Capturing intent and rationale for Linked Science: design patterns as a resource for linking laboratory experiments

Cameron McLean[1], Mark Gahegan[1], and Fabiana Kubke[2]

[1]University of Auckland, Centre for eResearch & Department of Computer Science
{ca.mclean,m.gahegan}@auckland.ac.nz
[2]University of Auckland, School of Medical Sciences
f.kubke@auckland.ac.nz

Abstract. The notion of design patterns, after architect Christopher Alexander, provides a powerful way to capture and describe reusable *design* knowledge for complex domains. In this position paper we present the idea of design patterns for molecular biology experiments, and discuss how they may be utilized to support experimental design reuse, reproducibility, and a platform for linking experiments. Design patterns provide an alternate terminology and interpretive framework that can capture expert experience and intent that is critical yet missing from current representations of lab methods that utilize web ontologies and computational workflows. We outline an approach to making design patterns a first class entity in support of linked experiments on the web and provide a glimpse of potential applications of laboratory design pattern knowledge.

Keywords: Linked Science, design patterns, semantics, ontology, workflows.

1 Introduction

While there has been much focus on the description and linkage of scientific datasets using web ontology languages, there has been less attention on how to describe and then link the surrounding laboratory methods that are an important step in generating such data. The use of biomedical ontologies[1] to annotate laboratory methods descriptions is a much needed and necessary first step to integrating laboratory experiments, however, current ontologies alone cannot always provide sufficient knowledge to support all the human reasoning and situated understanding one may need to act with such knowledge [1]. Traditionally, ontology takes in its remit the specification of domain semantics and hierarchical decomposition, while workflow representations encode processes - each effectively supporting the "what" and "how" of experiments respectively. Yet reusing an experimental design requires understanding of intent and rationale in addition to merely procedural facets - especially where they are to be executed in heterogeneous, non-computational (wet-lab) environments. This position paper introduces the concept of design patterns for laboratory experiments which can act as both container and notation to admit design rationale in a linked science setting.

---

[1] e.g. http://www.bioontology.org/

[2] For examples of lab patterns and use cases see http://goo.gl/D5RZsQ

2 Design patterns for laboratory experiments

Design patterns were first introduced in the domain of architecture by Christopher Alexander [2], as a way of encapsulating experts' knowledge and externalizing it to enable the generalization and communication of design. As a container for knowledge, patterns are realized as structured documents centered on problems, solutions, and the invariant "forces" that exist in a specified context. Through the invariant "forces", patterns identify, name, and abstract common themes in good design solutions that are gained from experience [3]. The pragmatic nature of design patterns, and their focus on expressing experience (rather than just domain concepts as for ontologies, or processes as for workflows) provides an architecture that can facilitate the adaption and reuse of laboratory experimental designs. Patterns provide a shared vocabulary and extensionally defined examples of solutions to complex problems and relate them back to an explicit rationale of why they are good.

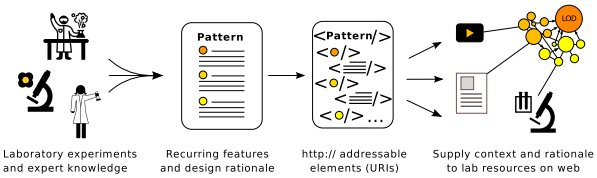

Fig1. Design patterns to capture laboratory knowledge.

Patterns not only give a set of concepts and vocabulary for a domain, but do so in a way that tells us what to do.[2] The problem/solution orientation of patterns gives us metaphorical dials to the domain, and tells us how to control and operate effectively with them. Unlike ontologies which aim to be capable of expressing any valid domain knowledge, patterns act more like recipes (yet more abstract and general than typical workflow representations) and give us a map[3] of the model space that tells us what parts we can vary, and what should remain invariant to achieve the desired outcome [4]. Considered as a form of knowledge management, we believe patterns offer additional advantages and add powerful metadata alongside traditional ontological and workflow approaches.

Other patterns exist in the context of the semantic web *e.g.* workflow[4] or ontology design patterns[5], which aim to provide usually *domain independent* constructs for normalizing and specifying knowledge modeling problems. In contrast, our notion of laboratory patterns as knowledge acquisition abstracts over laboratory procedures directly (*cf.* the modeling of them) to provide reusable design solutions for *scientific experiments* anchored in domain context - they supply us with domain concepts and relationships across diverse experiments gathered around a specified design intent.

---

2 For examples of lab patterns and use cases see `http://goo.gl/D5RZsQ`

3 A map analogy of patterns at `http://hillside.net/plop/2010/papers/kohls.pdf`

4 `http://www.workflowpatterns.com/`

5 `http://ontologydesignpatterns.org/wiki/Main_Page`

While we accentuate differences in representational approaches here, in reality we recognize the boundaries between ontologies, workflows, and design patterns are fuzzy as each tries to incorporate aspects of the other. For the purpose of discussion we make some general distinctions between the traditional forms of the three in Table 1 below.

Table 1. Some general distinctions between the traditional forms of workflows, ontologies, and design patterns for representing knowledge.

Property	Workflows	Ontologies	Design Patterns
Mode	Descriptive	Descriptive	Prescriptive/Instructive
Degree of formality	Formal	Formal	Typically not formal
Concepts defined	Procedurally	Intensionally	Extensionally
Focus	Specific procedure	General	Specific Problem
Utility	Replication and automation of processes	Model all feasible cases and compute inferences	Understand, adapt and reuse design solutions
Formal semantics	Yes	Yes	No
Models	Processes	Knowledge /Facts	Implementations

3 Making design patterns and their vocabularies web addressable entities.

The traditional form of design patterns are structured documents written in natural language. Thus, in order to transform them into a resource for linked science, a mechanism for publishing patterns and their vocabulary as defined web addressable entities is desired.

We view design patterns as data and ask how we may publish pattern knowledge following linked data practices. To begin we have developed a method for capturing pattern knowledge utilizing social methods adapted from other domains where design patterns are valid entities. The structured documents that result from "pattern mining" are collaboratively transferred to a semantic wiki based on the OntoWiki Application framework [5]. OntoWiki and its extensions enable the direct semantic content authoring of a knowledge base expressed in RDF, and provide for simple human and machine accessible interfaces for publishing linked data. Patterns entered by users become instances of a pattern model with defined syntax and semantics for pattern elements such as title, problem description, forces, context *etc*. The structure and URIs provided by the semantic wiki present an important first step in extending the form of design patterns from paper to a web based resource and supports the reuse of pattern content. Additionally, the wiki captures provenance, enables peer review, and serves attribution and credit for design pattern authors.

The challenges to this approach consist of specifying and refining the semantic formalization of pattern level concepts and their relations using RDFS, OWL and appropriate logics, and subsequently tailoring the OntoWiki Application framework. This work is non-trivial as patterns have complex, interrelated internal and external structures, and remains the current focus of our efforts.

4 A vision for the application of laboratory pattern knowledge

The annotation of lab procedural descriptions with vocabulary and context supplied by patterns is an obvious use of patterns as metadata. Currently, this coupling must be created manually due to the implicit nature of many pattern concepts, but the markup of existing documents or the authoring of future ones can be facilitated by adapting existing annotation tools such as Rightfield [6]. Laboratory methods and other data on the web indexed to patterns can provide an additional handle to browse, search, or filter methods at a granular level, across domains, and at the level of design intent – one which current semantics do not adequately provide

Patterns name invariant forces that exist in recurring lab scenarios and provide a valuable step towards the specification of minimal information reporting guidelines for diverse laboratory processes. Indeed, the need for "high-level abstractions of the components of experimental workflows" has been noted [7]. Furthermore, patterns resemble a wet-lab equivalent of abstract computational workflows described by [8].

Our vision is the creation of a laboratory pattern catalogue and web resource, providing scientists assistance in understanding, reusing, and adapting the diversity of published laboratory methods to their own needs. Principal in our approach is the publication of pattern content and vocabulary as linked data, such that it may be available for use anywhere on the semantic web.

We believe the problem/solution orientation of design patterns fits well with the cognitive processes of laboratory scientists when engaged with methods knowledge. In combination with workflows and ontologies, the pragmatic aspects of pattern knowledge can help provide a type of balancing – filling a representational gap in our methods descriptions somewhere between axiomized ontologies and workflows that can improve the epistemological adequacy of our scientific record.

5 References

1. Pike, W., and Gahegan, M. Beyond ontologies: Toward situated representations of scientific knowledge. *Int. Journal of Human-Computer Studies*, *65*(7), 674-688. (2007)
2. Alexander, C. The timeless way of building. New York: Oxford University Press. (1979)
3. May, D., & Taylor, P. Knowledge management with patterns. *Communications of the ACM*, *46*(7), 94–99. (2003)
4. Gamma, E., Helm, R., Johnson, R., & Vlissides, J.. Design patterns: Abstraction and reuse of object-oriented design. Springer Berlin Hiedelberg. (1993)
5. Heino, N., et al. Managing Web Content Using Linked Data Principles-Combining Semantic Structure with Dynamic Content Syndication. *Computer Software and Applications Conference, IEEE 35th Annual*. 245-250. (2011)
6. Wolstencroft, K., et al. RightField: embedding ontology annotation in spreadsheets. *Bioinformatics* 27(14), 2021-2022. (2011)
7. Taylor, C. F., et al. Promoting coherent minimum reporting guidelines for biological and biomedical investigations: the MIBBI project. *Nature biotechnology* 26:8, 889-896. (2008)
8. Garijo, D., and Gil, Y. A new approach for publishing workflows: abstractions, standards, and linked data. *Proceedings of the 6th workshop on Workflows in support of large-scale science*. *ACM*. (2011)