BUILDING AN ENTERPRISE ONTOLOGY OF BUSINESS PROCESS CROWDSOURCING: A DESIGN SCIENCE APPROACH

Nguyen Hoang Thuan, School of Information Management, Victoria University of Wellington, Wellington, New Zealand, Thuan.Nguyen@vuw.ac.nz

Pedro Antunes, School of Information Management, Victoria University of Wellington, Wellington, New Zealand, Pedro.Antunes@vuw.ac.nz

David Johnstone, School of Information Management, Victoria University of Wellington, Wellington, New Zealand, David.Johnstone@vuw.ac.nz

Ha Xuan Son, Can Tho University of Technology, Can Tho, Vietnam, Hxson@ctuet.edu.vn

Abstract

Crowdsourcing is an emerging strategy for enterprises to harvest information, labour, expertise and innovation from a wide, anonymous crowd. To fully adopt this strategy, enterprise systems should not only extend existing frameworks with new services and users, but also efficiently integrate crowdsourced activities with, usually more stable, organisational processes. However, an ontological structure for this integration is still missing. Although a few lightweight ontologies have been proposed in the crowdsourcing domain, they seem to focus on clarifying concepts rather than providing a structure for modelling, designing, and developing an enterprise system. The current study fills the gap by building an enterprise ontology of business process crowdsourcing. Our results identified the main business processes, data entities, data attributes, and their hierarchy relationships, which were structured into a lightweight ontology. We then added decision-making relationships and business rules that turn a lightweight into a heavyweight ontology. The built ontology is evaluated by triangulation when comparing with two ontology versions generated by automated tools. The current study is significant as it is the first effort to build an enterprise ontology of crowdsourcing, which provides a blueprint for integrating business process crowdsourcing into enterprise systems.

Keywords: Business Process Crowdsourcing, Crowdsourcing, Design Science, Enterprise System, Enterprise Ontology.

1   INTRODUCTION

Networked information technology enables organisations to employ a global, online, and anonymous workforce (Kittur et al. 2013; Saxton et al. 2013). Crowdsourcing utilises this workforce to accomplish specific tasks, which are usually published online in the form of an open call (Howe 2006; Zhao and Zhu 2014). This strategy has been gaining popularity in large enterprises like Threadless, Boeing, Procter and Gamble, Colgate-Palmolive, Lego, Nike, BMW, Starbucks, and Netflix (Howe 2006; Rosen 2011). These enterprises have used crowdsourcing for various purposes, like gathering expertise and creative ideas, marketing campaigns, and many other business endeavours. Further evidence of this popularity is the variety of crowdsourcing projects being launched. Some projects involve micro tasks with very short timeframes, while others concern more complex and long-lived tasks like product development and software testing (Djelassi and Decoopman 2013; Zogaj et al. 2014). These projects also concern a wide range of information processing activities, including data gathering and classification, ideation, problem solving, and even collaboration (Kittur et al. 2013; Saxton et al. 2013; Zou et al. 2014).

Enterprise systems (ES), which support and coordinate almost all information flows and business processes within organisations (Volkoff et al. 2005), are expected to evolve to meet the wide applications of crowdsourcing. The integration of the crowdsourcing strategy with internal business processes becomes critical since ES always strive for effectiveness and efficiency, and therefore seek a more integrated, structured and repeatable utilisation of crowdsourcing resources (Lopez et al. 2010; Satzger et al. 2012). In other words, ES could benefit by embracing crowdsourcing in their value proposition and effectively integrating the crowd in their capabilities and services. Aligning with this view, the current paper focuses on how an enterprise system can support business processes based on crowdsourcing, which we call *Business Process Crowdsourcing (BPC)* (Vecchia and Cisternino 2010).

Although various aspects of crowdsourcing have already been studied, an ontological structure that can be used to integrate BPC into an enterprise system is still missing. This can be seen through (at least) three challenges. First, existing crowdsourcing studies mostly address some particular aspects of BPC (Geiger and Schader 2014; Man-Ching et al. 2011), whereas ES require a comprehensive knowledge on the whole integrated business process (Dietz 2006; Volkoff et al. 2005). Second, to successfully establish BPC, ES have to understand all its related aspects, including activities, components, information flows, and data entities. These elements have not been specified, or to some extend have been examined only through an abstract level in the existing literature, e.g. the twelve components of complex-task crowdsourcing (Kittur et al. 2013) and the conceptual model of BPC (Thuan et al. 2014). Thirdly, the ES view is also expected to contribute to the success of the crowdsourcing strategy. For instance, it is already known that increasing task meaningfulness increases the participation rate in certain crowdsourcing platforms (Chandler and Kapelner 2013). However, how to coordinate it into the enterprise system is unknown.

Addressing these challenges, the current study aims at building an enterprise ontology that can provide a sound foundation supporting the integration of BPC. The ontology, which explicitly defines concepts and relationships in the domain of interest (Corcho et al. 2003) and enhances reasoning knowledge (Valaski et al. 2012), can provide the ontological structure necessary to understand the main constituents of BPC. This structure is necessary to model, design, and develop an enterprise system in aligning with BPC processes, a similar outcome of the work by Osterwalder (2004). From an academic point of view, the ontology contributes to solidify the domain knowledge of BPC, together with recent studies that start exploring and structuring concepts in the domain (Hetmank 2014; Luz et al. 2014). More precisely, our work extends these studies by further analysing diverse relationships, especially the decision-making relationships that can support the establishment of BPC. This enables casual-based explanations and predictions in the domain (Fonseca and Martin 2007).

To construct the ontology, we adopted a Design Science paradigm (Hevner and Chatterjee 2010; Hevner et al. 2004). In particularly we followed a two-step process. The first step extracted a set of ontological elements from the knowledge base, using a prior literature review. The second step synthesises the extracted elements into an enterprise ontology of BPC. This paper is specifically centred on the ontology building, since the literature review has been published elsewhere (Thuan et al. 2014). The built ontology is then evaluated by comparing with two ontological versions automatically generated by two software tools (Cimiano and Völker 2005; Fortuna et al. 2007).

2 BACKGROUND

2.1 The Value of Ontologies

Ontologies have been used to conceptualise and structure knowledge in different domains, including Information Systems (IS) with the popular usage of the Bunge-Wand-Weber ontology (Wand and Weber 1993). Guarino et al. (2009) characterised ontologies by three features: conceptualisation, explicit specification, and shared understanding. Conceptualisation refers to an abstract and simplified representation of the domain of interest, mainly through objects, concepts, and relationships. Explicit specification means that these objects, concepts, and relationships are explicitly defined to prevent any ambiguous meaning (Corcho et al. 2003). These two features are highly consistent to Gruber (1993) who defined ontologies as "explicit specifications of conceptualisations" (p. 199). The third feature suggested by Guarino et al. (2009) is shared understanding, which helps the ontology to be widely understood and applied. As all of the three mentioned features are important in enterprise ontologies, our perception of ontology is aligned to the characterisation proposed by Guarino et al. (2009).

Given this perception, we can further distinguish between different types of ontologies. The related literature has proposed a few schemas and dimensions to classify ontologies (Fonseca and Martin 2007; Guan et al. 2013). Among them, two widely adopted dimensions are level of generality (Guan et al. 2013; Sharman et al. 2004) and structural complexity (Corcho et al. 2003; Valaski et al. 2012). Considering the level of generality, Sharman et al. (2004) classify ontologies into top-level, domain, and application. In top-level ontologies, concepts and relationships can be used for multiple domains, whereas in domain and application ontologies, the ontological elements focus more on targeted domains and restricted applications. Another popular dimension to classify ontologies concerns the structural differences of knowledge they represent (Corcho et al. 2003; Valaski et al. 2012). More precisely, an ontology that only captures concepts and relationships is a lightweight ontology, while adding axioms turns lightweight into heavyweight ontologies. When positioning our to-be-built ontology in these classification dimensions, our work is aligned to the heavyweight/domain ontologies since we strictly focus on BPC and examine not only concepts but also decision-making relationships and business rules in this domain.

The literature has suggested several values that an ontology may bring to the domain of interest, like formulising the domain, easing communication among different parties, and allowing computer interpretability (Guarino et al. 2009; Valaski et al. 2012). While agreeing with these, we suggest also considering the contributions of ontologies for enterprise modelling. From this perspective, ontologies or enterprise ontologies provide structured means for modelling and managing the domain knowledge. Corcho et al. (2003) refer to an ontology as not only the collection of terms, but the knowledge that can be inferred from it. Recently, Wong et al. (2012) recommend that "every knowledge base has to be committed to a conceptualisation […]. This conceptualisation is what we refer to as ontologies" (p. 2). Furthermore, Ahmad et al. (2011) highlight the roles of ontologies in managing ES knowledge, including clarifying knowledge structure, reducing conceptual ambiguity, sharing knowledge, facilitating communication, and supporting operating of ES. As a result, an enterprise ontology can support the modelling, designing, deploying, integrating, and maintaining ES. Given the important role of enterprise ontologies, we expect that building an enterprise ontology of BPC can help consolidating the domain knowledge and enabling the integration of crowdsourcing into ES.

2.2   Crowdsourcing

Although the use of the crowd to solve difficult problems can be traced back to the 18th century, the term 'crowdsourcing' was only recently coined by Howe (2006). The recent emergence of crowdsourcing can be explained by the dominance of Web 2.0, which provides infrastructures for organisations to approach a wider workforce (Saxton et al. 2013; Zhao and Zhu 2014). Although it is only one decade since the term was first introduced, crowdsourcing has raised interest in organisations operating in multiple fields like marketing, education, software and medicine (Rosen 2011; Zhao and Zhu 2014). This wide adoption, on the one hand, enables crowdsourcing to become a rapidly-growing research field. In our previous study (Thuan et al. 2014), we found about a double increase on the number of articles published per year from 2008 to 2013. A similar trend has also been reported by other studies, such as the work by Tarrell et al. (2013).

On the other hand, research in multiple fields brings a variety of epistemologies, views, concerns and research methods (Geiger and Schader 2014; Man-Ching et al. 2011). When analysing the crowdsourcing literature (Thuan et al. 2014), we could not find any dominant theories or models. Instead, we could find a wide range of research styles, including case studies, proofs-of-concept, design studies, usability studies, tool developments, experiments, and other engineering contributions. Since these studies are highly focused on some isolated concerns and use different research theories and methods, their findings tend to be ad-hoc and sometimes conflicted. For instance, conflicting results on whether increasing payments can increase crowdsourcing output quality have been reported by different researchers (Liu et al. 2014; Mason and Watts 2009). All in all, these characteristics of crowdsourcing research indicate that it is an emerging area with scattered knowledge, rather than strong ontological foundations.

From an organisational perspective, crowdsourcing needs to evolve rapidly from this conflicted situation towards a more stable repeatable process. This requires an integration of crowdsourcing into organisational business processes (Lopez et al. 2010; Satzger et al. 2012). Lopez et al. (2010) state that "organisations require integration of [the] crowdsourced tasks with the rest of the business process" (p. 539). This integration tightens and streamlines the internal and crowdsourced activities, and allows ES establishing business processes on top of crowdsourcing tasks. In spite of this importance, the integration of BPC into ES faces several challenges. These challenges originate from the aforementioned lack of ontological soundness, and at least two significant differences between crowdsourcing and regular ES processes. First, ES usually involve repeatable business processes, while crowdsourcing deals with volatile and unique tasks (Geiger and Schader 2014). Second, ES tend to be associated with clear roles, responsibilities and activities, while BPC often distributes tasks to anonymous people.

Since enterprise ontologies can address these challenges by strengthening and standardising the knowledge bases, a few crowdsourcing ontologies have already been suggested (Hetmank 2014; Luz et al. 2014). Analysing the state of the art, Luz et al. (2014) proposed a taxonomic or hierarchical ontology of task-oriented crowdsourcing systems identifying nine structural dimensions, including nature of collaboration, architecture, worker selection, quality control, worker motivation, task design and configuration, task management, execution management, and result aggregation. More closer to the current study, Hetmank (2014) proposed a lightweight ontology for enterprise crowdsourcing providing "controlled vocabulary to communicate specific details about their crowdsourcing activity" (p. 2). The ontology defined 24 classes, which were grouped into seven components: user, project, crowdsourcing task, requirements, reward mechanism, evaluation mechanism, and contribution.

Although these ontologies help clarifying the field, they seem to focus more on defining concepts rather than building an enterprise ontology that can support the establishment of BPC. This enterprise ontology should be built, not only on well-defined concepts and hierarchy relationships but also on reasoning knowledge, including structural relationships and business rules. An example includes the impact of task/process design on the quality of the crowdsourcing outcomes (Eickhoff and De Vries 2013; Hoßfeld et al. 2013). These structural relationships and business rules play a crucial role on BPC

and the enterprise system supporting it, but have not been captured by the existing ontologies. Additionally, none of the existing ontologies suggest an integrated framework for organisational BPC. In short, an enterprise ontology of BPC still needs to be developed.

3 METHOD

To build the ontology, we followed a Design Science paradigm (Hevner and Chatterjee 2010; Hevner et al. 2004), which views the to-be-built ontology as a Design Science artefact. In this paradigm, artefacts must be founded on a rigorous knowledge base. Some popular Design Science methods require using extant theories as the knowledge base (e.g. Carlsson et al. 2011; Pries-Heje and Baskerville 2008). However, crowdsourcing is an emerging domain and extant theories seem not exist yet (Zhao and Zhu 2014). Thus, other sources of knowledge have to be considered for the ontology building. Hevner and Chatterjee (2010) suggest that besides theories, Design Science research can also be founded on experience and expertise, which is highly available in crowdsourcing literature. In a prior study (Thuan et al. 2014), we identified a large collection of "sources" reporting experience and expertise from a diversity of case studies, experiments, tool developments, etc. Following the Hevner and Chatterjee (2010)'s assertion, the current study uses that raw information to construct the enterprise ontology.

We have also reviewed the ontology engineering literature to identify and justify the activities of ontology development. We selected two activities which are commonly used in ontology engineering. They are *ontology capture* (Uschold and King 1995) and *knowledge organisation* (Küçük and Arslan 2014) (also called conceptualisation (López et al. 1999)). In the first activity, we derived ontological elements by a scoping literature review. This review systematically searched and selected 238 articles related to BPC from eight popular online bibliographic databases: ACM, EcoHost, IEEE, Emerald, Sage, Science Direct, Springer Link, and Wiley, which is described in another paper. Interested readers are pointed to Thuan et al. (2014). These articles were used as raw data in the current study. We then analysed the reviewed articles to identify concepts, hierarchical relationships, decision-making relationships, and business rules. This analysis was conducted using both inductive and deductive approaches. In the deductive approach, we derived a conceptual model reflecting the predominant view of the crowdsourcing literature, which is presented in Figure 1. This approach, similar to Osterwalder (2004), ensures that the captured elements are tightly focused and address core concepts of the crowdsourcing domain. On the other hand, some ontological elements were also inductively captured. Reasons for using the inductive approach include the emerging nature of the crowdsourcing field, where we can expect new concepts and relationships that are not aligned with the predominant view. To support both approaches, we developed an ontology schema to guide the ontology capture process (details in Section 4.1).
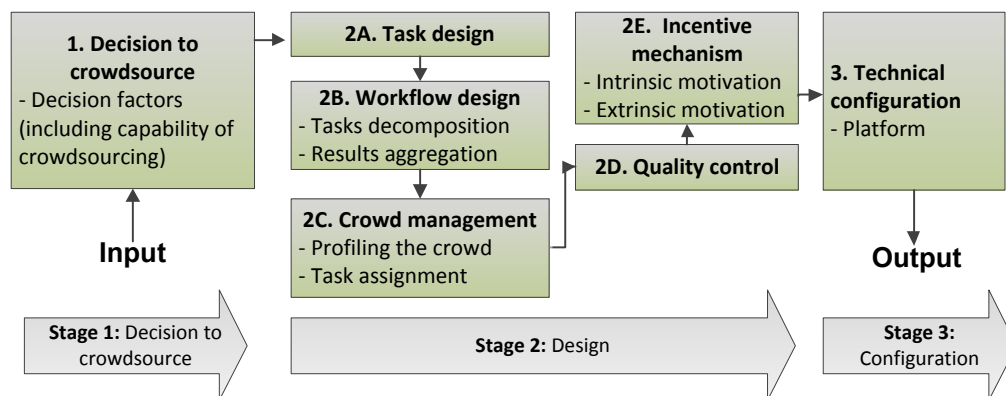

Figure 1.   A conceptual model of BPC (Thuan et al. 2014)

The second activity synthesised and organised the ontological elements into the enterprise ontology (Küçük and Arslan 2014). The synthesis was processed for each ontological element: distilling

concepts, hierarchy relationships, decision-making relationships, and business rules. We synthesised these ontological elements mainly based on the 'wisdom of the researchers' (Thuan et al. 2014), where we focused on elements supported by multiple studies. This 'wisdom of researchers' is particularly useful when the extracted elements are not consistent. For instance, different hierarchical relationships related to quality control mechanisms have been proposed in the reviewed articles. Some articles categorise quality control into design-time and run-time mechanisms (Allahbakhsh et al. 2013). A slightly different categorisation defines before-task, during-task, and after-task mechanisms (Alonso 2013). Other authors propose completely different categories, including supervised and unsupervised mechanisms (Baba and Kashima 2013). In these cases, by applying the 'wisdom of researchers' we chose the elements that were suggested by the majority of reviewed articles.

The synthesised elements were then organised into the enterprise ontology. Since the relationships revealed the fundamental structure of the BPC domain, the ontology ended up being organised around them. It is nevertheless important to note that the organisation process was highly iterative, where we extend, clean up, and update the ontology several times. The organisation process was also performed using inductive and deductive strategies. This can be exemplified with the process of obtaining hierarchical relationships. On the one hand, we relied on guidance from the reviewed articles, e.g. suggesting that quality control can be categorised into design-time and run-time mechanisms (Allahbakhsh et al. 2013), which suggests the adoption of a deductive strategy. On the other hand, no guidance was found for some groups of related concepts. For instance, even though several factors influencing the decision to crowdsource have suggested by the related literature, no schema structuring these concepts was actually found. In line with Nickerson et al. (2012), in these cases we inductively proposed a classification schema based on the common characteristics of the (sub) concepts.

4   ONTOLOGY BUILDING

4.1   Ontology Schema

This section presents a schema used to analyse the ontological elements for capturing ontological element. This schema is critical because it provides a structured approach (or meta-model) for data analysis and ontology capture (Levy and Ellis 2006; Okoli and Schabram 2010). Figure 2 graphically presents this schema consisting of two sections: *ontological representation* (left-hand side) and *knowledge representation* (right-hand side), operationalising the deductive and inductive approaches described in the previous section.

The ontological representation, which was mainly used for deduction, includes four elements: concepts, hierarchy relationships, decision-making relationships, and business rules. While concepts and hierarchy relationships are important to structure a domain knowledge (Corcho et al. 2003; López et al. 2004), decision-making relationships and business rules provide reasoning knowledge and thus are also critical when developing an enterprise system. When deducting, we analysed concepts and sub-concepts according to the left side of the model shown in Figure 2. For each extracted concept/sub-concept, we specified its name, synonym, and description. Since ontologies include both the concepts and the linked extensions between concepts (Corcho et al. 2003; López et al. 2004), the next considered element was the relationship. In this element, we analysed not only hierarchical relationships but also decision-making relationships. The former refers to taxonomic structures in the domain (López et al. 2004), where we adopted five hierarchy relationships commonly used in ontology engineering, including 'is a', 'include', 'categorise', 'instance of', and 'based on'. Regarding the decision-making relationships, we chose the following ones: 'positively influence', 'negatively influence', and 'associate', which are popularly suggested in the literature (e.g. Chandler and Kapelner 2013; Hoßfeld et al. 2013). The last considered element concerns business rules, which add constraints to the concepts and relationships.

When applying the ontological representation to the knowledge sources, we faced an issue that some emerging ontological elements did not align with our pre-defined codes. This is logical since the

adoption of diverse views and methods is typical of crowdsourcing research (as discussed in the background section). Addressing this issue requires an inductive approach that allows analysing and characterising knowledge from the bottom up. Thus, we adapted the knowledge representation approach proposed by Rockwell et al. (2010) (right-hand side of Figure 2). More precisely, we applied the following questions when analysing the knowledge sources: what are the main issues related to BPC? How these issues can be defined, i.e. characterising by what factors? What alternatives can be chosen to address the issues? And how to evaluate the proposed alternatives?
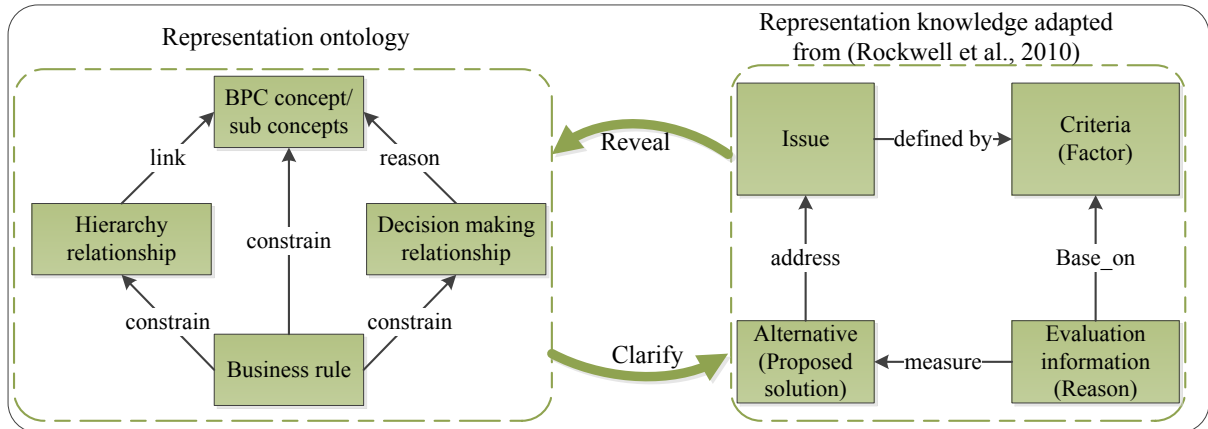

Figure 2.   Ontology schema (adapted from (Rockwell et al. 2010))

We note that the two parts of the ontology schema support each other. While the ontological representation section helps clarifying existing knowledge, the knowledge representation section allows further analysing the knowledge gaps, thus strengthening the approach. This schema was applied to every piece of information contributing to the ontology. As a result, the ontological elements: concepts, hierarchy relationships, decision-making relationships, and business rules were captured. The results are presented in the next sections.

4.2   An Enterprise Ontology of BPC

4.2.1   Concepts

We now report the results from the ontology building, starting with the most popular concepts and sub-concepts of BPC. As stated earlier, we focused on concepts and sub-concepts suggested by multiple sources. This is aligned and at more detailed-level in comparison to our previous work (Thuan et al. 2014). We also note that the number of sources (or articles published in the research outlets) supporting these concepts is largely different, ranging from two sources suggesting 'autonomy' as a property of 'task design' to 69 sources discussing the topic of 'quality control'. To provide an overview of the most important concepts within the BPC domain, Table 1 presents 39 (sub) concepts that were supported by at least 10 sources.

At a high level, Table 1 reveals several components of BPC that were presented in bold. Within these components, the results are that 'quality control', 'incentive mechanism', and 'crowdsourcing output' are the top-three most popular concepts. The popularity of these concepts has been noted by other researchers (Kittur et al. 2013; Zhao and Zhu 2014). Besides these well-accepted concepts, Table 1 also reveals several emerging components, including 'crowdsourcing task', 'characteristics of the crowd' and 'control and feedback'. At a more detailed level, Table 1 clarifies these components with their categories and sub-concepts. This clarification suggests that the conceptualisation in the current study has a more detailed level of abstraction compared to the conceptual model in Figure 1.

Since concepts presented in Table 1 are supported by at least ten sources, they are important in the domain, and should be seen as core elements of the ontology. However, we note that, besides these concepts, other identified (sub) concepts are still considered in the ontology construction. This is

because we believe that the importance of a concept should be seen from both the number of supporting sources and its relationships with other concepts, given the important roles of relationships in ontologies (Guarino et al. 2009; Sánchez and Moreno 2008).

Concept	No. of supporting sources
Quality control	**69**
Design-time	11
worker selection	16
Run-time	13
identifying malicious behaviour	19
gold standard	16
output agreement	12
Incentive mechanism	**46**
monetary reward	29
fun	11
Crowdsourcing output	**38**
output quality	36
Task design	**37**
Task description	10
Crowd management	**34**
Task assignment	20
Profiling the crowd	10
worker profile	10
worker reputation	10
Crowdsourcing task	**34**
simple task	13
complex task	12
Decision to crowdsource	**26**
Decision factor	19
Task characteristic	30
ease of task delineation	13
partitioned task	11
Availability of the crowd	19
Risk & Challenge	16
Availability of crowdsourcing platform	10
Characteristics of the crowd	**23**
Type of worker	12
Motivation of the crowd	10
Workflow design	**21**
Results aggregation	29
Task decomposition	10
Control and feedback	**17**
Technical configuration	**14**
Platform (intermediary)	13

Name conventions: components are in **Bold**; categories are in *Italic;* parent concepts are in Capital-first-letter, and sub concepts or attributes are in all-lower-letter

Table 1.   Main (sub) concepts supported by at least 10 sources

4.2.2   Hierarchy Relationships

Different from the previous section that identified individual concepts related to BPC, this section hierarchically structures these concepts to provide a holistic view of the BPC domain. Our analysis reveals a diversity of hierarchy relationships in the field of BPC. This diversity can be seen by counting each type of relationship: 'is a' (19 sources), 'include' (78 sources), 'categorise' (22 sources), 'instance of' (30 sources), and 'based on' (22 sources). These relationships and concepts identified in the previous section need to be organised in a manageable way. In the current study, such organisation is a trial-and-error process where several structures of the ontology were proposed, including tree structure and network structure. However, these structures seem not to be aligned with the current work. While the tree structure offers a good overview on the hierarchy relationships, it implies a division concern on each of its branches, and thus is more suitable for an individual view rather than our integrating approach. Although the network structure can support the integrated view, it makes the ontology representation becomes very complex with many relationships. Given that, the current study adopts a layer structure, which is suitable to our holistic view and reduces complexity by arranging concepts and relationships into layers. Furthermore, the layer view is totally suitable to the integration of crowdsourcing into enterprise systems, as enterprise integration includes different levels of concerns (Giachetti 2004; Hasselbring 2000).

The organisation results are presented in Figure 3. We note that in this figure, the 'is-a' relationship has been transferred to 'include' and 'categorise' relationships, and the 'instance of' relationships were not presented due to the limited space of the paper.
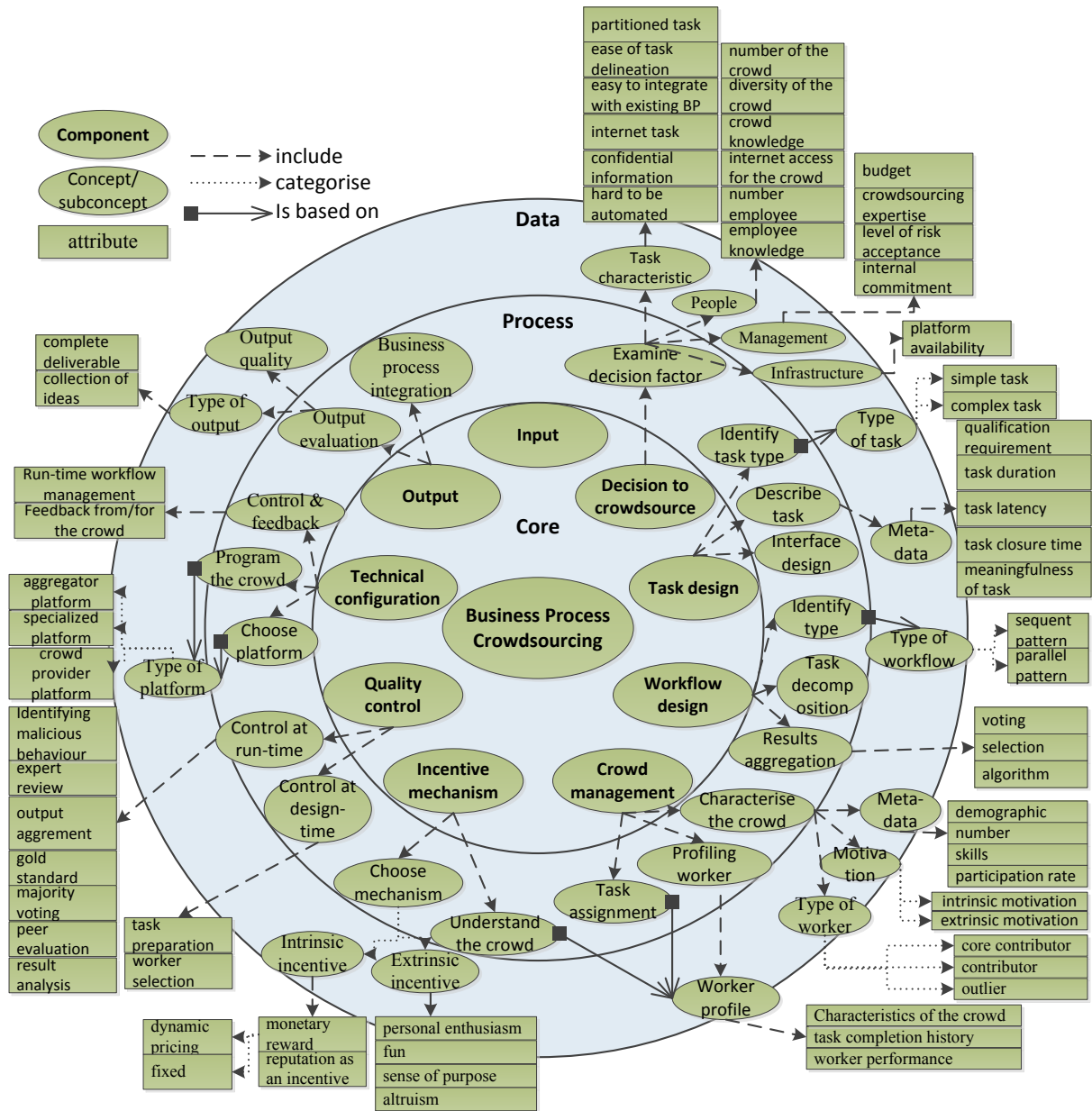

Figure 3.   A lightweight enterprise ontology of BPC

Figure 3 represents a lightweight ontology of BPC, which is structured into four layers. Overall, the ontology should be viewed from-inner-to-outer and clockwise direction. *The core layer* represents the main components of BPC that were captured in section 4.2.1. These components cover the whole BPC process as an input-process-output structure. To clarify these components, the other identified concepts in section 4.2.1 were further analysed and organised. From an ES perspective, processes and data (or information) are crucial elements in enterprise integration (Giachetti 2004), including the integration between crowdsourcing and ES. Given that, we classified and positioned the remaining concepts in the next layers as: process layer, data layer, and data attribute layer.

The process layer captures the activities performed in a particular component, and thus the main relationships among concepts in the process layer and components are 'include' relationships. For

instance, workflow design includes three activities: identifying type of task (Dai et al. 2013), task decomposition, and results aggregation (Kittur et al. 2013). *The data layer* represents the data entities that are used by the activities. In terms of relationships, this usage is represented mainly through two relationship types: 'include' and 'based on'. Some activities clarify (or include) data entities, e.g. 'describing task' includes clarification of 'meta-data'. Other activities are based on pre-defined data, e.g. 'identify type of task' is based on 'task type'. The final outermost layer then represents *the attributes* of each data entity. This representation is expressed through 'include' and 'categorise' relationships. The 'include' relationship shows the connection between an entity and its parts. For instance, the meta-data for task description consists of qualification requirement and task duration (Chilton et al. 2010), and may also consists of other attributes. Also presenting the whole-part relationship, the 'categorise' relationship further requires that all of the attributes make up an exhaustive decomposition of the whole concept. For instance, the reviewed articles suggest three categories of workers: core contributor, contributor, and outlier (Chanal and Caron-Fasan 2010; Stewart et al. 2010).

We note here three interesting points of the hierarchy relationships. First, these relationships enable explicitly structuring the related (sub) concepts in the domain. For instance, Figure 3 shows that the decision to crowdsource should be made based on examining the decision factors, including task characteristic, people, management, and infrastructure, which is consistent to (Thuan et al. 2013). Second, through structuring the hierarchy relationships, some interesting links that were not shown through individual studies were revealed in Figure 3. The link between incentive mechanism and crowd management can be seen as an example. In particular, organisations should understand the targeted workers when designing incentive mechanism (Chanal and Caron-Fasan 2010). This understanding can be achieved through worker profiles built by the 'crowd management' component (Khazankin et al. 2011). This suggests a close link between 'incentive mechanism' and 'crowd management', which is presented via the concept of 'understand the crowd' in Figure 3. Third, our results suggest that the number of articles supporting a relationship is far less than the number of articles discussing the related concepts within this relationship. This confirms the ad-hoc issue in the crowdsourcing literature in the sense that many reviewed sources only focus on individual concepts, rather than their links and interaction with others.

4.2.3   Decision Making Relationships and Business Rules

This section adds decision-making relationships and business rules into the lightweight ontology to build a heavyweight ontology of BPC. In particular, we find 89 'positive influence', 17 'negative influence', and 10 'association' relationships. Following the "wisdom of the researchers", we chose the relationships that were either suggested by multiple sources or that link popular concepts identified in section 4.2.1. To organise the chosen relationships, we based on the lightweight ontology but removed the process layer as we found only a few decision-making relationships in this layer. To simplify the presentation, Table 2 summarises the association relationships, while the positive and negative influences are showed in Figure 4.

Concept 1	Relationship	Concept 2
Output quality	associate	Worker profile; task design; quality control; task complexity; monetary reward
Type of task	associate	Type of workers; incentive mechanism; benefits for organizations; task design; results aggregation
Incentive mechanism	associate	Type of worker
Task design	associate	Quality control

Table 2.   Association relationships

As summarised in Figure 4, our analysis reveals three important trends. First, although the reviewed information sources indicate diverse relationships, many of them focus on how to influence crowdsourcing 'output quality' (Archak 2010; Chandler and Kapelner 2013). This is because 'output

quality' is the main indication for the success of crowdsourcing projects, and crowdsourcing can become an effective organisational strategy only when it can achieve high quality output. As a result, we allocated output quality in the centre of Figure 4. Second, some conflicting relationships can be found in the domain. For instance, seven sources suggest that monetary rewards positively influence output quality, whereas three sources do not find significant results to support the influence. Thus, further studies are needed to confirm the relationship. Finally, we find that the number of sources supporting a particular decision-making relationship is usually low (mainly 1 to 3), which contrasts with the large number of sources supporting hierarchy relationships. This is logical as our sources consist of academic articles in the IS field, where "IS have not been interested in publishing replications of prior studies" (Dennis and Valacich 2014 p. 1).
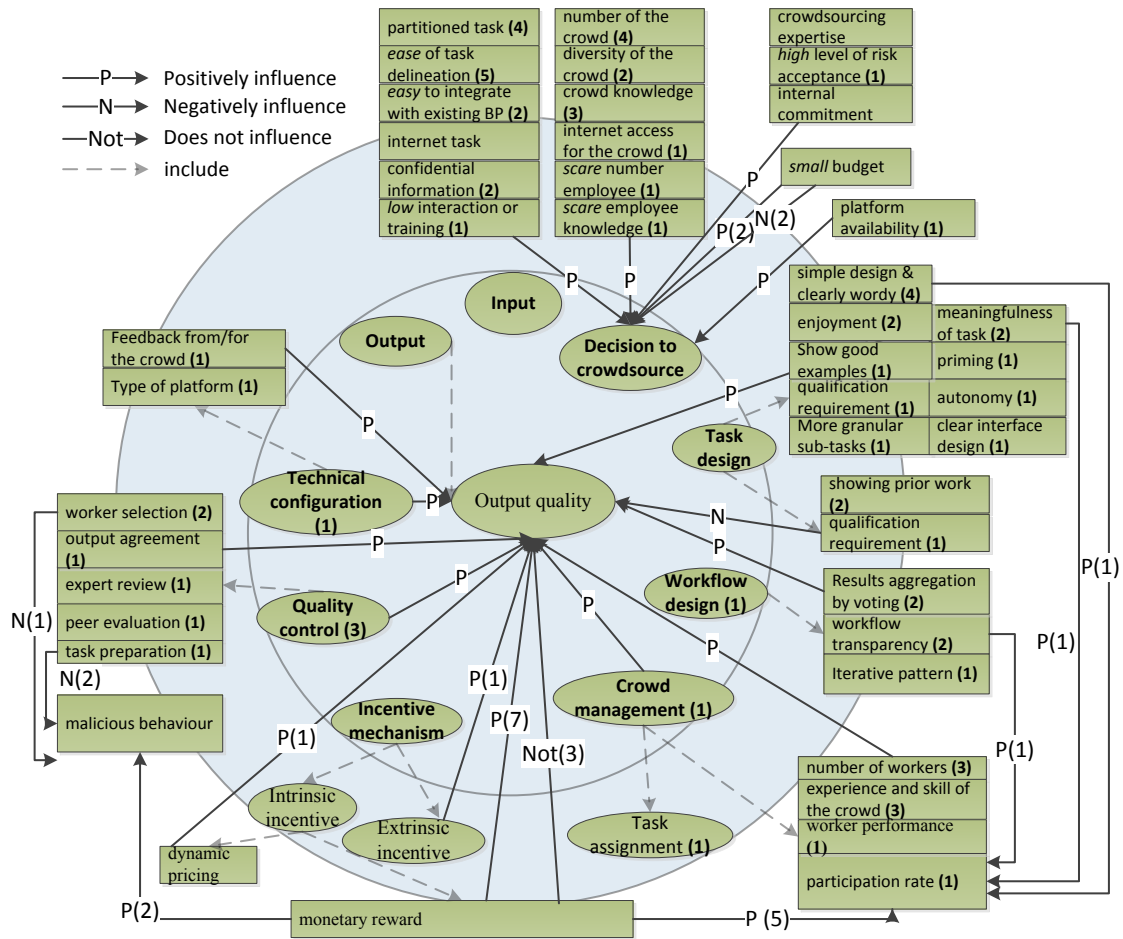

Figure 4.   Positive and negative influence relationships

Based on the structures provided by the hierarchy and decision making relationships (Figure 3 and 4), the knowledge base has also provided some business rules related to the establishment of BPC. Due to space constraints, Table 3 only presents some examples of the business rules related to 'output quality'. These rules, together with the decision-making relationships, provide guidance for organisations when integrating business process crowdsourcing into their ES.

Business rules
For taxonomy creation, *output quality* of crowdsourcing is equivalent to 80-90% of expert output
The more redundancy in performing a task (in iterative workflow), the better *outcome* of task results
Without quality control, more crowd workers are needed to achieve the same level of *output quality*

Table 3.   Some examples of business rules

5   EVALUATION

Design Science research involves two main activities: building and evaluating IS artefacts (Hevner and Chatterjee 2010; Hevner et al. 2004). After building the enterprise ontology of BPC, we now discuss its evaluation. Venable et al. (2012) suggest that Design Science evaluation helps: establishing artefacts' utility for achieving its stated purpose, comparing the artefact to other designed artefacts, and considering the side-effect and weaknesses of the artefact for future improvements. Although the first approach evaluates whether the artefact works and achieves its goals, and should probably be the most important criterion, this needs considerable efforts on practical, long-run applications, which according to Gregor and Hevner (2013) may not be feasible due to the constrained resources of most research projects. In the current study, we chose the second type of approach where we compared different versions of the enterprise ontology.

More precisely, we developed two automated versions of the ontology using the same sources of information. The two versions of ontologies were built using two software tools that generate ontologies from text: OntoGen (Fortuna et al. 2007) and Text2Onto (Cimiano and Völker 2005). While our ontology was built from a detailed review of a set of scientific articles (Thuan et al. 2014), the automated ontologies were built from the same sources but only using the articles' abstracts as input, similar to the approach in (Vogrinčič and Bosnić 2011). By comparing the automated ontologies with our own ontology (Section 4.2), we found high consistency on main ontological elements. For instance, the outcomes generated by OntoGen are presented in Figure 5. Comparing Figures 3 and 5, we note a strong match between the core components (i.e. tasks, quality control, incentive mechanism, technical configuration, and the crowd). Several detailed concepts are also similar, e.g. intrinsic motivation and extrinsic motivation, though few differences can be found. The comparison validates our ontology building process through triangulation as suggested by Carlsson et al. (2011) that "to strengthen the validity of design [theories], test triangulation may be beneficial" (p. 117).
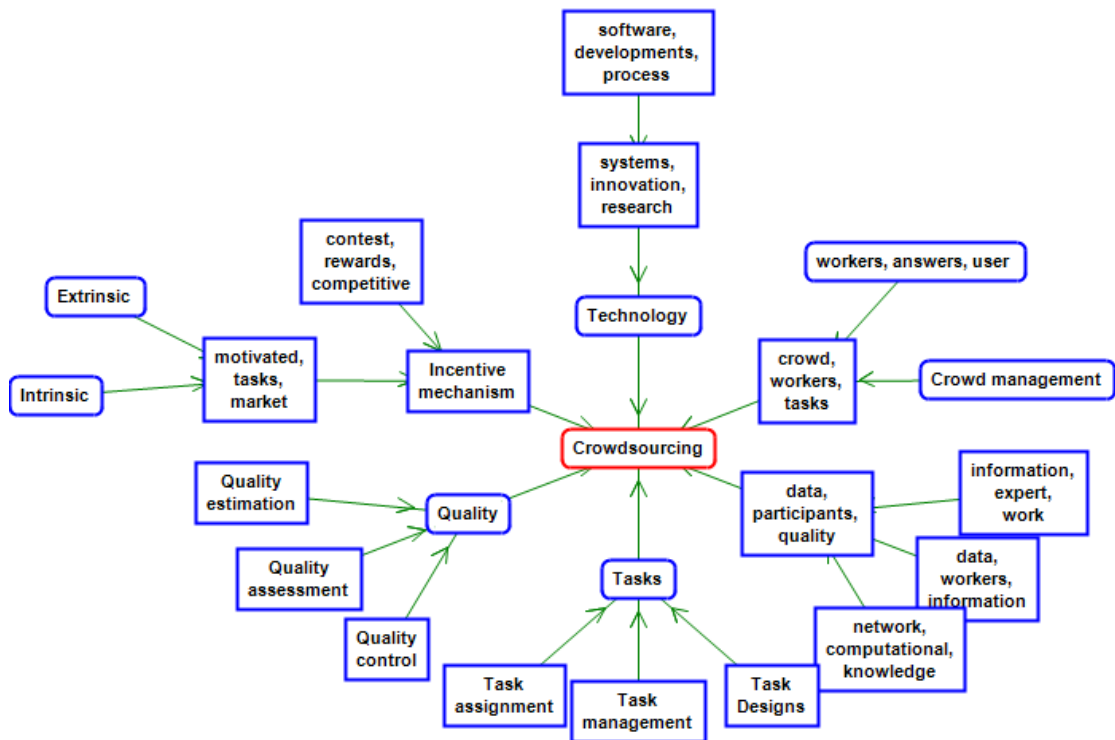

Figure 5.   Outcomes of OntoGen for comparing to the enterprise ontology

Closer examining the ontologies generated by the tools, we further note three important points. First, although the automatic approaches took less time and effort, and can capture main concepts and some

hierarchical relationships (Cimiano and Völker 2005; Fortuna et al. 2007), they are quite limited on the types of relationships, and cannot capture the non-hierarchical relationships. Second, as the automated ontologies define concepts based on the frequency of occurrence, rather than meaning, several extraneous composite concepts emerged, e.g. the combination of systems, innovation, and research in Figure 5. Third and finally, enterprise ontologies should distinguish different types of concepts, e.g. activities, data, and attributes, which currently cannot be supported by these tools. In short, we still believe that our approach is suitable to develop enterprise ontology of BPC.

6   CONCLUSION AND FUTURE WORK

There is a growing interest by enterprises in adopting crowdsourcing for their business processes (Djelassi and Decoopman 2013; Zogaj et al. 2014), which requires integrating crowdsourcing with enterprise systems. However, this integration has been difficult. Adopting the Design Science paradigm (Hevner and Chatterjee 2010), the current study structured the existing knowledge based on the crowdsourcing literature to identify the main concepts, hierarchical relationships, decision-making relationships, and business rules defining BPC. These elements were organised into a lightweight ontology first, and a heavyweight enterprise ontology later. The ontology was evaluated by triangularly comparing with the two automated ontologies generated by the software tools, which strengthens the validity of the design enterprise ontology (Carlsson et al. 2011).

This study is relevant for both organisations and academics. From an organisational perspective, the proposed ontology provides a synthesised BPC process, including the decision-making relationships and business rules, needed to integrate this process with ES. Thus, it can be used as an ontological structure for ES when analysing, planning and deploying crowdsourcing projects. This is important to allow organisations taking advantages by integrating crowdsourcing into their business processes. Although some previous studies have already discussed and defined concepts in the crowdsourcing domain (Luz et al. 2014; Zhao and Zhu 2014), the current study is the first proposing an enterprise ontology, which serves as a blueprint supporting organisations in their BPC integration. Furthermore, from the Design Science paradigm the BPC ontology serves as a knowledge base enabling artefact development (Ostrowski et al. 2014). This includes the development of decision tools that support organisations in their decision-making process related to BPC integration.

From a more academic perspective, an important contribution of this study is a broader view on the integrated crowdsourcing process, which follows the call for a more integrated and holistic view in crowdsourcing studies (Geiger and Schader 2014; Man-Ching et al. 2011). We reach this broader view by synthesising scattered findings in the research literature and providing a common understanding of the crowdsourcing domain. Another contribution is the possible complement of our work to the existing crowdsourcing ontologies (Hetmank 2014; Luz et al. 2014). More precisely, our work extends these ontologies by revealing additional concepts necessary for the integration between crowdsourcing and enterprise systems, and capturing important decision-making relationships thus building a heavyweight ontology. Our work also extends the conceptual model proposed in our previous work (Thuan et al. 2014). While the model represents the process of BPC (as seen via Figure 1), the ontology details the concepts, relationships, and business rules in the process.

Through a critical lens, there are some possible improvements that could be considered in the future. First, we understand the risk of synthesising ontological elements from very diverse sources of information. However, we believe that the ontology schema developed in the current study mitigated this risk by providing a consistency to data analysis and synthesis. Furthermore, we believe that gathering data from multiple sources also benefits from the 'wisdom of the crowd', utilising diverse points of view and opinions that help developing a more comprehensive perspective on a particular phenomenon (Surowiecki 2004). From an ontology-engineering point of view, our ontology can only be seen as an informal ontology, rather than a formal one that should be defined using an ontology language. Future work could formalise the ontology. We nevertheless note that developing an informal ontology before transferring it into a formal one is a common acceptable practice (Wong et al. 2012).

References

Ahmad, M. N., Zakaria, N. H., & Sedera, D. (2011). Ontology-based knowledge management for enterprise systems. *International Journal of Enterprise Information Systems, 7*(4), 64-90.

Allahbakhsh, M., Benatallah, B., Ignjatovic, A., Motahari-Nezhad, H. R., Bertino, E., & Dustdar, S. (2013). Quality control in crowdsourcing systems: Issues and directions. *Internet Computing, IEEE, 17*(2), 76-81.

Alonso, O. (2013). Implementing crowdsourcing-based relevance experimentation: an industrial perspective. *Information retrieval, 16*(2), 101-120.

Archak, N. (2010). Money, glory and cheap talk: analyzing strategic behavior of contestants in simultaneous crowdsourcing contests on TopCoder. com. *Proceedings of the 19th international conference on World wide web*, 21-30.

Baba, Y., & Kashima, H. (2013). Statistical quality estimation for general crowdsourcing tasks. *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, 554-562.

Carlsson, S. A., Henningsson, S., Hrastinski, S., & Keller, C. (2011). Socio-technical IS design science research: developing design theory for IS integration management. *Information Systems and e-Business Management, 9*(1), 109-131.

Chanal, V., & Caron-Fasan, M. L. (2010). The difficulties involved in developing business models open to innovation communities: the case of a crowdsourcing platform. *M@n@gement, 13*(4), 318-340.

Chandler, D., & Kapelner, A. (2013). Breaking monotony with meaning: Motivation in crowdsourcing markets. *Journal of Economic Behavior & Organization, 90*, 123-133.

Chilton, L. B., Horton, J. J., Miller, R. C., & Azenkot, S. (2010). Task search in a human computation market. *Proceedings of the ACM SIGKDD workshop on human computation*, 1-9.

Cimiano, P., & Völker, J. (2005). Text2Onto. In A. Montoyo, R. Muńoz & E. Métais (Eds.), *Natural language processing and information systems* (Vol. 3513, pp. 227-238): Springer.

Corcho, O., López, M. F., & Gómez-Pérez, A. (2003). Methodologies, tools and languages for building ontologies. Where is their meeting point? *Data & Knowledge Engineering, 46*(1), 41-64.

Dai, P., Lin, C. H., & Weld, D. S. (2013). POMDP-based control of workflows for crowdsourcing. *Artificial Intelligence, 202*, 52-85.

Dennis, A. R., & Valacich, J. S. (2014). A Replication Manifesto. *AIS Transactions on Replication Research, 1*(1), 1.

Dietz, J. L. (2006). The deep structure of business processes. *Communications of the ACM, 49*(5), 58-64.

Djelassi, S., & Decoopman, I. (2013). Customers' participation in product development through crowdsourcing: Issues and implications. *Industrial Marketing Management, 42*(5), 683-692.

Eickhoff, C., & De Vries, A. (2013). Increasing cheat robustness of crowdsourcing tasks. *Information Retrieval, 16*(2), 121-137.

Fonseca, F., & Martin, J. (2007). Learning the differences between ontologies and conceptual schemas through ontology-driven information systems. *Journal of the Association for Information Systems, 8*(2), Article 2.

Fortuna, B., Grobelnik, M., & Mladenic, D. (2007). OntoGen: Semi-automatic Ontology Editor. In M. Smith & G. Salvendy (Eds.), *Human Interface and the Management of Information. Interacting in Information Environments* (Vol. 4558, pp. 309-318): Springer.

Geiger, D., & Schader, M. (2014). Personalized task recommendation in crowdsourcing information systems—Current state of the art. *Decision Support Systems, 65*, 3–16.

Giachetti, R. E. (2004). A framework to review the information integration of the enterprise. *International Journal of Production Research, 42*(6), 1147-1166.

Gregor, S., & Hevner, A. R. (2013). Positioning and presenting design science research for maximum impact. *MIS Quarterly, 37*(2), 337-355.

Gruber, T. R. (1993). A translation approach to portable ontology specifications. *Knowledge acquisition, 5*(2), 199-220.

Guan, J., Levitan, A. S., & Kuhn Jr, J. R. (2013). How AIS can progress along with ontology research in IS. *International Journal of Accounting Information Systems, 14*(1), 21-38.

Guarino, N., Oberle, D., & Staab, S. (2009). What is an Ontology? In S. Staab & R. Studer (Eds.), *Handbook on ontologies* (pp. 1-17): Springer.

Hasselbring, W. (2000). Information system integration. *Communications of the ACM, 43*(6), 32-38.

Hetmank, L. (2014). *Developing an Ontology for Enterprise Crowdsourcing.* Paper presented at the Multikonferenz Wirtschaftsinformatik, Paderborn.

Hevner, A., & Chatterjee, S. (2010). *Design Research in Information Systems: Theory and Practice. Integrated Series in Information Systems* (Vol. 22): Springer Berlin Heidelberg.

Hevner, A., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS Quarterly, 28*(1), 75-105.

Hoßfeld, T., Keimel, C., Hirth, M., Gardlo, B., Habigt, J., Diepold, K., & Tran-Gia, P. (2013). Best Practices for QoE Crowdtesting: QoE Assessment with Crowdsourcing. *IEEE Transactions on Multimedia, 16*(2), 541-558.

Howe, J. (2006). The rise of crowdsourcing. *Wired Magazine 2006, 14,* 1-4.

Khazankin, R., Psaier, H., Schall, D., & Dustdar, S. (2011). QoS-Based Task Scheduling in Crowdsourcing Environments. In G. Kappel, Z. Maamar & H. Motahari-Nezhad (Eds.), *Service-Oriented Computing* (Vol. 7084, pp. 297-311): Springer.

Kittur, A., Nickerson, J., Bernstein, M., Gerber, E., Shaw, A., Zimmerman, J., . . . Horton, J. (2013). The Future of Crowd Work. *Proceedings of the 2013 Conference on Computer Supported Cooperative Work.*

Küçük, D., & Arslan, Y. (2014). Semi-automatic construction of a domain ontology for wind energy using Wikipedia articles. *Renewable Energy, 62*, 484-489.

Levy, Y., & Ellis, T. J. (2006). A systems approach to conduct an effective literature review in support of information systems research. *Informing Science: International Journal of an Emerging Transdiscipline, 9*, 181-212.

Liu, T. X., Yang, J., Adamic, L. A., & Chen, Y. (2014). Crowdsourcing with all-pay auctions: a field experiment on taskcn. *Management Science, 60*(8), 2020-2037.

Lopez, M., Vukovic, M., & Laredo, J. (2010, 5-10 July 2010). *PeopleCloud Service for Enterprise Crowdsourcing.* Paper presented at the 2010 IEEE International Conference on Services Computing (SCC), Miami, FL

López, M. F., Gómez-Pérez, A., & Corcho, O. (2004). *Ontological Engineering: with examples from the areas of Knowledge Management, e-Commerce and the Semantic Web*: Springer.

López, M. F., Gómez-Pérez, A., Sierra, J. P., & Sierra, A. P. (1999). Building a chemical ontology using methontology and the ontology design environment. *IEEE intelligent Systems, 14*(1), 37-46.

Luz, N., Silva, N., & Novais, P. (2014). A survey of task-oriented crowdsourcing. *Artificial Intelligence Review*, 1-27.

Man-Ching, Y., King, I., & Kwong-Sak, L. (2011). *A Survey of Crowdsourcing Systems.* Paper presented at the Privacy, security, risk and trust (passat), 2011 IEEE third international conference on social computing (socialcom), Boston, MA

Mason, W., & Watts, D. J. (2009). Financial incentives and the "performance of crowds". *Proceedings of the ACM SIGKDD Workshop on Human Computation*, 77-85.

Nickerson, R. C., Varshney, U., & Muntermann, J. (2012). A method for taxonomy development and its application in information systems. *European Journal of Information Systems, 22*(3), 336-359.

Okoli, C., & Schabram, K. (2010). A guide to conducting a systematic literature review of information systems research. *Sprouts: Working Papers on Information Systems, 10*(26).

Osterwalder, A., & Pigneur, Y. (2004). An ontology for e-business models. *Value creation from e-business models*, 65-97.

Ostrowski, L., Helfert, M., & Gama, N. (2014). Ontology engineering step in design science research methodology: a technique to gather and reuse knowledge. *Behaviour & Information Technology, 33*(5), 443-451.

Pries-Heje, J., & Baskerville, R. (2008). The design theory nexus. *MIS Quarterly, 32*(4), 731-755.

Rockwell, J. A., Grosse, I. R., Krishnamurty, S., & Wileden, J. C. (2010). A semantic information model for capturing and communicating design decisions. *Journal of Computing and Information Science in Engineering, 10*(3), 1-8.

Rosen, P. A. (2011). Crowdsourcing Lessons for Organizations. *Journal of Decision Systems, 20*(3), 309-324.

Sánchez, D., & Moreno, A. (2008). Learning non-taxonomic relationships from web documents for domain ontology construction. *Data & Knowledge Engineering, 64*(3), 600-623.

Satzger, B., Psaier, H., Schall, D., & Dustdar, S. (2012). Auction-based crowdsourcing supporting skill management. *Information Systems, 38*(4), 547–560.

Saxton, G. D., Oh, O., & Kishore, R. (2013). Rules of crowdsourcing: Models, issues, and systems of control. *Information Systems Management, 30*(1), 2-20.

Sharman, R., Kishore, R., & Ramesh, R. (2004). Computational Ontologies and Information Systems II: Formal Specification. *Communications of the Association for Information Systems, 14*(1), 9.

Stewart, O., Lubensky, D., & Huerta, J. M. (2010). Crowdsourcing participation inequality: a SCOUT model for the enterprise domain. *Proceedings of the ACM SIGKDD Workshop on Human Computation*, 30-33.

Surowiecki, J. (2004). *The Wisdom of Crowds: Why the Many Are Smarter Than the Few and How Collective Wisdom Shapes Business*. New York: Doubleday.

Tarrell, A., Tahmasbi, N., Kocsis, D., Tripathi, A., Pedersen, J., Xiong, J., . . . de Vreede, G.-J. (2013). Crowdsourcing: A Snapshot of Published Research. *Proceedings of the Nineteenth Americas Conference on Information Systems (AMCIS)*.

Thuan, N. H., Antunes, P., & Johnstone, D. (2013). Factors Influencing the Decision to Crowdsource. In P. Antunes, M. Gerosa, A. Sylvester, J. Vassileva & G. J. De Vreede (Eds.), *CRIWG 2013. LNCS* (Vol. 8224, pp. 110-125): Springer Berlin Heidelberg.

Thuan, N. H., Antunes, P., & Johnstone, D. (2014). Toward a Nexus Model Supporting the Establishment of Business Process Crowdsourcing. In T. K. Dang, R. Wagner, E. Neuhold, M. Takizawa, J. Küng & N. Thoai (Eds.), *FDSE 2014. LNCS* (Vol. 8860, pp. 136–150): Springer, Heidelberg.

Uschold, M., & King, M. (1995). Towards a methodology for building ontologies. *IJCAI95 Workshop on Basic Ontological Issues in Knowledge Sharing, Montreal*.

Valaski, J., Malucelli, A., & Reinehr, S. (2012). Ontologies application in organizational learning: A literature review. *Expert Systems with Applications, 39*(8), 7555-7561.

Vecchia, G., & Cisternino, A. (2010). Collaborative Workforce, Business Process Crowdsourcing as an Alternative of BPO. In F. Daniel & F. M. Facca (Eds.), *ICWE2010. LNCS* (Vol. 6385, pp. 425-430): Springer.

Venable, J., Pries-Heje, J., & Baskerville, R. (2012). A comprehensive framework for evaluation in design science research. *Design Science Research in Information Systems. Advances in Theory and Practice*, 423-438.

Vogrinčič, S., & Bosnić, Z. (2011). Ontology-based multi-label classification of economic articles. *Computer Science and Information Systems, 8*(1), 101-119.

Volkoff, O., Strong, D. M., & Elmes, M. B. (2005). Understanding enterprise systems-enabled integration. *European Journal of Information Systems, 14*(2), 110-120.

Wand, Y., & Weber, R. (1993). On the ontological expressiveness of information systems analysis and design grammars. *Information Systems Journal, 3*(4), 217-237.

Wong, W., Liu, W., & Bennamoun, M. (2012). Ontology learning from text: A look back and into the future. *ACM Computing Surveys (CSUR), 44*(4), Article 20.

Zhao, Y., & Zhu, Q. (2014). Evaluation on crowdsourcing research: Current status and future direction. *Information Systems Frontiers, 16*(3), 417-434.

Zogaj, S., Bretschneider, U., & Leimeister, J. M. (2014). Managing crowdsourced software testing: a case study based insight on the challenges of a crowdsourcing intermediary. *Journal of Business Economics, 84*(3), 375-405.

Zou, L., Ke, W., Zhang, J., & Wei, K. K. (2014). User Creativity in Crowdsourcing Community: From Extrinsic Motivation Perspective. *PACIS 2014* Paper 45.