



Towards Collaborative and Reproducible Scientific Experiments on Blockchain

Dimka Karastoyanova¹(✉) and Ludwig Stage²

¹ Information Systems Group, University of Groningen, Groningen, The Netherlands
d.karastoyanova@rug.nl

² SySS GmbH, Tübingen, Germany
bpm@stage.syss.de, bpm@ludwig-stage.de

Abstract. Business process management research opened numerous opportunities for synergies with blockchains in different domains. Blockchains have been identified as means of preventing illegal runtime adaptation of decentralized choreographies that involve untrusting parties. In the eScience domain however there is a need to support a different type of collaboration where adaptation is essential part of that collaboration. Scientists demand support for trial-and-error experiment modeling in collaboration with other scientists and at the same time, they require reproducible experiments and results. The first aspect has already been addressed using adaptable scientific choreographies. To enable trust among collaborating scientists in this position paper we identify potential approaches for combining adaptable scientific choreographies with blockchain platforms, discuss their advantages and point out future research questions.

Keywords: Flexible scientific choreographies · Reproducibility
Trust · Blockchain · Collaboration · Adaptive smart contracts

1 Introduction

Currently the blockchain technology has a significant impact on Business Process Management (BPM) research and is considered to be the main disruptor in this field. Challenges and opportunities of blockchains for BPM [3] have been identified and abundant research work has been reported towards identifying the best use of blockchains for enabling decentralized collaborative processes. Initial results have been demonstrated towards bridging the gap between the convenient process modeling provided by BPM systems and the “possibilities opened by blockchain platforms” [2], in particular related to the charm of keeping immutable trace of transactions without the need of a third party. The major opportunity to exploit is therefore the promise to enable trusted collaborations between “mutually untrusting parties” [2].

In this position paper we focus on only one of the aspects of BPM, namely *runtime adaptation of processes*. The discussion in [3] about how blockchain

relates to the BPM life cycle identifies the opportunity to utilize blockchains as one means of preventing illegal adaptation in order to ensure correct process execution, and ensuring the conformance with a model and rules defined in the contract among parties.

In this work we focus our research on the *synergies of the fields runtime adaptation of choreographies, blockchains and eScience*. Our motivation comes from the fact that in eScience, and in particular scientific workflows, there is a need for *adaptable or flexible choreographies* to support scientists in their trial-and-error manner of scientific exploration. We claim that scientists need enabling systems for a completely different type of collaboration when modeling their in-silico experiments. We identify the need for trusted, reproducible, collaborative adaptation of the in-silico experiments. Our position is that this need can be attended to by adaptable blockchain-based choreographies that allow collaborating scientists to track the provenance of the adaptation steps made in addition to the provenance of data, analyses and results. The other opportunity we identify is that adaptable blockchain-based choreographies can provide the means towards both RARE research (Robust Accountable Reproducible Explained) [1,4] and FAIR publishing (Findable Accessible Interoperable Reusable results).

With this position paper we want to identify the possible approaches to employ blockchain platforms for collaborative, adaptable and reproducible in-silico experiments. In Sect. 2 we will provide background information about the eScience requirements and the “Model-as-You-Go for Choreographies” approach that addresses only some of these requirements. In Sect. 3 we identify potential solutions, discuss their capabilities and identify open research questions to be overcome in future research on the synergies of BPM and blockchains in the field of eScience. We conclude the paper in Sect. 4.

2 Flexible Choreographies in eScience

Here we only discuss the two aspects of scientific experiments which are influencing our envisioned research the most: (1) the need to enable collaborative explorative research allowing scientists to interleave modeling and execution of experiment steps and (2) the aspect of reproducibility of experiments necessary in order to establish trust in the research method, data and obtained results.

Workflow technology offers a design and implementation approach for in-silico experiments and recent research results evidence considerable developments and broad acceptance of the concept scientific workflows. Scientists use scientific workflows to specify the control and data flow of experiments and orchestrate scientific software modules and services. The use of workflow technology in eScience fosters improvements in scientific collaboration through software services reuse. However, scientists have additional requirements on workflow modeling and enactment to ones of users in the business domain. Scientists often demand support for *trial-and-error experimentation* where (a) experiments are being modeled, started, paused, extended and resumed and (b) parts of the experiment are created and executed by different scientists on their own execution infrastructure. On the one hand scientists want to be able to start executing

incomplete, partially defined workflow models; add, remove and skip experiment steps to complete the model while it is being executed; reverse and repeat the execution of experiment steps with different parameters [7]. On the other hand all these operations are required to be performed in collaboration. Here, natural scientists are both the designers and users of a workflow model [8]. In our recent work we address these requirements with an approach called *Model-As-You-Go for Choreographies* [7]. The approach is based on runtime adaptation of processes, an interactive scientific Workflow Management System (sWfMS) [7] and a special middleware (called ChorMiddleware) coordinating the adaptation of choreographies [9] (see Fig. 1). The system supports the life cycle of scientific workflows. A modeling and monitoring environment is used to: (a) model collaborative experiments using choreographies with the ChorDesigner, (b) generate the visible interfaces of all participating workflows in the collaboration (Transformer component), (c) refine the internal workflow logic of the individual participant workflows and (d) serve as a monitoring tool. Scientists use the modeling tools to perform adaptation steps on the choreography that models the overall experiment or on the individual workflows, or both, and these changes are propagated to the running process instances on the workflow engines. In answer to the demand of scientists to monitor the state of the experiment that is currently being modeled and executed, we show the monitoring information directly in the modeling tools, as well as all adaptation steps. The workflow execution is performed by a set of sWfMS. The coordination is done by the ChorMiddleware implementing corresponding algorithms and coordination protocols.

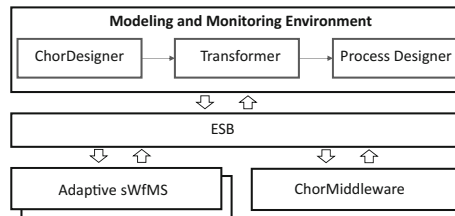


Fig. 1. Architecture of the Model-as-You-Go system (adapted from [9])

Another critical eScience requirement is *provenance*, which is the basis for *reproducible research* [4]. Computing environments used for scientific research are required to track the provenance of data, analyses and results with the purpose of ensuring reproducibility and repeatability of research, comparison of results and methods, preservation of the whole experiment and peer review [1]. In eScience this implies that all new tools and systems must enable provenance and need to expand recording, reporting, and reproduction of methods, data and results [4]. To enable *trust* among scientists in collaborative work on experiments, tracking provenance becomes even more important [1]. Consequently, the Model-as-You-Go approach has to ensure that the adaptive nature of all experiments is captured and reproducible on the level of both the individual adaptive

workflows and the choreography. Establishing provenance is not only necessary for the data used but also for the changes made by each of the scientists in the collaboration and the adaptation. Therefore there is a need to capture the changes made that have led to the final choreography model and that would help scientists understand what combination of software services and data have been used, in what sequence, thus document all their steps. This need could be addressed in a traditional way using an audit trail component of sWfMS, however the trend in scientific research towards more trusted scientific results calls for an approach more suitable for collaborative environments where no single party should have control over the adaptation ledger. As indicated above and in literature, blockchain could be the technology suitable to provide a solution to establish trust and support provenance and reproducibility of research [5].

3 Approaches for Reproducible, Collaborative and Adaptable Experiments

Considering the original focus of our work, namely the use of flexible choreographies in support of collaborative experiment modeling, and the available, standard-based system realization, we envision two approaches of employing blockchain.

The *first approach* (see Fig. 2, left) would be to reuse as much as we can from our existing realization system and combine it with a blockchain platform purely as a ledger. Supporters of blockchain for research suggest that it could improve “reproducibility and the peer review process by creating incorruptible data trails and securely recording publication decisions” [5]. Realizing this approach would mean that the audit trail (i.e. the history of workflow and choreography execution) is stored on a blockchain. The issue here is that typical audit trails are huge amounts of data, and in eScience by default the amounts of data we deal with is big anyhow. Storing data on the blockchain is very expensive, so it remains to be investigated how much of the historical data should be stored on the blockchain and how much on some other storage so that the reproducibility of the experiment can be guaranteed. Note that the history of all adaptation steps that is produced by our system has to be recorded too, which means that all the information we currently collect from workflow engines, the coordination middleware and the modeling tools that are the interface of scientists to the system has to appear on the audit trail.

In order to enable the FAIR publishing of the research results, which should also demonstrate the reproducibility of the experiment, the audit trail on the blockchain and the rest of the information necessary for that, but not more than the scientists would like to disclose, has to be read out from the system and presented appropriately. The visualization techniques necessary for that have to be delivered, too. The advantages of this approach are that we can reuse as much as possible of the existing sWfMS and because of this fact we would have a system capable of recording the trace of adaptations in place much faster. Such an approach may be appropriate enough for some scientific research efforts [5].

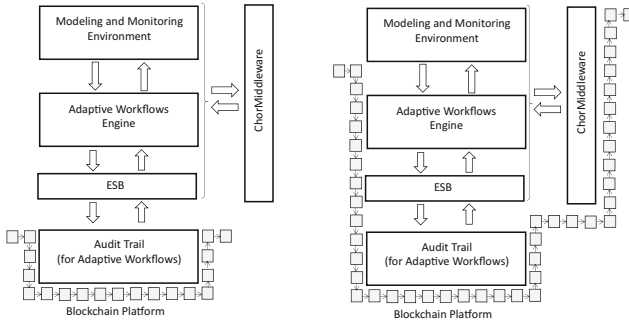


Fig. 2. Possible approaches for using blockchain with the Model-as-You-Go system

The disadvantage that we foresee from the current stand point is the fact that smart contracts, which enable blockchains to be more than just logs, would not be utilized and hence the capabilities they possess would remain unexplored.

The *second approach* is to investigate how a blockchain BPM system, such as [2], can be used instead of the workflow engine that is in place now and the adaptation mechanisms it implements, together with the middleware coordinating the adaptation of choreographies (see Fig. 2, right). This approach requires a new concept of *adaptable smart contracts*, because processes on a blockchain-based BPM system are deployed as smart contracts. Research in adaptable smart contracts will have to focus on the following activities: (a) Define the concept of adaptable smart contracts and identify the mechanisms of how smart contracts can be adapted; abundant research in process adaptation like [6] can be used as a systematic guideline to address this issue. (b) Investigate how adaptable collaborative scientific choreographies are represented on a blockchain BPM system using smart contracts. (c) As smart contracts stand for a “transaction protocol that executes the terms of a contract” [2], it has to be evaluated if the coordination protocols for choreography adaptation need to be designed, deployed and run as smart contracts as well. The system architecture of an enabling system may have different topologies featuring different functional components or parts of components on a blockchain. Investigation and evaluation of what the best architectural topology is for certain scientific domains must be carried out and at the same time consider the placement of data on the blockchain and the user’s security and privacy preferences. The consideration about how the design of the audit trail should look like is the same as with the first approach, however the monitoring capability may require a more complex architecture to be realized, since the monitoring information has to be made available directly into the choreography and workflow modeling tools. Publishing of the experimental results needs to be enabled with this approach, too. Advantages of this approach are that all steps in experiments and all adaptations performed will be stored in an immutable trace and that the coordination of adaptation will be a trusted protocol execution. For collaborative scientific explorations where reproducibility and trust are of utmost importance, this approach has a huge potential.

A disadvantage is the admittedly higher integration effort and complexity of the system and of the adaptation mechanisms.

4 Conclusions

In this paper we state our position that the combination of collaborative adaptable choreographies and the blockchain technology is a very promising one and qualifies as a solution for enabling trusted collaboration in eScience. We identified two possible courses of action for future research: the first approach uses blockchain platforms as a ledger only to store information relevant for the reproducibility of collaborative experiments and their results, and their publishing, whereas the second approach proposes using blockchain platforms for the execution of adaptive scientific choreographies and workflows through the notion of adaptive smart contracts. We have also identified the open research questions both approaches are facing and indicated their advantages and disadvantages. Admittedly, there are more open questions for future research. Some examples are the user friendliness of the potential realizations of either approach, their performance characteristics, and the access control mechanisms that will satisfy the demands of scientists to disclose only the minimum of information allowing for reproducibility.

References

1. Goble, C.: Results vary: the pragmatics of reproducibility and research object frameworks, keynote. iConference (2015). <https://www.slideshare.net/carolegoble/i-conference2015-goblefinalupload>
2. López-Pintado, O., et al.: Caterpillar: a blockchain-based business process management system. In: Proceedings of the BPM Demo Track and Dissertation Award (BPM 2017) (2017). http://ceur-ws.org/Vol-1920/BPM_2017_paper_199.pdf
3. Mendling, J., et al.: Blockchains for business process management - challenges and opportunities. CoRR abs/1704.03610 (2017). <http://arxiv.org/abs/1704.03610>
4. Mesirov, J.P.: Accessible reproducible research. *Science* **327**(5964), 415–416 (2010). <http://science.sciencemag.org/content/327/5964/415>
5. van Rossum, J.: Blockchain for research. *Science*, November 2017. https://figshare.com/articles/Blockchain_for_Research/5607778
6. Weber, B., et al.: Change patterns and change support features - enhancing flexibility in process-aware information systems. *Data Knowl. Eng.* **66**(3), 438–466 (2008)
7. Weiss, A., et al.: Model-as-You-Go for Choreographies: rewinding and repeating scientific choreographies. *IEEE Trans. Serv. Comput.* **PP**(99), 1 (2017)
8. Weiß, A., Karastoyanova, D.: Enabling coupled multi-scale, multi-field experiments through choreographies of data-driven scientific simulations. *Computing* **98**(4), 439–467 (2016)
9. Weiß, A., Andrikopoulos, V., Sáez, S.G., Hahn, M., Karastoyanova, D.: ChorSystem: a message-based system for the life cycle management of choreographies. In: Debruyne, C., et al. (eds.) OTM 2016. LNCS, vol. 10033, pp. 503–521. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-48472-3_30