

# I Think Therefore I Am Someone Else: Understanding the Confusion of Granularity with Continuant/Occurrent and Related Perspective Shifts

James D. Myers

National Center for Supercomputing Applications,  
1205 W. Clark St, Urbana, IL 61801  
jimmyers@illinois.edu

**Abstract.** Managing multiscale and multi-witness provenance is often assumed to involve relatively straight-forward matters of matching identifiers and recognizing composite processes and aggregate artifacts. However, the issue is much more complex and related to millennia of debate over the nature of objects and processes in the world. This work develops a set of concrete examples where such issues arise in provenance, discusses the core conceptual distinctions involved, and postulates a basic mechanism for extending provenance models to enable integration across granularities and process types, recognizing the OPM ‘agent’ concept as a special case.

**Keywords:** provenance, multiscale, semantic integration.

## 1 Introduction

Over the past few years, there has been a broad effort to define common requirements for provenance, to outline real-world use cases, to define core models of provenance, and to assess interoperability of existing systems. In these discussions, there has been recognition that there are a variety of levels of granularity and a variety of types of processes for which provenance is important, as well as recognition that many use cases of interest require integration of provenance information across these dimensions. The issues involved in such integration have mostly been viewed as simple matters of aggregation, i.e. requiring concepts such as ‘collections’ of artifacts and composite processes. However, the need for constructs such as agents (as in the Open Provenance Model) hint at deeper issues related to the concepts of identity and distinctions between continuant and occurrent (or endurant and perdurant respectively), and of versions and replicas.

Consider the use case hinted at in the title: a person engages in a thought process (or attends a class or participates in a discussion to include cases where the process is external to the person). In OPM and similarly expressive languages, one could model this as

NaiveArtifact  $\leftarrow$ (used) EduProcess  $\leftarrow$ (wasGeneratedBy) InformedArtifact,  
or

Person  $\leftarrow$ (wasControlledBy) EduProcess

but neither is satisfactory. The first ignores the persistent identity of the person involved (and indeed only talks about what we could call the frozen state of a person) while the later fails to describe the effect of the process (why pay for education if you are not changed by it in explicit ways?). One can see versions being edited (that we would like to consider as being states of an evolving book), bits being moved through storage hierarchies and network locations (that we would like to consider as a single persistent logical ‘file’), and even workflows producing files (which we would like to consider as discoveries made by a team, financed by sponsors, and leading to Nobel prizes) as additional examples where the nominal notion of granularity of description in fact obscures a perceptual shift in the nature of the artifacts (and processes) involved. It is conceptually incorrect, for example, to consider a person and their naïve state as aliases: combined with provenance that the person was issued a driver’s license one would erroneously conclude that driving home from class would be illegal.

The core of this issue has been debated for thousands of years and the question is basically one of identity – what constitutes an object (artifact) – and the relationship between state and change, object and process, continuant (a thing outside time) and occurrent (an event in time). Much of the philosophical debate appears to center on how far one can drive either perspective as a way to model the world, but at least some groups [1,2] have attempted to define a practical view that is very relevant for the development of a provenance model capable of handling multi-granularity, multi-process-type use cases. Specifically, Galton and Mizoguchi [1] propose a definition of objects (object identity) in terms of which processes are considered internal and which external to that object; the chemical processes holding a rock together are internal to it whereas the processes that move it (falling under gravity, throwing) are external to it, and, critically, it is this separation of processes that enables us to identify what we mean by the rock.

In this view, it is easy to recognize that the examples above, and many of the issues involved in integrating accounts from different witnesses, involve perspectives where the process being described in the provenance record is internal to objects in one view and external in the other. This is not an issue of granularity in the sense of simply ignoring intermediate states and processes. Consequently, the relationships needed to provide crosswalks between these views are ones that connect one type of objects (A) with another type (B) than can be considered as aspects/states of an A object that can be used and generated by processes considered internal to A objects. Examples of such relationship vocabularies include versioning (versions are states, connected by editing processes, of a mutable (editable) book), copying (copies are physical instantiations, connected by replication processes, of logical entities), and, less obviously, collecting (a specific set of pictures, created by insert/delete processes, can be considered part of a living collection).

How does this affect provenance models? It suggests that a minimal model capable of integrating accounts from witnesses observing at different granularities or who are concerned with different processes, should include a construct to identify artifacts as ‘aspectsOf’ other types of artifacts that are ‘engagedIn’ a given process. I.e. a witness could add to a statement that

B was ‘usedBy’ process P  
that

B is an ‘aspectOf’ a (different type of) artifact A that is ‘engagedIn’ P

where A is a type of artifact for which processes of type P are part of their normal lifecycle.

In the spirit of profiles in OPM, these relationships could be specialized to address versioning, copying, etc. The current notion of Agent in OPM can be seen as a start in this direction in which ‘engagedIn’ is specialized to represent control. In addition to clarifying the need for Agent (Agents are minimally a different type of artifact that has a different relationship with a process than its inputs and outputs, whose identity is unaffected by participation in the process, and for whom participation in such processes is part of their normal lifecycle). The line of argument presented here suggests that Agents in OPM should be allowed to have an additional relationship with artifacts that are ‘aspectsOf’/states of the Agent that enable a finer level of detail on how Agents control processes.

A scalable provenance model sufficiently powerful to model real-world use cases involving multiple type of processes and multiple granularities must include a mechanism that recognizes when the notion of identity has shifted due to assumptions about the set of processes that can change identity. Such a mechanism could be along the lines of the “aspectOf”/“engagedIn” bridging vocabulary given here, which defines artifact-to-artifact and artifact-to-process relationships but may also need to include process-to-process relationships (Agents ‘live’ and processes they engageIn are part of their overall ‘life’ process) and could potentially include inference rules (in the spirit of OPM’s use of timestamps to validate/refute causality claims) that would support assessment of the validity of integrated provenance claims.

## References

1. Galton, A., Mizoguchi, R.: The water falls but the waterfall does not fall: New Perspectives on objects, processes, and events. *Applied Ontology* 4, 71–107 (2009)
2. Grenon, P., Smith, B.: SNAP and SPAN: Towards Dynamic Spatial Ontology. *Spatial Cognition & Computation: An Interdisciplinary Journal* 4(1), 69–104 (2004)