Eero Hyvönen

Cultural Heritage and Digital Humanities on the Semantic Web

Developing Linked Data Infrastructures and Intelligent Applications for a Web of Wisdom

November 16, 2026

Springer Nature

This book is dedicated to my colleagues at the Semantic Computing Research Group (SeCo), Aalto University, Department of Computer Science, Finland, and at the Helsinki Centre for Digital Humanities (HELDIG) and Department of Computer Science, University of Helsinki, Finland.

Preface

Study the past if you would define the future.

Confucius (Kong Qiu) (c. 551–479 BC)

The Internet, World Wide Web (WWW), and Artificial Intelligence (AI) have become three foundational pillars of the digital age. Founded on the Internet's global communication network, the WWW gives everyone, both organizations and the general public, the democratic possibility to communicate, publish, and use data openly with billions of fellow users at nominal costs. Based on open WWW infrastructure and web content, new business models and applications have changed society and research in radical ways. Given the breakthroughs of the Semantic Web (SW) since the 2000's and Artificial Intelligence in the 2020's, the WWW has become a huge repository of data and content to be used by both data providers, application developers, and end users with intelligent User Interfaces (UI) and agents. [1]

The Semantic Web, Web of Data, is now transforming into a kind of Web of Wisdom (WoW) [2] where web documents are not only searched for by search engines and browsed, but the WoW is also used for knowledge discovery, to formulate research questions and answer them, for enhanced productivity and automation, decision making, and for generating new content using generative AI. AI means arguably the fourth step and revolution in our scientific thinking after the "three great humiliations" to human narcissism: the Copernican revolution started by Nicolaus Copernicus (1473–1543) took us away from the center of the world, Charles Darwin's (1809–1882) evolution theory reduced us as just another species of animals, and Sigmund Freud's (1856–1939) theories of mind indicated that our behavior is driven by the unconscious mind rather than rational thinking. AI displaces humanity from its exclusivity as the sole source of intelligence, creativity, and reasoning.

The technical progress behind digitalization has been phenomenally rapid, sped by Moore's Law on exponentially growing computational efficiency of computers, Big Data available, and new innovative algorithms, but has also introduced serious challenges and concerns. [3] Early WWW adapters included criminals using the new

platform for, e.g., fraud. Unethical content could now be easily published in addition to ethical. Social networks and media are widely used to publish disinformation and misinformation for political, financial, and other goals. AI algorithms that provide information and solve problems on the Web are based on maximizing income for advertisers whose goals are often different from the common good. A great concern of MyData movement is how to control our data and privacy; web platforms know more about us than we know and use the data for their own purposes, which can be harmful. Using personal data can facilitate better services on the Web, e.g., regarding insightful recommendations, well-targeted health services, etc., but controlling privacy is often needed, too. The dark side of the WWW is studied in the field of Digital Humanism [4]. However, the focus in this book is on the bright side of the WWW revolution: how the Web can be used for the good from a technical point of view when publishing and using Cultural Heritage (CH) data.

This book gives a vision and bird-eye view of how the WWW is evolving, especially pertaining to Cultural Heritage (CH) content publishing, Digital Humanities research, and application development. It is discussed how CH contents and data can be published on the Web in maximally useful ways from the point of view of three stakeholder groups: 1) The memory and other organizations who produce, curate, and publish the data. 2) Data consumers, including the general public and researchers in humanities in the field of Digital Humanities [5]. 3) The application developers who make the data useful to the consumers in different ways.

The book is aimed at computationally oriented CH professionals, museum curators, librarians, archivists, DH researchers, and computer scientists interested in Linked Data and CH applications on the Semantic Web. The text is focused on practice and applications, making it suitable for students, researchers, and practitioners developing web services and applications of CH, as well as to CH managers willing to understand the technical issues, opportunities, and challenges involved in linked data publication.

CH data are syntactically and semantically heterogeneous, multilingual, semantically rich, and highly interlinked. The data is produced in a distributed and open fashion by museums, libraries, archives, and media organizations, as well as individual persons in Citizen Science and other collaborative efforts. Managing publication of such richness and variation of content on the Web, and at the same time supporting distributed, interoperable content creation processes poses challenges, where traditional publication approaches need to be re-thinked. It is shown in this book that application of the principles and technologies of Linked (Open) Data and the Semantic Web is a promising approach to address these problems. The development is leading to the creation of large national and international CH portals, such as Europeana, to large open data repositories, such as the Linked Open Data Cloud, and massive publications of linked library data in Europe, the U.S., and Asia. Cultural Heritage has become one of the most successful application domains of the Linked Data and Semantic Web technologies. Semantic web infrastructures shared by stakeholders will be shown to be a key ingredient supporting this kind of work.

This book tells why, when, and how Linked (Open) Data and Semantic Web technologies can be employed in practice in publishing CH collections and other

contents on the Web. The text motivates and presents a general semantic portal model and publishing framework as a solution approach to distributed semantic content creation and publication, based on a Semantic Web infrastructure. Such an infrastructure, based on the global standards (recommendations) and best practices of the World Wide Web Consortium (W3C), includes shared metadata models and Knowledge Organization Systems (KOS) to populate them, supported by shared ontology and data services, as well as tools that alleviate the use of new technology and linked data in legacy cataloging systems. The goal of all this is to provide layman users and researchers with new, more intelligent and usable web applications that can be utilized by other web applications, too, via well-defined Application Programming Interfaces (API). At the same time, it is possible to provide publishing organizations with more cost-effective solutions and tools for content creation and publication.

The book begins with a topical introduction to Cultural Heritage (CH) data, Digital Humanities (DH) research based on it, and using Semantic Web (SW) technologies for publishing and using CH content on the Web. It is then explained in Part I on page 10 how CH content publishing is developing in successive generations towards a Web of Wisdom based on semantic web computing and Artificial Intelligence. Part II on page 32 gives a gentle introduction to the technologies needed for this. Using these technologies, it is possible to create a semantic web infrastructure that makes it arguably substantially easier to find, access, and (re)use DH data in DH research and applications. Part III on page 130 discusses how such an infrastructure can be built cost-effectively and in a sustainable way and presents lessons learned in Finland based on a long-standing national initiative. During this work, a model for the use of the infrastructure for applications, called the Sampo Model with related software tools, has gradually evolved, as presented in Part IV on page 170. Lessons learned in developing a semantic web infrastructure for CH in Finland (2003–2026) and beyond are presented. This infrastructure and the Sampo Model have been applied, as presented in Part V on page 196, to develop a series of tens of interlinked data services and semantic portals with up to tens of data providers and millions of users in some cases, suggesting the feasibility of the adopted approach. With each new application, a new piece into the underlying open infrastructure can be added to be re-used in the next applications.

The over-arching idea is to encourage developers to create a sustainable, ever-growing open infrastructure to pave the way towards a Web of Wisdom.

In Espoo, Finland, November 16, 2026 *Eero Hyvönen*

Acknowledgements

The book contents are based on collaboration with some 100 researchers and visitors in the Semantic Computing Research Group (SeCo)[1] at Aalto University, Department of Computer Science, Finland, and University of Helsinki, Helsinki Centre for Digital Humanities (HELDIG) and Department of Computer Science, Finland during 2001–2026. Thank you! The list is too long to be included here, but SeCo contributors to the research reported in the book are listed on the Web[2]. Furthermore, focused acknowledgments to co-authors and co-developers will be given in the different chapters of this book by references to underlying original papers based on joint work. On behalf of myself and SeCo, I would like express our thanks to the various supervisors and managers in our home universities, as well as the controllers and administrators who have made our life easier in the hectic bureaucratic academic world.

As important as the research group in our work has been our collaborators in memory organizations, four Finnish ministries, public sector organizations, fellow universities, and companies. They have not only provided the data and related domain expertise in our projects, but also funded it in many cases. For example, the national FinnONTO project 2003–2012 alone was supervised and funded by a consortium that finally included some 50 Finnish organizations and their representatives. A full class room was needed for the project board meetings. Again, it is not possible to list and thank every organization and representative here. Information about the project consortia is available at the project homepages to be referred to in the forthcoming chapters.

The most important funding organizations for our research have been the National Funding Agency for Technology and Innovation (Tekes), known today as Business Finland, the Academy of Finland (known today as the Research Council of Finland), and the European Commission. Several foundations have supported our research, too: The Finnish Culture Foundation, Memory Foundation for the Fallen, Jenny

[1] SeCo research group homepages: `https://seco.cs.aalto.fi/`

[2] SeCo people, alumni, and visitors: `https://seco.cs.aalto.fi/people/`

and Antti Wihuri Foundation, KAUTE Foundation, Foundation for Aalto University Science and Technology, and Nokia Foundation.

In addition to our home universities Aalto University and University of Helsinki, we have been happy to receive funding also from fellow universities, such as the University of Oxford, UK, University of Arts Helsinki, Finland, and Geneva Graduate Institute, Switzerland.

Special thanks to the Finnish Cultural Foundation for a special Eminentia Grant for writing this book.

This book has its foundations on my earlier book by Springer-Verlag "Publishing and Using Cultural Heritage Linked Data on the Semantic Web", published originally by Morgan & Claypool, 2012, Palo Alto, US, as the second volume in the "Synthesis Lectures on the Semantic Web" series. Thanks to professors Jim Hendler and Ying Ding for their invitation to contribute and to editor Michael Morgan for suggesting making a second edition of it. I am grateful to the current editors of the Synthesis series, professors Paul Groth and Ying Ding, for the opportunity to contribute again. Collaborations with the Springer Nature editors Susanne Filler, Melanie Rowan, and Prasanna Kumar Narayanasamy have been very smooth and productive.

Finally, I would like to thank my parents Kirsti and Matti for their constant support for my academic activities, my wife Suvi Kettula, with whom we managed to initiate the first major project of Seco in 2002 on *MuseumFinland – Finnish Museums on the Semantic Web*, and my children Hanna ja Jussi with apologies for spending so much time in the Academia.

Contents

1 Introduction 1
1.1 Cultural Heritage Areas 1
1.2 Digital Humanities and Humanism 2
1.3 Publishing and Using Cultural Heritage on the Web 3
1.4 Challenges of Publishing and Using Cultural Heritage Data 6
1.5 Outline of the Book 7
1.6 Bibliographical Remarks 7

Part I Vision: Towards a Web of Wisdom

2 Semantic Computing for Digital Humanities 11
2.1 The Idea of the Semantic Web 11
2.2 Why Semantic Web for Digital Humanities 13
2.2.1 Web of Linked Data 13
2.2.2 Alternative Approaches: Topic Maps and Property Graphs . . 17
2.3 Artificial Intelligence 18
2.3.1 Symbolic AI 18
2.3.2 Sub-symbolic AI 18

3 Paradigm Changes in Publishing Cultural Heritage Content 21
3.1 Introduction 21
3.2 Media Paradigms for Publishing CH Content 22
3.3 First Generation: Searching and Browsing for Close Reading 23
3.4 Second Generation: Portals with Tools for Distant Reading 25
3.5 Third Generation: Serendipitous Knowledge Discovery 26
3.6 Towards a Web of Wisdom 27

Part II Semantic Web Technologies

4 Linked Open Data: Principles and Data Model 33
4.1 Foundations of Linked Data 33
4.1.1 Linked Data Principles 33
4.1.2 Five-star Model for Linked Data 34
4.1.3 Extending the Five-star Model to Eight Stars 37
4.1.4 FAIR Data Principles 39
4.2 Representing Semantic Nets on the Web 40
4.2.1 Resource Description Framework RDF 41
4.2.2 RDF Schema 43

5 Publishing and Using Linked Data 47
5.1 Serialization Formats of RDF Graphs 47
5.1.1 RDF/XML 47
5.1.2 N-Triples 48
5.1.3 Turtle Notation 49
5.1.4 N-Quads and TriG for Named and Default Graphs 49
5.1.5 RDFa: Embedding RDF in Web Pages 50
5.1.6 JSON-LD format 51
5.2 Using HTTP URIs and IRIs 51
5.2.1 Content Negotiation and Redirection 53
5.2.2 Minting URIs 56
5.2.3 Linking Data Externally 57
5.3 Interfacing Linked Open Data 58
5.3.1 Linked Data Browsing 59
5.3.2 Querying Linked Data Using SPARQL 59
5.3.3 Download Facility 61
5.3.4 Linked Data Search Engines 62
5.4 User Interfaces for Linked Data Portals 62
5.4.1 Requirements for Semantic Portal User Interfaces 62
5.4.2 Search Paradigms for Semantic Portals 64
5.4.3 Analyzing and Visualizing Linked Data 69
5.5 Semantic Web Services and Mash-ups 69

6 Metadata Schemas 73
6.1 Metadata Types 73
6.2 Dublin Core and Application Profiles 75
6.3 Cataloging and Publishing Schemas 78
6.3.1 Categories for the Description of Works of Art (CDWA) . . . 79
6.3.2 Metadata Formats in Libraries 79
6.3.3 Metadata Formats in Archives 80
6.4 Harmonization Schemas 81
6.4.1 Europeana Metadata Models 82
6.4.2 CIDOC Conceptual Reference Model (CRM) 82
6.4.3 Linked Art 85
6.4.4 IFLA LRM and BIBFRAME 85

6.5 Harvesting Schemas and Protocols 88
6.6 Discussion: Object-, Event- and Process-centric Models 89

7 Knowledge Organization Systems 91
7.1 Approaches to Knowledge Organization Systems 91
7.1.1 Philosophy 92
7.1.2 Terminology 93
7.1.3 Lexicography and Linguistics 94
7.1.4 Information and Library Science 96
7.1.5 Computer Science 98
7.2 Semantic Web Ontology Languages 98
7.2.1 RDF Schema for Lightweight Ontologies 99
7.2.2 Simple Knowledge Organization System SKOS 99
7.2.3 Web Ontology Language OWL 100
7.3 Knowledge Organization System Types 102
7.3.1 Classifications, Thesauri, and Ontologies 102
7.3.2 Ontology Types by Major Domains 103
7.3.3 General Concept Ontologies 104
7.3.4 Actor Ontologies 105
7.3.5 Place Ontologies 109
7.3.6 Time Ontologies 112
7.3.7 Event Ontologies 114
7.3.8 Nomenclatures and Taxonomies 115

8 Reasoning Using Cultural Heritage Data 117
8.1 The Idea of Logic 117
8.2 Logical Interpretation of RDF(S) and OWL 118
8.3 Rules for Reasoning 120
8.3.1 Horn Logic vs. Description Logics 121
8.3.2 Closed World Assumption 122
8.3.3 Unique Name Assumption 123
8.4 Use Cases for Rules in Cultural Heritage 123
8.5 Large Language Models and Generative AI 125

Part III Semantic Web Infrastructure

9 Creating Cultural Content as Linked Data 131
9.1 Vocabulary and Ontology Creation 131
9.1.1 Conceptual Levels of Ontology Creation 132
9.1.2 Transforming Legacy Thesauri into Ontologies 133
9.1.3 Terminology Creation 138
9.1.4 Ontology Alignment 139
9.1.5 Ontology Evolution 140
9.2 Transforming Local Content into Linked Data 142
9.2.1 Transformation Process 142
9.2.2 Data Cleaning 144

9.2.3 Transforming Relational Data into RDF 144
9.3 Content Aggregation and Integration 147
9.3.1 Schema mapping 148
9.3.2 Identity resolution 148
9.3.3 Data fusion 149
9.4 Quality of Linked Data 149
9.4.1 Data Quality of Primary Sources 150
9.4.2 Metadata Quality 151
9.4.3 Maintaining Linked Open Data and Data Services 152

10 Building a Semantic Web Infrastructure 155
10.1 Extending the Layer Cake Model 155
10.2 Elements of a Semantic Web Infrastructure 156
10.3 Knowledge Organization System Infrastructure 157
10.3.1 FinnONTO Ontology Initiative for Ontology Creation 157
10.3.2 A Cloud of Linked Ontologies: KOKO 158
10.3.3 Ontology Library Services 160
10.3.4 Lessons Learned 163
10.4 Linked Open Data Services 165

Part IV Sampo Model and Sampo-UI Framework

11 Sampo Model 171
11.1 Introducing the Sampo 171
11.2 Sampo Model Principles 173
11.2.1 Support Collaborative Data Creation and Publishing 175
11.2.2 Use a Shared Open Ontology Infrastructure 176
11.2.3 Make Clear Distinction between the LOD Service and the User Interface (UI) 177
11.2.4 Provide Multiple Perspectives to the Same Data 178
11.2.5 Standardize Portal Usage by a Simple Filter-analyze Two-step Cycle 179
11.2.6 Support Data Analysis and Knowledge Discovery in Addition to Data Exploration 179
11.3 User Interfaces for Digital Humanities Research 180
11.4 LOD Maintenance 180

12 Sampo-UI Framework 183
12.1 Sampo-UI Interface Model 184
12.2 Creating a Portal Using Sampo-UI 185
12.3 Example Use Case: Nobel Prize Sampo 188
12.4 Extending Sampo-UI by the User Community 190
12.5 Discussion 192
12.6 Bibliographical Notes on Faceted Search 193

Part V Building Linked Data Services and Semantic Portals

13 Museums on the Semantic Web 197
13.1 Motivation: Why Museums on the Semantic Web? 198
13.2 Federated vs. Centralized Search 199
13.2.1 Distributed Strategy: Federated Search 200
13.2.2 Centralized Strategy: Aggregating Global Data 200
13.3 Data Model, KOS Ontologies, and Reasoning 201
13.4 Portal Design 202
13.5 Lessons Learned for Further Development 202

14 Cross-cultural Data on the Semantic Web 205
14.1 Why cross-cultural CultureSampo was needed? 205
14.2 Data Model and KOS Ontologies 206
14.3 Data Reconciliation 207
14.4 Portal Design 209
14.4.1 Maps Search and Browse Views 212
14.4.2 Relational Search 212
14.4.3 Collections 213
14.4.4 Finnish History 214
14.4.5 Skills and Cultural Narratives 214
14.4.6 Biographies 214
14.4.7 Semantic Kalevala 214
14.4.8 Karelian Culture 215
14.4.9 Using CultureSampo by Machines 215
14.5 Lessons Learned for Further Development 215

15 Literature on the Semantic Web 219
15.1 BookSampo Knowledge Graph and Data Service 220
15.2 Sampo-UI Interface for BookSampo Knowledge Graph 220
15.3 Utilizing BookSampo for Digital Humanities Research 223
15.4 Discussion 225

16 Medieval Manuscripts on the Semantic Web 227
16.1 Introduction 227
16.2 Modeling Manuscript Data 229
16.3 Using MMM Sampo 231
16.4 Discussion 234

17 Biographies on the Semantic Web 237
17.1 Introduction 237
17.2 Generations of Publishing Biographies 238
17.3 Norssi Alumni Biographical Dictionary 240
17.3.1 Extracting Structure from Text 241
17.3.2 Portal for Prosopograhical Research 243
17.4 U.S. Congress Prosopographer 244
17.4.1 Data Model and Linked Data Service 245
17.4.2 Using the Portal 245

17.5 BiographySampo 248
17.5.1 National Biography of Finland 248
17.5.2 Transformation into Linked Data 249
17.5.3 New Ways for Studying Biographies 250
17.5.4 Using BiographySampo for Historiographical Research . . . 253
17.6 AcademySampo 255
17.6.1 Primary Data and Knowledge Graph 256
17.6.2 Using AcademySampo Portal for Data Analysis 257
17.6.3 Using the SPARQL Endpoint 262
17.7 Discussion 263

18 Letters on the Semantic Web 265
18.1 Republic of Letters on the Semantic Web 265
18.1.1 LetterSampo Framework for Epistolary Data 267
18.1.2 LetterSampo – Historical Letters on the Semantic Web . . . 269
18.1.3 Discussion 275
18.2 LetterSampo Finland Data Service and Portal 275
18.2.1 Use Cases and Research Questions 276
18.2.2 Creating the Data Service 277
18.2.3 LetterSampo Finland Portal 279
18.2.4 Analyzing the Knowledge Graph for Better Data Literacy . . . 284
18.3 Bibliographical Notes 287

19 Opera Performances on the Semantic Web 291
19.1 Data and Ontologies 291
19.2 Using OperaSampo 292
19.3 Discussion 294

20 Art on the Semantic Web 297
20.1 Introduction 297
20.2 ArtSampo Knowledge Graph and User Interface 298
20.3 Using Generative AI to Enrich Metadata 299
20.4 Enriching the User Interface with AI-generated metadata 302
20.5 Discussion 304

21 Archaeology, Places, and Citizen Science on the Semantic Web 307
21.1 Reporting and Publishing Archaeological Finds 307
21.1.1 FindSampo Reporter 308
21.1.2 FindSampo Portal 309
21.1.3 Data and Ontologies 312
21.2 CoinSampo – Finnish Numismatic Public Finds 2013–2023 312
21.2.1 CoinSampo Data Service and Portal 313
21.2.2 Application to British Museum and Nomisma.org Data 315
21.2.3 Discussion 316
21.3 NameSampo – Placenames for Toponomastic Research 316
21.3.1 Research Questions 317

21.3.2 Data Sources . . . 317
21.3.3 Harmonizing Data Model for Places . . . 320
21.3.4 An Ontology for Classification of Geographic Features . . . 321
21.3.5 Using NameSampo for Toponomastic Research . . . 322
21.4 Discussion . . . 324

22 Military History on the Semantic Web . . . 327
22.1 Introduction . . . 327
22.2 LOD Services and Semantic Portals for Military History . . . 329
22.2.1 WarSampo – Finnish WW2 on the Semantic Web . . . 329
22.2.2 WarVictimSampo 1914–1922 – Casualties of Finnish Wars 1914–1922 . . . 330
22.2.3 WarMemoirSampo – Memoir Interviews of WW2 Veterans . 333
22.3 Discussion and Related Works . . . 334

23 Academic Cultural Intangible Heritage on the Semantic Web . . . 339
23.1 Publishing and Studying Intangible Cultural Heritage . . . 339
23.2 Representing Intangible Cultural Heritage Events . . . 340
23.3 Conferment Ceremonies at the University of Helsinki . . . 341
23.4 Creating the Knowledge Graph . . . 342
23.4.1 Data Models for Conferments and People . . . 342
23.4.2 Data Transformation, Enrichment, and Linking . . . 345
23.5 Using the Portal . . . 345
23.6 Conclusions . . . 348

24 Legislation and Case Law on the Semantic Web . . . 349
24.1 Introduction . . . 349
24.2 Finnish Legislation and Case Law on the Semantic Web . . . 350
24.2.1 LawSampo Linked Open Data . . . 350
24.2.2 Using LawSampo Portal . . . 354
24.2.3 Using LawSampo Data Service for Data Analysis . . . 356
24.2.4 Using Legislative Change History of Statutes . . . 357
24.2.5 Discussion . . . 357
24.2.6 Related Works . . . 358
24.3 Transnational Legislation on the Semantic Web . . . 359
24.3.1 Introduction . . . 359
24.3.2 Creating the Knowledge Graph and LOD Service . . . 360
24.3.3 Using the FINEST-LawSampo Portal . . . 362
24.3.4 Comparison with N-Lex . . . 364
24.3.5 Discussion . . . 365

25 Parliaments and Multilateralism on the Semantic Web . . . 367
25.1 Parliament of Finland on the Semantic Web . . . 367
25.1.1 Data Models and Knowledge Graphs . . . 369
25.1.2 Data Transformation . . . 370
25.1.3 Use Cases of ParliamentSampo Data . . . 372

25.1.4 ParliamentSampo Portal in Use 376
25.1.5 Bibliographical Notes 378
25.2 Minutes of Multilateralism on the Semantic Web 380
25.2.1 League of Nations Sampo 381
25.2.2 Using League of Nations Sampo for DH Research 383
25.2.3 Generalizing the Portal for Cross-institutional Research 384

26 Connecting Everything to Everything Else 387
26.1 Enriching and Harmonizing Data by Linking 387
26.2 SampoSampo Data Service 388
26.3 SampoSampo Portal 389
26.4 Use Cases of SampoSampo 389
26.4.1 Enriching Data for New Applications 391
26.4.2 Providing Global Functionalities 391
26.4.3 Checking Correctness of Data between Datasets 391
26.4.4 Knowledge Discovery of Connections with Explanations . . . 393
26.5 Enhancing Data Literacy of a Linked Data Cloud 395
26.6 Discussion 396

27 Conclusions 399

References 454

Index 455