth costly corporate habits; they can be manipulated to divine future trends. Just one problem: Those huge databases may be full of junk.” (cited after Wand and Wang (1996, pp. 86–87))

Replace “computers” with “social information systems” in the Journal quote, and it would sound quite familiar. This is especially the case now that social information systems (social IS) have become extremely successful in terms of user numbers and quantity of content, but are poorly understood with respect to data and information quality.

In essence, social IS “are information systems based on social technologies and open collaboration” (Schlagwein et al. 2011, p. 2). They include, for example, the various forms of social media. Many people use social IS to obtain and share general information, advice, or gossip, and for communication, entertainment, socializing, or political mobilizing (Kaplan and Haenlein 2010; Parameswaran and Whinston 2007a, 2007b; Schlagwein et al. 2011). Issues of data and information quality potentially affect all these aspects: Are users interested in and do they actually talk about the same phenomena? What motivates users to engage in social IS and to contribute content? Does the social medium allow producers of data to express their perceptions so consumers of data 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 data and information quality? These and other issues cannot necessarily be resolved successfully in social IS. For example, Denning and colleagues (2005) warned that developing an encyclopedia in an open and collaborative mode such as the Wikipedia project, rather than through the traditional editorial mode, might negatively affect quality of information. This has been proven to be partly correct, depending the topic domains or dimensions of quality (Mesgari et al. 2015).

Given the past decades of research on data quality (DQ) and information quality (IQ) for traditional information systems (Lee et al. 2002; Madnick et al. 2009; Sadiq et al. 2011; Xiao et al. 2014)—including the important role of IQ for IS success (DeLone and McLean 1992, 2003; Petter et al. 2008)—one might assume that understanding issues of DQ/IQ in social IS is merely a question of transferring existing definitions, frameworks, and measures to a new domain. In fact, several approaches have aimed at applying traditional DQ/IQ concepts to social IS (for an overview, see Chai, Potdar, and Chang (2009); Chai, Potdar, and Dillon (2009); Mesgari et al. (2015)). Also, with respect to IS success, IQ has been included in applications of the DeLone-McLean-model (DeLone and McLean 1992, 2003) to explain the success of social IS such as corporate intranets (Barnes and Vidgen 2009), online communities (Lili and Rong 2013; Zheng et al. 2009), and social micro-blogging services (Ou et al. 2011).

However, in line with other scholars (Lukeyanenko et al. 2014a; Lukeyanenko and Parsons 2015), we argue that traditional definitions of DQ/IQ are inappropriate to capture fully the characteristics of social IS, presumably because those definitions have been developed for traditional IS in an organizational context. For example, traditional DQ and IQ are often defined in terms of “fitness for use,” that is, whether data and information are fit to be used by data/information consumers (Ballou et al. 2003; Lukeyanenko et al. 2014a; Madnick et al. 2009; Strong et al. 1997; Wang and Strong 1996; Zhu and Wu 2011). Since “use” of data/information is supposed to serve as a benchmark to assess the level of DQ/IQ, this conceptualization implies three aspects: (1) user, task, and context of data/information use are known; (2) data/information use is prioritized over data/information production; and (3) development and usage of the IS are controlled by an IS sponsor that can enforce prioritization according to use to achieve “fitness for use” of data/information. While “fitness for use” is thus well suited for an organizational, utilitarian context (Parsons and Wand 2014), we argue it is inappropriate for social IS with highly variable, non-enterprise, voluntary, informal uses; users that produce and consume data/information in social interaction; and a central sponsor with little (if any) control over content and how the social IS is used. Similar problems arise when confronting other DQ/IQ paradigms with the concept of social IS, namely, the “intrinsic” DQ paradigm (Orr 1998; Wand and Wang 1996), the semiotic framework of DIQ paradigms (Price and Shanks 2005; Shanks and Darke 1998), including its social DQ paradigm (Shanks and Corbitt 1999), or the strategic DIQ paradigm (van der Pijl 1994).
We use the distinctive characteristics of social IS as described by Schlagwein et al. (2011) to establish a comprehensive understanding of the issues that arise when applying the traditional DQ/IQ paradigms in the social IS context. We thereby identify quality issues related to data and information in social IS that need systematic examination. Hence, the aim of this study is first to develop definitions of DQ and IQ in social IS that abstract from specific uses (i.e., user, task, context) of data/information, and that equally include the perspectives of consumers and producers of information, henceforth referred to as “prosumers” (Toffler 1980). For that purpose, we propose to view DQ/IQ in social IS as “matching” demand and supply of information between prosumers. “Matching” is seen as a socio-technical process of interaction, communication, and negotiation between prosumers enabled and mediated by technological means through social software and data. Certain socio-technical mechanisms may help promote social interaction and matching. So, to increase matching is both a social and technological challenge. We discuss how DQ/IQ in social IS can be improved, based on the proposed notion of matching supply and demand. Further, our definition can provide a new angle to evaluate existing mechanisms and foster the design of new mechanisms. Hence, this study should improve the understanding of DQ/IQ in social IS and guide further systematic research into this and related topics such as social IS success.

The remainder of this article is organized as follows. In the next section we lay the groundwork for developing a new definition of social DQ/IQ by introducing definitions of traditional IS, traditional DQ/IQ, and social IS, and by contrasting traditional IS to social IS. In the third section, we describe the literature review we conducted to compile a comprehensive set of existing DQ/IQ paradigms. In the fourth section, we discuss and evaluate existing paradigms of DQ/IQ with respect to their applicability to social IS. Further, we present our new “matching” paradigm for DQ/IQ in social IS and propose the ontology of DQ/IQ in social IS. In the fifth section, we discuss how this new paradigm corresponds to the specific characteristics of social IS, and illustrate how the rather abstract paradigm can inform the design of mechanisms to promote DQ/IQ in social IS. The paper concludes with a short summary of our work as well as a discussion of limitations and areas for further research into DQ/IQ in social IS.

Related Work

Traditional IS, DQ, and IQ

Questions regarding the quality of data and information processed by IS are not new to IS research (Lee et al. 2002; Sadiq et al. 2011; Wand and Wang 1996; Wang and Strong 1996; Xiao et al. 2014). In fact, researchers and practitioners alike have undertaken several approaches to define, measure, and improve DQ and IQ (for an overview see, for example, Knight and Burn (2005); Lee et al. (2002); Madnick et al. (2009)). Hence, before we turn to a rather new family of IS—namely social IS—and propose a new definition of data and information quality specific to them, it is advisable to point out what we understand about traditional IS, DQ, and IQ.

We share the long-held view of IS as socio-technical systems (Boell and Cecez-Kecmanovic 2015; Lee 2010) comprising both social (humans and groups) and technical (hardware and software) parts that interact to generate, process, and store information and data. The “role of an information system is to provide a representation of an application domain (also termed the real-world system) as perceived by the user” (Wand and Wang 1996, pp. 87–88). With respect to traditional IS, we think of IS primarily in an organizational context in which the system is supposed to support certain users in performing certain tasks and thus “aim to provide instrumental value to the user” (van der Heijden 2004, p. 696). The nature of these IS are hence classified as utilitarian, in contrast to IS of a hedonic nature that “aim to provide self-fulfilling value to the user” (van der Heijden 2004, p. 696). 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 have 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).

The quality of data and information in traditional IS is an important aspect for organizations, especially with respect to their success (Lee et al. 2002; Madnick et al. 2009; Petter et al. 2008; Price and Shanks 2005). Though the ideas of DQ/IQ are intuitively comprehensible, defining them and establishing how they can be measured and managed has long been challenging and changing (Madnick et al. 2009; Wand...
What makes it even more challenging and complicated is that studies dealing with DQ/IQ do not always differentiate between data and information and often do not share a similar understanding regarding these and other related terms (Kahn et al. 2002; Madnick et al. 2009; Price and Shanks 2005). Hence, to advance the understanding of DQ/IQ to social IS requires first laying out our understanding of data and information as well as DQ and IQ in the context of traditional IS.

Basically, data are what are stored in a database and processed by an IS: signs that are used according to certain syntactical rules, are objective, and represent facts about relevant phenomena external to the IS, that is, in the real/physical world (English 1999; Price and Shanks 2005; Wand and Wang 1996). 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; Glowalla and Sunyaev 2014; Price and Shanks 2005).

This distinction between data and information is also reflected in definitions of traditional DQ and IQ. A common definition of traditional DQ is the degree of integrity and correspondence of data to external phenomena, which comprises, for example, completeness, unambiguity, meaningfulness, and correctness (for example, in Orr 1998; Price and Shanks 2005; Wand and Wang 1996). Conversely, a common definition of traditional IQ is “fitness for use,” that is, the extent to which information can easily be perceived, interpreted, and applied to a task by the consumer of that information, based on data s/he receives (see, for example, Ballou et al. 2003; Madnick et al. 2009; Strong et al. 1997; Wang and Strong 1996). This may include dimensions such as accessibility, suitability of presentation, understandability, security, and flexibility.

Both concepts, DQ and IQ, have a long history in IS research. IQ is an antecedent of use and user satisfaction (Doll and Torkzadeh 1988; Petter et al. 2008; Wixom and Todd 2005) and is included in the well-known DeLone-McLean-model of IS success (DeLone and McLean 1992, 2003). Further, several studies have taken on the idea of DQ and IQ as multidimensional concepts, yielding a plethora of DQ/IQ frameworks with overlapping sets of dimensions (for an overview, see Jayawardene et al. (2013); Knight and Burn (2005); Lee et al. (2002)). Corresponding to the “fitness for use” notion of DQ/IQ, which dimensions are relevant and which levels of quality should be attained depends on the use (Kahn et al. 2002; Lee et al. 2002). Also in line with “fitness for use” and the importance of DQ/IQ for IS success, development of traditional IS typically begins with an analysis of the users’ information needs and other formal specifications (together referred to as “requirements”) before an IS is designed and implemented to fulfill these requirements to the extent possible (Hirschheim et al. 1995; Lukyanenko et al. 2014a). Requirements are used in this process as benchmarks to define the required levels of DQ/IQ prior to IS development, and to evaluate DQ/IQ when the system is in use. Thus, a traditional IS is usually fitted to the users' requirements to achieve high levels of (traditional) DQ/IQ. While the focus of traditional IS is on the information consumer, humans are also involved in the production of information (Kahn et al. 2002).

Though DQ and IQ have a history in IS research, differentiations between both concepts in the literature are inconsistent. Some studies focus only on DQ, excluding the perceptual transformation of data to information (Glowalla and Sunyaev 2014; Wand and Wang 1996). Others define “data quality” to cover both DQ and IQ, as defined above (Madnick et al. 2009). In this article, we distinguish between data, information, DQ, and IQ as mentioned above. However, to avoid misunderstandings, we use “DIQ” (“data and information quality”) in the remainder of this article when referring to both DQ and IQ.

Social IS

Since this article strives to develop a definition of DIQ for social IS, it must first be clear what a social IS is and how it differs from traditional IS. We build on an earlier definition, according to which a social IS is, essentially, an IS that is (1) based on social technologies and (2) enables or promotes open collaboration (Schlagwein et al. 2011, p. 2). Social technologies include basic Web technologies that enable the dynamic

---

1 The terminology in the domain of social IS is by no means consistent across the scientific community. Hence, we try to clarify—from our point of view—what is synonymous. What we call “social IS” may also be referred to as “social computing” (Parameswaran and Whinston 2007a, 2007b), “social-software
distribution, display, and (collaborative) creation of content, but also social software that more or less builds upon these technologies, such as wikis, social media/networking sites, platforms for content creation, and gaming (Kaplan and Haenlein 2010). Open collaboration enabled by these technologies can take various forms, such as socializing and communication (e.g., www.facebook.com), exchange of media content (e.g., www.youtube.com), collective creation of information repositories (e.g., www.wikipedia.org), or citizen science (e.g., www.galaxyzoo.org).

Social IS are still covered by the broader definition of IS as being socio-technical systems that acquire information (in this case, from humans/groups), store and process data, and present information to humans. However, applications of social IS expand beyond organizational contexts and use cases of traditional IS (Parameswaran and Whinston 2007a, 2007b). They “shift[s] 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 2007b, p. 763). In other words, when moving from traditional IS to social IS, the focus widens from concrete task support to all kinds of social interaction. This means one cannot generally classify the nature of social IS as exclusively utilitarian or hedonic. Some social IS may be more utilitarian (e.g., social bookmarking sites), others more hedonic (e.g., video sharing platforms), and some a mixture (e.g., social networking sites). The dominant nature will also be subject to prosumers’ perspective and usage of the social IS (Wang et al. 2009; Wu et al. 2011). Likewise, the widening from task support to social interaction should not be misunderstood as if social IS cannot be implemented and used in an organization; in fact, many companies already do so (Ellison et al. 2015; Leonardi et al. 2013; Wagner and Majchrzak 2006). But enterprise wikis, enterprise social media/networking sites, and other enterprise social IS still differ from traditional IS because they rely on social technologies and promotion of open collaboration, whereas traditional IS are designed to support specific tasks.

To illustrate how (enterprise and non-enterprise) social IS differ from traditional IS, we briefly summarize characteristics from Schlagwein et al. in which both IS types (ideally) differ from each other (for a more detailed discussion, see Schlagwein et al. (2011)).

**Sociability.** Central affordances of social IS are social interactions such as “friending,” “liking,” “sharing,” and “commenting,” while those of traditional IS are information and business processing. The user base of a social IS is often referred to as a “community” or “network” that has no predefined structures and roles. Instead, position, power, and control are distributed bottom-up, according to transparent social measures based on community interactions. In contrast, roles and credentials in traditional IS typically adhere to formal top-down, hierarchical structures of the organizational context (see Agarwal et al. 2008; Kohler et al. 2011; Xu et al. 2014).

**Openness.** Traditional IS typically have predefined user groups, at least to a certain degree, that must use the traditional IS either because they need information for their tasks (e.g., decision making) or because they are required to enter certain information (e.g., reporting). Traditional IS also explicitly exclude some people as users, whereas social IS, in contrast, are open participation systems with heterogeneous, voluntary, and possibly large user groups that are neither predefined nor restricted (see Howison and Crowston 2014; Parameswaran and Whinston 2007a, 2007b).

**User role.** For a traditional IS, there usually is a clear mapping of roles to people: developers (contractors or employees in the IT unit) create and maintain traditional IS; users (employees in functional units or customers) utilize it to enter or retrieve information. Social IS are also developed and maintained and capture and retrieve information, but users can engage in each of these activities and play the related roles voluntarily and based on motivation. Hence, social IS are “co-created” or “secondarily designed” by their users (see Germonprez et al. 2011; Millerand and Baker 2010).

---

systems” (Germonprez et al. 2011), or “social media” (Kaplan and Haenlein 2010), though we view “social media” as a certain subclass of social IS. Our understanding of “social technologies” is comparable to what has also been called “Web 2.0 technologies” (O’Reilly 2005). What we term “social software” is similar to “social computing tools” (Ali-Hassan and Nevo 2009).
**Content.** Content (i.e., stored data) in social IS is often termed “user-generated content” (UGC). UGC is typically less structured and more heterogeneous than content in traditional IS because of the openness of social IS and the multifunctional role of its users. Traditional IS, with their predefined roles and purposes as well as rules governing their use, exhibit higher levels of structuring and homogeneity of content (see Fader and Winer 2012; Ghose et al. 2012).

**Technology.** Social IS build on the aforementioned social technologies that are typically easier to interact with, be a user of, adapt to, or develop than are traditional IS. The source code of social IS can be open (e.g., www.wikipedia.org, www.joindiaspora.com), while that of traditional IS is typically closed. Development and deployment of traditional IS typically follows staged lifecycle models, with major and minor releases, whereas social IS are often in a perpetual beta state and continuously deployed. Further, and related to openness in a technological sense, social IS often provide Web interfaces to connect to other IS (see Kane and Fichman 2009; Prasarnphanich and Wagner 2009; Vaast et al. 2013).

**Location.** Social IS are inherently Internet-based online systems, accessed conveniently through, for example, a Web browser on various devices, and (almost) ubiquitously available to their users. The online aspect of social IS also implies that its community is continuously connected. Though traditional IS can be made accessible over the Internet, online access is not typically among their integral features. Instead, they are understood as local systems, available in a local network or installed on a local computer (Hirschheim and Klein 2012). Continuously connecting a community is not typically among their objectives (see Parameswaran and Whinston 2007a, 2007b).

Since participation in social IS is mostly voluntary, prosumers’ motivations to engage in social IS—for example, to use the social software and consume and produce content—play an important role. Though we cannot cover the wide array of literature on prosumer motivation in social IS, we offer a brief overview. Different motivations have been identified to affect engagement in various forms of social IS. Ridings and Gefen (2004) investigated prosumers’ reasons for joining a virtual community across various online communities. They found information exchange and making/cultivating friendships to be most important, followed by exchanging social support, recreation, and technical features of the community website. Wasko and Faraj (2005) found that a desired increase in reputation was positively associated with content contribution to an electronic community of practice. In a field experiment in an online movie recommendation system, Chen and colleagues (2010) found that social comparison—that is, providing information about how one’s level of contributions (in this case, movie ratings) compares to the median prosumer’s contributions—motivates prosumers to increase content contribution. Following a uses and gratification perspective (Blumler et al. 1974) to explain why people choose certain media, Shao (2009) proposed structuring different motivations according to three basic types of interactions of users with user-generated media: consuming (e.g., reading/watching content); participating (e.g., rating/liking/commenting content; or liking/friending other users), and producing (e.g., substantively creating own content). He associated consuming with the needs for information and entertainment, participating with social interaction and community development, and production with self-expression and self-actualization. Further, though analytically distinct, interactions and motivations are interdependent in reality (Shao 2009). Heinonen (2011) took up this differentiation between consumption, participation, and production, and investigated empirically how three basic motivational factors (entertainment, social interaction, information) may drive activities in each of the three interactive types.

**Structured Literature Review**

To develop a definition and ontology of DIQ in social IS, we first conduct a structured literature review of the DIQ domain in general. We identify relevant DIQ paradigms and definitions that we use to develop a taxonomy. We build on our results to distinguish different understandings of the concept and discuss them in the context of social IS. Based on our findings, we develop and propose our own definition and ontology of DIQ in social IS that accounts for the unique characteristics of this specific type of IS. In this section, we first describe the structured literature review process and then develop the taxonomy and categorize extant work on DIQ.
**Search Process**

Our literature review follows the best-practice approaches of the IS discipline (Kitchenham and Charters 2007; Webster and Watson 2002). We identified relevant articles by systematically searching the titles, keywords, and abstracts of all articles published in the Senior Scholars’ Basket (“Senior Scholars’ Basket of Journals” 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 that matched these keywords published before 20 April 2015. We manually screened the results, removed duplicates, and excluded articles that did not cover at least one of the concepts DQ, IQ, and DIQ. We included only articles that explicitly stated or referred to a definition of DQ and/or IQ.

**Coding Scheme**

From this screening of DIQ literature, we made two observations that guided the development of our taxonomy.

First, definitions of DIQ can exist on or comprise at least two different conceptual levels, namely, a DIQ paradigm and DIQ dimensions. The paradigm is a rather abstract and general statement about how quality should be conceptualized, while the dimensions detail the (most) relevant facets of quality according to and shaped by the paradigm. Three prominent examples of DIQ paradigms illustrate the paradigm-dimension distinction. The “intrinsic” DQ paradigm has been put forward by, among others, Orr (1998) and Wand and Wang (1996). These authors root their definition of DIQ in the “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, pp. 87–88), and hence define the DQ paradigm 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, p. 67). From this paradigm, Wand and Wang (1996) derive four dimensions of “intrinsic” DQ: completeness, unambiguity, meaningfulness, and correctness.

The well-known “fitness for use” paradigm by Wang and Strong (1996) provides another example. Acknowledging that product quality is ultimately based on consumers’ evaluation (Juran 1989; Juran and Gryna 1980), they define high-quality data as “data that are fit for use by data consumer” (Wang and Strong 1996, p. 6). Guided by this paradigm, they survey empirically how data consumers define DQ on a more specific level. From their findings, they develop a framework of 15 “data quality dimension” as a set of data quality attributes that represent a single aspect or construct of data quality” (Wang and Strong 1996, p. 6) and thus determine the “fitness for use” of data. Dimensions are, for example, accuracy, completeness, believability, relevancy, timeliness, ease of understanding, interpretability, and accessibility.

The paradigms in both definitions, while different, share several dimensions (e.g., completeness and correctness/accuracy). In fact, the comprehensive framework by Wang and Strong (1996) can be seen as *including* the intrinsic DQ definition and complementing it with aspects of data and information quality that are related to the use of data/information by consumers. This corresponds nicely to the distinction between data and information we made earlier: Since data are signs stored in a database that are supposed to represent external phenomena, data quality is determined by how good this representation is (in terms of, for example, the dimensions from Wand and Wang 1996)). Since information is data received and interpreted by humans, IQ is determined by the quality of the underlying data plus how well (in terms of the dimensions in the framework by Wang and Strong 1996)) data can be accessed, understood, and thus translated into information by humans. (Hence, it is more appropriate, in our terminology, to speak of “fitness for use” as a paradigm for DIQ.)

A last example is the integrated perspective on data, information, and their respective quality provided by the semiotic framework of DIQ by Shanks and Darke (1998) and Price and Shanks (2005). The framework builds upon the distinction between (1) the sign (e.g., a character, word, icon), (2) its referent or (intended) meaning (what the sign is supposed to refer to), and (3) its use or interpretation (how the sign
is understood and used by the interpreter). The authors borrow this from semiotics, which is the study of signs (Morris 1938; Peirce 1931; Price and Shanks 2005). 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 a human receiver). Quality of data and information can then be mapped to these relationships. DQ includes syntactic quality (“conformance to database rules”) and semantic quality (“correspondence to external phenomena”) (Price and Shanks 2005, p. 91). IQ comprises the receivers’ perception of syntactic and semantic quality (i.e., perception of DQ, since information is data received by humans) as well as pragmatic quality (“suitability for use” (Price and Shanks 2005, p. 91)).

For each of these three semiotic relations (or categories, as in Price and Shanks (2005)), quality may be defined and operationalized using various dimensions (criteria): for example, certain integrity rules (syntactic level); complete, unambiguous meaningful, correct mapping (semantic level); accessibility, suitable and flexible presentation, timeliness as well as perceptions of syntactic and semantic quality dimensions (pragmatic level) (Price and Shanks 2005). The work by Price and Shanks (2005) and Shanks and Darke (1998) underlines the importance of the paradigm-dimension distinction (category-criteria distinction), and how dimensions are shaped by and detail a paradigm. It also integrates three paradigms theoretically.

Our second observation from the screening of DIQ literature is that DIQ is typically seen as a multidimensional concept (Knight and Burn 2005; Lee et al. 2002; Zhang et al. 2014), that is, definitions include more than one dimension and the literature suggests a multitude of dimensions. Though we argued earlier that dimensions are shaped by and detail a rather abstract paradigm, it is also a common practice in the literature to define DIQ only in terms of its dimensions and not to follow a theoretical paradigm explicitly. Such studies typically combine several dimensions cited from the literature or developed ad hoc to define and operationalize DIQ. See, for example Blanco et al. (2010), McKinney et al. (2002), or Sanghoon et al. (2009), who used or combined existing DIQ scales (i.e., dimensions and items) from the literature to measure the construct of DIQ.

These observations informed the coding insofar as we focused on which paradigms were used in the studies, because the paradigm is what defines the perspective on DIQ, selection and definition of dimensions, and so forth. Further, the observations provided us with an initial set of categories, namely, intrinsic paradigm, fitness for use paradigm, semiotic framework, or no paradigm but only dimensions.

Analysis and Review

In this section, we present the results of our literature review process, apply the coding scheme to existing work on DIQ, and develop a taxonomy of DIQ definitions. We use the taxonomy to distinguish different understandings of DIQ and discuss them in the context of social IS. Finally, we develop our own definition of DIQ in the context of social IS, present a corresponding ontology of DIQ in social IS, and discuss implications of the new paradigm for socio-technical mechanisms to improve DIQ in social IS.

Review and Taxonomy of DIQ Definitions

Our search process resulted in a set of 730 articles. After removing duplicates and all articles that did not 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. In this section, we categorize and discuss the definitions and present our results as a taxonomy. Our observations from screening the literature provided an initial set of categories (see above), namely, definitions directly following the “intrinsic” paradigm; directly following the “fitness for use” paradigm; including both the “intrinsic” and “fitness for use” paradigms; employing the integrated semiotic framework; or using a combination of dimensions without referring to or defining a paradigm. We grouped articles not matching one of these categories into new categories, described below.

We assigned articles that do not cite one of the defined references (namely, Orr (1998) and Wand and Wang (1996)), but define a DIQ paradigm similar to the intrinsic to the new category “intrinsic—
indirect.2 For example, Link and Memari (2013) define DIQ in terms of data meeting referential integrity constraints. Bardaki et al. (2013) define DIQ for an object tracking system in terms of correspondence of information to real-world objects, further defined by the dimensions of completeness and accuracy. Some of these articles cite the work of Ballou and Pazer (1985), who define DQ as “accuracy (the recorded value is in conformity with the actual value), timeliness (the recorded value is not out of date), completeness (all values for a certain variable are recorded), and consistency (the representation of the data value is the same in all cases)” (Ballou and Pazer 1985, p. 153). This definition reflects the “intrinsic” paradigm as defined by Wand and Wang (1996) and Orr (1998).

Some definitions do not cite one of the defined references for “fitness for use” but are similar to this paradigm because they state a similar paradigm (e.g., “information quality is mainly to be evaluated from the user’s point of view” (Prestipino et al. 2006, p. 2), or “fitness for purpose” (Cheong and Chang 2007, p. 1000)), and/or because they cite similar paradigms (e.g., Bailey and Pearson (1983), who surveyed DIQ dimensions from managers, or the PSP/IQ model (Kahn et al. 2002), which builds on the “fitness for use” paradigm). We grouped these studies as “fitness for use—indirect.”

We grouped studies that borrowed the definition and/or measurement instrument for DIQ from one of four studies in which DIQ has been investigated as an antecedent of user satisfaction (DeLone and McLean 1992, 2003; Doll and Torkzadeh 1988; Wixom and Todd 2005) into a new category “user satisfaction.” One article (van der Pijl 1994) provides a paradigm of strategic DIQ in organizations and was assigned its own category (“strategic”). We grouped three articles into a new “social” category because they investigate social IS and provide DIQ definitions that do not fit into any of the other categories.

Table 1 is an overview of the categories, coding criteria, and number of articles in each category.

<table>
<thead>
<tr>
<th>Category</th>
<th>Criteria</th>
<th>Hits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intrinsic</td>
<td>Direct: Definition of DIQ as integrity or correspondence to external phenomena and citing/referring to at least one of the definitions from Orr (1998) or Wand and Wang (1996).</td>
<td>6</td>
</tr>
<tr>
<td>Fitness for use</td>
<td>Direct: Definition of DIQ in relation to use (user, task, or context) and citing/referring to at least one of the definitions from Strong et al. (1997), Wang (1998), or Wang and Strong (1996).</td>
<td>72</td>
</tr>
<tr>
<td>Intrinsic and fitness for use</td>
<td>“intrinsic—direct” and “fitness for use—direct”</td>
<td>2</td>
</tr>
<tr>
<td>Semiotic</td>
<td>Explicitly citing the semiotic DIQ framework from Price and Shanks (2005) and Shanks and Darke (1998) (although some studies also cite Wand and Wang (1996) and/or Wand and Strong (1996)).</td>
<td>10</td>
</tr>
<tr>
<td>Multidimensional</td>
<td>Combining dimensions from frameworks to create ad-hoc constructs of DIQ, but without stating or referring to a higher-level paradigm.</td>
<td>102</td>
</tr>
<tr>
<td>User satisfaction</td>
<td>Borrowing the definition and/or measurement instrument of DIQ from one of four studies in which DIQ is investigated as an antecedent of user satisfaction, namely DeLone and McLean (1992, 2003), Doll and Torkzadeh (1988), or Wixom and Todd (2005).</td>
<td>25</td>
</tr>
<tr>
<td>Strategic</td>
<td>New paradigm of strategic DIQ in organizations (van der Pijl 1994).</td>
<td>1</td>
</tr>
<tr>
<td>Social</td>
<td>Investigating social IS (mostly, social media), and DIQ definition did not qualify for one of the categories above.</td>
<td>3</td>
</tr>
</tbody>
</table>

Table 1. Categories, Criteria, and Number of Articles

2 Though integrity of data (syntactic quality) and correspondence of data to external phenomena (semantic quality) can be seen as two conceptually different paradigms (Price et al. 2008; Shanks and Darke 1998), we combine them in one category (“intrinsic”) because they both concern the quality of stored data.
Existing DIQ Paradigms versus Social IS

In the following section, we assess the applicability of the paradigms we found in the literature to social IS, given its characteristics. We discuss the “indirect” categories with their respective “direct” categories. We omit discussion of the “intrinsic and fitness for use” category because it has no additional conceptual content beyond the separate categories “intrinsic” and “fitness for use.” We also omit the “multidimensional” category from the discussion because studies in the category (by definition) do not employ a higher-level paradigm.

Intrinsic. The “intrinsic” category comprises two paradigms: integrity of data (syntactic level) and correspondence of data to external phenomena (semantic level). Part of the critique regarding definitions that restrict DQ to syntactic and semantic criteria has already been voiced earlier, and actually motivated the proposition of the use-centric “fitness for use” paradigm. As Wang and Strong argued, “it is the consumer who will judge whether or not a product is fit for use” (Wang and Strong 1996, p. 6). Strong, Lee, and Wang (while promoting “fitness for use”) further urged that to “solve organizational DQ problems ... one must consider DQ beyond the intrinsic view [and] move beyond stored data to include data in production and utilization processes” (Strong et al. 1997, p. 104). This argument underlies the insight that while necessary, syntactic and semantic quality alone are not sufficient to capture what data/information consumers who are supposed eventually to use the IS and data/information might think of DQ.

Obviously, this argument still holds for social IS in which prosumers are supposed to participate. In fact, the problems when focusing only on intrinsic DQ are intensified (compared to traditional IS) because prosumers in social IS not only receive data but also produce them (cf. user role and content in social IS). Further, prosumers are typically not professionally trained to produce the content in social IS, and enforcement of syntactic and semantic quality on behalf of the producers is hard to achieve because content production is voluntary and outside of professional, predefined business routines (cf. openness and content).

Surely high levels of intrinsic DQ may also be, ceteris paribus, desirable in social IS. However, prioritizing only this paradigm would come at the cost of other dimensions of quality that are important for prosumers, or might even discourage production. For example, in the context of content production by users/customers (e.g., citizen science, open innovation, social media), Lukyanenko, Parsons, and Wiersma (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, p. 93; as cited in Lukyanenko et al. (2014b, p. 4)) underrepresents the important role of the UGC producer. Specifically, completeness is typically determined with respect to data/information required by consumers, or with respect to a specific formal data model (which should, ideally, reflect the requirements). However, voluntary, heterogeneous UGC producers may not be willing or able to provide data that are complete in this sense. Nevertheless, consumers (e.g., companies, researchers, or a virtual community) 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). Hence, Lukyanenko and colleagues propose thinking of completeness in UGC contexts as “the extent to which a database captures all records of phenomena of potential interest to data consumers that data contributors are willing to provide” (Lukyanenko et al. 2014b, p. 4), and further propose using instance-based rather than class-based data models. They tested their proposition in a citizen science project on biodiversity in which voluntary users were asked to record observed species. They found that an IS that used an instance-based data model yielded significantly more contributions (observations) than one that used a traditional class-based data model (Lukyanenko et al. 2014b). Following these insights, in a different study (Lukyanenko et al. 2014a) the authors coined a new DIQ paradigm, “crowd IQ,” for applications similar to citizen science (discussed here under the category “social”).

Fitness for use. While the “fitness for use” paradigm is commonly applied to traditional IS, several studies also apply it to DIQ in social IS. For instance, Arazy and colleagues (2011), studying antecedents of IQ in Wikipedia articles, explicitly adapt the “fitness for use” perspective on IQ and employ the dimensions of accuracy, objectivity, completeness, and representation from Lee et al. (2002) to

We argue, however, that there are several problems when applying the “fitness for use” notion of DIQ to social IS. First, what “use” means is usually unknown ex ante because the prosumer group of a social IS is usually open, possibly large, heterogeneous, and changing (cf. sociability and openness of social IS). Hence, it is difficult to design the system according to specific information needs, tasks, and contexts given the variety of consumers, and to accommodate the best data production method for the different abilities, motivations, and knowledge among producers. In contrast, information needs in traditional IS are derived from and bound to specific business process tasks known throughout system development and use, and that are more stable. Further, the user base of a traditional IS is known and can be asked for quality criteria. Second, prioritization of consumption is inappropriate because production and consumption are mutually dependent in social IS with social interaction and open collaboration or prosumers (cf. user role). Further, content can be evaluated only when the system is in use, and adherence to requirements on behalf of the voluntary producers cannot be enforced as it can be in traditional IS by a central sponsor, who can purposefully select and design data/information sources and processes according to users’ requirements during system development and adapt during system usage (cf. content).

Third, since social IS are “located” online and, hence, potentially available ubiquitously, the use of a social IS is also subject to the users’ changing contexts (i.e., observable features such as location and occupation) and situations (i.e., social and emotional). This is different from traditional IS in which the “use” is also rather stable because traditional IS are typically designed to be used locally/offline in an organizational context (cf. location). Finally, technology in social IS must accommodate data consumption by unknown/heterogeneous data consumers and hence provide more flexible or adaptable mechanisms to select and present data. It 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 social IS in capturing data and bringing together prosumers who wish to collaborate (cf. technology).

In summary, the “fitness for use” notion of DIQ is well suited for the organizational context in which users, tasks, and context are known and rather stable (at least to a certain degree); in which the prioritization of user requirements is justified by the purpose of IS to support users in accomplishing their tasks; and in which adherence to user requirements in IS development and use can be enforced by a sponsor (i.e., management). However, it does not capture characteristics of social IS and how DIQ is constituted in social interaction. One might apply the “fitness for use” paradigm in a descriptive/ex-post evaluation of existing social IS by certain prosumers, but it cannot, for example, guide the design of social IS.

**Semiotic.** With respect to the syntactic, semantic, and pragmatic level of the semiotic DIQ paradigm, most of what has been written above about the “intrinsic” and “fitness for use” paradigm applies, and hence does not need to be repeated here. Nevertheless, since the semiotic DIQ paradigm is explicitly theory-based (namely, in semiotics (Price and Shanks 2005; Shanks and Darke 1998)), it would be interesting to investigate how DIQ could be extended to social IS based on semiotics. In fact, Shanks and Corbitt (1999) proposed to add a social level of DQ “on top” of the other three levels (syntactic, semantic, pragmatic), building upon the semiotic DIQ definition by Shanks and Darke (1998) and an extended taxonomy of semiotic levels by Stamper (1992). Shanks and Corbitt thus define (semiotic) social DQ as “the shared understanding of the meaning of symbols. The goals for social data quality are an understanding of different stakeholder viewpoints and an awareness of any biases and other cultural and political issues involved” (Shanks and Corbitt 1999, p. 789; emphasis in original).

While shared understanding of the meaning of symbols (i.e., UGC) is an important aspect of DIQ in social IS as well, the definition does not capture that it is not defined ex ante in social IS which content there will be (cf. openness, user role, content, location). Further, assuming that one social IS might potentially have a very large number of prosumers, “shared understanding” need 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. In other words, a definition of DIQ in social IS should incorporate the ideas...
of a partially shared understanding between prosumers with respect to which content is or should be in the social IS, and what the content means.

**User satisfaction.** Articles in this category refer either to the IS success model by DeLone and McLean (1992, 2003), in which IQ is an antecedent of system use, user satisfaction, individual impact, and organizational impact; or the user computing satisfaction measurement instrument developed by Doll and Torkzadeh (1988), which covers many aspects similar to common DIQ dimensions (e.g., precise information, needed content, presentation format, timeliness of information); or the integrated model of user satisfaction and technology acceptance by Wixom and Todd (2005), which has confirmed IQ to be related to information satisfaction. However, none of these studies state or cite a higher-order paradigm for DIQ, but rather focus on developing measurement instruments. Hence, there are no paradigms to be discussed with respect to social IS.

**Strategic.** An article by van der Pijl (1994) is the only study in this category. He proposes a strategic paradigm of DIQ in which “quality of information in the organization is understood as the degree of fit between the goals and targets of the organization and the information systems supporting those” (van der Pijl 1994, p. 185). In other words, the information needed to achieve organizational goals, business process targets, users' and providers' targets, and personal interests (teleological perspective) and how information is produced by IS (design) and data processing (causal perspective) should correspond (van der Pijl 1994). Van der Pijl (1994) writes explicitly of “fitness for use” as one important perspective on DIQ, but has a broader view that reaches beyond individual use to include organizational goals. Further, DIQ conceptualized as the fit between teleological and causal perspectives on data/information can provide a new angle on how to assess and manage DIQ. However, this paradigm is also not well suited for application to social IS, partly because of what has already been said on “fitness for use” and partly because in social IS 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.

**Social.** Finally, three studies are particularly interesting because they deal with DIQ in social IS and provide new approaches to DIQ. Valecha, Onook, and Rao (2013) study contributions and IQ in a collaborative crisis response IS (named “Ushahidi”) during the aftermath of the 2010 Haiti earthquake. Using Ushahidi, victims could make requests for aid, which were then categorized, mapped, visualized, and made available to aid agencies and the public. Requests could also be discussed in threads. Valecha and colleagues (2013) focus on two IQ dimensions they deem to be of paramount importance in crisis response: uncertainty reduction and urgency. The study’s empirical evaluation of Ushahidi aid-requesting threads with respect to these dimensions is reminiscent of the “fitness for use” paradigm, but the authors highlight explicitly the essential role of users/victims and their respective contributions (i.e., UGC) without which crisis response through Ushahidi would not have worked.

For such an application of DIQ in which content is primarily or exclusively user-generated, Lukyanenko, Parsons, and Wiersma (2014a) propose the DIQ paradigm of “crowd IQ” (their empirical work is situated in the area of citizen science; see above). Specifically, they define “crowd IQ” as “the extent to which stored information represents the phenomena of interest to data consumers (and project sponsors), as perceived by information contributors” (Lukyanenko et al. 2014a, p. 3). The crowd IQ paradigm highlights the importance of user contributions and the need for the IS to provide ways of information capturing suitable for the contributors, while acknowledging that this may come at the cost of fitness for use. More specifically, it is explicitly use-agnostic, that is, it allows for known as well as future uses of information. However, while crowd IQ explicitly acknowledges the important role of data/information producers and their perceptions, it is still specific to the crowd-sourcing context. It does not treat producers and consumers of data equally, since what are “phenomena of interest” are defined only by the consumers (and project sponsors). Data from producers may fit more or less with the interests of consumers. Further, the definition implies that the roles of producers and consumers are fixed. In fact, crowd IQ does not consider social interaction and open collaboration (at least not explicitly), which are essential for social IS (cf. sociability, openness, and user role). Hence, the definition of crowd IQ makes sense in the context of crowd-sourced data collection for flexible purposes (in which it has been proposed by Lukyanenko et al. 2014a), but it cannot serve as a paradigm of DIQ in social IS.

An empirical study by Kane and Ransbotham (2012) investigates IQ of articles in Wikipedia’s Medicine project and finds that quality is associated with the article-consumer-network’s structure. As a measure of the dependent variable (i.e., IQ), Kane and Ransbotham (2012) use the quality rating that has been
assigned to each article by the Wikipedia community during a process of nomination, review, and voting of articles. In addition to its interesting empirical results, the study demonstrates a way in which prosumers of a social IS prosumers negotiate and define, but also improve, quality of content. Prosumers can contribute to the articles (by writing new parts or editing existing parts), discuss issues with other prosumers, and nominate and vote for articles they evaluate as being of high quality, thus defining what high quality is and should be. Though Kane and Ransbotham (2012) do not define a DIQ paradigm, one could say that in the case of Wikipedia, quality is what the community judges to be of quality.

From these three studies of DIQ in social IS we learn three things that should inform an explicit paradigm of DIQ in social IS. First, UGC contributions and producers are vital for social IS; they decide which data they actually (want to) contribute. Hence, second, the social IS (most important, the social software, including interfaces, algorithms, and storage) should be able to accept data/information in ways as flexible and adaptable to the contributors as possible, while expecting a variety of content. It must also provide means by which consumers can find and receive data/information they need. Third, to improve the match between what is produced and what is/would be consumed, the social IS should provide means by which its prosumers can negotiate what quality of data/information means to them, thus constituting a normative understanding of DIQ through a socio-technical process.

**Definition of DIQ in Social IS**

To resolve the issues the existing paradigms of DIQ raise when applied to social IS, we propose a new paradigm of DIQ based on the previous discussion of existing DIQ paradigms and that follows the notion of “matching” the demand for and supply of information about certain phenomena.

Recall that social IS are meant to be places where individuals meet and interact in a variety of heterogeneous contexts. Meeting and interaction take place through reciprocal production and consumption of data/information regarding specific phenomena, mediated by the social software. More precisely, information provided by the producers are transformed into data by human-computer interaction (HCI) with the social software; data are stored, processed, and transmitted by the social software; and data are requested, received, and interpreted by consumers (in HCI), thus becoming information. Since people may simultaneously take the roles of information consumers and producers, they are more appropriately characterized as being prosumers in the social IS.

Which information about which phenomena a prosumer could supply to the social IS (hereinafter termed “information potential”) or might be searching for in the social IS (“information gap”) will be determined by her/his individual interest in and subjective perception of phenomena. However, which information s/he actually supplies (“supply”), thus producing data, and which data s/he actually requests, receives, and interprets (“demand”), thus becoming information, will also depend on her/his individual motivation to participate in the social IS.

Based on this conceptualization, we propose to view a social IS as a place where demand for information by some prosumers meets the information supplied by others, and may be (more or less) satisfied through information exchange in social interaction and open collaboration. We argue further that the paradigm of social DIQ should be conceptualized as the matching of the demand for information by some prosumers and the information supplied by others, determined by interest, perception, motivation, and produced/consumed data, as described above. Since information exchange in social IS is mediated by the social software, the actual level of matching is also affected by HCI during production/consumption of data (e.g., how easy, flexible, intuitive is the interface), and technical issues of data storage and processing by the social software.

Specifically, we define social IQ as

> the extent to which information supplied by prosumers (determined by their interest in and perception of phenomena as well as motivation and production of data) matches information demanded by other prosumers (likewise determined by interest in and perception of phenomena as well as motivation and consumption of data), mediated by social software and data through human-computer interaction;
and social DQ as

the extent to which processing and storage of data by social software facilitates representation of information and matching of information supply and demand.

Matching of information supply and demand has at least two facets: allocation and negotiation. Allocation means that matching can be achieved by finding appropriate content or counterparts for interaction for a given demand (or vice versa), given that social IS are characterized by heterogeneous content and a large, heterogeneous, changing user base. Negotiation means that the information supplied and demanded is not static in a social IS, but can be negotiated among its prosumers to increase the matching. Both allocation and negotiation involve prosumers and technical features of the social software. For example, recommendation algorithms may identify suitable content for consumers (allocation), or producers react and adopt their production to receive positive feedback from consumers (negotiation). Thus, defining, maintaining, and improving DIQ in social IS is essentially a socio-technical process enabled and supported by technical features of the social software as much as it relies on prosumers’ motivation and activity.

**Ontology of DIQ in Social IS**

Approaching DQ or IQ from an ontological (or theoretical) perspective is seen as an alternative to empirical, intuitive/ad-hoc, or literature-based approaches (Price and Shanks 2005; Wand and Wang 1996; Wang and Strong 1996). Based on the definition of social DIQ, we identify five central entities relevant to social DIQ: (1) the user of the social IS who acts as information producer (“producer” for short); (2) the user of the social IS who acts as information consumer (“consumer”); (3) the social software; (4) data; and (5) the phenomena of interest (“phenomena”). With respect to the definition above of an IS as a representation of a real-world system to the user (Wand and Wang 1996, pp. 87–88), we view entities 1 through 4 as parts of the social IS, while entity 5—phenomena—is external to the social IS. The five entities are related to each other as follows:

1. **Phenomena and producer** A producer is interested in certain phenomena, of which s/he has a subjective perception. Information thus mentally created by the producer constitutes her/his information potential.

2. **Producer and social software** A producer is motivated to share her/his information with others who share her/his interest. S/he engages the social IS to share her/his information with a wider audience, using the interface of the social software. Thus, data are produced in HCI.

3. **Social software and data** The social software retrieves and stores abstract data, processes and transforms them, and presents them through its user interface.

4. **Consumer and phenomena** A consumer is interested in certain phenomena, of which s/he may already have a subjective perception. However, s/he has a demand for further information regarding the phenomena (information gap).

5. **Consumer and social software** A consumer is motivated to engage in a social IS to fill her/his information gap by requesting, receiving, interpreting, and thus consuming data from the social software through HCI.

Figure 1 summarizes our ontology of social DIQ (relations are numbered according to the list above, and the names of entities are in bold).

**Discussion**

Confronted with the problem that traditional conceptualizations of quality of data and information are inappropriate for social IS, we develop a new paradigm of social DIQ based on the notion of “matching” information supply and demand. Our definition incorporates specifics of social IS as follows.
Sociability. Our definition of social DIQ highlights the importance of matching prosumers’ interests and respective exchanged information to each other, rather than supporting specific tasks. Specifically, the production of data suitable to the producer should match their consumption by a consumer with respective requirements, rather than focusing only on “fitness for use” by the consumer.

Openness. Since social IS are open and hence dynamic and somewhat unpredictable with respect to their user groups, the notion of matching in our definition of social DIQ has three implications. First, matching of production and consumption is something that can be assessed ex post rather than ex ante, but can probably be promoted by socio-technical means. Second, matching may change due to changes in the supply of information (i.e., different information) because the producers have changed, and changes in the demand for information because of a change in consumers. Third, it is less appropriate to assess matching on a global level than to do so partially, that is, whether subgroups of prosumers (and their respective supply and demand) match.

User role. The matching paradigm includes that prosumers of social IS can be producers and consumers at the same time. Hence, we use the term prosumer and mention both production and consumption to include that users of social IS can take different roles and that roles are interchangeable. Further, our definition speaks of a matching of supply and demand, which is meant to exclude any normative prioritization or hierarchy among the respective roles in a social IS with respect to DIQ. Specifically, we argue that while consumers want to find suitable content from producers, producers may also want to find an audience, depending on their motivation for engagement in the social IS. Our definition is thus different from “intrinsic” DQ, which excludes users’ perspectives, or “fitness for use” DQ, which prioritizes data consumption.

Content. Since data in social IS are mainly produced voluntarily and often by non-professionals, one must be aware that, unlike in traditional IS, social DIQ is subject to perception, motivation, abilities, knowledge, and interests of data producers and consumers—which our definition includes.

Technology. Social technologies (basic Web technologies as well as more complex social software) are the technical means for connecting prosumers and enabling open collaboration. They are different from technologies in traditional IS because they must accommodate not only data consumers but also data producers and their perceptions, motivations, abilities, knowledge, and interests. This would be ignored in a notion of DIQ that focuses primarily on the use of data/information. Therefore, similar to crowd IQ, our definition of social DIQ “recognizes the pivotal role of information contributors and motivates an effort to design systems sensitive to their points of view” (Lukyanenko et al. 2014a, p. 3). Further, social technologies are not only a means to convey data, but should also actively promote the matching of supply and demand in the social IS. This includes suitable and convenient mechanisms to gather, store, filter, transform, and present data, but also the possibility for prosumers to adapt social technologies to their preferences for a better human-computer interaction.
Location. The phenomena in which prosumers of social IS are interested, and how they perceive them, also includes questions of prosumer context and situation, which are important to social IS. Further, to match prosumers in a social IS appropriately also implies considering which contexts and situations match. Matching can be similarity, complementarity, or something else.

As already mentioned, defining, maintaining, and improving social DIQ (thus defined as matching) is a socio-technical process. This process can take the form of allocation (finding matching supply and demand) or negotiation (adopting supply and/or demand to match). Both forms can be supported by socio-technical mechanisms. Aggregation, filtering, search, and recommendation algorithms implemented in the social software can, for example, facilitate allocation. Meta-information (e.g., location) attached to stored data can even improve recommendations. Prosumers can be given the possibility to maintain digital profiles and allow tracking of activities in the social IS to enable improved personalized suggestions of content and other prosumers (Lu et al. 2015). Negotiation means prosumers voice their opinions about, for example, what is “good or bad” content, and which kinds of content they want to see in the social IS, but also that they listen to others’ opinions. Successful negotiation can result in either adjustment of content or opinions, or both. A straightforward example is the common rating/linking of content from which prosumers can tell what the community evaluates as being “good” content (e.g., Chen et al. 2011). If altruism is an important motivation among prosumers, simply rating content would possibly be sufficient for making prospective producers adapt their production. However, if production is driven by the desire to gain reputation, negotiation could be improved by an additional mechanism in which a prosumer’s reputation in terms of ratings is somehow signaled to others through her/his digital profile or avatar (e.g., Wasko and Faraj 2000, 2005). These and other mechanisms all rely on a combination of technical features as well as prosumers’ motivation and activities. Implementing and evaluating mechanisms is the task of the provider of the social IS. The matching paradigm of social DIQ can thus not only provide an analytical lens to understand better prosumer satisfaction and dissatisfaction, but also offer practical guidance through design, implementation, and operation of social software and social IS.

How can social DIQ be measured? Social DIQ is defined, maintained, and improved by the prosumers through social interaction. Hence, social DIQ is negotiated socially through social and technical mechanisms, and measuring social DIQ must therefore take into account the prosumers’ understanding and perception of social DIQ. It will depend on which mechanisms are in place. Wikipedia’s nomination, review, and voting process for articles, in which the community continuously measures social DIQ, provides a good example (see our section on “Existing DIQ Paradigms versus Social IS” and Kane and Ransbotham (2012)).

How are interaction and collaboration in a social information (meaning IT-supported) system different from direct (i.e., face to face) human–human interaction (HHI) in a social system? Actually, interaction and collaboration in both systems are similar with respect to our normative model of information potentials and gaps of actors constituted by their interest in and perception of phenomena, information supplied and demanded during interaction, and probably also with respect to the DIQ paradigm of matching. However, direct HHI and social IS differ with respect to their communication media characteristics. HHI has its advantages such as, for example, transmission of facial expressions, but limitations in terms of time (synchronous interaction), space (people have to come together, potentially traveling long distances), and reach (the number of people able to interact simultaneously is limited). The unique characteristics of social IS allow individuals to overcome some of these disadvantages. In enabling the collaborative and open generation, exchange, consumption, evaluation, and improvement of information in a self-organizing, socio-technical process, social IS can help increase the DIQ of information on phenomena that are of interest to their users. This, in turn, can affect the users’ motivation to use the system (DeLone and McLean 1992, 2003; Petter et al. 2008).

---

3 Ownership, development, and provision of a social IS can take different forms. Social IS may be owned, developed, and/or provided, for example, by a commercial provider to the public (e.g., www.facebook.com), by an organization and its IT department for internal use (e.g., enterprise social IS), or by the social IS’ prosumers themselves (e.g., www.joindiaspora.com).
Is it possible to link traditional definitions of DIQ and social DIQ? If so, how? Consider, for example, a researcher investigating data in social IS, a provider of a public social media site, or a company using enterprise social media, all wanting to assess the quality of data/information in a social IS. In fact, these actors do have a certain use in mind when asking for DIQ in social IS. Hence, in this case, one may apply traditional definitions and frameworks of DIQ to data/information from social IS, for instance, “fitness for use” or “intrinsic.” For example, one can assess empirically how accurately, timely, and comprehensively social media data on travel reviews reflect real-world travel and tourism (Tilly et al. 2015). Such an assessment of whether DIQ is sufficient for data from social IS to serve as observations in a research study explicitly takes care of the fact that these “digital trace data” (Howison et al. 2011) produced through a social IS are subject to perception, interest, motivation, social interaction, and other factors. Applying these data in “reality mining” (Eagle and Pentland 2006; Pentland 2009), “computational social science” (Lazer et al. 2009), or “big data analysis” (George et al. 2014) must be done with careful consideration of the social-IS context in which they have been produced (Howison et al. 2011; Kane et al. 2014) and should be accompanied by appropriate evaluations and tests (Ruths and Pfeffer 2014; Tufekci 2014).

Conclusion and Limitations

Starting from the observation that social IS are very different from traditional IS, we reviewed existing paradigms of DIQ with respect to their suitability to define DIQ in social IS. Since we identified several issues with existing DIQ paradigms when applied to social IS, we developed a new paradigm of social DIQ that incorporates the specific characteristics of social IS, namely, purpose of sociability, openness, users as prosumers, UGC, social technologies, and online availability. Basically, we define social DIQ as the matching of information supply and demand in a social IS. Further, our paradigm proposes to see DIQ in social IS as a socio-technical process in which DIQ is defined, maintained, and improved through social interaction and open collaboration. This process is enabled and supported by socio-technical mechanisms.

However, we could provide only a few examples to illustrate such mechanisms. Future research should conduct a comprehensive review of relevant literature and/or social IS in practice to get an overview of what kinds of mechanisms already exist and how these can be used to improve social DIQ. Further, mechanisms should be evaluated systematically.

Space limitations preclude us from discussing in more detail the possible settings in which social IS can occur and the implications those settings might have on the social IS, its data/information, and DIQ. For example, if social IS are used in organizations/enterprises (e.g., enterprise social media), prosumers’ engagement could be driven by different motivations, and prosumers might be more cautious about what to contribute (DiMicco et al. 2008; e.g., Ellison et al. 2015). Further research should investigate how different settings affect social DIQ.

IS research has proven traditional DIQ to be an antecedent of IS success (DeLone and McLean 1992, 2003; Petter et al. 2008). It seems natural to assume a similar link between social DIQ and social IS success, although the concept of social IS success has yet to be established. Wagner et al. recently drafted the concept “online community health” (our emphasis) with a focus on the prosumers’ evaluation of the online community. It is meant to “prioritize(s) effective inner workings ... over targeted output for externals” (Wagner et al. 2014, p. 3). They explicitly reject applying traditional notions and models of IS success to online communities because “success implies that the online community somehow contributes to the performance of the host organization” (Wagner et al. 2014, p. 4). While we think that speaking of either “health” or “success” is actually a matter of terminology, we agree that success in online communities and social IS in general should be defined from the perspective of the prosumer. However, the concept of social IS success should be elaborated, and the proposed link between social DIQ and social IS success should be investigated.

Finally, several frameworks of traditional DIQ and respective dimensions have been long established, operationalized, and tested (Knight and Burn 2005; Lee et al. 2002). It was beyond the scope of this study to operationalize and test empirically the concept of social DIQ in these comprehensive ways. Further research is needed to develop measures of social DIQ according to our paradigm and evaluate them in social IS.
References


*Thirty Sixth International Conference on Information Systems, Fort Worth 2015* 21


