1. Record Nr. UNINA9910823184603321

Autore	Berman Jules J
Titolo	Principles of big data : preparing, sharing, and analyzing complex information / / Jules J. Berman, Ph. D., M.D
Pubbl/distr/stampa	Amsterdam, Netherlands, : Elsevier, c2013 Waltham, MA : , : Morgan Kaufmann, , 2013
ISBN	0-12-404724-6
Edizione	[1st edition]
Descrizione fisica	1 online resource (xxvi, 261 pages) : illustrations
Collana	Gale eBooks
Disciplina	005.74
Soggetti	Big data Database management
Lingua di pubblicazione	Inglese
Formato	Materiale a stampa
Livello bibliografico	Monografia
Note generali	Description based upon print version of record.
Nota di bibliografia	Includes bibliographical references and index.
Nota di contenuto	Front Cover; Principles of Big Data: Preparing,Sharing,and Analyzing Complex Information; Copyright; Dedication; Contents; Acknowledgments; Author Biography; Preface; Introduction; Definition of Big Data; Big Data Versus Small Data; Whence Comest Big Data?; The Most Common Purpose of Big Data is to Produce Small Data; Opportunities; Big Data Moves to the Center of the Information Universe; Chapter 1: Providing Structure to Unstructured Data; Background; Machine Translation; Autocoding; Indexing; Term Extraction; Chapter 2: Identification, Deidentification, and Reidentification; Background Features of an Identifier System Registered Unique Object Identifiers; Really Bad Identifier Methods; Embedding Information in an Identifier: Not Recommended; One-Way Hashes; Use Case: Hospital Registration; Deidentification; Data Scrubbing; Reidentification; Lessons Learned; Chapter 3: Ontologies and Semantics; Background; Classifications, the Simplest of Ontologies; Ontologies, Classes with Multiple Parents; Choosing a Class Model; Introduction to Resource Description Framework Schema; Common Pitfalls in Ontology Development; Chapter 4: Introspection; Background; Knowledge of Self eXtensible Markup Language Introduction to Meaning; Namespaces and the Aggregation of Meaningful Assertions; Resource Description Framework Triples; Reflection; Use Case: Trusted Time Stamp;

Summary; Chapter 5: Data Integration and Software Interoperability; Background; The Committee to Survey Standards; Standard Trajectory; Specifications and Standards; Versioning; Compliance Issues; Interfaces to Big Data Resources; Chapter 6: Immutability and Immortality; Background; Immutability and Identifiers; Data Objects; Legacy Data; Data Born from Data; Reconciling Identifiers across Institutions
Zero-Knowledge Reconciliation The Curator s Burden; Chapter 7: Measurement; Background; Counting; Gene Counting; Dealing with Negations; Understanding Your Control; Practical Significance of Measurements; Obsessive-Compulsive Disorder: The Mark of a Great Data Manager; Chapter 8: Simple but Powerful Big Data Techniques; Background; Look At the Data; Data Range; Denominator; Frequency Distributions; Mean and Standard Deviation; Estimation-Only Analyses; Use Case: Watching Data Trends with Google Ngrams; Use Case: Estimating Movie Preferences; Chapter 9: Analysis; Background; Analytic Tasks
Clustering, Classifying, Recommending, and Modeling Clustering Algorithms; Classifier Algorithms; Recommender Algorithms; Modeling Algorithms; Data Reduction; Normalizing and Adjusting Data; Big Data Software: Speed and Scalability; Find Relationships, Not Similarities; Chapter 10: Special Considerations in Big Data Analysis; Background; Theory in Search of Data; Data in Search of a Theory; Overfitting; Bigness Bias; Too Much Data; Fixing Data; Data Subsets in Big Data: Neither Additive nor Transitive; Additional Big Data Pitfalls; Chapter 11: Stepwise Approach to Big Data Analysis; Background
Step 1. A Question Is Formulated

Sommario/riassunto

Principles of Big Data helps readers avoid the common mistakes that endanger all Big Data projects. By stressing simple, fundamental concepts, this book teaches readers how to organize large volumes of complex data, and how to achieve data permanence when the content of the data is constantly changing. General methods for data verification and validation, as specifically applied to Big Data resources, are stressed throughout the book. The book demonstrates how adept analysts can find relationships among data objects held in disparate Big Data resources, when the data objects are