Preface

This book is concerned with the processes for development of metadata application profiles, in an increasingly complex world of data, information standards and requirements for data exchange and interoperability.

The appearance of this book is timely. The rapid rise in the importance of data science is driving an urgent need for better metadata in order to manage and exploit the 'data deluge'. As interdisciplinary research becoming more prevalent, the issue of interoperability (or lack thereof) between contexts comes into sharp focus. Furthermore, the concept of 'metadata' has, in recent years, entered into mainstream consciousness. No longer an obscure concern for a small group of information professionals and researchers, metadata is fashionable and the average citizen now has at least heard the term and has some awareness that metadata is important.

Metadata application profiles equip systems - almost always based in software - to manage and exchange data. Application profiles allow us to gain some purchase on the complexity and variety of data, so that we might manage, use and re-use those data. However, as we will see in the chapters of this book, metadata application profiles are not easily constructed. Moreover, they cannot be developed in isolation. As Karen Coyle in the opening chapter says, issuing what might be heard as a 'call to arms':

The days in which one could invent their own metadata schema without concern for existing standards or data exchange are long over.

The chapters in this book cover a range of approaches to the development of metadata application profiles. Drawing on research and development work in disparate domains, taken together they illustrate how development methods must be tailored to fit the context. From the variety presented in these pages, it becomes clear that no single method can possibly solve every metadata application profile development use case.

The concept of the metadata application profile is defined by Heery and Patel (2000) as "data elements drawn from one or more namespace schemas combined

together by implementers and optimised for a particular local application". This admirably neat and succinct definition accommodates a large range of possible types of implementation, the heterogeneity of which is apparent in the descriptions in the chapters to follow.

This book, then, is by and for those people who work with metadata schemas and application profiles. Once this audience might have been limited to a group consisting of librarians and related information professions, and a handful of researchers working in some branch of information science. However, the new importance of data and its management has meant that a much wider range of people have a stake in metadata, from software engineers to scientists working with research datasets in just about any discipline, to bureaucrats needing to account for public spending, to the growing army of 'data journalists'. Whether it is realised or not, metadata is everywhere, and a growing number of people need to have some understanding of how to work with it in a way that maximizes interoperability for use and re-use. And even if they are not actively employed in its development, if they truly want to understand metadata and how to maximise its interoperability, then they need to appreciate how it is constructed.

The authors of the different chapters in this book represent a mixture of research and practical interests and, as such, the book should have something to offer researchers and practitioners alike.

In Application Profiles - An Overview, the author offers an overview of current practice in metadata application profile development. Ranging from the relatively simple approach of the 'data dictionary', through to the more sophisticated Singapore Framework, this chapter reveals in general terms not only what has been achieved so far, but also how much is left to do if we are to realise the promise of fully interoperable metadata. As in several other chapters in this book, great emphasis is placed on developing approaches to creating 'machine-readable' profiles. Application profiles which are machine readable and even, in the author's words, "to the extent possible, machine-actionable" create opportunities for automated or, at least, semi-automated processing of various kinds. Standard modelling languages such as UML become useful in this context, while automated testing and validation become viable too. The Singapore Framework's 'Description Set Profile' is recommended as a "strong theoretical and practical" foundation for this kind of development, and the community is encouraged to continue development in this space.

The Development of an Optimised Metadata Application Profile describes a development methodology which departs from convention by largely ignoring concerns of wider interoperability in favour of a narrow focus on implementation (although, of course, the pragmatic decision to re-use existing namespaces - notably DCMI Metadata Terms - does introduce a degree of interoperability into the described metadata application profile, whether or not this is a priority for the developer).

Tied to a specific use-case, the methods used here are borrowed, in part, from the 'agile' school of software development practice. From this point of view, a metadata application profile becomes just one more component in an application which is optimised early in the development process. While opportunities for wider interoperability may present themselves, these are rejected "ruthlessly" in a single-minded effort to satisfy only the specific requirements being addressed by the project. Other chapters in this book have focused on developing documentation for application profiles which is machine-readable, promoting 'semantic interoperability' between documents. By contrast, the approach to documentation described in this chapter is to focus almost entirely on the needs of the developers of software. This is in keeping with the general thrust of this chapter which is to promote implementation above all other considerations.

The Minimum Mandatory Metadata Sets for the KIM Project and RAIDmap focuses on the development of M3S - Minimum Mandatory Metadata Set, for the KIM and RAIDMap project records. The authors notably present the historical context of the domain of digital preservation metadata focusing all evolution from the development of the OAIS Reference Model to that of schemas such as PREMIS, which became the basis for the KIM Minimum Mandatory Metadata Sets - M3S. All the stages of the process of deriving the KIM M3S from PREMIS are then explained and the decisions taken are conveniently justified. The same process was used to derive the RAIDmap M3S, albeit it had different requirements from those of the KIM M3S. Once again, all the stages and decisions are explained and justified. The results of the application of this process in these two distinct settings become apparent as the authors explain it step by step. The tables with the final results thus appear to the reader as a welcome evidence of the application of the process previously explained. At the end of the chapter, the authors provide valuable recommendations on the development of M3S.

In A Methodology For Effective Metadata Design In Earth Observation, the focus is squarely on interoperability. Having established the degree of "technical diversity" in instruments and sensors found on satellites, and the corresponding variety of data formats produced by them, the author goes on to describe the challenge of designing systems to discover images in this heterogenous mass of data and "domain specific" metadata. This chapter is notable for its application of a "model-driven" approach borrowed from software engineering, using this to complement DCMI's Singapore Framework (which makes an appearance in other chapters), with the result that functional requirements are able to be formally defined. In common with some other chapters, the work here is underpinned by the use of linked open data on the Web.

In The Development process of a Metadata Application Profile for the Social and Solidarity Economy, the authors show how a community specific application profile was built using the Me4MAP method. The development of this application

profile has the singular characteristic of having been performed iteratively with the development of the Me4MAP method, which provided several feedback cycles between one development and the other. The chapter explains how the team was built and how each planned activity was carried on, and points to external sources for further information about the results.

In Developing Metadata Application Profiles for Open Educational Resources Federated Repositories - The Case of the Open Discovery Space Metadata Application Profile, the focus is on a method for developing application profiles federated Learning Object Repositories and an application scenario. The method consists of four steps targeted to define the metadata elements, their 'obligation status' and their 'value space'. The choice of metadata elements and value spaces is done using formulas that take into account several factors, including frequency. The chapter explains each of the steps and shows their application on a real world scenario, the Open Discovery Space, a federated repository of sixteen (16) learning object repositories. At the end of the chapter the reader may find a reflection over some aspects of the method.

In Using reverse engineering to define a Domain Model - The case of the development of a metadata application profile for European poetry, the focus is on the development of an application profile for a European poetry platform using the Me4MAP method. This platform aims to aggregate poetry metadata from various European platforms and provide services on this set of aggregated data, as well as make the data available as Linked Open Data. Of special interest is the way in which the authors used the Me4MAP method in a case where some of the activities provided by the method were already performed or partially performed. Thus, the focus is essentially on the design of the application profile from the analysis of the models of the various platforms / databases that provide metadata and that will inform the construction of the application profile at a later stage. At the end of the chapter the authors reflect on what went well and on which aspects to improve from the point of view of those who develop application profiles.

In Involving Data Creators in an Ontology-based Design Process for Metadata Models, the focus is in the process of determining and modeling the metadata elements and values for describing research data in specific domains. The resulting models are represented as lightweight ontologies which are used in the scope of the Dendro platform. The authors present the processes they used for different kinds of data through the presentation of four scenarios from four research domains: vehicle simulation, biological oceanography, hydrogen production and social sciences. The chapter emphasizes the user studies that the team conducted to determine how different types of data are produced and how they should be described to optimize retrieval. As in other chapters, at the end of this chapter the reader finds a section about the lessons learned.

By showing various approaches to the development of metadata application profiles, this book provides a starting point for those who want to develop this artefact and do not know where to begin. Each chapter focuses on the development process used and many of them reflect on the lessons learned, identifying aspects of methods which have yielded good results, together with those which need improvement, providing valuable reflection as a basis for future development. Together, the chapters provide, to the newcomer, a range of perspectives which will widen their understanding and better equip them to undertake development in this field.

This book also paves the way to an open, serious and in-depth discussion of methods for designing metadata application profiles. Until very recently the methods used in the development of metadata application profiles were based on those from other disciplines (especially computer and library sciences). In recent years, some tailored approaches to this subject have emerged, sometimes in distinct, but at other times in similar scenarios. The reader of this book is thus presented with some of these approaches, inviting them to reflection and action. It is then up to the metadata community to test, reflect and discuss the different approaches, the types of scenarios in which they may be applied and their validity and feasibility in a Linked Open Data environment.

REFERENCES

Heery, R., & Patel, M. (2000). Application profiles: Mixing and matching metadata schemas. *Ariadne*, *25*, 27–31.