

The representation of biomedical protocols

Larisa N. Soldatova✉, Ross D. King, Piyali S. Basu, Emma Haddi, Nigel Saunders

Brunel University, London, United Kingdom

Received 15 July 2013; Accepted 10 August 2013; Published 14 October 2013

Competing interests: the authors have declared that no competing interests exist.

Abstract

An explicit and logically consistent model for the representation of biomedical protocols would enable researchers in the Life Sciences to better record, execute, share, and report experimental procedures and results. The model we propose is based on the ontology of EXperimental ACTIONS (EXACT). The EXACT model is designed to define typical actions performed by biologists in labs and their essential attributes to enable recording of biomedical protocols in a computer processable form. This oral communication aims to report on the recent progress with the EXACT-based representation of biomedical protocols.

Motivation and Objectives

An explicit and logically consistent model for the representation of biomedical protocols would enable researchers in the Life Sciences to better record, execute, share, and report experimental procedures and results. The model we propose is based on the ontology of EXperimental ACTIONS (EXACT) (Soldatova et al., 2008). The EXACT model is designed to define typical actions performed by biologists in labs and their essential attributes to enable recording of biomedical protocols in a computer processable form.

EXACT was originally developed to support the protocols executed by the Robot Scientist "Adam". This is a physically implemented robotic system that applies techniques from artificial intelligence to execute cycles of automated scientific experimentation (King et al., 2009). A Robot Scientist can in a fully automatic manner: originate hypotheses to explain observations, devise experiments to test these hypotheses, physically run the experiments using laboratory robotics, interpret the results, and then repeat the cycle. Adam is capable of running in parallel thousands of experiments with yeast strains. The first ever fully automated scientific discovery made by Adam has captured public imagination and was listed by the *Times Magazine*¹ as one of the most important scientific discoveries of 2009. While the EXACT approach has been proved successful to represent and record experiments with yeast, it is not sufficient to support the representation and recording of a wider range of biomedical protocols. The proposed EXACT model is built on the success of EXACT ontology and extends its repre-

sentations to support a wide range of biomedical protocols.

Related works

Typically, ontological representations are focussed on modelling declarative knowledge about principal physical objects, and their qualities and relations with other objects. Process entities are included to represent the processes in which physical objects participate, e.g. gene-gene interactions. Representations where procedural knowledge plays the central role are rare. *The Ontology for Biomedical Investigations*² (OBI) includes both existential and procedural knowledge, but the main focus is on the representation of entities participating in biomedical investigations. For example, OBI Core contains only 17 procedural entities (occurrences), and about 100 continuants. OBI is sufficient to formally capture information about typical assays. However, standard operating procedures remain largely non-formalised, and are usually in the form of natural language with links to ontological classes to specify participating entities.

Initially the EXACT ontology contained only 45 experiment actions limited to the representation of biomedical lab automation protocols in yeast biology, and not all of them had well defined properties. Moreover, some of the defined experiment actions are not suitable for most of biomedical laboratories. For example, it is important to instruct a robot to remove a lid from a plate. However, such an action would be implicitly understood by a human researcher. At a laboratory standards workshop in Stockholm in December 2011 it was decided to modify the EXACT approach to suit the needs of the

¹ http://www.time.com/time/specials/packages/article/0,28804,1945379_1944416_1944423,00.html

² <http://obi-ontology.org/>

[Molecular Methods database](#)³ (MolMeth) for the recording of protocols (Klingström *et al.*, 2013).

This oral communication aims to report on the recent progress with the EXACT-based representation of biomedical protocols.

Methods

Analysis of biomedical protocols

We are manually inspecting thousands of published and also commercial biomedical protocols from several areas of biomedicine, including neurology, epigenetics, metabolomics, stem cell biology. We are analyzing instructions, what properties an experiment action has, what conditions are required and what goals are specified. We are also populating the EXACT ontology by newly identified experiment actions. We are modifying existing EXACT classes by specifying their properties. For example, the class *mix* has been defined as “to put together or combine (two or more substances or things) so that the constituents or particles of each are interspersed or diffused more or less evenly among those of the rest” (the Oxford English Dictionary, 1989). However, only one property equipment has been specified. The following new information about the experiment action *mix* has been added to the EXACT model:

```
has-participant (mix, entity) AND min
cardinality = 2
```

```
has-participant (mix, container)
```

to specify that at least two entities have to participate in the experiment action *mix*, and this action has to be carried out in some container. If a user while entering a lab protocol to a system would miss any of these properties, then the system would request to specify the missed properties of the action *mix*.

While (semi-) automated text mining methods are available, we judged that expert analysis of how Life Science practitioners express their procedural knowledge would output a higher quality knowledge model. We will use text mining tools to check if our model covers at least 95% of domain procedural knowledge. We will continue to analyze protocols till the coverage is sufficient.

Assessment of biomedical protocols by experts

Unfortunately, as often happens with natural language, the instructions in biomedical protocols are not always consistent or complete,

and therefore do not always guarantee full re-usability of the protocols (Soldatova *et al.*, 2008). For example, based on the analysis of existing biomedical protocols we have identified that the following attributes of the action store are typically recorded:

- an entity (what will be stored),
- duration (for how long it will be stored),
- condition (e.g. humid air),
- a location and/ or a container (where it should be stored).

However, it is not obvious what attributes are essential and must be recorded for each action store, and what attributes are optional. Some statements in published protocols, e.g. “store working solution at -20°C until use”, specify the entity and the condition, but not a duration or a location. There are also some statements about the action store in other protocols that specify locations and durations, but not conditions. We aim to capture all essential information about typical experiment actions, and also what information is optional and useful to record. We wish to strike the right balance between ensuring that all the essential information is recorded, and at the same time not requiring unnecessary or optional information from our users. Therefore we are consulting with experts in Life Sciences in order to define what properties of experiment actions are essential and what are optional.

Observation of the execution of experiment actions

Much procedural knowledge is implicit and difficult to verbalize, and therefore hard to capture and model. Therefore, a high quality representation of biomedical protocols can only be achieved if knowledge engineers directly observe how Life Sciences practitioners perform experiment actions in their labs. So far we have observed the execution of experiment actions in two different labs, one in the University of Aberystwyth (Wales), and the other in Brunel University (London). We are negotiating with three further labs to provide us with an access to their lab facilities and also to interview their biologists in order to capture implicit procedural knowledge.

Knowledge re-use

Previously defined relevant classes will be imported to the EXACT model. For example the OBI classes as *cell fixation* (definition: a protocol

³ <http://www.molmeth.org>

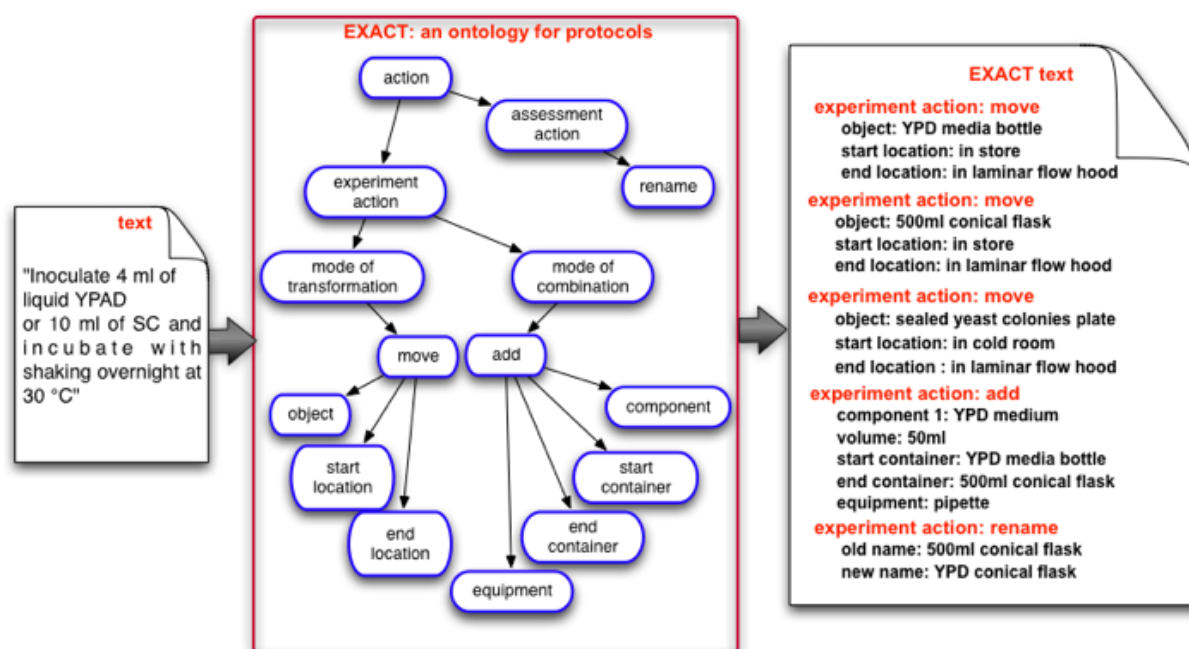


Figure 1. An example of the EXACT representation of biomedical protocol instructions.

application to preserve defined qualities of cells or tissues (sample) which may otherwise change over time), *decapitation* (definition: decapitation is a process by which the head of a living organism is physically removed from the body, usually resulting in rapid death), *labeling* (definition: the addition of a labeling reagent to an input biomaterial in order to detect the labeled material in the future) will be re-used in EXACT with the OBI URIs (Unique Resource Identifiers).

Results and Discussion

The number of experiment actions in the EXACT model has been increased significantly, and new properties had been defined. The EXACT model has been harmonised with the OBI representations. Currently the EXACT model is being verified by experts, and we are checking how well it covers the domain (see the methods section). By the end of this process we will deposit an updated EXACT model to [BioPortal](http://biportal.bioontology.org/)⁴.

We aim to provide an intuitive and easy representation of biomedical protocols and ensure that experimental procedures are fully reproduc-

ible (see Fig. 1). Through the use of the EXACT model reporting tools we will be able to provide biologists with more intelligent support. It will be possible to check if a protocol contains all the required information about experiment actions, suggest how to fill in any identified gaps or remove inconsistencies, provide templates for typical experiment actions, and help to re-use already recorded protocols.

Acknowledgements

This work has been partially funded by the Brunel University BRIEF award and a grant from Occams Resources.

References

- King RD, Rowland J, *et al.* (2009) The Automation of Science. *Science* **324**, 85-89. doi:10.1126/science.1165620
- Klingström T, Soldatova L, *et al.* (2013) Workshop on laboratory protocol standards for the molecular methods database. *New Biotechnology* **30**(2), 109-113.
- Soldatova LN, Aubrey W, *et al.* (2008) The EXACT description of biomedical protocols. *Bioinformatics* (Special issue ISMB) **24**, i295-i303. doi:10.1093/bioinformatics/btn156
- The Oxford English Dictionary (1989). Oxford University Press, 2nd ed.

⁴ <http://biportal.bioontology.org/>