Chapter 3
Business Process Crowdsourcing: Building Blocks

Specifying building blocks is like giving a process designer a box of Lego stones. He can play around with these stones and create completely new processes, limited only be his imagination and the pieces of stones supplied.
—Adapted from Osterwalder (2004)

This chapter analyses existing knowledge sources for synthesising the main building blocks of BPC. The synthesis is based on what we name the 'wisdom of the researchers' where a collection of researchers is wiser than single experts, similar to the wisdom of the crowd (Surowiecki, 2004). That is, the synthesis focuses on the building blocks that have been suggested by multiple researchers.

For this purpose, we adopt *a scoping literature review* as the main technique of this activity. A scoping review enables a comprehensive view on a particular topic (Paré et al., 2015), and thus is highly suitable for the emerging nature of BPC. More precisely, the 'scoping' review refers to a comprehensive sample strategy, which covers the breadth of knowledge sources existing in the domain. Further, scoping review is explicit in terms of how the search, selection, and data extraction are conducted. This increases the level of transparency and rigour of the research. We note that parts of the scoping literature review have been presented in our conference paper by Thuan et al. (2014).

3.1 Scoping Knowledge Sources

To begin the review, this research established a systematic process to ensure the rigour of the review results. We based the review process on the recommendations of how to conduct a good IS literature review, and especially, a good scoping review (Okoli, 2015; Paré et al., 2015). Following Okoli's (2015) recommendations, we adopted the five steps, including selecting sources, filtering sources, classifying sources, data extraction, and data synthesis. Figure 3.1 summarises the five steps of the scoping review, which are specified below.

N. H. Thuan, *Business Process Crowdsourcing*, Progress in IS,
https://doi.org/10.1007/978-3-319-91391-9_3

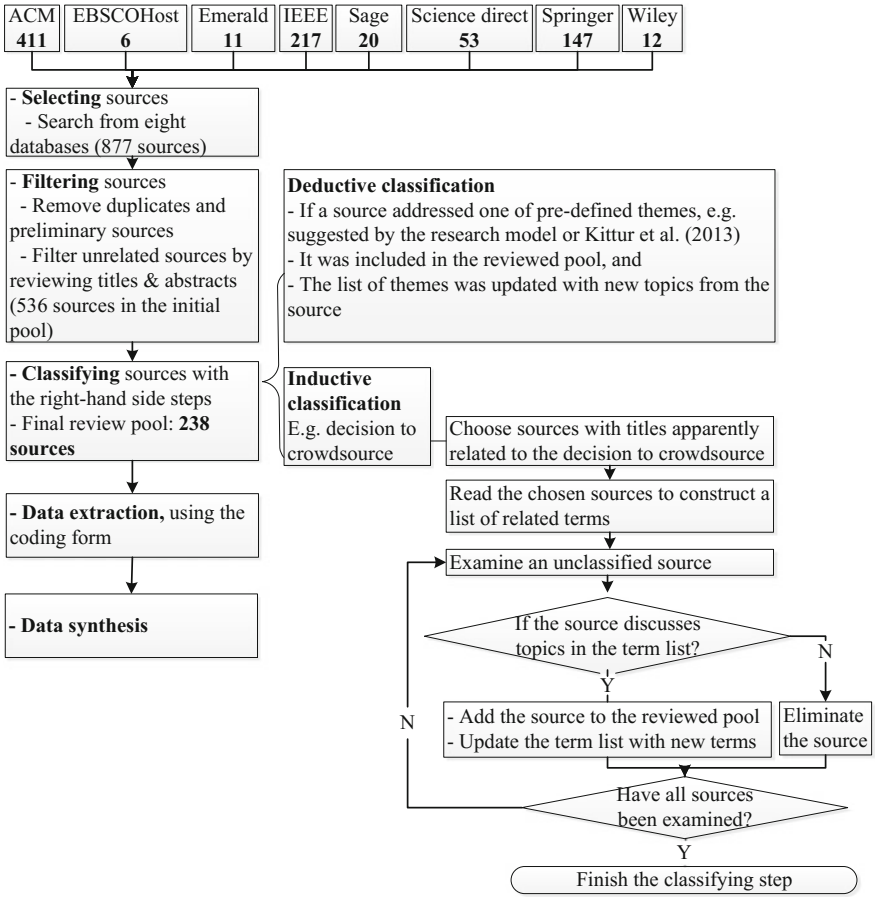

Fig. 3.1 Steps and summary of results of the scoping knowledge sources

3.1.1 Scoping Review Process

Selecting Sources

This initial step searched for the relevant sources about crowdsourcing. Following the scoping approach that highlights the comprehensiveness, the search was opened to multiple knowledge databases. More precisely, it relied on eight popular bibliographic databases: Association for Computing Machinery (ACM), Business Source Elite (EBSCOHost), Emerald Insight (Emerald), IEEE, Sage, Science Direct, Springer Link, and Wiley. In 2013, we searched for papers using the following keywords 'crowdsourcing', 'crowdsource', 'crowdsourced', 'crowdsourcer', and 'crowdsources' (the keyword 'crowdsourc*' was used to replace all the aforementioned keywords in certain allowable databases like Emerald, IEEE, and Sage). This choice of keywords was based on the perception that they are

representative and have been popularly used by other reviewers in the domain (Estellés-Arolas & González-Ladrón-de-Guevara, 2012; Hossain et al., 2015). As a result, we identified 877 knowledge sources, consisting of 667 conference papers and 210 journal articles. The search results are represented in the top parts of Fig. 3.1 and detailed in Table 3.1.

Filtering Sources

Although hundreds of sources were identified by the keyword search, many of them were clearly irrelevant to the subject of the book. Following a screening technique suggested by Okoli (2015), this step filtered out the irrelevant sources using the two following actions. First, we excluded posters, tutorials, extended abstracts, and work in progress, papers which are normally too preliminary to be considered as knowledge sources. In this process, we found 22 duplicates that were stored or indexed by more than one bibliographic databases. They were also removed from the pool. Additionally, we eliminated conference papers that had been extended into journal versions to prevent duplication. Second, we eliminated sources applying crowdsourcing to education, medical research, and games with a purpose because these sources have quite a different focus compared to our organisational view. We also eliminated crowdfunding sources, in which organisations raise capital for investments, and thus are distinct from our BPC definition. This elimination was based on the sources' titles, keywords, and abstracts. Through the filtering steps, the list of sources was sharpened into a pool of 536 sources.

Classifying Sources

After excluding irrelevant sources, this step included sources closely related to the research problem. To check whether a source focuses on BPC and thus keep it in the pool, we analysed the source topics. However, codifying topics was not a straightforward task as there was no complete classification frame specifically relevant to BPC. To address this challenge, we conducted a deductive and an inductive classification. In the deductive approach, we generated a list of pre-defined themes, based on the three stages of the research model (Sect. 2.3.3)

Table 3.1 Crowdsourcing sources on the eight bibliographic databases

Bibliographic databases	Number of conference papers	Number of journal articles	Total number of sources
ACM	408	3	411
EBSCOHost	0	6	6
Emerald	0	11	11
IEEE	170	47	217
Sage	0	20	20
Science Direct	0	53	53
Springer Link	89	58	147
Wiley	0	12	12
Total	667	210	877

and research foci suggested by Kittur et al. (2013), like workflow design, task assignment, task design, and quality control. If a source addressed one of the themes, it was kept in the pool and the list of themes was updated with new topics form the source.

The inductive approach was applied in cases where no classification schema could be found. For instance, there was no schema to classify sources related to the decision whether to crowdsource or not. In these cases, we followed the procedure described on the right-hand side of Fig. 3.1. The procedure is illustrated through the following example. First, we started by scanning the pool to choose sources whose titles apparently related to the decision to crowdsource, i.e. a source entitled 'To crowdsource or not to crowdsource?'. Second, we reviewed the chosen sources for identifying the relevant keywords, terms, and themes related to the crowdsourcing decision. This formed a list of terms, which was iteratively updated. Third, every unclassified source was checked to see whether it related to the term list. If so, the source was kept in the pool and the topics addressed by the source were used to update the term list. Otherwise, the source was eliminated. As a result of this procedure, we ended up adding related sources to the reviewed pool and building a term list for the codification.

Overall, we classified *238 sources* related to BPC. We noted that during the classification process there were many cases where sources would broadly refer to the list of terms but present indirect links to BPC. In these cases, a decision to include the sources rather than exclude them was made in order to keep the scoping review comprehensive. Making such decision was also a part of the 'wisdom of researchers', which suggests including diverse opinions that can latter on be collectively aggregated into stronger positions.

Data Extraction

This step extracted and identified building blocks, decision factors, and activities of BPC from the reviewed sources. For this purpose, we developed a coding form. To test the form, a PhD student was asked to code 20 random sources and the results were compared with the researcher's coding of the same sources. This led to small modifications of the coding form. The form codified four dimensions: general information, topic, findings, and application context. First, the first dimension was general information about the source, e.g. reference, year of publication, and whether it is a conference paper or a journal article, which is typically extracted by other reviews (Okoli & Schabram, 2010). Second, we codified the topics using the three stages of the research model and the term list, which was iteratively updated as described in the above section. Another considered dimension was the research findings, which are necessary to understand the BPC process. A part of this dimension included whether findings can be generalised to other situations or are limited to particular contexts (Mingers, 2003). Finally, the last considered dimension codified the practical outcomes of the sources, focussing on useful recommendations about BPC establishment. We also extracted to whom the recommendations were targeted and the crowdsourcing contexts where the outcomes could be applied to.

Data Synthesis
This step aggregated the data extracted by the coding forms. We reviewed the extracted data for building blocks, processes, decision factors, and activities that guided BPC establishment. This was a four-phase procedure. First, we analysed extracted topics and findings for these elements, which were compared and aggregated. Second, we merged the 'conceptually similar' elements. For instance, quality estimation (Baba & Kashima, 2013) and quality control (Allahbakhsh et al., 2013) were merged. Furthermore, many elements were linked to each other, e.g. expert evaluation is a technique to ensure quality control (Allahbakhsh et al., 2013). To rationalise the relationships among them, we mapped some sub-elements into more generic ones. Finally, we synthesised the sources' recommendations that were related to particular elements. As a result, elements extracted from individual sources were synthesised and transformed into thematic elements related to BPC establishment. They are discussed in the next sections.

3.2 Findings

This section reports results from the scoping review. As a result of the previous steps, we identified 238 sources related to BPC. The demographic information of these sources shows that 71% of them are conference papers and 29% are journal articles, which is consistent with the significant role of conference publications in IS and computer science (Freyne, Coyle, Smyth, & Cunningham, 2010). The number of publications per year are presented in Fig. 3.2, which shows a steady increase on the number of crowdsourcing studies published since 2008. This reflects the increasing maturity of the crowdsourcing field. This review also confirms the ad hoc nature of the crowdsourcing field as a large part of the reviewed sources (65%) provide findings that can only be generalised to a similar situation (the bottom parts of the columns in Fig. 3.2). Regarding to whom the implications of the reviewed sources are targeted to, the three most popular ones are managers, process designers, and programmers, which are essentially aligned to the three-stage model discussed in Sect. 2.3.3.

We now report the results of the scoping review in more detail. Considering the purpose of this chapter, we analysed the reviewed sources for building blocks of BPC.

3.2.1 Building Blocks of BPC

Our analysis revealed a diversity of building blocks, which are abstract elements of BPC. In particular, our review identified more than 20 building blocks and their sub-elements. However, the number of sources supporting each of them was highly different. For instance, 'quality control' was supported by more than 40 sources,

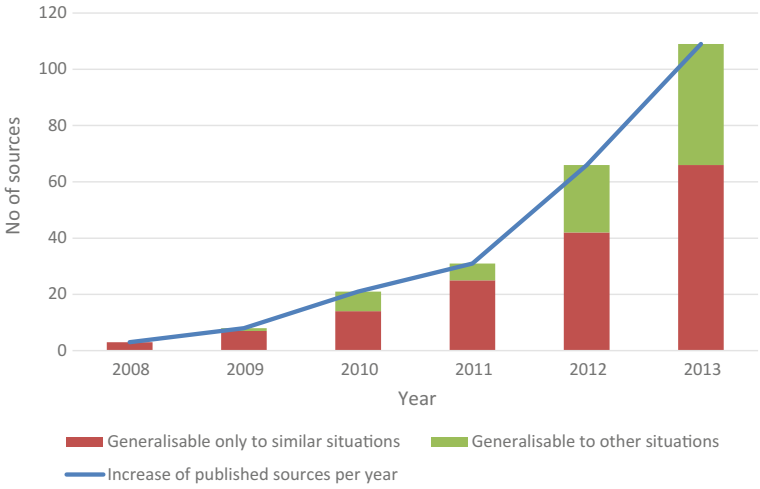

Fig. 3.2 Reviewed sources per year

while some sub-elements like 'guide crowdsourcing with artificial intelligent' were supported by only a few sources. Based on the 'wisdom of researchers' suggesting that aggregating results from groups of researchers outperform individual ones, we refined this list by concentrating only on building blocks supported by more than a certain number of reviewed sources. Choosing this number was quite sensitive. If the number was small, we might include too many building blocks, which unnecessarily increases the complexity of the analysis at this early stage. On the other hand, if the number was large, we might include only a few building blocks and thus might not represent the domain diversity. Given that, we selected a cut-off of 10 sources to balance between representation and complexity. Another reason for choosing this value was that there was a gap between the numbers of supporting sources before (e.g. 16 sources supporting 'circumstance to crowdsource and decision factors') and after the cut-off value (nine and eight sources supporting 'real-time response' and 'benefit & opportunity of crowdsourcing' respectively). As a result, Table 3.2 summarises the main BPC building blocks that are supported by at least ten sources.

From this table, the most popular building block is quality control, which has been suggested by 42 sources. Quality control refers to several techniques ensuring that the "[crowdsourcing] outcome fulfils the requirements of the requester [organisation]" (Allahbakhsh et al., 2013, p. 77). As crowdsourcing workers are voluntary, and thus it is hard for organisations to control their performance, quality control techniques are strongly relevant in a crowdsourcing strategy (Kittur et al., 2013; Zhao & Zhu, 2014). Moreover, incentive mechanisms and crowd management are also popular, being suggested by 37 and 32 sources respectively. To a lesser extent, Table 3.2 also indicates other relevant building blocks of BPC and

Table 3.2 Main building blocks of BPC

Building blocks of BPC	Number of supporting sources (n >= 10)
Quality control	42
Incentive mechanism	37
Crowd management	32
Task design	29
Result aggregation	26
Workflow design	25
Capability and characteristic of crowdsourcing	23
Task assignment	21
Output	17
Circumstance to crowdsource and decision factors	16
Platform	16
Technical configuration	16

their supporting sources, such as task design, result aggregation, workflow design, etc.

Overall, all building blocks identified in Table 3.2 emerge as key elements of BPC. These building blocks indicate repeatable activities within crowdsourcing processes, which backs the BPC concept that considers crowdsourcing as a repeated business process rather than an one-off activity. Further, as the identified building blocks are salient building blocks of BPC, we suggest using them to model and structure BPC, which is the focus of the next chapter. In short, our analysis has identified a set of common building blocks, serving the basic structure of BPC.

3.2.2 Factors Influencing the Decision to Crowdsource

We also identified the important role of the decision to crowdsource in BPC establishment. This important role is partly empirical, given the building block 'circumstance to crowdsource and decision factors' in Table 3.2, and partly theoretical, based on its starting position in the BPC process (discussed in Sect. 2.3.3).

Given the important role, we further analysed factors and sub-factors influencing the decision to crowdsource. The analysis followed the aforementioned review procedure, with two extensions. First, to keep the research up-to-date, we conducted forward searches based on the pool of sources. More precisely, we used the 'cited by' function in Google Scholar to identify the recent publications that cited the sources. The results from these searches increased the number of relevant sources on the decision to crowdsource to *50*. Second, the coding process was slightly modified for identifying directions of influence on the decision to crowdsource. We added quotes on the potential factors and marked '+' for factors that positively influence the decision and '−' for the ones that negatively influence the decision,

similar to the method used by Smith et al. (2008). The analysis results, part of which have been presented in Thuan et al. (2016), are now reported in more detail.

Table 3.3 highlights the set of factors that influence the decision to crowdsource. In particular, we found nine main factors, which were then decomposed into sixteen sub-factors or properties. We show how much the knowledge sources support them by presenting the number of supporting and non-supporting sources (the last two columns). We note that the number of supporting sources on a generic factor may be different with the sum of the corresponding references in its sub-factors. This is because in some cases, a source may concern several sub-factors and thus is coded multiple times, while in other cases, some sources study a generic factor as a whole without concerning its sub-factors.

The results from Table 3.3 indicate that 'task' is the most salient factor influencing the decision to crowdsource. 60% of the sources suggest this factor, sometimes under different names such as challenges (Seltzer & Mahmoudi, 2013), problems (Brabham, 2008a; Muhdi et al., 2011), and crowd work (Kittur et al., 2013). This salience is because the task factor is where the substantive decision starts from. It is the first-order question that has to be answered when crowdsourcing (Malone et al., 2010). This factor is also important as it determines several aspects of a crowdsourcing strategy, including the targeted crowd that has the ability to perform the task, the chosen platform for publishing the type of tasks, and the internal experts supporting crowdsourcing activities. Table 3.3 also presents seven sub-factors of tasks. Four of them positively influence the decision to crowdsource: whether tasks are easy to delineate (10 sources), to partition (8 sources), to integrate with existing business processes (7 sources), and to be done through the Internet (5 sources). Three other sub-factors negatively influence the decision: whether the task includes confidential information, needs high interaction, or can be automated.

Besides the task, two factors that are most addressed by the reviewed sources are the availability of the crowd and risk. The crowd, which comprises who will perform a task, was found in 38% of the reviewed sources. These results are not surprising because the crowd is one of the three key underpinnings behind the crowdsourcing concept, as shown in Chap. 2. Out of 50, 14 sources suggest the risk factor, which has a negative impact on the decision to crowdsource (i.e. more risk means less opportunity to crowdsource). To a lesser extent, other factors like infrastructure availability, availability of crowdsourcing experts to manage tasks, budget, internal human resources, and internal commitment also seem to influence the decision to crowdsource. Lastly, the level of organisations' technology adoption is the least addressed factor.

In summary, the review allowed us to systematically identify a set of factors that influence the decision to crowdsource. Using the identified factors, we can evaluate whether BPC is a suitable approach for a particular organisational context. Yet the relationships, similarity, and disparity among these factors still need to be examined and structured, which will be examine in the next chapter.

Table 3.3 Factors that influence the decision to crowdsource

Decision factors	Factor's properties/ sub-factors	Number of supporting sources	Number of non-supporting sources
Task		30	1
	Ease of delineation	10	
	Partitionable	8	
	Ease of integration with existing business processes	7	
	Done through the Internet	5	
	Confidential information (–)	3	1
	High interaction or requiring training (–)	2	
	Hard to be automated	1	
Availability of the crowd to perform the task		19	
	Number of members	9	
	Diversity	6	
	Knowledge	5	
	Internet access	3	
Risks (–)		14	1
	Low quality results (–)	8	
	Loss of intellectual property (–)	4	1
Infrastructure		12	
	Availability of crowdsourcing platform	10	
Expertise to manage the crowdsourcing activity		6	
Small budget		4	4
Lack of internal human resources to accomplish the task		3	
	Number of employees	3	
	Knowledge	2	
Lack of internal commitment (–)		3	
Slow in technology adoption (–)		1	

3.3 Summary and Discussion

This research conducted a scoping review of domain knowledge sources through a systematic process. The process retrieved 877 sources from eight bibliographic databases and finally considered 238 sources relevant to BPC. An overview on the reviewed sources confirmed the ad hoc nature of the BPC domain, which has supported the motivation of the book to study BPC. Analysing the sources in detail, the results revealed and synthesised the major building blocks of BPC. Of them, there were twelve most salient BPC building blocks supported by at least ten reviewed sources (Table 3.2). The analysis also identified factors influencing the decision to crowdsource. It revealed nine factors and sixteen sub-factors that should be considered in the crowdsourcing decision (Table 3.3). The identified building blocks, decision factors, and synthesised knowledge provide raw materials for the next research stages.

Overall, the scoping review offers accumulated knowledge of what the literature has reported in the domain. It has confirmed that there are repeatable processes of crowdsourcing strategies, through the identification of building blocks repeatedly suggested by the knowledge sources. The repeatable processes are the important antecedent of BPC and properly constitute business processes of crowdsourcing. Regarding the nature of the review, since the review process was arranged systematically and presented explicitly, it is possible for the review process and its results to be reproduced. This increases rigour of the review process and adds confidence to the review results. All in all, a combination of knowledge accumulation and systematic-ness constitutes the value of the scoping knowledge sources.