

Challenges for Workflows Technology to Support Crowd-sourcing Activities

Maria E. Orłowska

Polish-Japanese Institute of Information Technology

Warsaw, Poland

e-mail: *omaria@pjwstk.edu.pl*

Abstract - Workflow systems deliver remarkable results in many domains of business. Deployment of workflows on a large scale set the standard of rigorous execution of complex business processes controlled by multiple constraints and often completed by physically distributed agents. The emerging trends of paid and volunteer crowd work offer potential opportunity for improving productivity and expanding creativity in performed tasks. Crowd-sourced components of a traditional organizational process can not be supported uniformly by current workflows management technology. This paper frames the major challenges that stand in the way to effectively incorporate crowd-sourcing into existing business environment to construct a new form of cooperation. We lay out research challenges in most important areas: the process design and verification, design of typical patterns to support business segments completed by asynchronous crowd involvement through unidentified partners, enhancement of workflows management system functionality to accommodate crowd-sourcing activities.

Keywords - crowd-sourcing; workflow systems

I. INTRODUCTION

There has been an essential transformation in the way the enterprise information technology is perceived by the business in current times. The period when business process management was just concerned with a strict scheduling of predefined tasks is over. With the expansion of mobile technology, social media and crowd sourcing, the expectation from the enterprise has grown-up. The nature of social media and crowd-sourced activities challenge the industry to facilitate a business process that can interact with individuals and assist them with their requests and contributions they make. Furthermore, from the workforce perspective, in the connected world, many activities usually performed by employees of a company or social organization may benefit from some assistance from the outside. Since, the concept of cooperating parties became common and effective, the form of outsourcing as the contracting out of a business process or function to a third-party became a strategy in many domains of business [10][18][25].

The concept of outsourcing is also used to exemplify the practice of delegating fragments of the overall activity on ad hoc bases to the third party without any obligation of persistence of such relationship. Typically, outsourcing is not limited to a single country; it includes both foreign and domestic contracting, and, recently, often includes relocation of a business function to another country. In all

such cases, the main service provider is aware of the subcontracting party, its competences and reputation, expected quality of work and associated costs. In outsourcing arrangements, the relationship between involved parties is covered by formal agreements. Often, the economic considerations are the driving force for such business strategy, but not always.

The concept of crowd-sourcing is one significant step further. It is the process of obtaining required services, ideas, or content by soliciting contributions from a large group of unidentified people, and especially from an online community, rather than from conventional employees or suppliers. It combines the efforts of many self-identified volunteers or part-time personnel, where each contributor of their own initiative adds a portion to the final result. Let us note that often many contributors perform the same task not knowing about each other. Consequently, selection of acceptable results is additional obligation of the owner of such out contracted process. The most natural way to differentiate crowd-sourcing from outsourcing is the fact that the completion of an individual task comes from an undefined public rather than being accomplished by a specific, named and restricted by initial agreements individual.

The most frequently used example of a crowd-sourced work is Wikipedia [4][17][26]. Other examples, but different in nature can be associated with a design and content evaluation [8][11]- [14][17][28], images labeling [1][7], extensive testing of publically accessible e-service functionality or a contest for general design of a physical entity [5][6]. The educational domain forms another group of applications with already well recognized impact [3][31].

In this paper, we look at the technology that could assist with effective and well-formed crowd-sourcing contributions to a larger business process. Established organizations have developed expertise in deployment of workflow technology for their core activities. The assignment of task to partners/workers, the methods and correctness of the process design, data flow, and time constraints for conventional workflows is extensively studied for a number of years [19]-[23]. Work on international standards constitutes the base for most products in the class of workflows allowing their cooperation and effective integrations [33]. However, there is no systemic support to incorporating publicly performed activities as a part of workflows supported processes. As initial exploitation of such mix of partners became an appealing and interesting option for some types of work, there is a well justified need for efforts towards

technological solutions. It is clear, that the concept of business workflows partially executed by public input must be appropriately supported with a new workflow capability [33]. The call for workflow services to accommodate crowd-sourced activities with an acceptable quality is a new concept requiring further research. The purpose of this paper is to identify the scope of required extensions of the existing workflows technology to become a new generation of business process enhancement tools.

The paper is organized as follows. In Section II, we look at the impact on the process design generated by required extensions, followed by the discussion in Section III where we specify the scope of necessary modifications to the standard management systems to accommodate both groups of partners. Finally, in Section IV we summarize the discussion and outline future research work.

II. PROCESS DESIGN

Designing a business process and its partition into separate but meaningful tasks, that later are allocated to dedicated staff members, is a complex preparation phase prior to any technological solution deployment [20][21]. Clearly, not each task is suitable for crowd work. The crowd work can be seen as an element of a partnership between task provider, and, by contrast to typical business workflow, undefined workforce. Thus, when designing a process with crowd work participation, it is important to develop tools to support not only the work itself, and those performing the work, but also enhance the technology to enforce its processing according to the design decisions.

It is characteristic that in a standard business organization, completion of successfully terminated tasks is not evaluated for the final acceptance. The complete responsibility of the correctness and the quality of the execution is assigned to the experienced and trained worker discharging the activity. This observation shows the major difference in the subsequent steps of workflow structure when crowd workers are involved.

Each crowd-sourced activity can then be executed by many players, with different background, different skills and ability to complete specified assignment, and, with not necessarily converging intentions of the task provider. This is the main reason, why the subsequent verification step, following the collection of submissions, must be performed by the professional and dedicated staff. The quality confirmation and the final selection of submitted work, as a part of our business process, are necessary. Since we consider the general case of the crowd-sourcing, where activities can be performed by unidentified individuals, we must have a mechanism to evaluate the submitted outputs. This observation impacts directly on the design phase of the process. Below we identify and briefly discuss major challenges in the design process of such a workflow.

The question of a direct replacement of an existing activity within the process by a crowd sourced action is not

a trivial one. The traditional business workflows have a pre-defined order of carrying out all pre-specified activities. This restriction impacts on the overall duration of the process execution and it is tightly coupled with the temporal constraints spanning segments of the process [19][22]. Allowing for a flexible engagement of the crowd-workers may force some relaxation of time constraints, due to inability to schedule such activities. Moreover, it is not easy to predict the number of the involved parties (players), but this in return, restricts corresponding definitions and the duration estimation of the evaluation procedures. Additionally, cross-checking the available data reflecting on the reputation of involved players complicates this segment of the process [29].

It is important to note, that if the crowd activity is not a stand-alone, but incorporated in the rigorous well defined process with obligation to deal with a large scale operation, the process design is a new phenomenon to be addressed.

A subsequent part of the design process is the consideration dedicated to the data flow between activities [21]. Typically, each task has its input and output data that must be maintained in complete integrity. The impact of deployment of the crowd force, to deal with the data flow cross the process, must be also carefully investigated.

An additional question relates to the fundamental workflow component – the work list. The traditional business systems supported by a workflows technology use sophisticated methods of allocating tasks to the individual and dedicated staff members. Typically, it is a small group of well trained and competent people. This condition can not be applied to the cases with unidentified partners.

Moreover, workflow patterns [9] could be examined again, at this time, for the case of a crowd-sourced execution. It seems to be rational to attempt construction of new workflow patterns to deal with these specially designed components, such as: multiple execution of a task, evaluation of task completion, concurrently performing checks of reputation [30] and reliability of a partner, the credibility of work completed, and, generation of potential rewards structure. The effectiveness of loading all those listed components into existing methodology of process design needs to be carefully examined. Another appealing option points toward keeping them rather separate, but with well-formed method of the integration into current workflows systems. Further fundamental issues are associated with workflow engine – the heart of the workflows management process [33]. Those are open questions at the current stage of the research work.

III. WORKFLOWS MANAGEMENT SYSTEM

A standard workflow management system allows the user to define multiple processes, often for different types of business activities or, recently, for scientific computation procedures [32]. At each stage of the process, one individual or a well identified group is fully responsible for

a specific task. Once the task is completed, the workflow software ensures that the individuals responsible for the next activity are promptly notified and they receive all the data required to execute a subsequent stage of the process [26]. Often, workflows can be spanned by more complex dependencies between process components. For example, if a document is to be translated into several languages, the translation manager could select required languages, and then each selection would be activated as a scheduled work for the translation, ready to be taken by the specific contracted translators. Only, when all the translators have completed their respective tasks, the next segment of the process could be scheduled. The above example might be considered as a suitable activity for the crowd sourcing. Let us look at the differences in that case by contrast to the standard workflow execution. Here, the system will not schedule those activities to interpreters. Instead, it shall advertise the required translation job by some form of an electronic communication to attract participation. Naturally, we can't predict how many translators will undertake the work, and, how many for each language, as well as when precisely they will deliver their work. Therefore, we can't expect completion of this activity within a particular identified time. To some extent, it depends from the translators interest in the call for this activity, the conditions presented etc. Finally, the quality of the submitted work also needs attention. We shouldn't automatically include the submitted documents into production instead, verification checks are necessary. Even if it is perceived as a two steps activity, direct deployment of a standard workflow system to manage such crowd-sourcing will not be possible. An additional and important question is related to the overall economic viable of such an approach. Here, one can attempt some reasonable estimation from the hand driven experimentations.

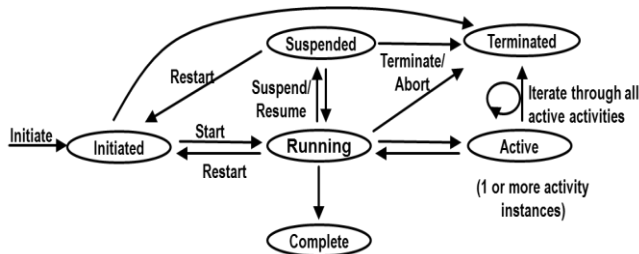


Figure 1. States transitions for a process instances

According to the Workflow Reference Model [33], the internal conditions defining the status of a process instance at a particular point in time, is maintained by the workflows system, following the transitions allowed by the defined finite state machine shown on Figure 1.

At the level of the individual process instance, the status of an activity, at a given point in time, executed by a given worker is represented by the other finite state machine as depicted in Figure 2 below.

This discussion already demonstrates that we must specify a number of novel concepts and necessary modifications to complement those already well established, to be able to address crowd sourcing incorporation into an existing process definition executed by the standard workflows technology. In summary, no direct allocation of activities indicates a necessary extension to the concept of work list handler, to release it from the organizational constraints, where competences are stored for all individuals capable to execute a given task. Further on, there should be no restrictions on the number of workers performing the job. Typical workflow systems do not allow undertaking of an activity instance for the concurrent execution by several workers. Furthermore, activations of multiple instances of the automata in Figure 1 and Figure 2 destroy the overall concept of rigid control over the process progress. This observation indicates definite loss of one of the most expected benefits of workflows deployment: a precise monitoring of the individual instances, availability of the progress development reporting, and setup alarms of any departure of the execution from the design parameters.

It is envisaged that even while not all jobs are being crowd executable, there are portions of almost any job that can be performed outside the organization.

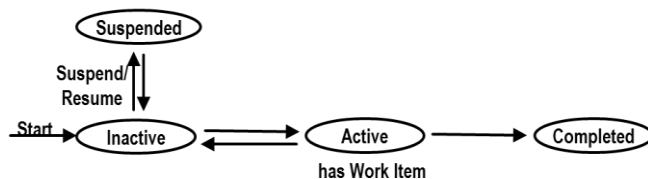


Figure 2. States transitions of activity instances

The question how we can move towards future of crowd work, that is more attractive for both requester and workers, must be addressed while building the next generation of cooperation support systems. Consideration of the worthiness of reconstruction of the current solutions versus building new from scratch to accommodate these new requirements is fundamental. This is another open question. Furthermore, it is vital to stress here that both styles of work, including the allocation of work segments relaxed from current constraints, should be offered and enforced by one pieces of technology.

IV. CONCLUSION AND FUTURE WORK

The crowd-sourcing can take place on many different levels and across a range of industries. This concept touches crosswise many social and business interactions. It can change the way we work, employ, do research and marketing. Leveraging collaborative practices and providing the tools may lead to significant business repercussion and probable restructuring.

However, crowd workflows are still quite undeveloped, and they are most successful with highly targeted single

tasks. Conversely, great potential hidden in the crowd abilities and wisdom should be better explored [28]. Unlocking this potential may improve productivity and creativity of the way the work is completed [16].

This paper lists only selected aspects of research to be undertaken to get to the next stage of a generic support for this type of crowd involvement. We pointed out that the current technology, build with overall intension of business process enforcement within a group of trusted partners, is not suitable for a simple extension.

The most promising research plan should include attempts to relax some of the constraints imposed on the standard concept of an activity. The minimum requirements for effective construction of a predefined process patterns should include crowd work progress reports, followed by a selection of the reputable partners and finally evaluation by the process owner. The new process segments should be easily customizable for the different types of activities. Further modification to the work list handler and the strict execution schema ordered by the steady finite state machines allows believing that perhaps "light" workflows solution is needed. Such a complete segment (including all the steps required to accommodate crowd work) could be perceived as a single activity of a standard workflow process.

All the considerations on the synthetic data have been promising and indicate sensibility of this approach.

To move ahead, through an iterative approach, we shall build essential components to accommodate all the crowd involvement requirements. An extensive experimentation is vital for the final usability assessment of the introduced ideas.

REFERENCES

- [1] L. Von Ahn and L. Dabbish, "Labeling images with computer game". Proceedings of the SIGC conference on Human factors in computing systems, ACM, 2004, pp. 319-326.
- [2] A. Anagnostopoulos, L. Becchetti, C. C. Castillo, A. Gionis and S. Leonardi, "Online team formation in social networks". Proceedings of the 21st international conference on World Wide Web, ACM, 2012, pp. 839-848.
- [3] M. Anderson, "Crowdsourcing Higher Education: A Design Proposal for Distributed Learning". MERLOT Journal of Online Learning and Teaching 7, 4, 2011, pp. 576-590.
- [4] S. L. Bryant, A. Forte and A. Bruckman, "Becoming Wikipedian: transformation of participation in a collaborative online encyclopedia". GROUP '05, ACM Press, 2005, pp. 1-10.
- [5] N. Archak and A. Sundararajan, "Optimal Design of Crowdsourcing Contests". ICIS 2009 Proceedings, 2009, pp. 1-16.
- [6] J. J. Chen, N. J. Menezes and A. D. Bradley, "Opportunities for Crowdsourcing Research on Amazon Mechanical Turk". Interfaces 5, 2011, pp. 1-3.
- [7] O. Dekel and O. Shamir, "Vox populi: Collecting high-quality labels from a crowd". Proc. 22nd Annual Conference on Learning Theory, 2009, pp. 377-386.
- [8] J. Chevalier and D. Mayzlin, "The effect of word of mouth on sales: Online book reviews". Journal of Marketing Research, 43(3). 2006, pp. 345-354.
- [9] W. M. Van Der Aalst, A. H. Ter Hofstede, B. Kiepuszewski and A. P. Barros, "Workflow patterns". Distributed and parallel databases 14, 1 2003, pp. 5-51.
- [10] C. Dellarocas, "Strategic Manipulation of Internet Opinion Forums: Implications for Consumers and Firms". Management Science, 52(10), 2006, pp. 1577-1593.
- [11] M. Franklin, D. Kossmann, T. Kraska, S. Ramesh and R. Xin, "CrowdDB: answering queries with crowdsourcing". Proc. SIGMOD '11, 2011, pp. 61-72.
- [12] K. Heimerl, B. Gawalt, K. Chen, T. S. Parikh and B. Hartmann, "Communitysourcing: Engaging Local Crowds to Perform Expert Work Via Physical Kiosks". Proc. CHI '12, 2012, pp.1539-1548.
- [13] A. Kittur, B. Smus, S. Khamkar and R. E. Kraut, "Crowdforge: Crowdsourcing complex work". Proceedings of the 24th annual ACM symposium on User interface software and technology, 2011, pp. 43-52.
- [14] M. I. Melnik and J. Alm, "Does a seller's ecommerce reputation matter? evidence from ebay auctions". Journal of Industrial Economics, 50(3). 2002, pp. 337-349.
- [15] A. Kittur, B. Suh and E. H. Chi, "Can you ever trust a wiki?: impacting perceived trustworthiness in Wikipedia". Proceedings of the 2008 ACM conference on Computer supported cooperative work, 2008, pp. 477-480.
- [16] A. Kittur, "Crowdsourcing, collaboration and creativity". XRDS 17, 2, 2010, pp. 22-26.
- [17] M. J. Metzger, "Making sense of credibility on the web: Models for evaluating online information and recommendations for future research". J. Am. Soc. Inf. Sci. Technol. (58), 2007, pp. 2078-2091.
- [18] P. Resnick, R. J. Zeckhauser, J. Swanson and K. Lockwood, "The value of reputation on ebay: A controlled experiment". Experimental Economics, 9(2), 2006, pp. 1-47.
- [19] S. W. Sadiq, O. Marjanovic and M. E. Orłowska, "Managing change and time in dynamic workflow processes". International Journal of Cooperative Information Systems, 2000, pp. 93-116.
- [20] S. W. Sadiq, M. E. Orłowska and W. Sadiq, "Specification and validation of process constraints for flexible workflows". Information Systems, 30, 2005, pp. 349-378.
- [21] S. W. Sadiq, M. E. Orłowska, W. Sadiq and C. Foulger, "Data flow and validation in workflow modeling". Proceedings of the 15th Australasian database conference-Volume 27, 2004, pp. 207-214.
- [22] S. W. Sadiq, W. Sadiq and M. E. Orłowska, "Pockets of flexibility in workflow specification". Conceptual Modeling-ER, 2001, pp. 513-526.
- [23] W. Sadiq and M. E. Orłowska, "Analyzing Process Models using Graph Reduction Techniques". Information Systems (25), 2000, pp. 117-134.
- [24] J. Schroer and G. Hertel, "Voluntary Engagement in an Open Web-Based Encyclopedia: Wikipedians and Why They Do It" Media Psychology 12, 1, 2009, pp. 96-120.

- [25] S. Senecal and J. Nantel, "The influence of online product recommendations on consumers' online choices". *Journal of Retailing*, 80(2), 2004, pp. 159-169.
- [26] M. Shen, G. H. Tzeng and D. R. Liu, "Multi-criteria task assignment in workflow management systems". *System Sciences, Proceedings of the 36th Annual Hawaii International Conference on Social Sciences*, 2003.
- [27] J. Surowiecki, "The wisdom of crowds: why the many are smarter than the few and how collective wisdom shapes business, economies, societies and nations". Boston, MA: Little, Brown, 2004.
- [28] Y. Tanaka and K. Yamamoto, "Enhancing credibility judgment of web search results". *CHI'11*. New York: ACM . 2011, pp. 1235-1244.
- [29] N. Thompson, "More companies pay heed to their "word of mouse" reputation". *New York Times*, 2003.
- [30] D. Wilson and R. Kreps, "Reputation and imperfect information". *Journal of Economic Theory* (27), 1982, pp. 253-279.
- [31] D. S. Weld, E. Adar, L. Chilton, R. Hoffmann and E. Horvitz, "Personalized Online Education—A Crowdsourcing Challenge". *Workshops at the Twenty-Sixth AAAI Conference on Artificial Intelligence*, 2012.
- [32] J. Yu and R. Buyya, "A taxonomy of scientific workflow systems for grid computing". *ACM Sigmod Record* 34, 3, 2005, pp. 44-49.
- [33] The Workflows Reference Model, Workflows Management Coalition, <http://www.wfmc.org/standards/docs/tc003v11.p> [retrieved: January, 2015]