Chapter 31
The Development of an Optimised Metadata Application Profile

Paul Walk
EDINA, University of Edinburgh, UK

ABSTRACT

This chapter describes an approach to the development of a metadata application profile. It is particularly concerned with the class of application profile which is optimised for a specific use-case, rather than those which are more concerned with supporting general interoperability in a broader domain. The example of the development of a particular application profile, RIOXX, is used to illustrate some of the methodology discussed. Much of the approach described in the chapter was designed during the course of the development of RIOXX. Issues which are given particular consideration include a focus on the close involvement of 'implementors' (normally software developers), the adoption of ideas from agile software development, continuous testing and open development, and ongoing maintenance and sustainability "on a shoestring".

INTRODUCTION

Application profiles consist of data elements drawn from one or more namespace schemas combined together by implementers and optimised for a particular local application. (Heery & Patel, 2000)

This is one of the earliest and, arguably, the best definition of the term application profile. The author prefers this definition because, while mentioning 'data elements', 'namespaces' and 'schemas' (all crucial aspects), it is also careful to emphasise the importance of the involvement of 'implementers' and 'optimisation' for a 'particular local application'.

While this definition suggests that application profiles ought to be focused on meeting particular requirements, the recent history of their varied development places them on a scale, from those which have been developed to enable broad interoperability in a general domain or 'application space', to

DOI: 10.4018/978-1-5225-3422-8.ch031

those which are, indeed, optimised to enable a particular application. The concept of interoperability is intrinsically bound up in the development of metadata standards and profiles. However, interoperability and optimisation do not always complement each other as features, and application profiles tend to emphasise one or the other. For example, the Scholarly Works Application Profile (SWAP) is described as "a DC Application Profile for describing an eprint, or scholarly work" ("Scholarly Works Application Profile", 2009). The documentation for this application profile says nothing about use-cases and it does not identify any 'particular local application'. SWAP, like plenty of other application profiles, is designed to facilitate as-yet-unknown applications by explaining how to describe a particular type of resource to an appropriate level of detail. Essentially, this kind of support for interoperability is about enabling the possibility of future interoperation, rather than being focused on any particular, intended implementation. As such, if SWAP were to be placed on the aforementioned scale, it would be positioned close to the 'interoperability' end of the axis, and some distance away from 'optimisation'.

This chapter is concerned with the other end of that scale, with application profiles which are optimised for particular local applications. The specific example of RIOXX is used to illustrate a methodology for the development of an application profile which focuses on supporting its intended implementation. RIOXX has been developed to address a specific use-case: this use-case together with the application profile itself are briefly introduced, while the bulk of this chapter is concerned with the methodology used in development, deployment and implementation.

BACKGROUND

The Use-Case

In the United Kingdom, research funded by national government is managed by seven Research Councils. With a growing desire to see publicly-funded research made more openly available to those who have funded it through their taxes, a policy for 'open access' to published research papers has been introduced by Research Councils UK (RCUK), the body which represents the Research Councils in such matters.

The RCUK Policy on Open Access[1] demands that any organisation which receives Research Council funding should make the outputs from that funding available. RCUK recommends the use of open-access repositories for this purpose. Most research-active institutions in the UK maintain an institutional repository, so this is the *de facto* standard way to make such outputs available. While there is growing interest in other types of research output—in particular data—the focus of this policy is on research *papers*. Publishers often place 'embargoes' on scholarly publications - preventing them from being shared, except through the publisher's own system, for a period (usually between 6-24 months) after publication.

An important detail is that RCUK requires that the institution provide assurance that the policy is being followed. They expect the institution to report that a paper exists, within a short period after it has been accepted for publication, regardless of any embargo. This requires that the institution make available a certain amount of information about the paper. In addition to the common metadata elements such as 'title' and 'author', the institution also needs to supply information about the source of the funding for the paper being described, as well as some indication of the license terms under which it is made available.

Because research-active institutions in the UK already manage papers in institutional repositories, RCUK worked with an organisation called Jisc to fund the development of a metadata application profile