Preface

Linking Enterprise Data is a new concept, based on an idea more than twenty years old. Tim Berners-Lee’s original proposal for the World Wide Web in March 1989 was based on a system of linked information systems. The early Web was intended to interlink information from various systems to solve organizational problems, such as the high turnover of people and the restriction of information to data silos. The hope was to create a distributed information system that would allow "a pool of information to develop which could grow and evolve with the organisation and the projects it describes."

The Web has grown into the world’s largest information system. By 2000, Web architecture had been dissected and described by Roy Fielding. Representational State Transfer (REST) was Roy’s answer to why the Web worked so well. In a world plagued by software problems, machine crashes, and network outages, the Web never fails. The Web is robust and resilient to change. The Web survives changing machinery, operating system updates, changes in the way we structure index and find information. No other software system provides the features and functions of World Wide Web.

Linked Data techniques have become interesting to organizations of every shape and size. The Linked Open Data (LOD) project began as a community effort of the World Wide Web Consortium’s Semantic Web Education and Outreach Group. The project has begun to turn the document-oriented Web into a database of global proportions. The ability of the modern Web to deal with both documents and data have shaped a general solution for information dissemination and integration. The time for linking enterprise data has come.

This book records some of the earliest production applications of linking enterprise data. Parts of it serve as a roadmap for those seeking to replicate their successes. Part I of this book attempts to answer the question why enterprise data should be linked. The chapters in Part I provide valuable guidance to those writing business cases, for those needing to justify internal development efforts, or for those writing requests for proposals to external vendors. Dean Allemang discusses why enterprises must adopt Web techniques for data integration and provides such techniques fit into enterprise systems. Dean makes a strong case that enterprises
must change the way they approach information technology systems. Indeed, since information systems have such a profound impact on the operational aspects of a business, he makes the case that enterprises need to change the way they approach their operations.

Edward Curry, Andre Freitas, and Sean O’Riain discuss the role of community-based data curation. Enterprises have become more distributed, less centrally managed and less integrated in their systems. The lessons Ed and his colleagues have captured from real-world attempts to curate distributed data for the purposes of ensuring data quality will apply to many enterprises. They provide some important best practices extracted from early adopters of Linked Data techniques.

Part II is short, but critically important. Part II provides material assistance for business managers seeking to propose Linked Data projects. Bernadette Hyland discusses the characteristics of enterprises ready to take on Linked Data projects and provides useful fodder for business cases. Her simple guidelines for getting a Linked Data project started have generally been lacking in the public discussion to date. Kristen Harris’ real-world experiences creating and managing the swOrDFish project at Sun Microsystems demonstrated the potential of linked enterprise data to integrate disparate systems in large enterprises. She provides guidance for the navigation of corporate management to approve and support projects with far-reaching infrastructural ramifications.

The techniques of Linked Data can be subtle and technical, although not out of reach for those with traditional enterprise skills. Part III provides three explanatory chapters that address different technical aspects of linking data. Alexandre Passant, Philippe Laublet, John G. Breslin and Stefan Decker present ways to integrate enterprise social networking solutions such as wikis and blogs. Initial enterprise adoption of new technologies can sometimes create new problems. Alexandre and his co-authors offer both insight and solutions to the integration of Web 2.0 and Web 3.0 techniques.

Roberto Garcia and Rosa Gil demonstrate how the translation of existing data sources may be brought to the Web of data. Reza B’Far and I offer technical approaches to enterprise problems of scale. Reza addresses logical reasoning techniques for enterprise-scale data and I present ways to ensure the long-term viability of Linked Data identifiers.

Part IV provides five success stories from the front line of enterprise adoption. Each story highlights a different aspect of Linked Data in an enterprise context. Thomas Baker and Johannes Keizer address standards for highly distributed operations developed for the Food and Agriculture Organization of the United Nations. Steve Harris, Tom Ilube and Mischa Tuffield show how Linked Data techniques are used as the basis for their Web-scale company, Garlik. Constantine Hondros of publisher Wolters Kluwer illustrates and presents several approaches for solving integration problems of textual content. Chimezie Ogbuji develops a new enterprise system using a Linked Data approach and Yves Raimond, Tom Scott, Silver Oliver, Patrick Sinclair and Michael Smethurst of the British Broadcasting Corporation present their innovative corporate treatment of the Web itself as their content management system.
We have been able to draw some tentative conclusions regarding success criteria for Linked Data projects in an enterprise. First and foremost may be from Jeff Pollock of Oracle Corporation when he said, "If information systems are to keep up with business, we need to change more than technology - we need to change how people deal with technology." Linked Data techniques offer us a means to do just that; we can radically change the interfaces to our existing systems while we build upon them. We can wrap and expose our silos in order to layer a Web-like distributed system over them.

Secondly, the lessons of the Web clearly apply to enterprises. The Web works for some very good, and very explainable, reasons. Those reasons transcend Representational State Transfer (REST), the architectural principles that underlay the idealized Web, and add the techniques of the Semantic Web, especially that subset being used by the Linked Data community. Individual technologies, though, are clearly less important than techniques that have proven their worth. Technologies continue to evolve; good techniques are more resilient and worth building upon.

Note how different the organizations in the success stories are from one another! A broadcasting company, a publishing company, a healthcare provider, a data security firm, an international policy organization. Other chapters referenced other types of organizations, including a utility company and library organizations. If Linked Data techniques work for all of them, those techniques are very likely to apply to others.

All of our success stories have some interesting commonalities: At least one expert in Semantic Web techniques was used by each organization. Each attacked a significant business problem instead of relying on the technologies to "sell themselves". Each leveraged significant existing investments, especially those with captured or implied semantics. Every success relied upon universal addressing of resources via the Web’s Uniform Resource Indicator (URI) scheme.

There were also some major differences between the success stories. Those differences define tools and techniques that are more situationally dependent. The most noteworthy is that very different degrees of data modeling were employed. Complete, top-down data modeling is expensive, difficult and should be undertaken only where it provides value. Specific technologies to describe data (OWL, SKOS, RDF serialization formats) varied widely, as did the use of the SPARQL query language.

Trust may be a larger issue in intra-business data than it is on the general Web if business decisions are being made based on the information. Issues of trust in large organizations may be facilitated by social considerations, e.g. via signing of work, taking credit for additions or edits, tying comments to logins. Many of today’s enterprises are large and distributed enough to make use of Web techniques for building and maintaining trust socially over a technical framework.

The four parts of this book are presented hierarchically, like most books in the last 2300 years of Western tradition. The material in this book should not be thought of as a hierarchy, but rather like a graph, like the Web itself. All of the chapters in this book contain nuggets of information useful to enterprise professionals looking to apply Linked Data techniques. The opening chapters do address technology and success stories as well as laying and conceptual foundation. The technique chapters
reference success stories of their own. The chapters addressing success stories are chock full of lessons learned in relation to management, approach and style. It is no more possible to fit these chapters into a strict hierarchy than it is to do so with content on the Web. Readers are encouraged to troll the index and review the notes at the beginning of each Part to find the information most relevant to themselves.

The enterprise application of Web architecture to business problems is in its infancy. We hope that this book can be used to assist those managers, data professionals and developers at the forefront of solving today’s formidable enterprise data challenges.

Observant readers may notice that any given chapter may use either American and British spelling. The use of mixed spelling systems represents the international nature of the contributing authors and, indeed, the international range of Linked Data research and deployment. We consider such diversity to be a feature, not a bug.

Nigam Shah of the National Center for Biomedical Ontology provided reviews and commentary on this book’s contents, as did most of the individual chapter authors. Ivan Herman of the W3C and Eric Miller and Uche Ogbuji of Zepheira provided introductions to prospective authors and suggested content. Our editors at Springer, Susan Lagerstrom-Fife and Jennifer Maurer helped to make the creation of this book much easier than it could have been. Thank you to all.

Fredericksburg, Virginia, USA, David Wood
June 2010
Linking Enterprise Data
Wood, D. (Ed.)
2010, XXVI, 291 p., Hardcover