
References

1. ABITEBOUL, S., BUNEMAN, P., AND SUCIU, D. *Data on the Web: From Relations to Semistructured Data and XML*. Morgan Kaufmann Publishers, 2000.
2. ADAM, N. R., AND WORTMANN, J. C. Security Control Methods for Statistical Databases: A Comparative Study. *ACM Computing Surveys* 21, 4 (1989), 515–556.
3. AGRAWAL, R., GUPTA, A., AND SARAWAGI, S. Modeling Multidimensional Databases. In *Proc. ICDE 2000* (Birmingham, UK, 1997).
4. AIMETTI, P., MISSIER, P., SCANNAPIECO, M., BERTOLETTI, M., AND BATINI, C. Improving Government-to-Business Relationships through Data Reconciliation and Process Re-engineering. In *Advances in Management Information Systems - Information Quality (AMIS-IQ) Monograph*, R. Y. Wang, E. M. Pierce, S. E. Madnick, and C. W. Fisher, Eds. Sharpe, M.E., April 2005.
5. AL-LAWATI, A., LEE, D., AND MCDANIEL, P. Blocking-aware Private Record Linkage. In *Proc. IQIS 2005 (SIGMOD Workshop)* (Baltimore, MC, 2005).
6. AMAT, G., AND LABOISSE, B. Une Gestion Operationnelle de la Qualite Donnees. In *Proc. 1st Data and Knowledge Quality Workshop (in conjunction with ECG)* (18th January 2005, Paris, France).
7. ANANTHAKRISHNA, R., CHAUDHURI, S., AND GANTI, V. Eliminating Fuzzy Duplicates in Data Warehouses. In *Proc. VLDB 2002* (Hong Kong, China, 2002).
8. ARDAGNA, D., CAPIELLO, C., COMUZZI, M., FRANCALANCI, C., AND PERNICI, B. A Broker for Selecting and Provisioning High Quality Syndicated Data. In *Proc. 10th International Conference on Information Quality (IQ 2005)*.
9. ARENAS, M., BERTOSSI, L. E., AND CHOMICKI, J. Consistent Query Answers in Inconsistent Databases. In *Proc. PODS'99*.
10. ATALLAH, M. J., KERSCHBAUM, F., AND DU, W. Secure and Private Sequence Comparisons. In *Proc. ACM Workshop on Privacy in the Electronic Society (WPES 2003)* (Washington, Washington DC, 2003).
11. ATZENI, P., AND DE ANTONELLIS, V. *Relational Database Theory*. The Benjamin/Cummings Publishing Company, 1993.
12. ATZENI, P., MERIALDO, P., AND SINDONI, G. Web Site Evaluation: Methodology and Case Study. In *Proc. International Workshop on Data Semantics in Web Information Systems (DASWIS 2001)* (Yokohama, Japan, 2001).

13. AVENALI, A., BERTOLAZZI, P., BATINI, C., AND MISSIER, P. A Formulation of the Data Quality Optimization Problem in Cooperative Information Systems. In *Proc. CAISE International Workshop on Data and Information Quality* (Riga, Latvia, 2004).
14. AYRES, I., AND NALEBUFF, B. Going Soft on Microsoft? The EU's Antitrust Case and Remedy. *The Economists' Voice*.
15. BALLOU, D. P., AND PAZER, H. L. Modeling Completeness versus Consistency Tradeoffs in Information Decision Contexts. *IEEE Transactions on Knowledge Data Engineering* 15, 1 (2003), 240–243.
16. BALLOU, D. P., AND TAYI, G. K. Enhancing Data Quality in Data Warehouse Environments. *Communications of the ACM* 42, 1 (1999).
17. BALLOU, D. P., WANG, R. Y., PAZER, H., AND TAYI, G. K. Modeling Information Manufacturing Systems to Determine Information Product Quality. *Management Science* 44, 4 (1998).
18. BARATEIRO, J., AND GALHARDAS, H. A Survey of Data Quality Tools. *Datenbank Spectrum* 14 (2005), 15–21.
19. BASEL COMMITTEE ON BANKING SUPERVISION. <http://www.ots.treas.gov>.
20. BATINI, C., CERI, S., AND NAVATHE, S. B. *Conceptual Data Base Design: An Entity Relationship Approach*. Benjamin and Cummings, 1992.
21. BATINI, C., AND MECELLA, M. Enabling Italian e-Government Through a Cooperative Architecture. *IEEE Computer* 34, 2 (2001).
22. BATINI, C., NARDELLI, E., AND TAMASSIA, R. A Layout Algorithm for Data Flow Diagrams. *IEEE Transactions on Software Engineering* (April 1986).
23. BELIN, T. R., AND RUBIN, D. B. A Method for Calibrating False Matches Rates in Record Linkage. *Journal of American Statistical Association* 90 (1995), 694–707.
24. BERTI-ÉQUILLE, L. Quality-Adaptive Query Processing over Distributed Sources. In *Proc. 9th International Conference on Information Quality (IQ 2004)*.
25. BERTI-ÉQUILLE, L. Integration of Biological Data and Quality-driven Source Negotiation. In *Proc. ER 2001* (Yokohama, Japan, 2001).
26. BERTOLAZZI, P., DE SANTIS, L., AND SCANNAPIECO, M. Automatic Record Matching in Cooperative Information Systems. In *Proc. DQCIS 2003 (ICDT Workshop)* (Siena, Italy, 2003).
27. BILKE, A., BLEIHOLDER, J., BÖHM, C., DRABA, K., NAUMANN, F., AND WEIS, M. Automatic Data Fusion with HumMer. In *Proc. VLDB 2005 Demonstration Program* (Trondheim, Norway, 2005).
28. BITTON, D., AND DEWITT, D. J. Duplicate Record Elimination in Large Data Files. *ACM Transactions on Databases Systems* 8, 2 (1983).
29. BOAG, A., CHAMBERLIN, D., FERNANDEZ, M. F., FLORESCU, D., ROBIE, J., AND SIMÈON, J. XQuery 1.0: An XML Query Language. W3C Working Draft. Available from <http://www.w3.org/TR/xquery>, November 2003.
30. BOUZEGHOUB, M., AND PERALTA, V. A Framework for Analysis of Data Freshness. In *Proc. IQIS 2004 (SIGMOD Workshop)* (Paris, France, 2004).
31. BOVEE, M., SRIVASTAVA, R. P., AND MAK, B. R. A Conceptual Framework and Belief-Function Approach to Assessing Overall Information Quality. In *Proc. 6th International Conference on Information Quality (IQ 2001)*.
32. BRAVO, L., AND BERTOSI, L. E. Logic Programming for Consistently Querying Data Integration Systems. In *Proc. IJCAI 2003*.

33. BRUNI, R., AND SASSANO, A. Errors Detection and Correction in Large Scale Data Collecting. In *Proc. 4th International Conference on Advances in Intelligent Data Analysis* (Cascais, Portugal, 2001).
34. BUECHI, M., BORTHWICK, A., WINKEL, A., AND GOLDBERG, A. ClueMaker: a Language for Approximate Record Matching. In *Proc. 8th International Conference on Information Quality (IQ 2003)*.
35. BUNEMAN, P. Semistructured Data. In *Proc. PODS 1997*.
36. BUNEMAN, P., KHANNA, S., AND TAN, W. C. Why and Where: A Characterization of Data Provenance. In *Proc. International Conference on Database Theory (ICDT 2001)* (London, UK, 2001).
37. CALI, A., CALVANESE, D., DE GIACOMO, G., AND LENZERINI, M. On the Role of Integrity Constraints in Data Integration. *IEEE Data Eng. Bull.* 25, 3 (2002), 39–45.
38. CALÌ, A., LEMBO, D., AND ROSATI, R. On the Decidability and Complexity of Query Answering over Inconsistent and Incomplete Databases. In *Proc. PODS 2003*.
39. CALÌ, A., LEMBO, D., AND ROSATI, R. Query Rewriting and Answering under Constraints in Data Integration Systems. In *Proc. IJCAI 2003*.
40. CALVANESE, D., DE GIACOMO, G., AND LENZERINI, M. Modeling and Querying Semi-Structured Data. *Networking and Information Systems Journal* 2, 2 (1999), 253–273.
41. CAPPIELLO, C., FRANCALANCI, C., AND PERNICI, B. Preserving Web Sites: a Data Quality Approach. In *Proc. 8th International Conference on Information Quality (IQ 2003)*.
42. CAPPIELLO, C., FRANCALANCI, C., PERNICI, B., PLEBANI, P., AND SCANAPIECO, M. Data Quality Assurance in Cooperative Information Systems: a Multi-dimension Certificate. In *Proc. DQCIS 2003 (ICDT Workshop)* (Siena, Italy, 2003).
43. CARUSO, F., COCHINWALA, M., GANAPATHY, U., LALK, G., AND MISSIER, P. Telcordia's Database Reconciliation and Data Quality Analysis Tool. In *Demonstration at VLDB 2000*.
44. CHARNES, A., COOPER, W. W., AND RHODES, E. Measuring the Efficiency of Decision Making Units. *European Journal of Operational Research* 2 (1978).
45. CHAUDHURI, S., GANTI, V., AND MOTWANI, R. Robust Identification of Fuzzy Duplicates. In *Proc. ICDE 2005*.
46. CHEN, Z., KALASHNIKOV, D. V., AND MEHROTRA, S. Exploiting Relationships for Object Consolidation. In *Proc. IQIS 2005 (SIGMOD Workshop)*.
47. CHITICARIU, L., TAN, W., AND VIJAYVARGIYA, G. An Annotation Management System for Relational Databases. In *Proc. VLDB 2004* (Toronto, Canada, 2004).
48. CHURCHES, T., AND CHRISTEN, P. Some Methods for Blindfolded Record Linkage. *BMC Medical Informatics and Decision Making* 4, 9 (2004).
49. CUI, Y., WIDOM, J., AND WIENER, J. L. Tracing the Lineage of View Data in a Warehousing Environment. *ACM Transactions on Database Systems* 25, 2 (2000), 179–227.
50. DASU, T., AND JOHNSON, T. *Exploratory Data Mining and Data Cleaning*. J. Wiley Series in Probability and Statistics, 2003.
51. DASU, T., JOHNSON, T., MUTHUKRISHNAN, S., AND SHKAPENYUK, V. Mining Database Structure or, How to Build a Data Quality Browser. In *Proc. SIGMOD 2002* (Madison, WI, 2002).

52. DATA WAREHOUSING INSTITUTE. Data Quality and the Bottom Line: Achieving Business Success through a Commitment to High Quality Data. <http://www.dw-institute.com/>.
53. DAVIS, G. B., AND OLSON, M. H. *Management Information Systems: Conceptual Foundations, Structure, and Development (2nd ed.)*. McGraw-Hill, 1984.
54. DAVIS, R., STROBE, H., AND SZOLOVITS, P. What is Knowledge Representation. *AI Magazine* 14, 1 (1993), 17–33.
55. DAYAL, U. Query Processing in a Multidatabase System. In *Query Processing in Database Systems*. Springer, 1985, pp. 81–108.
56. DE AMICIS, F., AND BATINI, C. A Methodology for Data Quality Assessment on Financial Data. *Studies in Communication Sciences* (2004).
57. DE GIACOMO, G., LEMBO, D., LENZERINI, M., AND ROSATI, R. Tackling Inconsistencies in Data Integration through Source Preferences. In *Proc. IQIS 2004 (SIGMOD Workshop)* (Paris, France, 2004).
58. DE MICHELIS, G., DUBOIS, E., JARKE, M., MATTHES, F., MYLOPOULOS, J., PAPAZOGLU, M. P., SCHMIDT, J., WOO, C., AND YU, E. Cooperative Information Systems: A Manifesto. In *Cooperative Information Systems: Trends & Directions*, M. Papazoglou and G. Schlageter, Eds. Academic-Press, 1997.
59. DE SANTIS, L., SCANNAPIECO, M., AND CATARCI, T. Trusting Data Quality in Cooperative Information Systems. In *Proc. 11th International Conference on Cooperative Information Systems (CoopIS 2003)* (Catania, Italy, 2003).
60. DEDEKE, A. Building Quality into Information Supply Chain. In *Advances in Management Information Systems - Information Quality (AMIS-IQ) Monograph*, R. Y. Wang, E. M. Pierce, S. E. Madnick, and C. W. Fisher, Eds. Sharpe, M.E., April 2005.
61. DEMPSTER, A., LAIRD, N., AND RUBIN, D. Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society Series B* 39, 1 (1977).
62. DONG, X., HALEVY, A. Y., AND MADHAVAN, J. Reference Reconciliation in Complex Information Spaces. In *Proc. SIGMOD 2005*.
63. DUBLIN CORE. <http://dublincore.org/>.
64. DUNN, H. L. Record Linkage. *American Journal of Public Health* 36 (1946), 1412–1416.
65. ELFEKY, M. G., VERYKIOS, V. S., AND ELMAGARMID, A. K. Tailor: A Record Linkage Toolbox. In *Proc. 18th International Conference on Data Engineering* (San Jose, CA, 2002).
66. ELMASRI, R., AND NAVATHE, S. *Foundamentals of Database Systems (5th ed.)*. Addison-Wesley Publishing Company, 1994.
67. ENGLISH, L. Process Management and Information Quality: How Improving Information Production Processes Improves Information (Product) Quality. In *Proc. 7th International Conference on Information Quality (IQ 2002)*.
68. ENGLISH, L. P. *Improving Data Warehouse and Business Information Quality*. Wiley & Sons, 1999.
69. EPPLER, M., AND MUENZENMAIER, P. Measuring Information Quality in the Web Context: A Survey of State-of-the-Art Instruments and an Application Methodology. In *Proc. 7th International Conference on Information Quality (IQ 2002)*.
70. EPPLER, M. J., AND HELFERT, M. A Classification and Analysis of Data Quality Costs. In *Proc. 9th International Conference on Information Quality (IQ 2004)*.

71. EUROPEAN PARLIAMENT. Directive 2003/98/EC of the European Parliament and of the Council of 17 November 2003 on the Re-use of Public Sector Information. Official Journal of the European Union, 2003.
72. EUROSTAT. <http://epp.eurostat.cec.eu.int/pls/portal/>.
73. FALORSI, P. D., PALLARA, S., PAVONE, A., ALESSANDRONI, A., MASSELLA, E., AND SCANNAPIECO, M. Improving the Quality of Toponymic Data in the Italian Public Administration. In *Proc. DQCIS 2003 (ICDT Workshop)* (Siena, Italy, 2003).
74. FALORSI, P. D., AND SCANNAPIECO, M., Eds. *Principi Guida per la Qualità dei Dati Toponomastici nella Pubblica Amministrazione (in Italian)*. ISTAT, serie Contributi, vol. 12. Available at: http://www.istat.it/dati/pubbsci/contributi/Contr_anno2005.htm, 2006.
75. FAN, W., LU, H., MADNICK, S., AND CHEUNG, D. Discovering and Reconciling Value Conflicts for Numerical Data Integration. *Information Systems* 26, 8 (2001).
76. FELLEGI, I. P., AND HOLT, D. A Systematic Approach to Automatic Edit and Imputation. *Journal of the American Statistical Association* 71, 353 (1976), 17–35.
77. FELLEGI, I. P., AND SUNTER, A. B. A Theory for Record Linkage. *Journal of the American Statistical Association* 64 (1969).
78. FISHER, C. W., AND KINGMA, B. R. Criticality of Data Quality as Exemplified in Two Disasters. *Information Management* 39 (2001).
79. FOWLER, M. *UML Distilled: A Brief Guide to the Standard Object Modeling Language*. Pearson Education, 2004.
80. FRATERNALI, P., LANZI, P. L., MATERA, M., AND MAURINO, A. Model-Driven Web Usage Analysis for the Evaluation of Web Application Quality. *Journal of Web Engineering* 3, 2 (2004), 124–152.
81. FUXMAN, A., FAZLI, E., AND MILLER, R. J. ConQuer: Efficient Management of Inconsistent Databases. In *Proc. SIGMOD 2005* (Baltimore, MA, 2005).
82. GALHARDAS, H., FLORESCU, D., SHASHA, D., SIMON, E., AND SAITA, C. A. Declarative Data Cleaning: Language, Model, and Algorithms. In *Proc. VLDB 2001* (Rome, Italy, 2001).
83. GLOVER, F., AND LAGUNA, M. *Tabu Search*. Kluwer Academic Publishers, 1997.
84. GOERK, M. SAP AG Data Quality@SAP: An Enterprise Wide Approach to Data Quality Goals. In *CAiSE Workshop on Data and Information Quality (DIQ 2004)* (Riga, Latvia, 2004).
85. GRAY, J. *The Benchmark Handbook for Database and Transaction Systems*. Morgan Kaufmann, 1993.
86. GRECO, G., GRECO, S., AND ZUMPARO, E. A Logical Framework for Querying and Repairing Inconsistent Databases. *Transactions on Knowledge and Data Engineering* 15, 6 (2003), 1389–1408.
87. GRECO, G., AND LEMBO, D. Data Integration with Preferences Among Sources. In *Proc. ER 2004* (Shanghai, China, 2004).
88. GU, L., BAXTER, R., VICKERS, D., AND RAINSFORD, C. P. Record Linkage: Current Practice and Future Directions. Technical Report 03/83, CMIS 03/83, Canberra, Australia.
89. GUPTIL, C., AND MORRISON, J. *Elements of Spatial Data Quality*. Elsevier Science Ltd, Oxford, UK, 1995.

90. HALL, P. A., AND DOWLING, G. Approximate String Comparison. *ACM Computing Surveys* 12, 4 (1980), 381–402.
91. HAMMER, M., AND CHAMPY, J. *Reengineering the Corporation: a Manifesto for Business Revolution*. 2001.
92. HAN, J., AND KAMBER, M. *Data Mining: Concepts and Techniques*. Morgan Kaufmann Publishers, 2000.
93. HERNANDEZ, M. A., AND STOLFO, S. J. Real-world Data is Dirty: Data Cleansing and The Merge/Purge Problem. *Journal of Data Mining and Knowledge Discovery* 1, 2 (1998).
94. INFOIMPACT. <http://www.infoimpact.com/iqproducts.cfm>.
95. INTERNATIONAL CONFERENCE ON INFORMATION QUALITY (IQ/ICIQ). <http://www.iqconference.org/>.
96. INTERNATIONAL MONETARY FUND. <http://dsbb.imf.org/>.
97. INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. <http://www.iso.org>.
98. INTERNATIONAL WORKSHOP ON DATA AND INFORMATION QUALITY (DIQ). <http://www.computing.dcu.ie/research/dataquality/diq/>.
99. INTERNATIONAL WORKSHOP ON INFORMATION QUALITY IN INFORMATION SYSTEMS (IQIS). <http://iqis.irisa.fr/>.
100. INTERNATIONAL WORKSHOP ON QUALITY OF INFORMATION SYSTEMS (QOIS). <http://deptinfo.cnam.fr/qois2006/>.
101. INTERPARES PROJECT. <http://www.interpares.org>.
102. ISAKOWITZ, T., STOHR, E. A., AND BALASUBRAMANIAN, P. RMM: a Methodology for Structured Hypermedia Design. *Communication of the ACM* 58, 8 (1995).
103. JARKE, M., JEUSFELD, M. A., QUIX, C., AND VASSILIADIS, P. Architecture and Quality in Data Warehouses: an Extended Repository Approach. *Information Systems* (1999).
104. JARKE, M., LENZERINI, M., VASSILIOU, Y., AND VASSILIADIS, P., Eds. *Fundamentals of Data Warehouses*. Springer Verlag, 1995.
105. JARO, M. A. Advances in Record Linkage Methodologies as Applied to Matching the 1985 Census of Tampa, Florida. *Journal of American Statistical Society* 84, 406 (1985), 414–420.
106. JOSANG, A., ISMAIL, R., AND BOYD, C. A Survey of Trust and Reputation Systems for Online Service Provision. *Decision Support Systems* (2005).
107. KAHN, B. K., STRONG, D. M., AND WANG, R. Y. Information Quality Benchmarks: Product and Service Performance. *Communications of the ACM* 45 (2002).
108. KATIFORI, V., POGGI, A., SCANNAPIECO, M., CATARCI, T., AND IOANNIDIS, Y. OntoPIM: how to rely on a Personal Ontology for Personal Information Management. In *Proc. Workshop on The Semantic Desktop 2005*.
109. KIM, W., AND SEO, J. Classifying Schematic and Data Heterogeneity in Multidatabase Systems. *IEEE Computer* 24, 12 (1991), 12–18.
110. KISSNER, L., AND SONG, D. Private and Threshold Set-Intersection. Tech. Rep. CMU-CS-05-113, Carnegie Mellon University, February 2005.
111. KRAWCZYK, H., AND WISZNIEWSKI, B. Visual GQM Approach to Quality-driven Development of Electronic Documents. In *Proc. 2nd International Workshop on Web Document Analysis (WDA2003)* (Edinburgh, UK, 2003).
112. LAKSHMANAN, L. V., NG, R. T., AND RAMESH, G. To Do or Not to Do: the Dilemma of Disclosing Anonymized Data. In *Proc. SIGMOD 2005*.

113. LARSEN, M. D., AND RUBIN, D. B. An Iterative Automated Record Matching using Mixture Models. *Journal of American Statistical Association* 79 (1989), 32–41.
114. LEE, Y. W., STRONG, D. M., KAHN, B. K., AND WANG, R. Y. AIMQ: A Methodology for Information Quality Assessment. *Information and Management* (2001).
115. LEHTI, P., AND FANKHAUSER, P. Probabilistic Iterative Duplicate Detection. In *Proc. OTM Conferences 2005*.
116. LENZERINI, M. Data Integration: A Theoretical Perspective. In *Proc. PODS 2002* (Madison, WI, 2002).
117. LEVY, A. Y., MENDELZON, A. O., SAGIV, Y., AND SRIVASTAVA, D. Answering Queries Using Views. In *Proc. PODS 1995* (San Jose, CA, 1995).
118. LIM, E. P., AND CHIANG, R. H. A Global Object Model for Accommodating Instance Heterogeneities. In *Proc. ER'98* (Singapore, Singapore, 1998).
119. LIN, J., AND MENDELZON, A. O. Merging Databases Under Constraints. *International Journal of Cooperative Information Systems* 7, 1 (1998), 55–76.
120. LIU, L., AND CHI, L. Evolutionary Data Quality. In *Proc. 7th International Conference on Information Quality (IQ 2002)*.
121. LOHNINGEN, H. *Teach Me Data Analysis*. Springer-Verlag, 1999.
122. LONG, J. A., AND SEKO, C. E. A Cyclic-Hierarchical Method for Database Data-Quality Evaluation and Improvement. In *Advances in Management Information Systems - Information Quality (AMIS-IQ) Monograph*, R. Y. Wang, E. M. Pierce, S. E. Madnick, and C. W. Fisher, Eds. Sharpe, M.E., April 2005.
123. LOSHIN, D. *Enterprise Knowledge Management - The Data Quality Approach*. Morgan Kaufmann Series in Data Management Systems, 2004.
124. LOW, W., LEE, M., AND LING, T. A Knowledge-based Approach for Duplicate Elimination in Data Cleaning. *Information Systems* 26, 8 (2001).
125. LYMAN, P., AND VARIAN, H. R. How Much Information.
126. MARTINEZ, A., AND HAMMER, J. Making Quality Count in Biological Data Sources. In *Proc. IQIS 2005 (SIGMOD Workshop)*.
127. MCKEON, A. Barclays Bank Case Study: Using Artificial Intelligence to Benchmark Organizational Data Flow Quality. In *Proc. 8th International Conference on Information Quality (IQ 2003)*.
128. MECCA, G., Merialdo, P., ATZENI, P., AND CRESCENZI, V. The (Short) ARAENEUS Guide to Web-Site Development. In *Proc. 2nd International Workshop on the Web and Databases (WebDB'99)* (1999).
129. MIKLAU, G., AND SUCIU, D. A Formal Analysis of Information Disclosure in Data Exchange. In *Proc. SIGMOD 2004*.
130. MISSIER, P., AND BATINI, C. A Multidimensional Model for Information Quality in Cooperative Information Systems. In *Proc. 8th International Conference on Information Quality (IQ 2003)*.
131. MISSIER, P., AND BATINI, C. A Model for Information Quality Management Framework for Cooperative Information Systems. In *Proc. 11th Italian Symposium on Advanced Database Systems (SEDB 2003)* (Cetraro, Italy, June 2003).
132. MISSIER, P., AND BATINI, C. An Information Quality Management Framework for Cooperative Information Systems. In *Proc. International Conference on Information Systems and Engineering (ISE 2003)* (Montreal, Canada, July 2003).

133. MISSIER, P., LACK, G., VERYKIOS, V., GRILLO, F., LORUSSO, T., AND ANGELETTI, P. Improving Data Quality in Practice: a Case Study in the Italian Public Administration. *Parallel and Distributed Databases 13*, 2 (2003), 135–160.
134. MONGE, A., AND ELKAN, C. An Efficient Domain Independent Algorithm for Detecting Approximate Duplicate Database Records. In *Proc. SIGMOD Workshop on Research Issues on Data Mining and Knowledge Discovery (DMKD'97)* (Tucson, AZ, 1997).
135. MOTRO, A., AND ANOKHIN, P. Fusionplex: Resolution of Data Inconsistencies in the Data Integration of Heterogeneous Information Sources. *Information Fusion* (2005).
136. MOTRO, A., AND RAGOV, I. Estimating Quality of Databases. In *Proc. 3rd International Conference on Flexible Query Answering Systems (FQAS'98)* (Roskilde, Denmark, 1998).
137. MUTHU, S., WITHMAN, L., AND CHERAGHI, S. Business Process Re-engineering: a Consolidated Methodology. In *Proc. 4th Annual International Conference on Industrial Engineering Theory, Applications and Practice* (San Antonio, TX, 1999).
138. NALEBUFF, B. Competing Against bundles. P. Hammond and G. Myles, Eds., Oxford University Press.
139. NAUMANN, F. *Quality-Driven Query Answering for Integrated Information Systems*. Springer Verlag, LNCS 2261, 2002.
140. NAUMANN, F., FREYTAG, J. C., AND LESER, U. Completeness of Integrated Information Sources. *Information Systems 29*, 7 (2004), 583–615.
141. NAUMANN, F., AND HÄUSSLER, M. Declarative Data Merging with Conflict Resolution. In *Proc. 7th International Conference on Information Quality (IQ 2002)*.
142. NAUMANN, F., LESER, U., AND FREYTAG, J. C. Quality-driven Integration of Heterogeneous Information Systems. In *Proc. VLDB'99* (Edinburgh, UK, 1999).
143. NAVARRO, G. A Guided Tour of Approximate String Matching. *ACM Computing Surveys 31* (2001), 31–88.
144. NEBEL, B., AND LAKEMEYER, G., Eds. *Foundations of Knowledge Representation and Reasoning*, lecture notes in artificial intelligence ed., vol. 810. Springer-Verlag, 1994.
145. NEILING, M., JURK, S., LENZ, H. J., AND NAUMANN, F. Object Identification Quality. In *Proc. DQCIS 2003 (ICDT Workshop)* (Siena, Italy, 2003).
146. NEWCOMBE, H. B., KENNEDY, J. M., AXFORD, S. J., AND JAMES, A. P. F. Automatic Linkage of Vital Records. *Science 130* (1959).
147. NIGAM, K., MCCALLUM, A., THRUN, S., AND MITCHELL, T. Text Classification from Labeled and Unlabeled Documents using EM. *Machine Learning 39* (2000), 103–134.
148. OBJECT MANAGEMENT GROUP (OMG). Unified Modeling Language Specification, Version 1.5. OMG Document formal/03-03-01, 2003.
149. OFFICE OF MANAGEMENT AND BUDGET. Information Quality Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility, and Integrity of Information Disseminated by Agencies. <http://www.whitehouse.gov/omb/fedreg/reproducible.html>.
150. OMG. Data Quality and the Bottom Line: Achieving Business Success through a Commitment to High Quality Data. <http://www.uml.org/>.

151. ORACLE. <http://www.oracle.com/solutions/business-intelligence>.
152. OSTMAN, A. The Specifications and Evaluation of Spatial Data Quality. In *Proc. 18th ICA/ACI International Conference* (Stockholm, Sweden, 1997).
153. OZSU, T., AND VALDURIEZ, P. *Principles of Distributed Database Systems*. Prentice Hall, 2000.
154. PAPAKONSTANTINOY, Y., ABITEBOUL, S., AND GARCIA-MOLINA, H. Object Fusion in Mediator Systems. In *Proc. VLDB 1996* (Bombay, India, 1996).
155. PARSSIAN, A., SARKAR, S., AND JACOB, V. Assessing Information Quality for the Composite Relational Operation Join. In *Proc. 7th International Conference on Information Quality (IQ 2002)*.
156. PARSSIAN, A., SARKAR, S., AND JACOB, V. Assessing Data Quality for Information Products: Impact of Selection, Projection, and Cartesian Product. *Management Science* 50, 7 (2004).
157. PARSSIAN, A., SARKAR, S., AND JACOB, V. Assessing Data Quality for Information Products. In *Proc. 20th International Conference on Information Systems (ICIS 99)* (Charlotte, NC, December 1999).
158. PASCHKE, A., DIETRICH, J., AND KULHA, K. A Logic Based SLA Management Framework. In *ICSW 2005 Workshop on Semantic Web and Policy Workshop (SWPW 2005)* (2005).
159. PERNICI, B., AND SCANNAPIECO, M. Data Quality in Web Information Systems. *Journal of Data Semantics* (2003).
160. PIERCE, E. M. Extending IP-MAPS: Incorporating the Event-Driven Process Chain Methodology. In *Proc. 7th International Conference on Information Quality (IQ 2002)*.
161. PIPINO, L. L., LEE, Y. W., AND WANG, R. Y. Data Quality Assessment. *Communications of the ACM* 45, 4 (2002).
162. PIRA INTERNATIONAL. Commercial Exploitation of Europe's Public Sector Information, Final Report for the European Commission, Directorate General for the Information Society, October 2000.
163. POIRIER, C. A Functional Evaluation of Edit and Imputation Tools. In *UN/ECE Work Statistical Data Editing* (Rome, Italy, 2-4 June 1999).
164. QUANTIN, C., BOUZELAT, H., ALLAERT, F., BENHAMICHE, A., FAIVRE, J., AND DUSSERRE, L. How to Ensure Data Security of an Epidemiological Follow-up: Quality Assessment of an Anonymous Record Linkage Procedure. *International Journal of Medical Informatics* 49, 1 (1998).
165. RAHM, E., THOR, A., AUMUELLER, D., DO, H. H., GOLOVIN, N., AND KIRSTEN, T. iFuice - Information Fusion Utilizing Instance Correspondences and Peer Mappings. In *Proc. 8th International Workshop on the Web and Databases (WebDB 2005)* (2005).
166. RAMAN, V., AND HELLERSTEIN, J. M. Potter's Wheel: An Interactive Data Cleaning System. In *Proc. VLDB 2001* (Rome, Italy, 2001).
167. REDMAN, T. C. *Data Quality for the Information Age*. Artech House, 1996.
168. REDMAN, T. C. The Impact of Poor Data Quality on the Typical Enterprise. *Communications of the ACM* (1998).
169. REDMAN, T. C. *Data Quality The Field Guide*. The Digital Press, 2001.
170. SAATY, T. L. *The Analytic Hierarchy Process*. McGraw-Hill, 1980.
171. SAMARATI, P. Protecting Respondents' Identities in Microdata Release. *IEEE Transactions on Knowledge and Data Engineering* 13, 6 (2001), 1010–1027.
172. SARAWAGI, S., AND BHAMIDIPATY, A., Eds. *Interactive Deduplication Using Active Learning* (Edmonton, Alberta, Canada, 2002).

173. SCANNAPIECO, M., AND BATINI, C. Completeness in the Relational Model: A Comprehensive Framework. In *Proc. 9th International Conference on Information Quality (IQ 2004)*.
174. SCANNAPIECO, M., PERNICI, B., AND PIERCE, E. M. IP-UML: A Methodology for Quality Improvement based on IP-MAP and UML. In *Advances in Management Information Systems - Information Quality (AMIS-IQ) Monograph*, R. Y. Wang, E. M. Pierce, S. E. Madnick, and C. W. Fisher, Eds. Sharpe, M.E., April 2005.
175. SCANNAPIECO, M., VIRGILLITO, A., MARCHETTI, C., MECELLA, M., AND BALDONI, R. The DaQuinCIS Architecture: a Platform for Exchanging and Improving Data Quality in Cooperative Information Systems. *Information Systems* 29, 7 (2004), 551–582.
176. SCHALLEHN, E., SATTLER, K. U., AND SAAKE, G. Extensible and Similarity-Based Grouping for Data Integration. In *Proc. of the ICDE 2002* (San Jose, CA, 2002).
177. SHANKARANARAYAN, G., WANG, R. Y., AND ZIAD, M. Modeling the Manufacture of an Information Product with IP-MAP. In *Proc. 5th International Conference on Information Quality (IQ 2000)*.
178. SHENG, Y. H. Exploring the Mediating and Moderating Effects of Information Quality on Firms? Endeavor on Information Systems. In *Proc. 8th International Conference on Information Quality (IQ 2003)*.
179. SHENG, Y. H., AND MYKYTYN JR., P. P. Information Technology Investment and Firm Performance: A Perspective of Data Quality. In *Proc. 7th International Conference on Information Quality (IQ 2002)*.
180. SMITH, T. F., AND WATERMAN, M. S. Identification of Common Molecular Subsequences. *Molecular Biology* 147 (1981), 195–197.
181. STOICA, M., CHAWAT, N., AND SHIN, N. An Investigation of the Methodologies of Business Process Reengineering. In *Proc. of Information Systems Education Conference* (2003).
182. STOLFO, S. J., AND HERNANDEZ, M. A. The Merge/Purge Problem for Large Databases. In *Proc. SIGMOD 1995* (San Jose, CA, 1995).
183. STOREY, V., AND WANG, R. Y. Extending the ER Model to Represent Data Quality Requirements. In *Data Quality*, R. Wang, M. Ziad, and W. Lee, Eds. Kluwer Academic Publishers, 2001.
184. STOREY, V. C., AND WANG, R. Y. An Analysis of Quality Requirements in Database Design. In *Proc. 4th International Conference on Information Quality (IQ 1998)*.
185. SU, Y., AND JIN, Z. A Methodology for Information Quality Assessment in the Designing and Manufacturing Processes in Mechanical Products. In *Proc. 9th International Conference on Information Quality (ICIQ 2004)*.
186. TAMASSIA, R., BATINI, C., AND DI BATTISTA, G. Automatic Graph Drawing and Readability of Diagrams. *IEEE Transactions on Systems, Men and Cybernetics* (1987).
187. TARJAN, R. E. Efficiency of A Good But Not Linear Set Union Algorithm. *Journal of the ACM* 22, 2 (1975), 215–225.
188. TATARINOV, I., AND HALEVY, A. Y. Efficient Query Reformulation in Peer-Data Management Systems. In *Proc. SIGMOD 2004*.
189. TEJADA, S., KNOBLOCK, C. A., AND MINTON, S. Learning Object Identification Rules for Information Integration. *Information Systems* 26, 8 (2001).

190. ULLMAN, J. D. *Principles of Database and Knowledge-Base Systems*. Computer Science Press, 1988.
191. ULUSOY, G., AND KARABULUT, K. Determination of the Bundle Price for Digital Information Goods. University of Sabanci, Istanbul, 2003.
192. U.S. NATIONAL INSTITUTE OF HEALTH (NIH). <http://www.pubmedcentral.nih.gov/>.
193. VAN DER AALST, W., AND TER HOFSTEDÉ, A. YAWL: Yet Another Workflow Language. *Information Systems* 30, 4 (2005), 245–275.
194. VASSILIADIS, P., VAGENA, Z., SKIADOPOULOS, S., KARAYANNIDIS, N., AND SELLIS, T. ARTKOS: Toward the Modeling, Design, Control and Execution of ETL Processes. *Information Systems* 26 (2001), 537–561.
195. VERMEER, B. H. P. J. How Important is Data Quality for Evaluating the Impact of EDI on Global Supply Chains ? In *Proc. HICSS 2000*.
196. VERYKIOS, V. S., ELMAGARMID, A. K., BERTINO, E., SAYGIN, Y., AND DASSEN, E. Association Rule Hiding. *IEEE Transaction on Knowledge and Data Engineering* 16, 4 (2004).
197. VERYKIOS, V. S., MOUSTAKIDES, G. V., AND ELFEKY, M. G. A Bayesian Decision Model for Cost Optimal Record Matching. *The VLDB Journal* 12 (2003), 28–40.
198. W3C. <http://www.w3.org/WAI/>.
199. WAND, Y., AND WANG, R. Y. Anchoring Data Quality Dimensions in Ontological Foundations. *Communications of the ACM* 39, 11 (1996).
200. WANG, R. Y., CHETTAYAR, K., DRAVIS, F., FUNK, J., KATZ-HAAS, R., LEE, C., LEE, Y., XIAN, X., AND S., B. Exemplifying Business Opportunities for Improving Data Quality from Corporate Household Research. In *Advances in Management Information Systems - Information Quality (AMIS-IQ) Monograph*, R. Y. Wang, E. M. Pierce, S. E. Madnick, and C. W. Fisher, Eds. Sharpe, M.E., April 2005.
201. WANG, R. Y., LEE, Y. L., PIPINO, L., AND STRONG, D. M. Manage Your Information as a Product. *Sloan Management Review* 39, 4 (1998), 95–105.
202. WANG, R. Y., AND MADNICK, S. E. A Polygen Model for Heterogeneous Database Systems: The Source Tagging Perspective. In *Proc. VLDB'90* (Brisbane, Queensland, Australia, 1990), pp. 519–538.
203. WANG, R. Y., PIERCE, E., MADNICK, S., AND FISHER, C. *Information Quality, Advances in Management Information Systems*. M.E. Sharpe, Vladimir Zwass Series, 2005.
204. WANG, R. Y., STOREY, V. C., AND FIRTH, C. P. A Framework for Analysis of Data Quality Research. *IEEE Transaction on Knowledge and Data Engineering* 7, 4 (1995).
205. WANG, R. Y., AND STRONG, D. M. Beyond Accuracy: What Data Quality Means to Data Consumers. *Journal of Management Information Systems* 12, 4 (1996).
206. WANG, R. Y., ZIAD, M., AND LEE, Y. W. *Data Quality*. Kluwer Academic Publisher, 2001.
207. WEIS, M., AND NAUMANN, F. DogmatiX Tracks down Duplicates in XML. In *Proc. SIGMOD 2005*.
208. WHITE, C. Data Integration: Using ETL, EAI, and EII Tools to Create an Integrated Enterprise. <http://ibm.ascential.com>, 2005.
209. WIEDERHOLD, G. Mediators in the Architecture of Future Information Systems. *IEEE Computer* 25, 3 (1992).

210. WINKLER, W. Improved Decision Rules in the Fellegi-Sunter Model of Record Linkage. In *Proc. of the Section on Survey Research Methods, American Statistical Association* (1993).
211. WINKLER, W. E. Using the EM Algorithm for Weight Computation in the Fellegi and Sunter Model of Record Linkage. In *Proc. of the Section on Survey Research Methods, American Statistical Association* (1988).
212. WINKLER, W. E. Matching and Record Linkage. In *Business Survey Methods*. Wiley & Sons, 1995.
213. WINKLER, W. E. Matching and Record Linkage. In *Business Survey Methods*. Wiley & Sons, 1995.
214. WINKLER, W. E. Machine Learning, Information Retrieval and Record Linkage. In *Proc. of the Section on Survey Research Methods, American Statistical Association* (2000).
215. WINKLER, W. E. Methods for Evaluating and Creating Data Quality. *Information Systems* 29, 7 (2004).
216. WINKLER, W. E. Quality of Very Large Databases. Technical Report RR-2001/04, U.S. Bureau of the Census, Statistical Research Division, Washington, Washington DC, 2001.
217. WISZNIEWSKI, B., AND KRAWCZYK, H. Digital Document Life Cycle Development. In *Proc. 1st International Symposium on Information and Communication Technologies (ISICT 2003)* (Dublin, Ireland, 2003).
218. YAN, L. L., AND OZSU, T. Conflict Tolerant Queries in AURORA. In *Proc. CoopIS'99* (Edinburgh, UK, 1999).
219. ZHOU, C., CHIA, L. T., AND LEE, B. S. QoS Measurement Issues with DAML-QoS Ontology. In *Proc. 2005 IEEE International Conference on e-Business Engineering (ICEBE'05)* (2005).

Index

- D*²*Q* model, 59
- accessibility, 4, 34, 234
- accident insurance registry, 72
- accuracy, 4–7, 15, 19, 20, 22, 23, 33, 37, 39, 40, 42, 52, 54, 59, 60, 64, 73, 76, 78, 89, 99, 103, 113–115, 141, 152, 162, 165, 170, 186, 189, 192, 193, 195, 201, 207, 216, 231
 - attribute accuracy, 22
 - attribute/thematic accuracy, 33
 - database accuracy, 22
 - positional accuracy, 33
 - relation accuracy, 22, 77
 - semantic accuracy, 21
 - strong accuracy error, 23
 - syntactic accuracy, 20
 - tuple accuracy, 77
 - weak accuracy error, 23
- active learning procedure, 125
- activity, 209
- adaptability, 217
- address, 13
- administrative flow, 13
- aggregated data, 65
- aggregation function, 67
- ambiguous representation, 36
- annotation, 56, 57
- anonymization, 225
- application domain
 - accident insurance, 72
 - addresses, 203
 - administrative processes, 99
 - archival, 33
 - bibliographic references, 203
 - biology, 14, 15
 - census, 99
 - chambers of commerce, 190
 - e-Government, 12, 13, 72
 - financial, 167, 203
 - geographical, 33
 - geospatial, 33
 - health, 99
 - life sciences, 12, 13
 - localization data, 177
 - medical data, 203
 - pension data, 203
 - public administration, 177
 - resident persons, 72
 - social insurance, 72, 190
 - social security, 190
 - statistical, 33
 - tax payers, 72, 203
 - world wide web, 12
- as-is process, 186
- Ascential Software, 2
- assessment, 221
 - objective assessment, 221
 - subjective assessment, 221
- assessment methodology, 164, 167, 178
 - analysis, 168
 - objective assessment
 - qualitative, 169
 - quantitative, 168
 - variables selection, 167
- attribute, 8, 52, 99
- attribute completeness, 26

- attribute conflict, 145
 - intolerable, 155
 - tolerable, 155
- attribute/thematic accuracy, 33
- autonomy, 9

- Basel2, 3
- Bayes theorem, 113
- Bayesian model, 111
- Bayesian network, 111
- belief function, 226
- Bellman, 203
- benchmarking, 217
- benchmarking methodology, 167
- benefit, 163, 198
 - intangible, 94, 185
 - monetizable, 94
 - quantifiable, 94
- best quality default semantics, 140
- best quality record, 140
- best quality representative, 140
- bi-gram, 105
- biology, 15
- Boyce Codd normal form, 5, 47
- bridging file, 121
- broker, 229
- bundle of data, 229
- business process, 51, 162, 230
- business process reengineering, 165, 167, 180, 186, 196

- cartesian product, 55, 73
- cause, 185
- CDQM, 181, 188, 190
- CDQM methodology, 181
- cell value, 8
- censored data, 85
 - left, 85
 - right, 85
- certain answer, 158
- certainty factor, 123
- Chambers of commerce, 190, 197
- check plan, 83
- civil status registry, 72
- closed world assumption, 25, 73, 75, 80
- clue, 210
- collections of data, 162
- common character, 105
- compactness, 45
- comparison function, 20, 104, 107, 127
 - 3-gram, 127
 - bi-gram, 127
 - bi-gram distance, 105
 - common character, 105
 - cosine similarity, 106
 - edit distance, 104, 127
 - Hamming distance, 105
 - Jaro algorithm, 105, 127
 - Jaro string comparator, 105
 - n-gram distance, 105
 - q-gram distance, 105
 - Smith-Waterman, 105
 - soundex code, 105
 - token frequency-inverse document frequency, 106
 - transposed character, 105
- comparison space, 107
- comparison vector, 107
- compatibility plan, 83
- complaint, 215
- completeness, 28
- complete view, 158
- completeness, 5, 6, 23, 38–40, 42, 44, 52, 54, 59, 60, 70, 72–79, 85, 138, 141, 152, 165, 166, 173, 181, 186, 192, 193, 195, 217, 231
 - attribute completeness, 26
 - completeness in the relational model, 24
 - completeness of Web data, 27
 - population completeness, 24
 - relation completeness, 26, 74
 - schema completeness, 24
 - tuple completeness, 74
 - value completeness, 25, 74
- completeness in the relational model, 24
- component data item, 7
- component model, 62
- composition algebra, 73
- compound item, 67
- computer science, 4
- conceptual perspective, 209
- conceptual relation, 74
- conceptual schema, 19, 51
- conceptual schema dimension, 39
- conditional independence assumption, 110, 111
- conflict

- attribute, 145
 - description, 144
 - heterogeneity, 144
 - instance-level, 145
 - key, 145
 - resolve attribute-level, 150
 - resolve tuple-level, 150
 - semantic, 144
 - structural, 145
- conflict-tolerant query model, 150
- connectivity infrastructure, 195
- consistency, 5, 6, 19, 20, 22, 30, 31, 38, 40, 42, 59, 60, 82, 141, 152, 231
- constraint
 - foreing key constraint, 31
 - referential constraint, 31
- consumer, 65
- content-based resolution, 153
- Control chart, 87
- cooperative information system, 10, 59, 65, 212
- cooperative infrastructure, 195
- coordinated spot market, 230
- corporate house-holding, 2
- correctness, 15, 20, 38, 192, 217
- correctness with respect to requirements, 43
- correctness with respect to the model, 43
- cosine similarity, 106
- cost, 71, 88, 163, 166, 188, 198, 216
 - data entry cost, 92
 - data processing, 92
 - data usage cost, 92
 - direct, 92
 - indirect, 92
 - information scrap and rework, 89
 - loss and missed opportunity, 89
 - of custom software, 188
 - of equipment, 188
 - of licences, 188
 - of personnel, 188
 - process failure cost, 89
- cost and benefit classification, 71
- cost classification, 89
 - English classification, 89
 - EpplerHelfert classification, 91
 - Loshin classification, 90
- cost of the improvement program, 186
- cost optimization, 70, 71
- cost optimization activity, 174
- cost trade-offs, 71
- cost-based integration of heterogeneous sources, 71
- cost-benefit analysis, 71
- coverage, 77
- critical area, 185
- critical tuple, 226
- cube, 67
- currency, 4, 5, 15, 19, 20, 28, 29, 39, 54, 59, 60, 124, 141, 152, 165, 166, 185, 189, 192, 193, 195
- custom propagation scheme, 57
- custom scheme, 58
- customermatching, 2
- DaQuinCIS, 143, 212
- DaQuinCIS architecture, 212
 - data processing, 214
 - data quality broker, 213
 - monitoring, 214
 - quality analyzer, 214
 - quality assessment, 214
 - quality brokering function, 213
 - quality factory, 214
 - quality improvement function, 213
 - quality notification service, 214
 - quality repository, 214
 - rating service, 214
- data
 - aggregated data, 65
 - elementary data, 65
- data and data quality (D^2Q) model, 59, 213
- data base management system, 9, 51, 68, 138
- data base management systems, 99
- data class, 59
- data cube, 67
- data description model, 51
- data edit, 30
- data editing, 32
- data entry cost, 92
- data envelopment analysis, 138
- data glitch, 86
- data integration, 3, 17, 56, 70, 133, 203, 227
 - instance conflict resolution, 203

- materialized, 134
- virtual, 10, 134, 145
- data integration system, 157, 213, 229
- data manipulation model, 51
- data mining, 16
- data model, 62, 68, 72, 222
 - semistructured, 140, 143
- data owner, 15
- data processing, 92
- data provenance, 51, 56
- Data quality Act, 3
- data quality activity, 12, 69, 98, 163, 202
 - cost optimization, 70, 71
 - cost and benefit classification, 71
 - cost trade-offs, 71
 - cost-based integration of heterogeneous sources, 71
 - cost-benefit analysis, 71
 - data integration, 17, 70, 203
 - data-driven, 195
 - deduplication, 100
 - duplicate identification, 100
 - error correction, 70, 82, 83, 186
 - error detection, 70
 - error localization, 70, 82
 - instance conflict resolution, 70
 - instance-level conflict resolution, 143
 - new data acquisition, 70, 71, 83
 - normalization, 70
 - object identification, 70, 97, 100, 186, 203
 - process-driven, 195
 - profiling, 203
 - quality composition, 70, 71
 - quality-driven query answering, 70
 - record linkage, 70, 100, 205
 - record matching, 189
 - schema cleaning, 71
 - schema matching, 71
 - schema profiling, 71
 - schema reconciliation, 104
 - source trustworthiness, 70, 71
 - standardization, 70, 71, 205
- data quality dimension entity, 52
- data quality improvement process, 12, 163
- data quality measure entity, 53
- data quality measurement process, 12, 163
- data quality methodology, 161
- data quality profile, 64
- data quality schema, 53
- data schema, 59, 66
- data set, 8
- data source, 224
 - contact, 224
 - mail, 224
 - picture, 224
 - spreadsheet, 224
- data steward, 173
- data transformation, 204
 - fold, 205
 - format, 204
 - merge, 205
 - split, 204
- data usage cost, 92
- data value dimension, 39
- data warehouse, 9, 67, 101, 174, 203
- Data Warehousing Institute, VII, 2
- data-driven strategy, 164
- data/activity matrix, 186, 187, 197
- database benchmark, 217
- database/organization matrix, 182, 190
- dataflow/organization matrix, 182, 190
- DBMS, 51, 68, 138
- decision model, 101
- decision rule, 108
- decision tree, 125
- decisional process, 65
- declarative language, 206
- deduplication, 100, 207
- default scheme, 57, 58
- default-all scheme, 57, 58
- definition domain, 8
- delivery time, 29
- density, 77
- dependency, 31
 - functional dependency, 31
 - inclusion dependency, 31
 - key dependency, 31
- description conflict, 144
- design deficiency, 36
- design time, 146
- difference, 55
- dimension, 1, 11, 51, 69, 163, 221
 - accessibility, 234

- accuracy, 78, 141, 152, 192, 195, 231
- availability, 142
- completeness, 76, 78, 79, 85, 138, 141, 152, 192, 195, 231
- conceptual schema dimension, 39
- consistency, 141, 152, 231
- coverage, 77
- currency, 141, 152, 192, 195
- data quality dimension, 20
- data value dimension, 39
- density, 77
- domain specific, 222
- reputation, 138
- semantic accuracy, 21
- soundness, 76
- source trustworthiness, 215
- syntactic accuracy, 20, 222
- timeliness, 138, 231
- tuple accuracy, 76
- tuple inaccuracy, 76
- tuple mismembership, 76
- usability, 234
- dimension metric, 193
- dimensional hierarchy of relations, 117
- disaster, 231
- distance function
 - global distance, 125
 - local distance, 125
- distributed information system, 9
- distribution, 9
- distributional outliers, 87
- document, 232
- domain, 8, 99
- domain constraint, 30
- Dublin core, 33
- duplicate identification, 100
- duplication, 22

- e-Business, 71
- e-Government, 12, 13, 71, 97, 212
- e-Government cooperative architecture, 195
 - connectivity infrastructure, 195
 - cooperation infrastructure, 195
 - event notification infrastructure, 195
- EDI, 231
- edit, 32, 82
 - implicit, 84
- edit distance, 20, 104
- edit-imputation problem, 32, 83
- efficiency, 127
- electronic data interchange, 231
- elementary data, 7, 65
- elementary item, 67
- elimination functions, 153
- EM algorithm, 111
- empirical approach for dimensions, 36, 38
- encryption scheme, 227
- entity, 67
- Entity Relationship diagram, 45
- Entity Relationship model, 45, 52
- entity relationship model, 67
- equational theory, 114
- error correction, 70, 82, 83, 186
 - of inconsistencies, 82
- error detection, 70
- error localization, 70, 82, 83
 - of inconsistencies, 82
- error type, 209
- European directive on reuse of public data, 4
- event notification infrastructure, 195
- event process chain diagram, 62
- exact view, 158
- expectation-maximization algorithm, 111
- exploratory data mining, 16
- extension, 19
- external source of data, 163
- extract, 203
- Extraction Transformation Load (ETL) process, 208

- false match, 109, 110, 126
- false negative, 102, 126
- false negative percentage, 126
- false non-match, 109, 126
- false positive, 102, 125
- false positive percentage, 126
- feature, 152
 - accuracy, 152
 - availability, 152
 - clearance, 152
 - cost, 152
 - timestamp, 152
- feature-based resolution, 153
- features, 142

- federated data, 7
- field, 8, 98
- file, 8, 98, 99
- financial domain, 167
- flexibility, 217
- foreign key constraint, 31, 101
- foreign key dependency, 117
- forward propagation approach, 57
- framework, 201, 212
 - DaQuinCIS, 152, 212
 - FusionPlex, 141, 215
- frequently-changing data, 8
- full outer join, 74
- full outer join merge, 80
- full outer join merge operator, 74
- functional dependency, 31, 47, 99
- fusion functions, 153
- fusion method, 142
- FusionPlex, 141, 143, 152, 215
- FusionPlex architecture
 - conflict detection module, 216
 - conflict resolution module, 216
 - fragment factory, 216
 - mapping database, 216
 - query parser and translator, 216
 - query processor, 216
- garbling, 36
- gene data, 14
- general-purpose methodology, 164
- generalization, 67
- GLAV, 135, 142
- global database, 158
- global schema, 135, 157, 229
- global-as-view, 135
- global-local-as-view, 135
- Google, 1
- group by, 149, 154
- Hamming distance, 105
- harmonic mean, 127
- heterogeneity
 - schema, 133
 - technological, 133
- heterogeneity, 9
- heterogeneity conflict, 144
- IBM, 2
- ideal relation, 74
- identifier attribute, 74
- improvement methodology, 232
- improvement process, 166, 198
- imputation, 32, 83
- inaccuracy, 37
- inclusion dependency, 31
- incomplete representation, 36
- incompleteness, 76
 - relation incompleteness, 76
- inconsistency, 38
- indirect cost, 92
- inductive learning technique, 125
- information, 6
- information flow, 65, 67, 68
 - input information flow, 65
 - internal information flow, 65
 - output information flow, 65
- information item
 - logical information item, 67
 - physical information item, 66
- information product, 7, 61
- information product map, 61
- information quality criteria, 138
- information quality measurement
 - methodology, 232
- information scrap and rework, 89
- information system, 9
 - cooperative, 65
 - distributed, 9
 - management, 64
 - monolithic, 9
 - peer-to-peer, 11
- input information flow, 65
- input time, 30
- instance conflict resolution, 70
- instance inconsistency assumption, 142
- instance-level conflict, 144
- instance-level conflict resolution, 134, 143
- instance-level heterogeneity, 133
- intangible benefit, 185
- integer linear programming, 229
- integer programming, 85
- integrability, 217
- integrity, 3, 4, 33
- integrity constraint, 30, 99, 158, 207
 - domain constraint, 30
 - interrelation constraint, 30
 - intrarelation constraint, 30

- key-foreign key, 101
- intension, 19
- intepretability, 33
- interaction model, 62
- internal information flow, 65
- interoperability
 - physical, 2
 - semantic, 2
- interrelation constraint, 30
- intersection, 73
- intra-organizational process, 65
- intrarelation constraint, 30
- intuitive approach for dimensions, 36, 39
- IP-MAP, 61, 172
- IP-MAP construct block
 - business boundary, 62
 - customer, 62
 - data quality, 62
 - data storage, 62
 - decision, 62
 - information system boundary, 62
 - processing, 62
 - source, 62
- IP-UML, 63, 172
 - data analysis model, 64
 - data quality profile, 64
 - intrinsic information quality category, 64
 - quality analysis model, 64
 - quality association, 64
 - quality data class, 64
 - quality design model, 64
 - quality requirement, 64
 - stereotyped activity, 64
 - stereotyped actors, 64
 - stereotyped dependency relationships, 64
- Istat methodology, 177, 181, 185
- Jaro algorithm, 105
- Jaro string comparator, 105
- join, 73, 149
- joint frequency distribution, 85
- key, 47, 99
- key conflict, 145
- key dependency, 31
- knowledge base, 16
- knowledge reasoning, 16
- knowledge representation, 16
- labeled data, 102
- last update metadata, 29
- learning
 - supervised learning, 103
 - unsupervised learning, 103
- left outer join merge, 75, 80
- life sciences, 12, 13
- local schema, 229
- local-as-view, 135, 229
- logical information item, 67
- logical perspective, 209
- logical plan, 206, 207
- logical schema, 19, 51
- long-term-changing data, 8
- loss and missed opportunity costs, 89
- m-probabilities, 108, 110
- machine learning, 111
 - supervised learning, 111
- macroprocess, 162, 183
- macroprocess quality, 162
- macroprocess/norm-service-process
 - matrix, 183
- management, 4
- management information system, 17
- management information systems
 - model, 64
- manifesto of cooperative information
 - systems, 10
- manipulation language, 51
- mapping database, 216
- mapping rule, 125
- mapping rule learner, 125
- marginal frequency distribution, 85
- markup language, 9
- match, 109
- matching function, 206
- materialized data integration, 134
- matrix
 - data/activity matrix, 186, 187, 197
 - database/organization, 182, 190
 - dataflow/organization, 182, 190
 - macroprocess/norm-service-process, 183
 - process/organization, 183, 191
- mean time between failuress, 85

- meaningless state, 36
- measurement method, 222
- mediator-wrapper architecture, 135
- metadata, 33, 56, 61, 153, 209
 - creator, 33
 - date, 33
 - description, 33
 - format, 33
 - language, 33
 - last update metadata, 29
 - publisher, 33
 - source, 33
 - subject, 33
- metadata layer, 209
- metamodel layer, 209
- metaschema, 67
- methodological phase, 171
 - assessment, 171
 - assign responsibilities on data, 171
 - assign responsibilities on processes, 171
 - benefits evaluation, 171
 - data analysis, 171
 - DQ requirements analysis, 171
 - find critical areas, 171
 - model the process, 171
 - non quality cost evaluation, 171
 - perform measurement, 171
 - improvement, 171
 - check effectiveness of improvements, 172
 - design improvement solutions on data, 171
 - design improvement solutions on processes, 172
 - establish process control, 172
 - find causes of errors, 171
 - manage improvement solutions, 172
 - re-design processes, 172
- methodology, 1, 11, 12
 - assessment methodology, 178
 - CDQM methodology, 181
 - for assessment, 164, 167
 - for benchmarking, 167
 - general-purpose methodology, 164
 - Istat methodology, 177
 - special purpose methodology, 164
 - TDQM methodology, 170, 172
 - TQdM methodology, 170, 174
- metric, 19, 125, 221
 - effectiveness, 127
 - efficiency, 127
 - false negative percentage, 126
 - false positive percentage, 126
 - harmonic mean, 127
 - precision, 126, 130
 - recall, 126, 130
- middleware, 9
- minimalization, 43
- minimum change principle, 83
- mismember tuple, 74
- model, 1, 11, 12, 51
 - data and data quality model, 213
 - data model, 68, 72
 - management information systems
 - model, 64
 - process model, 68
- monolithic information system, 9
- multi-attribute decision making
 - method, 141
- multidimensional cube, 67
- multidimensional database model, 67
- n-gram, 105
- nested join, 150
- new data acquisition, 70, 71, 83
- non identifier attribute, 74
- non-match, 109
- norm, 183
- normalization, 45, 70, 71, 202
- normalized entity relationship schema, 47
- null value, 24, 155
- object identification, 70, 97, 100, 203, 216, 222
 - complexity, 216
- object identification activity, 186
 - choice of comparison function, 102, 125
- decision method, 128
 - human interaction, 128
 - input, 128
 - metrics, 130
 - objective, 128
 - output, 128
 - selection/construction of a representative, 128

- size of data, 130
- types of data, 130
- decision model, 125
- preprocessing, 102
- search space reduction, 104, 119, 120, 125
 - blocking, 104, 120
 - filtering, 104
 - hashing, 104
 - pruning, 104
 - sorted neighborhood, 104
- searching method, 127
 - blocking, 127
 - pruning, 127
- verification, 102
- object identification problem, 22, 97
- object identification technique, 106, 147
 - 1-1 matching, 121
 - Atlas, 123
 - bridging file, 121
 - cost based, 106, 112, 130
 - Delphi, 106, 117, 130, 131
 - DogmatiX, 106, 119, 130, 131
 - empirical, 113, 127, 130
 - Fellegi and Sunter family, 106, 107
 - Fellegi and Sunter model, 130
 - Intelliclean, 122, 130
 - knowledge-based, 121, 130
 - priority queue algorithm, 116, 131
 - priority queue method, 130
 - probabilistic, 106, 130
 - sorted neighbor method, 130
 - sorted neighborhood, 127
 - sorted neighborhood method, 113, 130
 - incremental, 115
 - multi-pass approach, 115
- objective evaluation, 167
- objectivity, 3, 4
- Office of Management and Budget, 3
- ontology, 224
- open world assumption, 25, 73, 75, 80
- operating system, 9
- operation deficiency, 36
- operational process, 65
- Oracle, 3
- organization, 65, 67, 162
 - consumer, 65
 - producer, 65
- organization fusion, 2
- organization model, 62
- outlier, 86, 130
 - distributional, 87
 - time series, 88
- output information flow, 65
- peer data management system, 227
- peer-to-peer information system, 11, 33
- personal data registry, 72
- personal information management, 222
- pertinence, 44
- perturbation-based techniques, 225
- physical information item, 66, 68
- physical interoperability, 2
- physical perspective, 209
- physical plan, 206, 207
- policy, 209
- polygen algebra, 55
- Polygen model, 55
- polygen relation, 55
- polyinstance, 142, 216
- polytuple, 216
- population completeness, 24
- portable device, 224
 - mobile phone, 224
 - PDA, 224
- positional accuracy, 33
- possible match, 101, 109
- precision, 126, 130
- preservation, 233
 - dynamic model, 234
 - static model, 233
- preservation process, 233
- price, 138
- prime-representative record, 115
- privacy, 225
- privacy violation, 225, 226
- private record linkage, 226
- problem, 185
- process, 65, 67, 183
 - as-is, 186
 - decisional process, 65
 - intra-organizational process, 65
 - operational process, 65
 - to-be, 186
- process control, 165, 167, 180
- process driven methodology
 - process re-design, 180

- process driven strategy
 - process re-design, 167
- process failure costs, 89
- process model, 68, 209
- process quality, 162, 231
 - effectiveness, 231
 - efficiency, 231
- process quality robustness design, 231
- process re-design, 165, 167, 180
- process-driven methodology, 180
- process control, 180
- process-driven strategy, 165, 167
 - business process reengineering, 165, 167, 180, 196
 - process control, 165, 167
 - process re-design, 165
- process/organization matrix, 183, 191
- producer, 65
- profiling, 203
- programming language, 9
- project, 55
- projection, 73
- provenance, 33
- pruning, 138
- public administration
 - accident insurance agency, 190, 197
 - central agencies, 177
 - district, 177
 - local agencies, 177
 - municipality, 177
 - peripheral agencies, 177
 - province, 177
 - region, 177
 - social security agency, 190, 197
- public sector information, 228

- q-gram, 105
- quality
 - of macroprocesses, 162
 - of processes, 162
 - of service, 162
- quality association function, 59
- quality attribute model, 54
- quality class, 59
- quality composition, 70, 71
- quality correspondence assertions -
 - specific criteria, 138
- quality cube, 67
- quality demand profile, 66
- quality indicator, 54
- quality of documents, 232
- quality of schema, 19
- quality of schema dimension, 42
- quality of service, 222
- quality of technique
 - criteria, 217
- quality of technique criteria
 - adaptability, 217
 - completeness, 217
 - correctness, 217
 - cost, 217
 - flexibility, 217
 - integrability, 217
 - performance, 217
 - qualitative, 217
 - quantitative, 217
 - reliability, 217
 - robustness, 217
 - scalability, 217
 - transparency, 217
- quality of technique criteria
 - usability, 217
- quality offer profile, 66
- quality profile, 67
- quality schema, 59, 66
- quality selector, 60
- quality-driven query answering, 70
- quality-driven query planning, 138
- quality-driven query processing, 134
- quality/cost driven query answering, 228
- quasi-identifier, 226
- query correspondence assertions, 138
- query decomposition algorithm, 229
- query language, 51
- query model, 149
- query time conflicts, 146
- query-view security problem, 226
- questionnaire schema, 32

- raw data item, 7
- real relation, 74
- recall, 126, 130
- record, 8, 98, 107
 - prime-representative, 115
- record linkage, 70, 98, 100, 205, 222, 225
 - probabilistic, 202
- record linkage prevention, 225

- record linkage promotion, 225
- record matching, 189
- reference relation, 25, 73, 80
- referential constraint, 31
- registry
 - accident insurance, 72, 197
 - chambers of commerce, 197
 - civil status, 72
 - personal data, 72
 - social insurance registry, 72
 - social security, 197
 - tax payers, 72
- relation, 8, 99
- relation accuracy, 77
- relation completeness, 26, 74
- relational model, 52
- relational algebra, 142
- relational algebra operator
 - cartesian product, 55, 73
 - difference, 55
 - full outer join, 74
 - full outer join merge, 74, 80
 - intersection, 73
 - join, 73
 - left outer join merge, 75, 80
 - project, 55
 - projection, 73
 - restrict, 55
 - right outer join merge, 75
 - selection, 73
 - union, 55, 73
- relational hierarchy, 117
- relational model, 73
- relational operator
 - extended cartesian product, 142
- relational table, 8
- relationship, 67
- relative deviation, 88
- reliability, 38, 217
- repository, 209
 - of metadata, 209
- reputation, 138
- requirements, 51
- resolution function, 75, 148
 - ANY, 150
 - AVG, 148
 - CONCAT, 148
 - MAX, 148
 - MAXIQ, 148
 - MIN, 148, 150
 - SUM, 148
- resolution policy, 216
- restrict, 55
- reverse query approach, 57
- right outer join merge, 75
- robustness, 217
- rule, 122, 210
 - duplicate identification, 122
 - merge-purge, 122
- SAP, 3
- scenario, 209
- schema cleaning, 71
- schema completeness, 24
- schema consistency assumption, 142
- schema heterogeneity, 133
- schema level conflict, 144
- schema matching, 71
- schema profiling, 71
- search space, 101
- search space reduction, 120
- secure set intersection method, 226
- security, 4
- selection, 73
- semantic conflicts, 144
- semantic interoperability, 2
- semistructured data, 51, 223
- semistructured data model, 140, 143, 152
- semistructured elementary data, 8
- semistructured data, 6
- service, 162, 183
- service level contract, 222
- service quality, 162
- set covering problem, 85
- set of edit rules, 83
 - consistent, 83
 - non redundant, 83
 - valid, 83
- set of organizations, 162
- set relationship
 - containment, 74
 - disjointness, 74
 - independence, 74
 - quantified overlap, 74
- similarity function
 - co-occurrence, 118
- similarity measure

- co-occurrence, 118
 - textual, 118
- similarity score, 124
- Simple Additive Weighting method, 139
- size of a relation, 25
- sliding window, 131
- Smith-Waterman, 105
- social insurance registry, 72
- sound view, 158
- soundex code, 105
- soundness, 76
- source schema, 157
- source selection, 186, 205
- source trustworthiness, 70, 71, 215
- source-specific criteria, 138
- special purpose methodology, 164
- SQL, 149
- SQL operator
 - group by, 149, 154, 207
 - join, 149
 - nested join, 150
- stable data, 8
- standardization, 70, 71, 202, 205
- star schema, 67, 101
- statistics, 4, 16
- strategy
 - data-driven, 164
 - process-driven, 164, 165
- strategy for conflict resolution
 - HighConfidence, 150, 156
 - PossibleAtAll, 150, 156
 - RandomEvidence, 150, 156
- structural conflict, 145
- structure extraction, 203
- structured data, 6
- structured elementary data, 8
- structured file, 8
- subjective evaluation, 167
- superkey, 47
- supervised learning, 103, 111
- supplier, 229
- supply chain, 231
- suppression-based technique, 225
- syntactic accuracy, 222

- table, 8
- Tailor, 217
- Tailor functionality
 - comparison function, 217
 - decision model, 217
 - measurement, 217
 - searching method, 217
- tax payers registry, 72
- TDQM, 180
- TDQM methodology, 170, 172
- technique, 1, 7, 8, 11, 12, 15, 16, 18, 55, 68–71, 95, 161, 163, 164, 166, 168, 171, 175, 177, 179, 180, 187, 188, 193, 198, 200–202, 216–218
 - domain dependent, 102
 - domain independent, 102
 - empirical, 102
 - for instance conflict resolution, 144
 - for object identification, 99, 106, 147, 216
 - knowledge-based, 102
 - probabilistic, 102
 - sorted neighborhood, 208
- technique for instance-level conflict resolution
 - OO_{RA} , 147, 155
 - Aurora, 147, 150, 156
 - DaQuinCIS, 147, 152, 156
 - FraSQL-based, 147, 154
 - Fusionplex, 147, 152, 156
 - OO_{RA} , 156
 - SQL-based, 156
 - SQL-based conflict resolution, 147, 149
- technological heterogeneity, 133
- text equivalent content, 34
- theoretical approach for dimensions, 36
- threshold, 130
- time related dimension, 39
- time series outliers, 88
- timeliness, 29, 38, 40, 138, 231
- timestamp, 216
- to-be process, 186
- token frequency-inverse document frequency, 106
- tolerance strategies, 149
- tool, 1, 11, 171, 177–179, 187, 188, 201–203, 209, 217, 218
 - Ajax, 202, 203, 206
 - Artkos, 203, 208
 - Bellman, 203
 - Choice Maker, 202, 210
 - commercial, 187

- Intelliclean, 202, 203
- open source, 187
- Potter's Wheel, 203
- Telcordia's tool, 203, 205
- toolbox, 201, 216
- toolkit
 - Tailor, 217
- TQdM, 174, 180
- TQdM methodology, 170, 174
- transformation operator, 206
 - clustering, 207
 - mapping, 207
 - matching, 207
 - merging, 207
 - view, 207
- transitive closure, 114, 123
- transparency, 217
- transposed character, 105
- tree traversal, 103, 118
- threshold, 108
- true match, 110, 126
- true negative, 126
- true non-match, 126
- true positive, 126
- truncated data, 85
- trust, 228
- trustworthiness, 15, 56
- tuple, 99
- tuple accuracy, 76, 77
- tuple completeness, 25, 74
- tuple mismembership, 76
- type of data, 6
 - administrative data, 99
 - aggregated data, 7
 - complex structured data, 99, 103, 106
 - component data item, 7
 - dimensional hierarchy of relations, 117
 - elementary, 7
 - federated, 7
 - frequently-changing, 8
 - information product, 7
 - long-term-changing, 8
 - raw data item, 7
 - relational hierarchy, 117, 130
 - semistructured, 6, 99, 103, 119
 - semistructured elementary data, 8
 - simple structured, 99
 - stable, 8
 - structured, 6
 - structured elementary, 8
 - unlabelled, 111
 - unstructured, 6
 - web, 8
 - XML data, 119, 130
- type of information system, 6, 9
 - cooperative, 10
 - data warehouse, 9
 - distributed, 9
 - monolithic, 9
 - peer to peer, 11
- u-probabilities, 108, 110
- UML
 - class, 63
 - model element, 63
 - profile, 63
 - relationship, 63
 - specification, 63
 - stereotype, 63
 - tag definition, 63
 - tagged value, 63
- UML activity diagram, 64
- UML class, 63
- UML model element, 63
- UML object flow diagram, 64
- UML profile, 63
- UML relationship, 63
- UML stereotype, 63
- UML tag definition, 63
- UML tagged value, 63
- unfolding, 140
- union, 55, 73
- union-find data structure, 116
 - find, 116
 - union, 116
- universal relation, 74, 75
- unlabeled data, 102
- unstructured data, 6, 232
- unsupervised learning, 103
- usability, 217, 234
- user, 162, 184
- user-defined aggregation, 154
- user-defined grouping, 154
 - context aware, 154
 - context free, 154
- user-query specific criteria, 138
- utility, 3, 4

value completeness, 25, 74

view

complete, 158

exact, 158

sound, 158

virtual data integration, 10, 134, 145

volatility, 29

web data, 8

web information system, 15, 230

where provenance, 56

why provenance, 56

window size, 127

within deviation, 88

world wide web, 12

World Wide Web consortium, 34

XML, 100, 222

XML data model, 60

XML document, 106

XML schema, 51, 101

XQuery, 60

XQuery function, 60