Increasing Preservability of Research by Process Management Plans

Tomasz Miksa
SBA Research
Vienna, Austria
tmiksa@sba-research.org

Andreas Rauber
Vienna University of Technology
& SBA Research
Vienna, Austria
rauber@ifs.tuwien.ac.at

ABSTRACT

The reproducibility of research and possibility of reliable verification of results are key assumptions of credible data-driven science. In this position paper we present how the preservability of research can be improved by Process Management Plans, which are a novel process-centric approach to data management.

1. PROCESS MANAGEMENT PLANS

The advent of new means of performing research and sharing results in eScience settings poses new threats and sets new requirements to research procedures. One of them is providing long-term access and means for ensuring correct interpretation of research artefacts in the future. Preservability is defined as the degree to which these artefacts can be archived for as long as necessary ensuring its trustworthiness, redeployed and re-executed according to the expectations in a future environment that might potentially be different from the original. One of the attempts to increase preservability and thus the availability and reuse of scientific data are Data Management Plans. This data-centric approach aims to ensure that the scientific data is not only being properly managed during the course of a project, but also that it is available and preserved in the future. Unfortunately, the core of the research, which is the underlying process, is not in their focus. This results in a tremendous loss of valuable information needed to document, verify, preserve or re-execute the experiment. For this reason, in this position paper we propose a new concept of Process Management Plans (PMPs) which complements the description of scientific data taking a process centric view, viewing data simply as the result of underlying processes such as capture, (pre-) processing, transformation, integration and analyses. PMP is a living document, which is created at the time of process design and is maintained and updated during the lifetime of the experiment by various stakeholders. Contrary to current Data Management Plans, PMPs should be both human and machine readable. Their structure as out-

lined below necessarily appears relatively abstract due to a wide range of requirements and different practices in scientific domains in which PMPs should be used. The proposed structure of a PMP is as follows:

1. Introduction and context (human readable).
2. Description of the process, its context (using a formal context model [3]), as well as data used and produced by the process.
3. Preservation history (using preservation plans [2]).
4. Issues on sharing, reuse, verification, and legal aspects.
5. Monitoring and external dependencies [1].
6. Adherence and review.

PMPs increase preservability of research, because the comprehensive information on tools, actions and environment in which the results were obtained are documented. Furthermore, they enhance reproducibility and reusability of eScience experiments, because their implementation allows automatic collection of required information directly from the environment in which the scientific experiment takes place. Thus their actionability and enforceability is increased. They are more than just a paper report (or a digital version of it). This helps to ensure higher precision and coherence of information included in a PMP.

We are currently working on automation of PMP creation and verification by extraction of process characteristics automatically from its environment. Specific focus is on tool support to automate many of the various documentation steps, specifically capturing and monitoring of low-level process characteristics and performance aspects.

ACKNOWLEDGMENTS

This research was co-funded by COMET-K1 (FFG) and EC (ICT 269940/TIMBUS).

2. REFERENCES

- [1] BECKER, C., DURETEC, K., PETROV, P., FARIA, L., FERREIRA, M., AND RAMALHO, J. C. Preservation watch: What to monitor and how. In *Proc. of the 9th Int. Conf. on Preservation of Digital Objects* (2012).
- [2] KULOVITS, H., BECKER, C., KRAXNER, M., MOTLIK, F., STADLER, K., AND RAUBER, A. Plato: A preservation planning tool integrating preservation action services. In *ECDL* (2008), pp. 413–414.
- [3] MAYER, R., RAUBER, A., NEUMANN, M. A., THOMSON, J., AND ANTUNES, G. Preserving scientific processes from design to publication. In *Proc. of the 16th Int. Conf. on Theory and Practice of Digital Libraries* (2012), pp. 113–124.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

DPRMA '13 Jul 25–26 2013, Indianapolis, IN, USA

Copyright 2013 ACM 978-1-4503-2185-3/13/07 ...\$15.00.