Abstract. The Dublin Core Metadata Initiative (DCMI) maintains a vocabulary of several dozen metadata terms, notably the fifteen-element “Dublin Core.” These terms (and their historical versions) are identified with URI references, described in Web documents and machine-processable schemas, indexed in registries, cited in application profiles, and of course used in metadata records. Within DCMI, however, the emphasis has been less on growing this small vocabulary than on clarifying how the DCMI vocabulary can “play well” with other, complementary vocabularies in a Semantic Web that recombines semantics from multiple sources for specialized purposes. This priority has led DCMI to examine models for referencing other vocabularies and to clarify the modeling bases for interoperability among heterogeneous systems.

1 Maintaining a Small Vocabulary

Since the publication of Version 1.0 and RFC 2413 in September 1998, the fifteen elements of the “Dublin Core” — Title, Description, Date, and the rest — have become a familiar feature of the digital landscape [11]. They have been printed on tee shirts, used in millions of metadata records, and enshrined in an international standard, ISO 15836 [7]. Its user community, the Dublin Core Metadata Initiative (DCMI), has evolved from a workshop series into a maintenance network with an increasingly global basis of institutional stakeholders [8].

Since the publication of Version 1.1 in 1999 there has been little discussion in the Initiative about the Core itself. The fifteen elements have some long-recognized and well-understood flaws, but none so serious as to justify the disruption of a full-scale revision. Rather, discussion has focused more on the policies, processes, metadata principles, and data models that provide a context for the Core.

The problem as initially posed in 1994 — that of agreeing on a simple set of descriptors for embedding in Web pages — has evolved into a cluster of issues ranging from technical architecture to naming policies and maintenance procedures. Much of this discussion has aimed at clarifying how the elements can be used with an ever-evolving array of Web technologies, from the simple HTML of the early years through XML and XML Schemas to Resource Description
Framework and OWL Web Ontology Language. Much of the effort has been motivated by the vague but powerful vision of a Semantic Web in which well-defined data and metadata underpin applications that are increasingly automated and intelligent.

This paper summarizes the state of discussion as of October 2004 regarding the identification and referencing of metadata terms in DCMI– and related non-DCMI–maintained vocabularies. The paper reduces the problem of maintaining a vocabulary for the Semantic Web to its essence: the declaration of terms with persistent identifiers and the maintenance of assertions about those terms — i.e., about what they mean, how they change over time, and how they relate to terms in other vocabularies. Specifically, the paper examines several areas in which practical problems reveal issues relevant to the more fundamental problem of maintaining and using metadata vocabularies in a Semantic Web environment:

- persistence and semantic stability policies for term identifiers,
- the historical versioning of terms,
- namespace hosting,
- the identification of controlled vocabularies of values,
- the “re-use” of terms in application profiles,
- the etiquette of assertions about terms in other vocabularies.

Taking each of these areas in turn, this paper will reveal a basic tension between the need to solve problems in the short term and the sustainability of solutions for the long term. The question of when to create new metadata terms and when to cite or re-use existing terms maintained by others depends on the sustainability of the social and institutional processes by which the vocabularies are maintained. Section Two will lay the groundwork for this discussion by describing how DCMI forms identifiers for its own metadata terms and what social commitment is thereby implied. Section Three will then describe how practices for referencing terms in other, complementary vocabularies have emerged in response to various practical problems.

None of these issues are unique to Dublin Core, so the solutions devised for the DCMI context should be of interest to other vocabulary maintainers, especially as the solutions are tested over time in an evolving landscape of Semantic Web vocabularies.

2 Identifying Metadata Terms

2.1 DCMI Namespace Policy

The Internet was revolutionary because it made the resources of any connected server accessible via a single global address space. The vision of a future Semantic Web further generalizes this notion of a global space of addresses to that of a global space of identifiers. According to Tim Berners-Lee, “The most fundamental specification of Web architecture, while one of the simpler, is that of the Uniform Resource Identifier, or URI. The principle that anything, absolutely