SAP PRESS

Gatling, Brague, Champlin, Stefani, Weigel, Bryce, Dharwad, Engel, Gardella, Grewal, Felsheim, Haelterman, Hamer, Jackson, Keilen, Kuppe, McFadden, Seguin, Sinha, Stridinger, Waite

Enterprise Information Management with SAP®

www.sap-press.com

Galileo Press

Bonn • Boston

Contents at a Glance

PART I SAP's Enterprise Information Management Strategy and Portfolio

1 Introducing Enterprise Information Management 27
2 Introducing Information Governance 55
3 Big Data with SAP HANA and Hadoop 81
4 SAP's Solutions for Enterprise Information Management 103
5 Practical Examples of EIM 165

PART II Working with SAP's Enterprise Information Management Solutions

6 Getting to Know Your Data with SAP Information Steward 223
7 Introducing SAP Data Services 251
8 Managing Data Quality with SAP Data Services 289
9 Monitoring Data Quality with Scorecards in SAP Information Steward 339
10 Processing Text Data with SAP Data Services 357
11 Migrating Data with SAP Data Services 373
12 Introducing SAP Master Data Governance 385
13 Introducing SAP NetWeaver Information Lifecycle Management 423
14 Introducing SAP Extended Enterprise Content Management by OpenText 467

Contents

Foreword 17
Preface 19

PART I SAP's Enterprise Information Management Strategy and Portfolio

1 Introducing Enterprise Information Management 27

1.1 Defining Enterprise Information Management 28
1.1.1 Example of Information Flow Through a Company 30
1.1.2 Types of Information Included in Enterprise Information Management 33
1.2 Common Use Cases for EIM 36
1.2.1 Operational EIM 37
1.2.2 Analytical EIM 39
1.2.3 Information Governance 40
1.3 Common Drivers for EIM 40
1.3.1 Operational Efficiency as a Driver of EIM 41
1.3.2 Information as an Organizational Asset 43
1.3.3 Compliance as a Driver of EIM 44
1.4 Impact of Big Data on EIM 45
1.5 SAP's Strategy for EIM 48
1.6 Typical User Roles in EIM 50
1.7 Example Company: NeedsEIM, Inc. 51
1.7.1 CFO Issues 52
1.7.2 Purchasing Issues 52
1.7.3 Sales Issues 52
1.7.4 Engineering and Contracts Issues 52
1.7.5 Information Management Challenges Facing NeedsEIM, Inc. 53
1.8 Summary 53

2 Introducing Information Governance 55

2.1 Introduction to Information Governance 56

2.2 Evaluating and Developing Your Information Governance Needs and Resources 58
2.2.1 Evaluating Information Governance 58
2.2.2 Developing Information Governance 62
2.3 Optimizing Existing Infrastructure and Resources 63
2.4 Establishing an Information Governance Process: Examples 65
2.4.1 Example 1: Creating a New Reseller 67
2.4.2 Example 2: Supplier Registration 68
2.4.3 Example 3: Data Migration 71
2.5 Rounding Out Your Information Governance Process 75
2.5.1 The Impact of Missing Data 75
2.5.2 Gathering Metrics and KPIs to Show Success 76
2.5.3 Establish a Before-and-After View 80
2.6 Summary 80

3 Big Data with SAP HANA and Hadoop 81

3.1 SAP HANA 81
3.1.1 Basics of SAP HANA 82
3.1.2 How SAP HANA Works 83
3.1.3 SAP HANA for Analytics and Business Intelligence 85
3.1.4 SAP HANA as an Application Platform 85
3.2 SAP HANA and EIM 86
3.2.1 Using SAP Data Services to Load Data into SAP HANA 87
3.2.2 EIM Solutions and SAP HANA 88
3.3 Big Data and Hadoop 88
3.3.1 The Rise of Parallel Processing Architectures 88
3.3.2 NoSQL: A New Kind of Database 89
3.3.3 Hadoop Primer 90
3.3.4 Hadoop: Enterprise Use Cases 95
3.4 SAP HANA and Hadoop 99
3.4.1 The V's: Volume, Variety, Velocity 99
3.4.2 SAP HANA: Designed for Enterprises 100
3.4.3 Hadoop as an SAP HANA Extension 100
3.5 EIM and Hadoop 100
3.5.1 ETL—Data Services and the Information Design Tool 101
3.5.2 Unsupported: Information Governance and Information Lifecycle Management 101
3.6 Summary 102

4 SAP's Solutions for Enterprise Information Management 103

4.1 SAP Data Services as a Data Foundation 106
4.1.1 Basics of SAP Data Services 106
4.1.2 SAP Data Services Integration with SAP Applications 108
4.1.3 SAP Data Services Integration with Non-SAP Applications 111
4.1.4 Data Cleansing and Data Validation with SAP Data Services 112
4.1.5 Text Data Processing in SAP Data Services 115
4.2 SAP Information Steward 118
4.2.1 Data Profiling and Data Quality Monitoring 119
4.2.2 Metadata Analysis 122
4.2.3 Cleansing Rules 123
4.2.4 Business Glossary 125
4.3 SAP NetWeaver Master Data Management and SAP Master Data Governance 126
4.3.1 SAP NetWeaver Master Data Management 127
4.3.2 SAP Master Data Governance 127
4.4 SAP Solutions for Enterprise Content Management 130
4.4.1 Overview of SAP's ECM Solutions 133
4.4.2 SAP Extended Enterprise Content Management by OpenText 137
4.4.3 SAP Document Access by OpenText and Archiving by OpenText 140
4.5 SAP NetWeaver Information Lifecycle Management 142
4.5.1 Retention Management 146
4.5.2 System Decommissioning 147
4.6 SAP Solutions for Data Migration 150
4.6.1 SAP Data Migration Software 151
4.6.2 Predefined Content and Rules for Data Migration 153
4.6.3 Systems Migration 155
4.6.4 SAP Rapid Marts 155
4.7 Information Governance in SAP 156
4.7.1 Information Governance Use Scenario Phasing 157
4.7.2 Technology Enablers for Information Governance 159
4.8 NeedsEIM, Inc. and SAP's Solutions for EIM 162
4.9 Summary 163

5 Practical Examples of EIM 165

5.1 EIM Architecture Recommendations and Experiences by Procter and Gamble 165
5.1.1 Principles of an EIM Architecture 166
5.1.2 Scope of an EIM Enterprise Architecture 168
5.1.3 Structured Data 169
5.1.4 The Dual Database Approach 170
5.1.5 Typical Information Lifecycle 171
5.1.6 Data Standards 176
5.1.7 Unstructured Data 177
5.1.8 Governance 179
5.1.9 Role of the Enterprise Information Architecture Organization 184
5.2 Managing Data Migration Projects to Support Mergers and Acquisitions 184
5.2.1 Scoping for a Data Migration Project 185
5.2.2 Data Migration Process Flow 187
5.2.3 Enrich the Data Using Dun and Bradstreet (D&B) with Data Services 191
5.3 Evolution of SAP Data Services at National Vision 192
5.3.1 Phase 1: The Enterprise Data Warehouse 192
5.3.2 Phase 2: Enterprise Information Architecture—Consolidating Source Data 193
5.3.3 Phase 3: Data Quality and the Customer Hub 195
5.3.4 Phase 4: Application Integration and Data Migration 197
5.3.5 Phase 5: Next Steps with Data Services 198
5.4 Recommendations for a Master Data Program 198
5.4.1 Common Enterprise Vision and Goals 198
5.4.2 Master Data Strategy 199
5.4.3 Roadmap and Operational Phases 199
5.4.4 Business Process Redesign and Change Management 199
5.4.5 Governance 200
5.4.6 Technology Selection 200
5.5 Recommendations for Using SAP NetWeaver Process Integration and SAP Data Services 201
5.5.1 A Common Data Integration Problem 201
5.5.2 A Data Integration Analogy 202

5.5.3 Creating Prescriptive Guidance to Help Choose the Proper Tool 203
5.5.4 Complex Examples in the Enterprise 204
5.5.5 When All Else Fails… 206
5.6 Ensuring a Successful Enterprise Content Management Project by Belgian Railways 206
5.6.1 Building the Business Case 206
5.6.2 Key Success Factors for Your SAP Extended Enterprise Content Management by OpenText Project 213
5.7 Recommendations for Creating an Archiving Strategy 216
5.7.1 What Drives a Company into Starting a Data Archiving Project? 216
5.7.2 Who Initiates a Data Archiving Project? 217
5.7.3 Project Sponsorship 218
5.8 Summary 220

PART II Working with SAP's Enterprise Information Management Solutions

6 Getting to Know Your Data with SAP Information Steward ... 223

6.1 Cataloging Data Assets and Their Relationships 224
6.1.1 Configuring a Metadata Integrator Source 225
6.1.2 Executing or Scheduling Execution of Metadata Integration 227
6.2 Profiling Data 228
6.2.1 Configuration and Setup of Connections and Projects 229
6.2.2 Getting Basic Statistical Information about the Data Content 232
6.2.3 Identifying Cross-Field or Cross-Column Data Relationships 236
6.3 Monitoring and Assessing the Quality of Your Data 239
6.3.1 Defining Validation Rules Representing Business Requirements 241
6.3.2 Binding Rules to Data Sources for Data Quality Assessment 244
6.3.3 Executing Rule Tasks and Viewing Results 246
6.4 Summary 250

7 Introducing SAP Data Services 251

7.1 Data Integration Scenarios 252
7.2 SAP Data Services Platform Architecture 253
7.2.1 User Interface Tier 255
7.2.2 Server Tier 257
7.3 SAP Data Services Designer Overview 259
7.4 Creating Data Sources and Targets 263
7.4.1 Connectivity Options for SAP Data Services 263
7.4.2 Connecting to SAP 265
7.4.3 Connecting to Hadoop 268
7.5 Creating Your First Job 268
7.5.1 Create the Data Flow 268
7.5.2 Add a Source to the Data Flow 269
7.5.3 Add a Query Transform to the Data Flow 269
7.5.4 Add a Target to the Data Flow 270
7.5.5 Map the Source Data to the Target by Configuring the Query Transform 270
7.5.6 Create the Job and Add the Data Flow to the Job 271
7.6 Basic Transformations Using the Query Transform and Functions ... 272
7.7 Overview of Complex Transformations 275
7.7.1 Platform Transformations 275
7.7.2 Data Integrator Transforms 277
7.8 Executing and Debugging Your Job 280
7.9 Exposing a Real-Time Service 281
7.9.1 Create a Real-Time Job 282
7.9.2 Create a Real-Time Service 284
7.9.3 Expose the Real-Time Service as a Web Service 286
7.10 SAP Data Services and SAP HANA 287
7.11 Summary 288

8 Managing Data Quality with SAP Data Services 289

8.1 Data Cleansing 291
8.1.1 Data Parsing 292
8.1.2 Data Standardization 303
8.1.3 Data Correction 309
8.1.4 Data Validation 311

8.2 Data Enhancement 313
8.3 Data Matching 315
8.3.1 Rule-Based Matching Method 321
8.3.2 Weighted Scoring Matching Method 322
8.3.3 Combination Matching Method 323
8.3.4 Data Consolidation 330
8.4 Other Important Data Quality Capabilities 333
8.4.1 SAP Data Quality Management, Version for SAP Solutions 333
8.4.2 Data Quality Blueprints 335
8.4.3 Using Data Quality Beyond Customer Data 336
8.5 Summary 337

9 Monitoring Data Quality with Scorecards in SAP Information Steward 339

9.1 Components of a Data Quality Scorecard 340
9.2 Defining and Setting Up a Data Quality Scorecard 343
9.2.1 Modeling the High-Level Key Data Domains 343
9.2.2 Defining Data Quality Dimensions for the Key Data Domains 344
9.2.3 Adding the Validation Rules and Quality Dimensions for a Key Data Domain 346
9.2.4 Binding the Data Sources to the Data Quality Scorecard 348
9.3 Viewing the Data Quality Scorecard 349
9.3.1 Drilling Down into Data Quality Scorecard Details 350
9.3.2 Analyze Failed Record Sample Data 351
9.4 Identifying Data Quality Issue Impact or Data Quality Root Causes in the System Landscape 352
9.4.1 Reviewing Data Quality Issue Impact on Downstream Processes or Reports 353
9.4.2 Analyzing Data Quality Root Causes 355
9.5 Summary 355

10 Processing Text Data with SAP Data Services 357

10.1 Introduction to Text Data Processing Capabilities in SAP Data Services 359
10.1.1 Entity Extraction Transform Overview 361
10.1.2 How Extraction Works 362
10.2 Text Data Processing and NeedsEIM, Inc. 364
10.2.1 NeedsEIM, Inc. Pain Points 364
10.2.2 Using the Entity Extraction Transform 366
10.3 Summary 372

11 Migrating Data with SAP Data Services 373

11.1 Introduction to Data Migration 373
11.2 Data Migration Content Delivered by SAP Best Practices and Rapid Deployment Solutions 375
11.2.1 Data Migration Content 376
11.2.2 Architecture of SAP Best Practices and Data Services for Data Migration to SAP Applications 377
11.3 Migrating Customer Data to SAP ERP 379
11.3.1 Step 1: Download the Migration Content 379
11.3.2 Step 2: Install the Software 380
11.3.3 Step 3: Review the Jobs in Data Services 381
11.4 Systems Migration 383
11.5 Summary 384

12 Introducing SAP Master Data Governance 385

12.1 SAP Master Data Governance Overview 386
12.1.1 Deployment Options 388
12.1.2 Change Request and Staging 389
12.1.3 Process Flow in SAP Master Data Governance 391
12.2 Getting Started with SAP Master Data Governance 392
12.2.1 Data Modeling 393
12.2.2 User Interface (UI) Modeling 394
12.2.3 Data Quality and Search 395
12.2.4 Process Modeling 396
12.2.5 Data Replication 397

12.2.6 Key and Value Mapping 397
12.2.7 Data Transfer 399
12.2.8 Activities Beyond Customizing 399
12.3 Governance for Custom-Defined Objects: Example 400
12.3.1 Define New Models for Airline Entity and Currency Entity 401
12.3.2 Define Relationship Between the Airline and Currency Entity Types 402
12.3.3 Define User Interface 403
12.3.4 Link the User Interface Configuration to the Data Model 405
12.3.5 Create a Change Request Process 406
12.3.6 Assign Processors to Workflow 407
12.3.7 Test the New Airline Change Request 407
12.4 Rule-Based Workflows in SAP Master Data Governance 409
12.4.1 Classic SAP Business Workflow and Rule-Based Workflow using SAP Business Workflow and BRFplus 410
12.4.2 Designing Your First Rule-Based Workflow in SAP Master Data Governance 417
12.5 NeedsEIM, Inc.: Master Data Remediation 420
12.6 Summary 422

13 Introducing SAP NetWeaver Information Lifecycle Management 423

13.1 The Basics of Information Lifecycle Management 425
13.1.1 What Does ILM Mean? 425
13.1.2 Why Is ILM Needed? 425
13.2 Overview of SAP NetWeaver Information Lifecycle Management 426
13.2.1 Cornerstones of SAP NetWeaver ILM 428
13.2.2 Data Archiving Basics 429
13.2.3 ILM-Aware Storage 432
13.2.4 Architecture—What You Need to Run SAP NetWeaver ILM 434
13.3 Managing the Lifecycle of Information in Live Systems 436
13.3.1 Audit Area 437
13.3.2 Data Destruction 439

13.3.3 Legal Hold Management 440
13.4 Managing the Lifecycle of Information from Legacy Systems 441
13.4.1 Preliminary Steps 442
13.4.2 Steps Performed in the Legacy System 444
13.4.3 Steps Performed in the Retention Warehouse System 444
13.4.4 Handling Data from Non-SAP Systems During Decommissioning 446
13.5 System Decommissioning—Detailed Example 447
13.5.1 Data Extraction 448
13.5.2 Data Transfer and Conversion 452
13.5.3 Reporting 459
13.5.4 Data Destruction 463
13.6 Summary 466

14 Introducing SAP Extended Enterprise Content Management by OpenText 467

14.1 Capabilities of SAP Extended ECM 469
14.1.1 Data and Document Archiving 470
14.1.2 Records Management 471
14.1.3 Content Access 472
14.1.4 Document-Centric Workflow 472
14.1.5 Document Management 472
14.1.6 Capture 473
14.1.7 Collaboration and Social Media 473
14.2 How SAP Extended ECM Works with the SAP Business Suite 474
14.3 Integration Content for SAP Business Suite and SAP Extended ECM 476
14.3.1 SAP ArchiveLink 476
14.3.2 Content Management Interoperability Standard (CMIS) and SAP ECM Integration Layer 477
14.3.3 SAP Extended ECM Workspaces 479
14.4 Summary 485

The Authors 487
Index 495

This chapter introduces you to SAP's solutions for EIM, allowing you to judge which products are of the most interest to you. Based on what you learn in this chapter, you can determine which chapters in Part II you should read first and where you need to start learning more.

4 SAP's Solutions for Enterprise Information Management

SAP's solutions for EIM have both great breadth and depth, spanning the full capabilities for managing information from its acquisition, through active use, until retirement and destruction. This chapter discusses all of the currently available products and solutions SAP offers for EIM. The portfolio will continue to grow as SAP continues to make investments in this area.

Chapter 1 discussed EIM in SAP's overall portfolio; Figure 4.1 shows the same portfolio view but also lists the EIM products, which are explored further in this chapter.

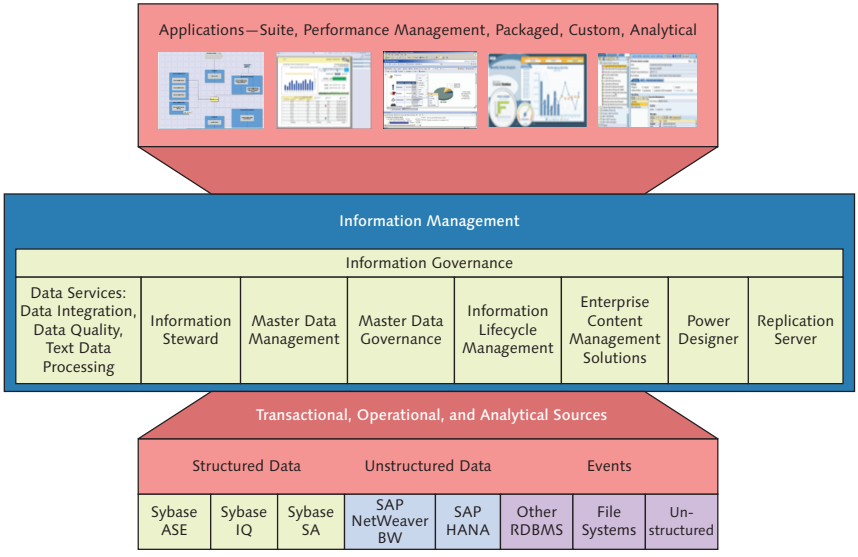

Figure 4.1 SAP's Solutions for EIM

Figure 4.1 shows SAP's solutions for EIM in schematic form. At the top of the graphic are applications that depend on the data that the SAP solutions manage; at the bottom are the data sources and the types of data, which both feed the information management products and are managed by them. The middle of the graphic is dominated by the set of enterprise information management solutions offered by SAP, all beneath a box labeled "Information Governance." (Information governance is not a product, but a discipline that is supported by multiple products. It is discussed in more detail in Chapter 3.) The following products are included in SAP's solutions for EIM:

- **SAP Data Services**
  Hereafter referred to as Data Services, this is a product that has combined SAP Data Integrator and SAP Data Quality Management into one, even though the solutions can still be delivered independently if needed. However, most customers require both products, so this book will focus on Data Services and refer to Data Integrator as the data integration capabilities delivered with Data Services. SAP Data Quality Management will be referred to as the data quality capabilities delivered with Data Services. A third major capability built into Data Services is called text data processing.
- **SAP Information Steward**
  Hereafter referred to as Information Steward.
- **SAP NetWeaver Master Data Management**
  Hereafter referred to as SAP NetWeaver MDM.
- **SAP Master Data Governance**
  Hereafter referred to as SAP MDG.
- **SAP NetWeaver Information Lifecycle Management**
  Hereafter referred to as SAP NetWeaver ILM.
- **Enterprise Content Management (ECM) by OpenText**
  ECM by OpenText is a group of related products. ECM products included in SAP's solutions for EIM are:
  - SAP Archiving by OpenText
  - SAP Document Access by OpenText (includes SAP Archiving by OpenText)
  - SAP Extended Enterprise Content Management by OpenText (hereafter referred to as SAP Extended ECM). SAP Extended ECM includes both SAP Document Access by OpenText and SAP Archiving by OpenText.

- SAP Invoice Management by OpenText
- SAP Document Presentment by OpenText

While there are many OpenText products included with SAP's solutions for EIM, this book will focus on SAP Extended ECM and its components: SAP Document Access by OpenText and SAP Archiving by OpenText.

- **Sybase Replication Server (hereafter referred to as Replication Server)**
  Replication Server supports the enterprise data management environment by replicating and synchronizing databases such as Sybase Adaptive Server Enterprise (ASE), Oracle, Microsoft SQL Server, and IBM DB2 Universal Database (UDB) database transactions.
- **Sybase Power Builder (hereafter referred to as Power Builder)**
  Power Builder is a modeling tool that offers a model-driven approach to improve business intelligence and information architecture. Enterprise architects use it to bring together many of the moving parts that make up the enterprise: systems, applications, business processes, and requirements.

In addition to the software products mentioned above, there is a solution called data migration that is built on top of existing EIM products. It is not listed in Figure 4.1 because it is not standalone software, but a solution built on Data Services.

This chapter will provide a brief introduction to SAP's capabilities for EIM, whether they be software products (such as Data Services, Information Steward, SAP MDG, etc.), solutions built on software products (such as data migration), or disciplines (such as information governance). Part II of the book will provide greater details on some of the capabilities; however, due to the extent of the capabilities, not all of them will be covered in depth.

Note

We do not focus on Sybase solutions in this book. For more information on Replication Server and Power Builder go to *http://www.sybase.com*.

Chapter 1 discussed the use of EIM for on-boarding, active use, and off-boarding of information. The following is a list of SAP's EIM capabilities for each area, all of which will be discussed in more detail in this chapter:

- **On-boarding**
  Includes Data Services, Information Steward, SAP NetWeaver ILM, SAP NetWeaver

MDM, and SAP MDG as well as the use of data migration and the on-boarding of content with SAP Extended ECM.

- **Active use**
  Includes Data Services, Information Steward, SAP NetWeaver MDM, SAP MDG, SAP Document Access by OpenText, and SAP Extended ECM.
- **Off-boarding**
  Includes Data Services, SAP Document Access by OpenText, SAP Extended ECM, and SAP NetWeaver ILM.

4.1 SAP Data Services as a Data Foundation

Data Services is the primary tool to extract, transform, and load data from one or more source systems into one or more target systems; Data Services lets you improve, integrate, transform, and deliver trusted data to critical business processes across the enterprise for both SAP and non-SAP systems. Data Services can be used in almost any scenario that requires you to move, enrich, transform, or cleanse data, and in this regard functions as the technology foundation for a coherent enterprise information management strategy. The following sections provide an overview of the capabilities delivered with Data Services.

4.1.1 Basics of SAP Data Services

Data Services is used for data migration, systems migration, data synchronization, application data cleansing, loading data warehouses and data marts, and for query, reporting, analysis, and dashboard data provisioning. The three major capabilities of Data Services are data cleansing, data validation, and text data processing. The core of Data Services is the data services engine (see Figure 4.2, which shows Data Services' capabilities and typical uses). As mentioned previously, Data Services combines two products: Data Integrator for extraction, transformation, and loading (ETL) and SAP Data Quality Management for data validation and data cleansing. The third major capability of Data Services is text data processing and is referred to simply as *text data processing*.

The left side of Figure 4.2 shows the data sources supported by Data Services. Data Services can access data from a wide variety of applications and file sources and can consume almost every type of data—structured, semi-structured, and

unstructured—from those sources. Data Services can be called in batch mode (for example, to extract and deliver data for reporting in the data warehouses) or via client applications such as SAP ERP, SAP Customer Relationship Management (CRM), or custom applications to perform data transformation and cleansing in real time.

Notice in the figure that Data Services shares a common technology layer with the SAP BusinessObjects Business Intelligence (BI) platform. This allows for common user provisioning; advanced user management; password and security policies; use of external authentication mechanisms such as Active Directory, lightweight directory access protocol (LDAP), or SAP NetWeaver Identity Management; and granular access control.

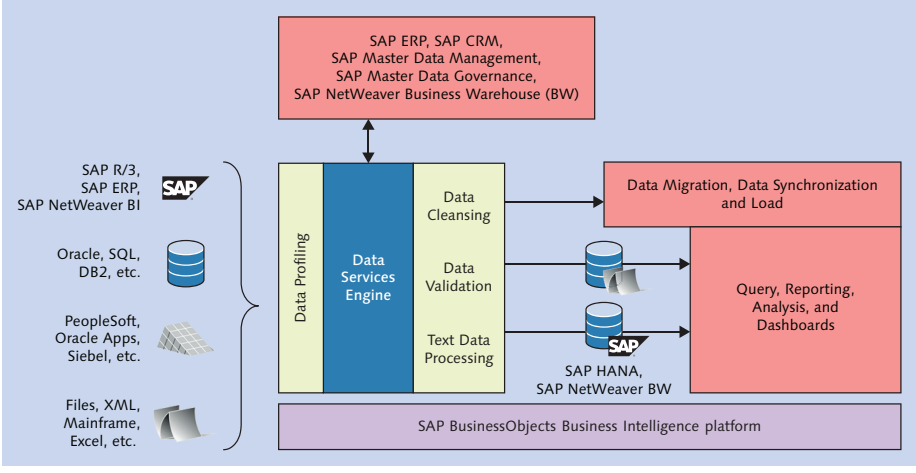

Figure 4.2 Data Services Architecture

You can see in Figure 4.2 how Data Services can access many types of systems and applications and can work with many kinds of data and how that data can be profiled. The connectivity options and profiling capabilities of Data Services are described below.

Connectivity Options for SAP Data Services

Data Services supports many connectivity options for structured and unstructured data. These include SAP applications, databases, other vendor applications, and pure files (Excel, mainframe, etc.). Details of the connectivity options are discussed in Chapter 7.

Data Profiling in SAP Data Services

Data profiling is the practice of determining the overall quality of data and finding data anomalies. While Information Steward is the primary tool for data profiling, technical data profiling can be done directly in Data Services; you can go beyond simple viewing of the data to conduct an analysis of your data. You can build better jobs in Data Services by understanding the following types of information:

- Frequency distribution
- Distinct values
- Null values
- Minimum/maximum values
- Data patterns (e.g., Xxx Xxxx99, 99-Xxx)
- Comparison of values between data sets
- Drill-down to view specific records

The profiling in Data Services allows you to quickly assess the source data to discover problems and anomalies, such as:

- Twenty-one percent of the employees in the Human Resources source system do not have a country associated with their ID.
- There are four genders entered: 60% are male, 30% are female, and the other 10% are either "unknown" or have a question mark in the gender field.
- Benefits are available for all new employees after one month of service, but 35% of three-month employees still have no benefit ID assigned to their HR records.

You can also quickly detect patterns, distinct values, and null values for zip codes, product codes, sales items, and other key data fields to better understand your data.

4.1.2 SAP Data Services Integration with SAP Applications

Data Services provides seamless integration with SAP applications as an integration and data quality tool. Specific examples include SAP CRM, SAP ERP, SAP NetWeaver MDM, SAP MDG, SAP NetWeaver Business Warehouse (BW), and SAP HANA. Data Services is used as a service to the applications, called only to perform a specific function when needed. It can also be used to load data into the applications. Next we discuss a few examples in a bit more detail.

SAP Data Services for SAP Data Quality Management with SAP ERP, SAP Customer Relationship Management, and SAP NetWeaver Master Data Management

Figure 4.3 shows common uses of Data Services with the SAP Business Suite. One common use of Data Services with the SAP Business Suite is the deep integration with Business Address Services (BAS). BAS is an SAP NetWeaver capability embedded in the ABAP application server. BAS provides flexible dialog integration for standard functions such as creating, changing, displaying, and finding addresses. It is a reusable component across the SAP Business Suite that is used heavily in SAP ERP (such as SAP ERP Central Component [ECC] 6.0) and SAP CRM. With Data Services integration to BAS, as addresses are updated, the data quality capabilities in Data Services are used to correct the addresses and check for duplicates. Figure 4.3 shows Data Services integration with BAS.

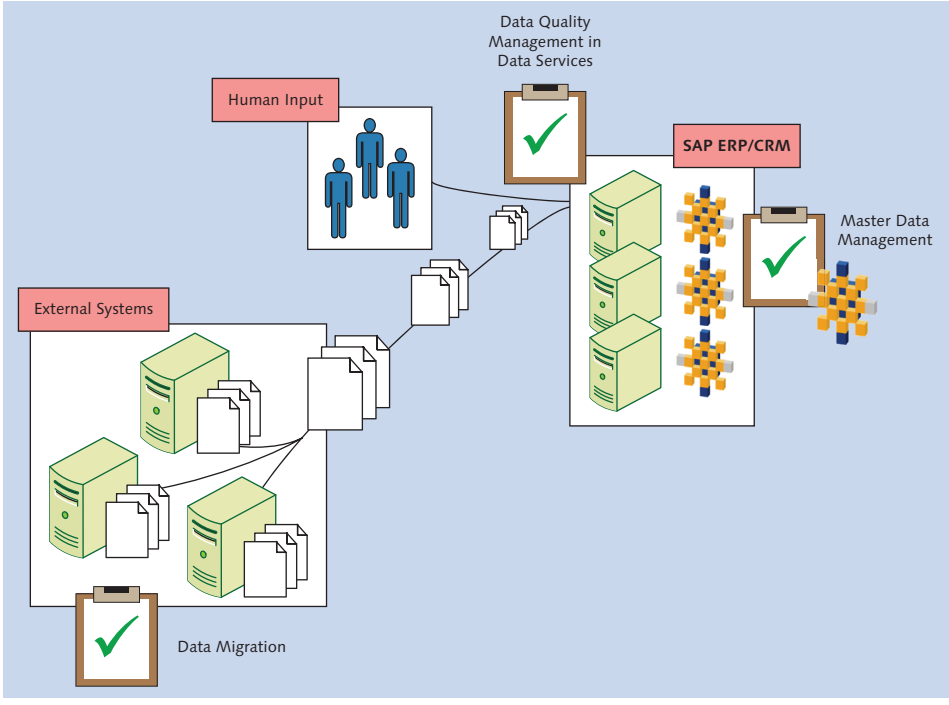

Figure 4.3 Data Services with SAP Applications

In Figure 4.3, notice that Data Services is used for data migration to the SAP Business Suite. SAP provides a robust data migration solution for mapping and validating

source data against the SAP target system using Data Services. Data Services is used to migrate each application object (for example, materials, sales orders, cost centers, etc.). For each object, the data is cleansed, validated against the required configuration in the SAP target system, and loaded into the SAP system. Reports are provided using SAP BusinessObjects Web Intelligence so users can monitor, remediate, and govern the data migration project.

Additional integration with Data Services for SAP applications includes integration with SAP NetWeaver MDM and SAP MDG. SAP NetWeaver MDM uses Data Services for loading data as well as cleansing names and addresses, de-duplication and automatic consolidation, and data validation. Data Services provides real-time cleansing, matching, and consolidation activities for SAP MDG, providing re-usability between SAP NetWeaver MDM and SAP MDG. Figure 4.4 shows Data Services integration with SAP MDG.

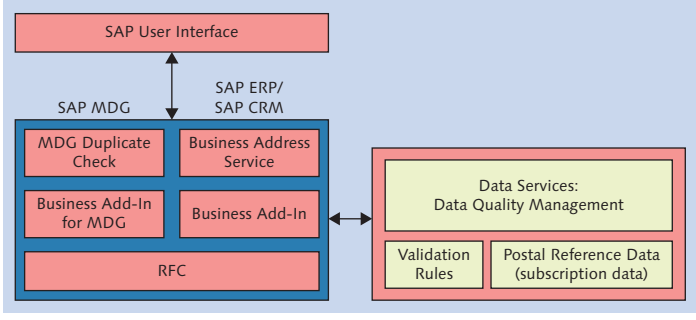

Figure 4.4 Data Services Integration with SAP MDG, SAP ERP, and SAP CRM

Next we will discuss how Data Services is used with SAP HANA, SAP NetWeaver BW, and the SAP BusinessObjects BI platform.

SAP Data Services for SAP HANA, SAP NetWeaver Business Warehouse, and the SAP BusinessObjects Business Intelligence Platform

A key strength of Data Services is its easy integration with data warehouses and databases for analytics. This is true for Data Services integration with SAP NetWeaver BW, SAP HANA, and the entire SAP BusinessObjects BI platform.

Data Services is the preferred tool for loading non-SAP data into SAP HANA. In fact, the data integration capabilities of Data Services are included with SAP HANA. Additionally, SAP is making major improvements for a seamless user interface

between Data Services and SAP HANA. This is explained in more detail in Chapter 3 and Chapter 7. Figure 4.5 shows the integration of Data Services with SAP NetWeaver BW and SAP HANA.

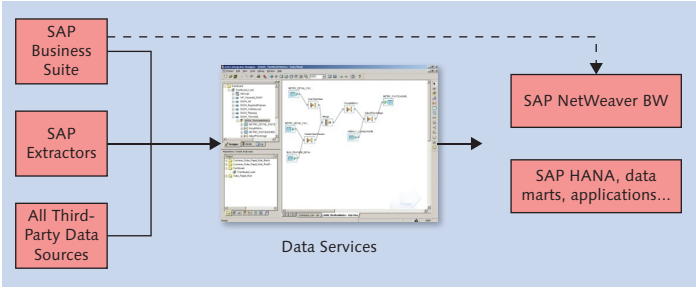

Figure 4.5 Data Services Integration with SAP NetWeaver BW and SAP HANA

Data Services can also be used for loading data into SAP NetWeaver BW. Existing SAP NetWeaver BW customers can now easily apply data quality transformations when loading data into SAP NetWeaver BW. They can use one tool to define all extraction, validation, and cleansing rules to load all data (SAP and non-SAP) into SAP NetWeaver BW. This is important to understand because everyone faces data quality issues, and you don't want to re-implement your ETL jobs to add data quality. Non-SAP NetWeaver BW customers can get *native* access to the SAP Business Suite, including delta changes, without going through SAP NetWeaver BW.

In addition to integration with databases and data warehouses, Data Services also works natively with the SAP BusinessObjects BI platform. Data Services is used for the provisioning of data for reports, dashboards, ad hoc queries, OLAP (online analytical processing) analyses, and data exploration. Data Services provides access and integration of disparate data from virtually any data source, structured or unstructured. This data can be built up in a data warehouse or data mart to provide historical trending and analysis for more accurate decision making. Data Services also enables the understanding of information context (data lineage) to help you make more confident decisions. The integration with the SAP BusinessObjects BI platform is standard functionality that comes out-of-the-box.

4.1.3 SAP Data Services Integration with Non-SAP Applications

Data Services has a much longer history of integration with non-SAP applications than it does with SAP applications. It started as a pure ETL tool that moved data

from any source to any source. It has grown into a full-scale data foundation that includes extraction, transformation, loading, data quality, and text data processing, with deep integration into SAP applications while maintaining the ability to integrate into any application. In your enterprise there could be few or many non-SAP systems where integration, data quality, and data validation are required. Data Services is well equipped for the requirement of moving and transforming data between diverse non-SAP systems. Figure 4.2, showed the connectivity options for Data Services, ranging from Microsoft Excel spreadsheets to Oracle applications to mainframe connectivity. Refer back to Figure 4.3: You can also see the reference to data migration, data synchronization, and data loading. In this book, we will focus on data migration to SAP target applications when we discuss data migration. When migrating to SAP applications, the target canonical data formats, field requirements, and so on, are delivered by SAP. However, Data Services can easily be used to migrate to a new home-grown application or some other niche application as well; in this case, however, you need to define the target structure.

Figure 4.3 also showed that external systems and human input are linked to the data quality capabilities in Data Services. This is very common for Data Services integration with non-SAP applications. One very specific example is the SAP Data Quality Management SDK (software development kit). SAP has many software partners who develop their own software but use part of SAP's solution within their solution. The SAP Data Quality Management SDK provides developers with a lightweight integration method to integrate the robust capabilities of the data cleansing and validation capabilities directly in their own custom applications.

Non-SAP integration with Data Services includes loading third-party data in SAP HANA and SAP NetWeaver BW, as well as extracting data to go in other data marts, data warehouses (such as Sybase IQ), and applications.

Data Services' native capabilities for dealing with both SAP and non-SAP data provide great flexibility so that Data Services can be embedded in SAP and non-SAP applications and used across the SAP family of solutions where data cleansing, validation, and integration are critical for the application.

4.1.4 Data Cleansing and Data Validation with SAP Data Services

SAP Data Quality Management is a key capability in Data Services and will be covered in detail in Chapter 8. Data quality capabilities include address cleansing, data

standardization, data validation, data correction, data enrichment, and matching. Figure 4.6 shows an example of the entire data quality process.

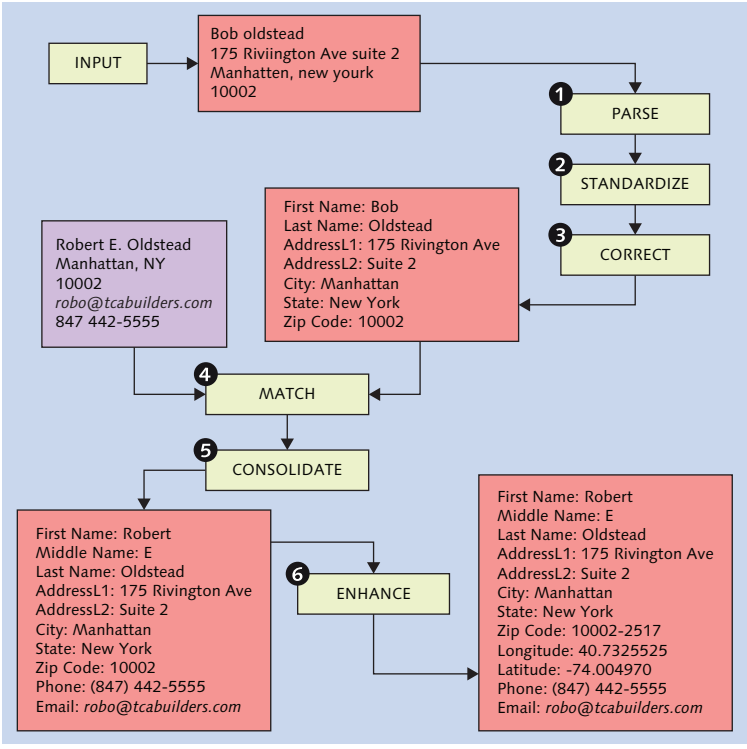

Figure 4.6 Example of the Data Quality Process

To begin, the input record is parsed into its component parts and standardized. Figure 4.6 shows the following example:

- First Name: Bob
- Last Name: oldstead
- Address1: 175 Riivington Ave
- Address2: Suite 2
- City: Manhatten
- State: new yourk
- Zip Code: 10002

Errors in the record are corrected; for instance, "Oldstead" is capitalized, and the street name, city, and state are edited. After this step, the record looks like this:

- First Name: Bob
- Last Name: Oldstead
- Address1: 175 Rivington Ave
- Address2: Suite 2
- City: Manhattan
- State: New York
- Zip Code: 10002

Data Services then searches for a matching record in the target application and finds the following match:

- Name: Robert E. Oldstead
- City, State: Manhattan, NY
- Zip Code: 10002
- Email: *robo@tcabuilders.com*
- Phone: 847 442-5555

The two records are consolidated:

- First Name: Robert
- Last Name: Oldstead
- Address1: 175 Rivington Ave
- Address2: Suite 2
- City: Manhattan
- State: New York
- Zip Code: 10002
- Phone: (847) 442-5555
- Email: *robo@tcabuilders.com*

Finally, the record is enhanced with a normalized zip code and geographical coordinates to provide a complete record:

- First Name: Robert
- Last Name: Oldstead
- Address1: 175 Rivington Ave
- Address2: Suite 2
- City: Manhattan
- State: New York
- Zip Code: 10002-2517
- Longitude: 40.7325525
- Latitude: -74.004970
- Phone: (847) 442-5555
- Email: *robo@tcabuilders.com*

While this example focused on cleansing a customer record, the data quality process can be applied to business partners, material products, services, and many other types of data that need parsing, standardization, and data cleansing.

4.1.5 Text Data Processing in SAP Data Services

The text data processing (TDP) feature in Data Services unlocks unstructured text sources. With advanced linguistic and semantic capabilities, TDP knows the elements of human language, how these elements differ from language to language, and how their dependencies express the underlying salient information. It understands the meaning and context of information, not just the words themselves. Through core natural-language processing of the entity extraction transform, TDP extracts key information—the "who, what, where, when, and how" in text—revealing important relations and events occurring between these entities. Ultimately, it gives structure to unstructured text sources and disregards the noise, allowing for business analytics and reporting. The following are some common use cases for text data processing:

- **Electronic discovery law**
  Discovery is a pretrial phase in a lawsuit in which each party makes information available to the other. TDP accelerates the process of analyzing documents for evidence.

- **Brand monitoring**
  A company wants to know what people are saying about its products and brands. Are they creating buzz? Is that buzz positive or negative, and how does it change over time?
- **Employee morale**
  Sentiment analysis of collaboration spaces and employee surveys can detect employee problems and feedback.
- **Customer service**
  Companies can optimize call centers by routing questions to the right person.

The following are several business examples of how text data processing is used today.

- **Media company**
  - Problem: Records containing artist names, songs, and movie titles are stored in legacy repositories where they can't be queried, analyzed, or aggregated.
  - Goal: Integrate legacy records into a warehouse for a single source of truth for reporting.
  - Solution: Text data processing identifies artist names, titles, dates, and so on. Data quality normalizes names and titles. Integrated, trusted data is loaded into a data warehouse.
- **Law enforcement**
  - Problem: Agents need to link the common characteristics of crime scenes with suspects. Databases are incomplete, and important clues may be lost in the volumes of textual information in field notes, case documents, IM logs, emails, and so on.
  - Goal: Integrate content from unstructured sources into agency databases so it is available for investigations. Via analytical application, the agent queries for suspects, things, organizations, and locations, as well as links or co-occurrences between these.
  - Solution: Text data processing extracts person names, objects, locations, and other data of interest using public sector extraction rules. The extracted entities are cleansed and normalized by data quality. Human review of the output occurs prior to database commit.

Optional Function for Law Enforcement

When joining results into the repository, the match function looks for similar persons, addresses, or other attributes that may resolve identities.

- **Maintenance and repairs**
  - Problem: A company doesn't know what repair issues are most common or why. Managers are under pressure to reduce maintenance-related costs. Technicians do not always select correct values in a database.
  - Goal: Analyze field technicians' repair logs and profile maintenance issues to drive predictive analytics for future maintenance problems. Identify root causes of costly warranty repairs and report back to engineering. Complete blank fields in structured data based on comments.
  - Solution: Text data processing recognizes concepts, part numbers, and problems. It matches cluster concepts, which reveal the most common issues. Models and part numbers can be normalized.

Text data processing extracts the text and clearly recognizes its meaning from a large number of words. One example is reporting on the most mentioned concepts. An output report created by analyzing comments about a car could result in the information in Table 4.1 (with drill-down capabilities to get more details).

Concept	Total Feedback	Negative Sentiments	Positive Sentiments	Requests
Handling	109	9	98	2
Maintenance	98	19	64	15
Brake	66	20	43	3
Cup holder	51	27	20	4
Interior	50	4	42	4
Mileage	37	4	30	3

Table 4.1 Car Sentiment Report

From the car sentiment report you can see that cup holder and brakes have the most negative sentiments and that handling has the most positive. This report can

be used to drill down to collect feedback that can be used to support marketing and advertising or provide input into future designs of the car.

Text data processing is covered in more detail in Chapter 10.

4.2 SAP Information Steward

Information Steward delivers a single environment to discover, assess, define, and monitor the quality of your enterprise data. It is positioned as a key user interface for governance with the ability to monitor the quality of your data over time and centrally catalog the different data assets and business term definitions in your organization. The combination of these disciplines allows companies not only to identify bad data, but also to see where the data is coming from (and where it is used in downstream processes).

Information Steward provides an environment that enables you to do the following:

- *Discover* and understand enterprise data. This includes data profiling and metadata management. Information Steward gives your company the means to discover its data assets, visually assess its quality, isolate issues, and understand the impact they are having on your critical business processes. And in doing so, it provides a starting point for both business and IT users to collaborate on and improve data quality. It is an essential starting point for any information governance initiative.
- *Define* the data quality requirements and data quality business rules, as well as the common business term glossary centrally in one application. Information Steward provides a means to define ownership and accountability of critical business data assets and the ability to relate them to business terms in a business glossary. This helps establish a common understanding of the data and its meaning for key stakeholders of a governance program. In addition, Information Steward provides users with the ability to define validations rules to further assess the quality of the data. And as data quality issues become apparent, both business and IT users are able to define cleansing packages to address them.
- *Monitor* data quality over time, provide data quality scorecards in business user applications, and resolve data quality issues over time. With Information Steward both business and IT users are empowered to monitor data quality throughout the enterprise via the use of standardized scorecards that are easy to access, intuitive to use, and allow the user to interactively drill down to the data sources

that have failed to meet defined validation rules. In addition, business users can also access these same data quality scores in the environments they already work in. For example, an analyst viewing a report has the ability to see where the data in the report originated and the data quality score associated with the report sources.

In this section we will provide an overview of each major capability in Information Steward. Chapter 6 and Chapter 9 will provide more details on the profiling and data quality scorecards. In Information Steward there are four major areas:

- **Data profiling and data quality monitoring**
  These are used to gain insight, assess data, and provide dashboards for monitoring data quality over time.
- **Metadata analysis**
  Metadata analysis integrates and relates metadata from multiple systems in the company and allows transparency of the existing assets in the company and how they are related to each other.
- **Cleansing rules**
  Cleansing rules enable the creation of specific rules for cleansing data according to company policies, industry standards, and so on.
- **Business term glossary**
  A business term glossary provides a central glossary for organizations to store business terms or definitions that have been approved by the responsible business experts.

Tip

The term *metadata,* when used with respect to Information Steward, refers to technical metadata regarding the structure of the data. The National Information Standards Organization defines metadata as "structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use or manage an information resource."

Each of these areas will be discussed next.

4.2.1 Data Profiling and Data Quality Monitoring

Data profiling and data quality monitoring are key capabilities of Information Steward. You can assess and monitor the data quality of a specific data domain for

a specific purpose (for example, customer quality assessment, sales system migration, or master data quality monitoring).

By measuring data quality along the core processes in your organization, from the point of entry down to the place of consumption, you can easily identify which business process is processing good data. Ongoing data quality monitoring allows you to respond early to slight dips in your data quality level. With early detection you can set up the necessary cleansing activities or processes, rather than identify data quality issues after they have negatively impacted your core business processes.

Figure 4.7 shows an example of a data quality scorecard that provides a quality score along specific data quality dimensions.

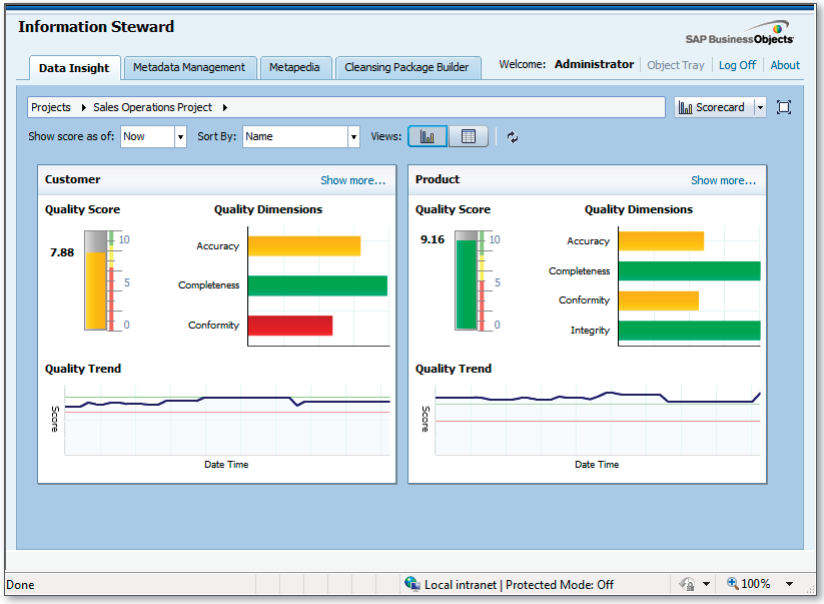

Figure 4.7 Example Data Quality Scorecard

The following are typical activities that are done for data profiling and data quality monitoring:

- **Adding tables to a project**
  For example, in a project to assess supplier master data, you would include tables and files that should be analyzed to assess supplier master data.

- **Profiling the data**
  You interactively execute data profiling steps to create statistical information on minimum values, maximum values, frequency distribution of values, pattern distribution of values, and so on. This is a first insight into the content of the data when the first outliers are identified. For example, with supplier master data, say you assess the data and notice that the pattern of the supplier name includes a high percentage of special characters. You can directly create validation rules from the profiling results area that the data steward later verifies and approves from a business perspective.
- **Setting up the data quality validation rules**
  You might have a great set of already documented or identified requirements based on policies, industry standards, or external regulation that you can transfer into validation rules. A typical validation rule for supplier master data validates that the name of the supplier does not contain any characters other than letters or number and blanks.
- **Binding the validation rules to your data**
  Since the central validation rules are generic and can be reused for multiple tables, they are bound to the multiple tables in your company after their approval by the data steward. For example, a definition for a valid telephone number format is done once centrally by the data steward but is then bound to your employee, customer, and supplier data tables.
- **Executing the rule tasks**
  Apply the defined validation rules to the existing data in the tables, files, or applications that you have bound to your project to calculate the number of failed records for each rule.
- **Analyzing or reviewing the current data quality level from different perspectives**
  From a technical perspective you can view the quality of the data based on specific tables or applications; for example, data quality in the completeness of sales orders tables for the sales application. Alternatively, the data quality level can be viewed from the perspective of a business user or data steward who is responsible for a specific data domain such as supplier master data. In this case, you are only interested in the overall score of your key data domain; in other words, drill down from the data quality scorecard domain (such as supplier, product, or customer) to the data quality dimensions, to the individual rules set up for your data domain, and down to the individual tables where your data is stored.

- **Scheduling rule tasks for ongoing monitoring**
  Since data quality measurement is not a singular activity, you can set up scheduled execution of your rule tasks using the internal scheduling capability.

Specific details of how to create a data quality dashboard are covered in Chapter 9.

4.2.2 Metadata Analysis

Metadata analysis enables users to know what data assets exist in the company and the relationships between those data assets. For example, in the business intelligence suite, users can track the source data used for a report, and they can see how a change in a source will impact a report. For example if you delete a column in one of your source systems, this might have a significant impact on your reports; it could result in missing data and erroneous calculations. Metadata analysis empowers users to do the following:

- Understand the entire business intelligence environment
- Trace a data lineage from a report to sources
- Identify a change impact from a source to reports and users
- Manage metadata from various data sources, data integration technologies, and business intelligence systems
- Lower the total cost of ownership by tracking usage and promote reuse of data and reports
- Improve decision making and regulatory compliance
- Provide root-cause and impact analysis of bad data in the system landscape

Metadata analysis consolidates metadata from various sources, stores it in a central metadata repository, and analyzes the dependencies and relationships between the different data assets. This consolidated and related metadata enables auditing, usage, change impact, and data lineage analysis. Metadata can be supplemented with custom attributes and annotations to add further information to the collected metadata to provide a full picture and understanding of the existing data assets. Figure 4.8 shows the integration with the central repository. The universe creation in the figure refers to the universe capability in the SAP BusinessObjects BI platform.

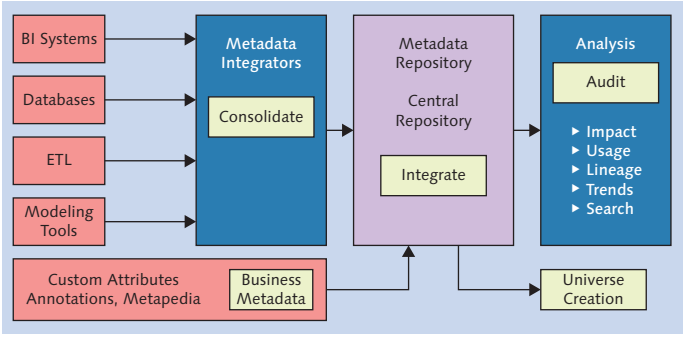

Figure 4.8 Consolidate, Integrate, and Audit Your Metadata

One of the most common use cases of the metadata management capabilities in Information Steward is the direct integration of the data lineage capabilities into the launch pad in the SAP BusinessObjects BI platform. In the launch pad, users can choose to view the lineage of a report they are working on. The lineage offers a graphical representation of what data from which sources a report uses and how the data is flowing into the report. In addition to knowing how much they can trust the data in the tables, users can also see the results of a data quality assessment on the source or data warehouse staging area tables. Leveraging the metadata analysis with the data quality assessment provides credibility to the report.

Another example of metadata analysis is to support data quality root-cause analysis and impact analysis. While you are reviewing the data quality results you can drill into the failed data. When you drill down you can directly call the metadata management functionality to see where the data in the assessed column is coming from. This enables you to evaluate the source of the error as well as where the data is used in downstream reports. The linkage goes all the way to the business terms and linkage to the business process. This allows you to drill down from your business process name to the associated tables and contextualize the quality score for the table and the individual columns.

4.2.3 Cleansing Rules

Information Steward also provides a capability called the cleansing package builder. This enables the creation of custom cleansing rules that Data Services uses as a transformation step to cleanse and standardize your data. You can create custom parsing and standardization rules that can be used with the data cleanse transform in Data Services. (Note: Chapter 8 covers the data cleanse transform.)

Custom cleansing rules can be created to parse and standardize account numbers, product codes, product descriptions, purchase dates, part numbers, SKUs, dates, names, social security numbers, and so on.

The cleansing package builder was created to empower the data steward or subject matter expert to develop a custom data cleansing solution, allowing the user to easily and quickly develop new data cleansing solutions for data domains not provided out-of-the-box (for example, product data in the pharmaceutical industry) and to customize the cleansing packages delivered by SAP.

The data steward defines how the data should be classified simply based on the desired output. Once the desired output is created, the cleansing package builder automatically creates the data dictionary, rules, and patterns that make up a cleansing package, which is then consumed by Data Services data cleanse transforms. The following example shows the input and output after using a custom cleansing rule developed using the cleansing package builder:

> *Glove ultra grip profit 2.3 large black synthetic leather elastic with Velcro Mechanix Wear.*

After the cleansing step executes, the parsed and standardized output would look like the information shown in Table 4.2.

Parsed Text Output	Standardized Text Output
Product Category	Glove
Size	Large
Material	Synthetic Leather
Trademark	Pro-Fit 2.3 Series
Cuff Style	Elastic Velcro
Palm Type	Ultra-Grip
Color	Black
Vendor	Mechanix Wear
Standard Description	Glove—Synthetic Leather, Black, size: Large, Cuff Style: Elastic Velcro, Ultra-Grip, Mechanix Wear

Table 4.2 Parsed and Standardized Output from the Data Cleanse Transform

In the cleansing package builder, the data steward can set up how to create a standard for parsed values, a standard description, and any punctuation. For example, the input includes the phrase "profit 2.3," but after the cleansing package executes, the description for the trademark attribute is "Pro-Fit 2.3 Series." You can also see that the standard description has appended multiple attributes together and added additional punctuation. Ultimately, the data steward can control how the data looks based on business requirements. The cleansing package builder can be used to create custom cleansing rules for specific industry standards such as GS1 (Global Standards 1) and UNSPSC (United Nations Standard Products and Services Code). It can be used to create company-specific cleansing rules so that data is uniformly cleansed according to company policies and compliance requirements in diverse application systems.

4.2.4 Business Glossary

If you ask five employees in your company what the term *revenue* means to them, you might get at least three different definitions. But for analysis and decision making, it is essential that everyone in the company be on the same page when talking about the same terms or policies. Information Steward provides *Metapedia* as a central location for the definition, maintenance, and approval of business terms; its goal is to promote proactive information governance with common understanding and agreement on business terms and concepts.

The application range for Metapedia is broad: from descriptive explanations of terms (like in a simple dictionary) to central storage of calculation formulas, to technical terms, to business process descriptions associated with tables and data assets in business processes. Metapedia also enables business users to understand data attributes used in business intelligence reporting environments with business user-oriented descriptions. In Metapedia, you can organize business terms into categories or hierarchies that align with business subject matter or lines of business.

Metapedia's searching capabilities let you search for not only the terms that have been created, but also synonyms that have been assigned to them. From the search results you can easily trace the associated objects or their details into the categories and hierarchies the terms are linked to. In this way, Metapedia terms can be associated and linked to other technical objects in your Information Steward system. This is used to extend the central data assets repository with business terms. For example, you can define business names in Metapedia and assign them to the

technical column names in your metadata management. If a business user searches for a business term such as *telephone number,* he can get a list of all table columns associated with that business term in the system landscape, including telephone number information, even if the column is labeled "PHNNR."

Metapedia is a helpful component to bridge the gap between IT and the business user, providing a common understanding of what information is available where in the company.

4.3 SAP NetWeaver Master Data Management and SAP Master Data Governance

Managing master data is critical to an enterprise-wide information management strategy. Every company, every line of business, is impacted by master data. From financial close, to product development, to order management, to help desk support, to supplier management, every organization is impacted by master data—especially when the master data is wrong, unclear, redundant, or conflicting. SAP's solutions for EIM include an integrated suite for managing and governing master data. These solutions include previously discussed capabilities, such as Data Services and Information Steward. Additionally, two other key capabilities for managing and governing master data include SAP NetWeaver MDM and SAP MDG. The focus of SAP NetWeaver MDM is to consolidate and harmonize master data in application- and system-agnostic contexts (for example, in analytical scenarios, to ensure trustworthy data for enterprise reporting). SAP NetWeaver MDM supports all domains for master data consolidation, harmonization, and stewardship, covering business initiatives such as mergers and acquisitions, divestitures, and supply chain optimization.

SAP MDG provides centralized master data creation, maintenance, and governance for the SAP Business Suite and beyond. In the standard version, this is pre-built for selected data domains, such as the financial, material, supplier, and customer domains. It can be extended for additional domains. SAP MDG is focused on operational master data management. It is a natural extension of the business processes running in the SAP Business Suite, adding prebuilt master data creation, change, and distribution processes and stewardship.

4.3.1 SAP NetWeaver Master Data Management

SAP NetWeaver MDM enables you to manage master data across multiple domains and multiple systems. It is primarily used for master data consolidation and master data harmonization; both are described below.

Master Data Consolidation

Master data consolidation aggregates master data across SAP and non-SAP systems into a centralized master data repository. Once data is consolidated, you can search for data across linked systems, identify identical or similar objects across systems, and provide key mapping for reliable companywide analytics and reporting.

Master Data Harmonization

Master data harmonization first consolidates the master data from heterogeneous systems and then harmonizes the data to meet global quality standards. It ensures high-quality master data by distributing harmonized data that is globally relevant using distribution mechanisms. Subscribing applications can enrich master data with locally relevant information.

For more details on SAP NetWeaver MDM see *SAP NetWeaver Master Data Management* (Heilig et al., SAP PRESS, 2007).

4.3.2 SAP Master Data Governance

SAP MDG provides centralized governance for selected master data domains based on SAP's standard data models and is extensible to custom-defined models. It performs native integration with the SAP Business Suite, using existing business logic and customer-specific configuration, for validation of master data while it is being created. In addition, SAP MDG integrates with Data Services for data quality and enrichment. It uses the data quality engine mentioned previously to inform the user about duplicate data sets. It also uses Data Services' rich address enrichment features to support users in entering correct addresses for their business partners.

SAP MDG uses SAP Business Workflow for the governance process and allows you to design your own workflows for the central master data creation or maintenance process. These workflows typically involve multiple people adding their expertise to complex master data (for example, one person edits a material's classification

information and another adds units of measurement and their conversions). The workflows also include workflow steps for approval of the changed data.

Until the approval, inactive data is kept in a staging area, separate from productive master data, and it is only after approval that the data is posted to the master data tables in the SAP system and distributed to systems that subscribe to these master data changes. Through this process, SAP MDG delivers consistent definition, authorization, and replication to corporate systems such as SAP Business Suite and non-SAP systems. By tracking all changes and approvals, SAP MDG supports compliance and a complete audit trail.

SAP MDG provides master data maintenance capabilities through a role-based, domain-specific user interface. In addition to central maintenance of master data in the SAP Business Suite, data from external sources is supported. For example, you can upload data into SAP MDG via a file or services and then further process (or enrich) the data in SAP MDG. SAP MDG provides governance for specific master data areas. The primary areas are discussed below.

SAP Master Data Governance for Financial Master Data

SAP MDG provides governance processes for a broad set of financial master data, including charts of accounts, general ledger accounts, companies and conglomerate companies, consolidation units, profit centers and cost centers including their hierarchies, and cost elements and cost element hierarchies.

SAP Master Data Governance for Supplier and Customer Master Data

SAP MDG supports the governance of customer master data creation and changes. The focus of the first release is generic business partner attributes, general customer data, and address information. For both customers and suppliers, SAP MDG supports the search, display, change, and creation of single master data sets as well as mass processing or changes to hierarchies.

Figure 4.9 is an example of the master data governance process for supplier data. A requestor enters the initial supplier data and submits the master data record for approval. A specialist can review the request and either return it to the requestor for additional information, reject the request, or approve the request. If additional data is required to enrich the record—for instance, purchasing data—then a domain specialist can perform this data enrichment. Once the record is complete and has received final approval, it is created in the SAP ERP system and, if applicable,

distributed to other systems that share the data. Prebuilt workflows are used to control the creation process at every step.

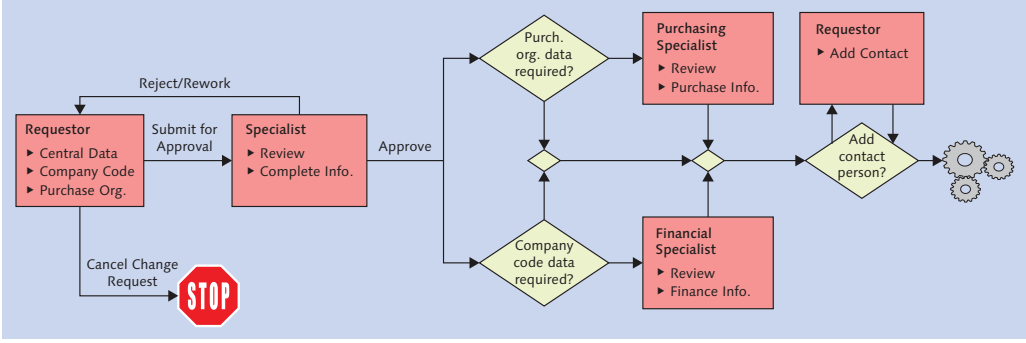

Figure 4.9 Example of SAP MDG Process for Supplier

SAP Master Data Governance for Material Master Data

SAP MDG allows for central creation and change processes for material master data. The focus is on identifying and describing attributes of the material master data, including basic data and classification. Again, you can search, display, change, and create a single material master data element, perform mass processing, and upload files.

In addition to the governance processes provided, users can also create customized governance processes. Business process experts can define enhancements and checks (for example, by enriching parts of data through assigning an enrichment spot, skipping unnecessary checks, or ensuring that validations occur by assigning checks). They can also set field properties for entity types and attributes by specifying which fields are relevant and which relevant fields are required through field properties. They can assign a different user interface in addition to the standard interface configured for the data model. All this allows a maximum of flexibility on the change request step level.

In addition to this flexibility at design time, power users or even end users can easily adapt the user interface to their specific needs at runtime. For example, they can rearrange the screen layout or set default values on selected fields for their most often-used screens. All this is based on the new user interface design paradigm that has been adopted across SAP MDG's master data domains.

SAP Master Data Governance Integration with SAP Data Services and SAP Information Steward

Native integration between Data Services, SAP MDG, and Information Steward is critical to ensuring ongoing data quality, data remediation, duplicate prevention, validations, and data enrichment. Combined, these three applications provide a holistic solution for governing master data in the SAP Business Suite.

SAP MDG focuses on comprehensive master data governance ready for use in SAP applications, through ready-to-run governance processes for specific master data domains. In addition, the SAP MDG application foundation allows for custom-defined master data objects, processes, and user interfaces. SAP MDG provides prebuilt validation against SAP business logic and customers' configuration settings and can distribute master data to SAP and non-SAP applications.

Further details of SAP MDG are covered in Chapter 12.

4.4 SAP Solutions for Enterprise Content Management

Enterprise content management (ECM) is a broad area that includes the capturing, managing, preservation, and delivery of content and documents to organizations and business processes. ECM focuses on unstructured information. Some consider ECM to be a completely separate discipline from EIM, but SAP takes a different view; documents and other content are critical information assets that must be managed in a holistic information management strategy. ECM covers a wide range of technologies and use cases, extending EIM to unstructured content that is critical to the business. ECM delivers solutions that focus on the following business pain points:

- Inability or excessive effort to find and reuse relevant content
- Exponential growth of unstructured content
- Content fragmentation across applications and content silos
- "Tribal knowledge" vs. enterprise information
- Need for long-term preservation and legal compliance
- Cost reduction and increased process efficiency via digitization of paper documents

These pain points reduce business efficiency as it relates to content and increase legal and compliance risk. *Reduced business efficiency* includes the inability to find what is needed to complete a business process. For example, before you sign a contract with a new vendor, a specific spreadsheet or email needs a final review, but where is it? Perhaps it is in an application that the user can't access or in a file share for which the user has no permissions, or the content has been fragmented and parts are stored among several applications. There could also be paper documents that are difficult to tie to a digital process.

Increased legal and compliance risk includes content that was not kept for the legally required duration (for example, someone got rid of an old file without realizing that some documents for a legal hold were inadvertently deleted as well), or content was kept longer than legally required. Additionally, regulation requires auditable content lifecycle records.

Table 4.3 depicts a list of concepts and activities typically considered to belong to the ECM solution space. SAP provides solutions for each of the ECM components listed in the table.

ECM Process	Key Features
Document Management	▶ Basic content handling capabilities such as retrieval, storage, versioning, and check-in/check-out ▶ Metadata handling ▶ Security handling ▶ Audit trails
Input Management	▶ Capture manually created or scanned information (office documents, forms, or rich media) ▶ Capture application-created information (XML, financial applications, or e-forms) ▶ Recognize printed or written information (optical character recognition, barcode, or hand-printed character recognition)
Output Management	▶ Assemble documents from text fragments, data, or a clause library (based on simple rules or complete business logic) ▶ Make transformation (conversion, syndication, and compression)

Table 4.3 Scope of Enterprise Content Management Solutions Provided by SAP

ECM Process	Key Features
Output Management (Cont.)	▶ Perform security (digital rights management, watermark, and digital signatures) ▶ Distribute (Internet, intranet, email, fax, CD/DVD, or paper)
Records Management	▶ Manage enterprise information as a business record ▶ Perform file plan classification (automatically and interactively) ▶ Fulfill compliant standards (such as U.S. Department of Defense 5015.2 certification and International Standard Organization 15489 standard) ▶ Map classifications to retention rules and support of disposition management
Web Content Management	▶ Create web content (manage content with template support) ▶ Include documents or social media/collaboration capabilities in web content ▶ Manage web content lifecycle and publish content to different channels
Digital Asset Management	▶ Support consuming, annotating, cataloging, storage, and retrieval of digital assets (photographs, videos, and music) ▶ Manage metadata for digital assets, copyrights, and distribution rights
Email Management	▶ Email classification ▶ Integrate emails into business processes or records

Table 4.3 Scope of Enterprise Content Management Solutions Provided by SAP (Cont.)

SAP's ECM solutions provide end-to-end support for processing unstructured content, including activities such as capturing, managing, storing, preserving, sharing, and distributing content. Capabilities for managing unstructured content in SAP's ECM solutions include:

- **Content creation and capture**
  This produces new content with editing capabilities and handling of inbound information as content (for example, content provided via email or scanning).

- **Content discovery, sharing, and collection**
  This covers all activities around finding and searching content and browsing through existing content repositories. Created content has to be shared and published. A typical example of sharing content is to create web content and publish the information in an intranet. Collaboration capabilities are crucial to working jointly on content.
- **Content governance and compliance**
  This is required to freeze content (for example, via generating stable outputs in PDF) or to treat content as a record, which means fulfilling predefined compliance rules for the content. An example of a compliance standard is the U.S. Department of Defense (DoD) 5015.2 certification.
- **Content archiving and destruction**
  This process ensures a proper, total cost of ownership-minimizing end of life for content.

This section discusses the actual SAP products that are associated with the tasks of ECM. We begin by offering an overview of all the ECM solutions, and then dive into a more specific discussion of the most commonly used solutions: SAP Extended Enterprise Content Management by OpenText, SAP Document Access by OpenText, and SAP Archiving by OpenText.

4.4.1 Overview of SAP's ECM Solutions

To deliver on the full breadth of functionality and to offer an enterprise grade ECM offering, SAP's extended and full-fledged ECM offerings include both SAP solutions and solutions provided by OpenText as solution extensions. The OpenText products have been tested to conform to the SAP product standards, can be purchased from SAP, and are maintained via SAP standard support mechanisms.

The following is a list of ECM solutions, divided into SAP solutions provided by OpenText and SAP solutions provided by SAP.

SAP's ECM Solutions Provided by OpenText

- **SAP Document Presentment by OpenText**
  Automates document generation and personalization. Examples include billing statements, campaign execution, and customer-focused marketing messages.

- **SAP Digital Asset Management by OpenText**
  Provides optimized management of rich media assets, as well as the ability to integrate media into marketing communications. For example, images for a new washing machine that might be used on the web, for mobile devices, or printed on a brochure. These washing machine images could be part of a marketing campaign for new washing machines, with all images managed by brand and model.
- **SAP Invoice Management by OpenText (including optical character recognition)**
  Minimizes capture effort and human intervention and streamlines accounts payable processes (for example, scanning invoices to create the invoice document in the SAP system).
- **SAP Travel Receipt Management by OpenText**
  Automates receipt handling and streamlines the expense management process.
- **SAP Employee Management by OpenText**
  HR file administration; enables self-service scenarios such as a requests to see personnel files.
- **SAP Portal Content Management by OpenText**
  Manages version control, sharing, and collaboration of documents in SAP NetWeaver Portal.
- **SAP Portal Site Management by OpenText**
  Manages web content in a company's SAP NetWeaver Portal installation.
- **SAP Archiving by OpenText**
  Storage of information, including scan and viewer technology. For example, it stores scanned images of invoices in the archive and links them to the invoice in the SAP ERP system.
- **SAP Document Access by OpenText (includes SAP Archiving by OpenText)**
  Offers 360-degree information access with SAP GUI context, meaning that from within SAP GUI, users can get a complete view of all content associated with a transaction, master data, or business process. For example, when you look at a plant asset, such as a machine, all manuals and pictures associated with a physical asset on the plant floor are available in the SAP asset transaction. This includes

the ability to check in and check out documents, get authorization and access rights, and search for documents.

- **SAP Extended ECM by OpenText (includes SAP Archiving by OpenText, SAP Document Access, and SAP Portal Content Management by OpenText)**
  Connects unstructured content with SAP business processes—including document and records management, collaboration, archiving, scanning, information retrieval—that interconnect with SAP software and support core business processes. For example, you can associate a picture, resume, or performance review with an employee.

Figure 4.10 shows several of SAP's ECM solutions provided by OpenText.

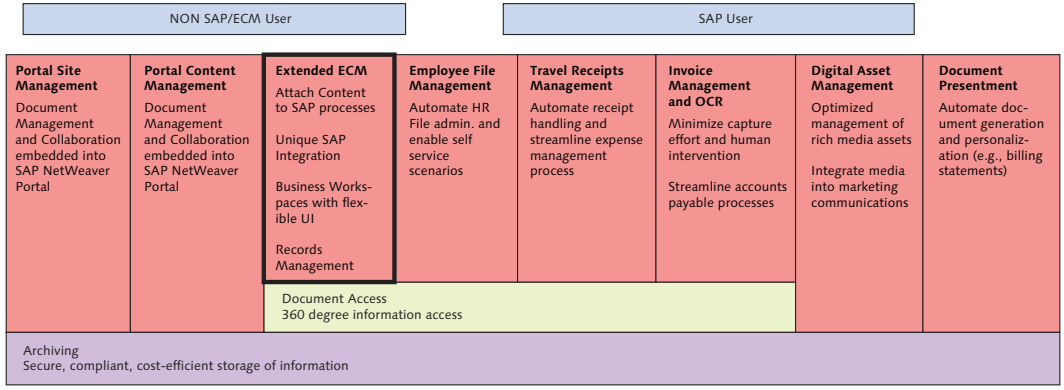

Figure 4.10 SAP's ECM Solutions Provided by OpenText

In Figure 4.10 notice that the bottom horizontal solution is archiving. All OpenText solutions use the archiving layer for secure, compliant, and cost-effective storage of all content. Notice Document Access above the archiving layer. Document Access is available in the SAP GUI to view content related to a business object; for example, a customer. Also notice at the top that all content is available to SAP and non-SAP users with security access to the content.

SAP's ECM Solutions Provided by SAP

In addition to the solution extensions provided by OpenText, SAP offers a range of products that provide content management capabilities:

- **SAP Product Lifecycle Management (PLM) Document Management System and SAP Easy Document Management**
  Supports the scenarios delivered with SAP Product Lifecycle Management, including integration with computer-aided design tools (which, for example, lets you store CAD files of a new engine in the SAP system and automatically create and update material master data in the SAP system based on the information maintained in CAD tools).
- **SAP cFolders**
  Supports seamless collaboration on documents and material master data with external parties such as suppliers. This can be done without giving them access to your corporate system within the firewall.
- **SAP NetWeaver Folders Management**
  Allows customers to build intuitive, structured views of diverse application data to boost business user productivity and support workflows on these folders and cases (for example, to manage dispute cases or change cases).
- **CRM Content Management in SAP CRM**
  Attachment service that allows users to attach documents to any CRM object (for example, a lead within the SAP CRM web interface).
- **Email Response System (ERMS) capability in SAP CRM**
  Manages automatic replies to emails (for example, a notification that your email was received and the recipient will get back to you as soon as possible).
- **SAP Enterprise Workspace for SAP NetWeaver Portal**
  Supports easy creation of collaboration spaces, workgroups, and teams via SAP NetWeaver Portal.
- **Knowledge Management in SAP NetWeaver Portal**
  Supports basis document management functionality in the SAP NetWeaver Portal frontend.
- **SAP NetWeaver Information Lifecycle Management (ILM)**
  Manages retention and retirement of content associated with SAP via SAP ArchiveLink business processes in system decommissioning and retention management scenarios.
- **SAP StreamWork**
  Content capability in the cloud for collaboration on documents and projects (for example, writing an EIM book by collaborating with SAP and non-SAP authors).

- **SAP NetWeaver Enterprise Search**
  Has search capabilities for structured and unstructured content.
- **SAP Content Server**
  Content server delivered by SAP and supporting SAP ArchiveLink to support document storage.

4.4.2 SAP Extended Enterprise Content Management by OpenText

SAP Extended ECM is a full ECM suite (including document management, collaboration, capture, workflow, content access, records management, and archiving) and combines content management capabilities with deep SAP integration to provide content-enriched SAP business processes. It provides a new paradigm where content is managed *within* the business processes instead of *outside* the business processes. This is shown in Figure 4.11.

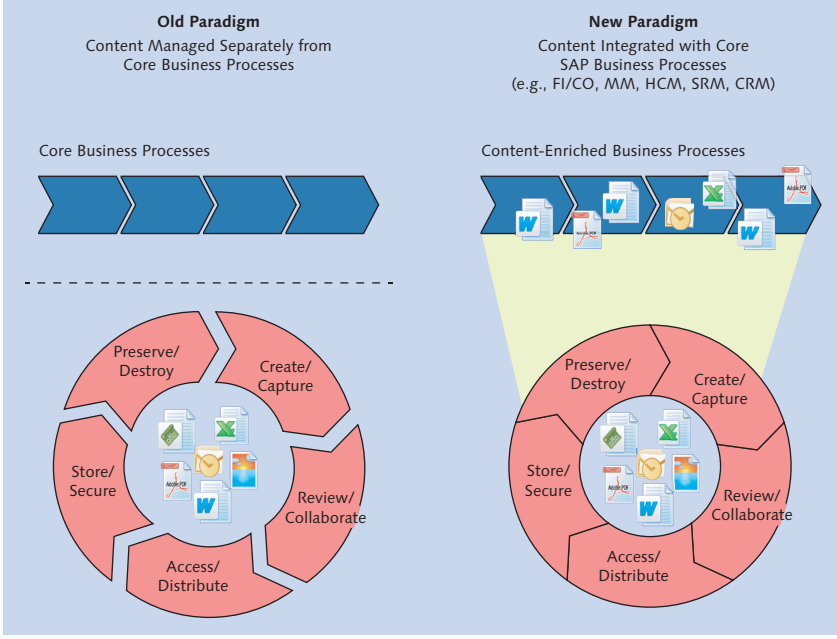

Figure 4.11 SAP Extended ECM Integrates Content with SAP Business Processes

Figure 4.11 depicts the integration of content and the business process with the content-enriched business processes on the right side of the figure. The integration between SAP systems and SAP Extended ECM is described next.

Deep Integration with SAP

The deep integration of SAP Extended ECM with other SAP systems includes work spaces that natively extend SAP processes. SAP Extended ECM is seamlessly integrated inside SAP applications such as the SAP Business Suite and SAP NetWeaver Portal to enable SAP users to easily access content. In addition to accessing content in SAP's native user interfaces, content can be accessed from an SAP Extended ECM web user interface, Microsoft Outlook, and Microsoft Windows Explorer. The support for many user interfaces allows for sharing of content between different stakeholders in an organization, independent from what applications the content originates or what a user's preferred user interface may be.

The deep connection with SAP extends beyond the user interface to data, structures, roles, and permissions. Basically everything customers have already standardized around SAP processes can be leveraged for ECM use cases. The impact is that SAP and ECM are always in sync. As shown in Figure 4.12, the SAP Business Suite is integrated not only in the frontend, but also in the backend by data referencing and data, user, and authorization synchronization, which supports single sign-on.

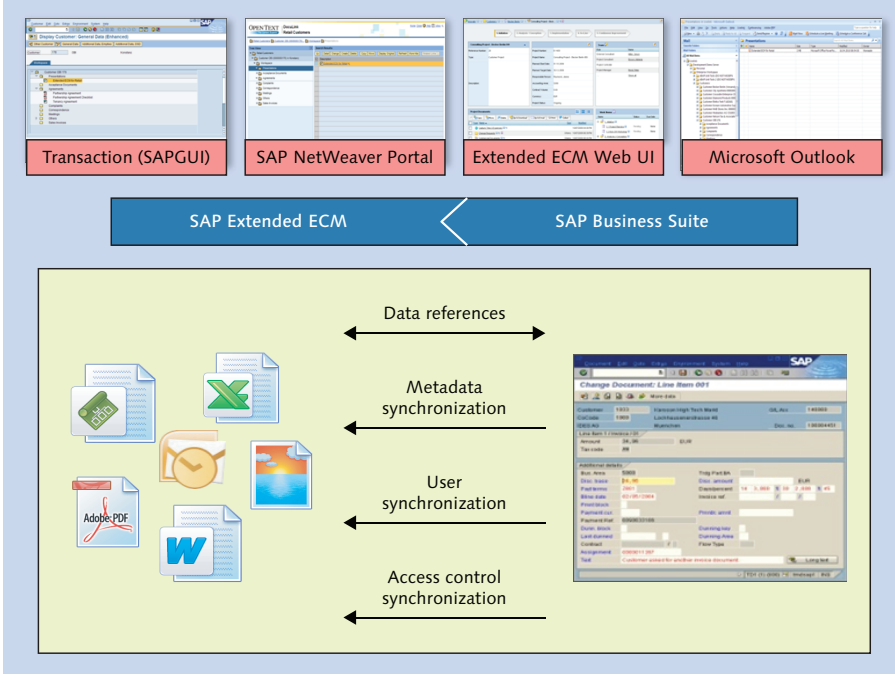

Figure 4.12 SAP Extended ECM Integration with SAP Business Suite

In Figure 4.12, notice the constant synchronization of data references between SAP Extended ECM and the SAP Business Suite. Metadata synchronization, user synchronization, and access controls are driven from the SAP Business Suite to SAP Extended ECM, meaning that the security setup in your SAP system is inherited and applied to SAP Extended ECM. Also notice the variety of user interfaces in the figure. The content can be accessed from SAP GUI, SAP NetWeaver Portal, an SAP Extended ECM user interface, Microsoft Outlook, Lotus Notes, web browsers, and Windows Explorer (not included in the figure).

Examples of Content-Rich SAP Business Processes

SAP Extended ECM's deep integration provides very content-rich SAP business processes in the SAP Business Suite. Some examples of those content-enriched processes are project management, product lifecycle management, procurement, quality management, case and claims management, asset management, and customer relationship management. In each of these examples the process is enriched due to the content.

One specific example is SAP Enterprise Asset Management, which is used heavily in SAP Plant Maintenance for processes such as equipment failure analysis, shutdown turnaround, and general change management. In this example, SAP Extended ECM leverages and extends SAP Enterprise Asset Management for plant maintenance, providing content management and collaboration capabilities in every step related to the analysis and reporting the results to various stakeholders. The following is the process in SAP Plant Maintenance:

1. **Create maintenance notification**
   When the maintenance notification is created, a work space is created in SAP Extended ECM that includes the notification documents. The work space allows you to add additional content for sharing.
2. **Create work orders**
   As the work order is created to investigate the problem, a failure analysis process is triggered in SAP Extended ECM. SAP Extended ECM collects failure data, scope, photographs, analysis, email correspondence with suppliers, and a final report for review.
3. **Schedule and fix**
   As the technician is assigned and works on documents, all photos, manuals, and

correspondence—that is, all content required for the repair—is provided via SAP Extended ECM.

4. **Completion**
   The root cause analysis can require updates to maintenance plans, equipment repair, and so on. All information is captured in the final analysis report managed by SAP Extended ECM.

Figure 4.13 shows an example of the content integrated in asset management.

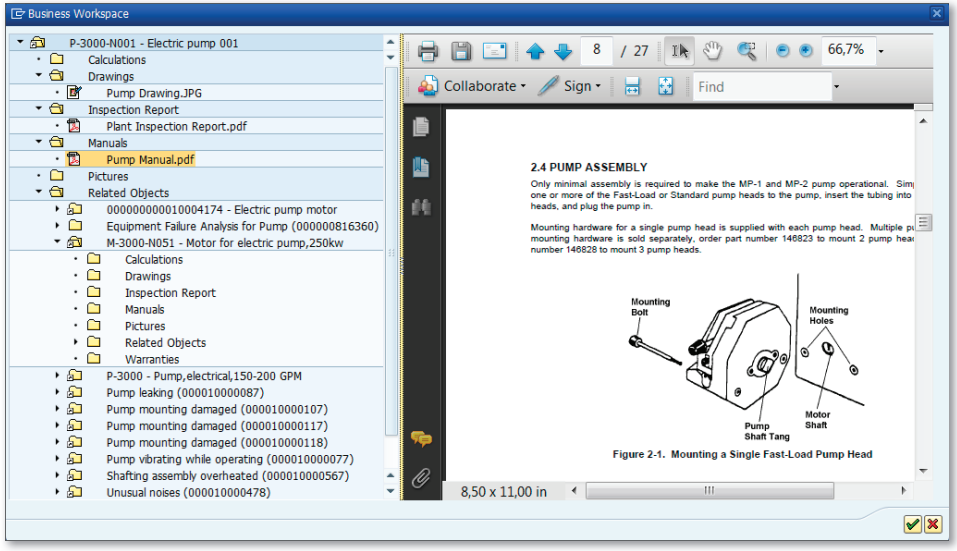

Figure 4.13 Example of Content Integrated in Asset Management

Figure 4.13 shows a workspace in SAP Plant Maintenance where the user can access all content regarding a specific piece of equipment. In the example it is an electric pump. The user can access the manuals, inspection reports, pictures, drawings, warranties, and so on required to maintain and fix the electric pump. (Note: Workspaces will be discussed in more detail in Chapter 14.)

4.4.3 SAP Document Access by OpenText and Archiving by OpenText

SAP Document Access by OpenText and SAP Archiving by OpenText are two key capabilities included in SAP Extended ECM that enable great flexibility when managing unstructured content. These capabilities are used heavily in both SAP Extended

ECM and SAP NetWeaver ILM (covered in Section 4.5). SAP Document Access by OpenText and SAP Archiving by OpenText are discussed in more detail next.

SAP Document Access by OpenText

SAP Document Access is powerful enterprise content management software that integrates fully with your SAP applications. It allows you to store, manage, and retrieve documents across SAP transactions or applications such as SAP ERP or SAP CRM. Content can also originate from various sources such as host systems, legacy software, or customer-specific applications. SAP Document Access provides three key features:

- **Ease of use**
  Enables even occasional users to quickly retrieve content by working with the intuitive, customizable navigation based on folder structures.
- **Ready access**
  Offers 360-degree views of SAP documents and SAP online and archived data.
- **Web interface**
  In addition to the SAP GUI interface, the web interface provides a convenient web user interface that can be integrated into a customer portal or in the SAP NetWeaver Portal component.

An example of using SAP Document Access is providing front office employees with a complete view of all customer interactions (for example, customer messages, emails, and letters written to the company), all brought together in one place so the employee can quickly gain insight into the customer's request.

SAP Archiving by OpenText

SAP Archiving by OpenText securely stores all types of documents and links them to SAP transactions via SAP ArchiveLink. SAP Archiving by OpenText also minimizes your database growth and improves performance, resulting in a lower total cost of ownership for your SAP software. Combined with SAP NetWeaver ILM, the archiving function lets you address the compliance requirements associated with electronic document retention regulations from government agencies (such as the U.S. Securities and Exchange Commission and the U.S. Food and Drug Administration) and mandates (such as the Sarbanes-Oxley Act and Germany's Principles of Due Operation of Electronic Bookkeeping Systems and Principles of Data Access and Auditing of Digital Documents).

What Is Archiving?

The term *archive* has two meanings:

- The relocation and removal of content from your systems (for example, removing all closed purchase orders over 10 years old and relocating them to a storage system).
- Associating content with data in your system (for example, scanning invoices or emails associated with a vendor).

SAP Archiving by OpenText helps with both processes.

As you can see, ECM is a very large area. Chapter 14 focuses on SAP Extended ECM and how to get started using SAP Extended ECM in SAP ERP.

4.5 SAP NetWeaver Information Lifecycle Management

Information has a predictable lifecycle. It is created, lives, and changes in databases, repositories, and software solutions. It is archived and eventually destroyed. A comprehensive strategy for information lifecycle management (ILM) focuses on maintaining a balance between total cost of ownership, risk, and legal compliance. It specifies steps to define, document, and introduce ways to better manage the information that exists across the organization. The ILM process includes knowing and categorizing your data, defining policies that govern what you do with the data, and setting up your system in such a way that you can apply these policies to your data. You then implement your information management strategy with the help of technology. The added value of ILM compared to more conventional information management strategies is automation and completeness.

Several factors have converged to bring about a major change in the way companies manage data. For example, awareness of the unforeseen events that could occur—such as natural disasters, terrorist attacks, or financial meltdowns—has set in motion new efforts to protect consumer identity and privacy. This awareness has led to an explosion in legal requirements. There are currently thousands of regulations worldwide pertaining to the handling of electronic data, and that number continues to grow. Another factor that is becoming increasingly urgent is the management of legacy systems in both SAP software and non-SAP software landscapes. Not only are the aged systems generating ongoing and unnecessary costs through the consumption of electricity and the need for administration and maintenance effort, but they pose a significant legal risk. The data they hold must

be accessible and ready to be presented at relatively short notice (for example, to an auditor in the event of a tax audit). However, as knowledge about the old systems gradually fades from the company and the old hardware is physically removed, it is difficult to predict whether that data will be accessible when needed. The management and retention of information has become so important that traditional data management approaches are no longer sufficient. An effective ILM strategy is an essential part of an enterprise's overall strategy for dealing with the challenges of cost, compliance, and risk. SAP has developed an approach to ILM that meets the complex information management needs of today's organizations.

SAP NetWeaver ILM comprises the policies, processes, practices, and tools used to align the business value of information with the most appropriate and cost-effective IT infrastructure from the time information is created through its final destruction. The complete, flexible, and automated approach to ILM advanced by SAP gives you the support you need to adapt to constantly changing regulations. It helps you accomplish the following steps:

- Identify and categorize your data
- Define policies that govern what you do with data
- Leverage application and storage technology in such a way that you can apply these policies to your data
- Implement your information management strategy
- Shut down your legacy systems in a legally compliant manner

Figure 4.14 shows the evolution from data archiving to retention management to systems decommissioning. This evolution is discussed in more detail next.

SAP Archiving by OpenText versus SAP Data Archiving

SAP Archiving by OpenText is a product used for the storage of content. For example, it is used to store scanned invoices in an SAP ERP system. SAP Archiving by OpenText is also used for the storage of those scanned invoices after they are retired from the SAP system. This product was discussed in Section 4.4.3, and is not the focus of this section.

In this section, we talk about *SAP Data Archiving.* This refers to the *relocation* of business-complete data from the SAP application database to the *storage system*. The archived data can be accessed from the original system at any time for read-only activities, such as reporting, for example, the archiving of closed invoices older than 18 months.

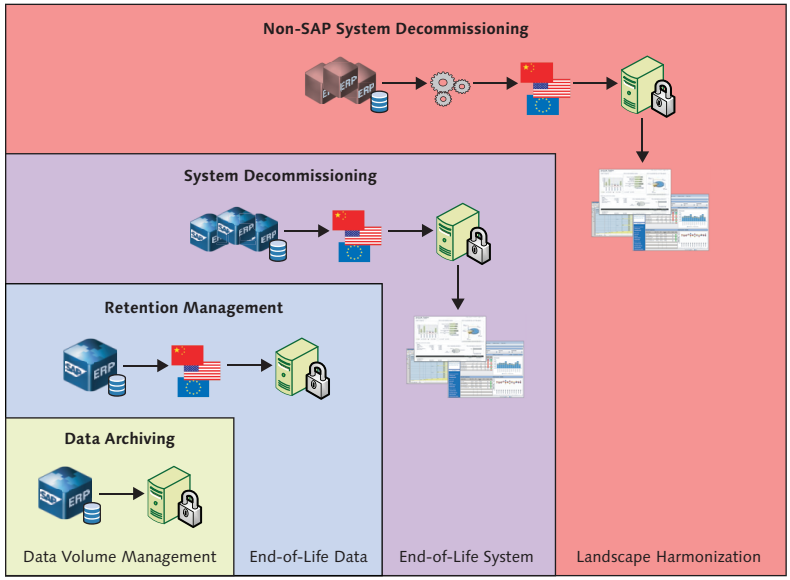

Figure 4.14 Evolution of Data Archiving

SAP NetWeaver ILM has evolved from a simple data archiving solution to a comprehensive solution for determining how long and in what ways data should be kept and used and the means for decommissioning SAP and non-SAP systems. It works with Data Services and SAP Archiving by OpenText for a complete ILM solution. SAP Data Archiving has been around for many years and lets you archive selected business objects from business-complete processes, relocating them from the database to an external storage system. Examples include closed purchasing documents, order history, and technical items such as IDocs and workflow work items. SAP Data Archiving has evolved into retention management, enabling you to proactively monitor how long you need to keep data in your system and placing legal holds as required on selected data. The further expansion of SAP NetWeaver ILM includes decommissioning of SAP and non-SAP systems that are no longer required.

SAP NetWeaver ILM provides three major use cases and scenarios for use:

- **SAP Data Archiving**
  Again, this data archiving should not be confused with SAP Archiving by OpenText. It is a product that enables you to associate images, such as scanned invoices, with SAP business objects. *SAP Data Archiving* is a core component of SAP

NetWeaver ILM that enables you to move business-complete data (transactional and master data) that is no longer required in everyday business from the database into long-term, less expensive storage. The data is still available to meet regulatory or other legal requirements when it is stored in the archive.

- **Retention management**
  SAP tools and technologies for SAP NetWeaver ILM provide retention policy management functions that support the complete information lifecycle from creation to retention to destruction. They enable you to enter different rules and policies reflecting various criteria, including where data is stored, the duration of data retention, and when data can be destroyed. The policies, usually based on external legal requirements or internal service-level agreements, can be applied to both structured and unstructured data stored on all types of media. Based on the rules, the data is automatically archived into the correct storage area and receives an expiration date. SAP NetWeaver ILM-aware storage integration means your storage systems understand and can act on the stored data based on rules you define. At the end of the lifecycle, an automated destruction function permanently destroys the archived data when the expiration date has been reached. SAP NetWeaver ILM includes functions that support legal-hold management and automated electronic discovery.

- **System decommissioning**
  SAP NetWeaver ILM provides a complete approach to shutting down legacy systems and bringing the data from both SAP and non-SAP software into a central SAP NetWeaver ILM retention warehouse. In the retention warehouse, the data is stored with the relevant expiration dates in SAP NetWeaver ILM-aware storage and can be viewed with different reporting options even after the original system has been shut down. Functions of SAP NetWeaver BW as well as local reporting functions are used to extend the SAP NetWeaver ILM retention warehouse with centralized auditing and reporting functions. Included are functions that enable you to create snapshots of master and customizing data and provide on-demand analysis of archived data.

The major SAP NetWeaver ILM components that support these use scenarios are retention management and system decommissioning capabilities within SAP NetWeaver ILM, which are discussed next. Both these components integrate natively with SAP Archiving by OpenText and SAP Document Access.

Note

SAP Archiving by OpenText is used in SAP Extended ECM by OpenText and SAP NetWeaver ILM. When removing data and content from SAP systems, SAP NetWeaver ILM manages all SAP-structured data and all associated content (for example, scanned invoices, employee pictures, etc.) via SAP ArchiveLink. SAP Extended ECM manages the archiving of all non-SAP–related content, such as collaborative documents, emails, physical boxes of old documents not in an SAP system, or web images (for example, used for a television advertisement) not associated with an SAP system.

4.5.1 Retention Management

Retention management provides a range of dedicated tools and functions to help you adequately control the retention of your data along its entire lifecycle. This holds true for structured information, such as a sales documents posted in the SAP application, as well as for unstructured information such as scanned invoices. Once a policy is set up and activated in the SAP NetWeaver ILM policy engine, data is protected against destruction until its expiration date has been reached. Retention management is based on SAP Data Archiving technology, so you do not have to start from scratch if you are already using SAP Data Archiving. You can use retention management functions directly within your SAP ERP application as of SAP ERP enhancement pack 4 or higher. Retention management includes the follow capabilities:

- Archiving data from open and completed business processes
- Storing archived information on an SAP NetWeaver ILM-certified server (to guarantee nonchangeability of the information and protect from premature destruction)
- Determining and managing retention rules
- Inclusion of retention rules for stored information so the destruction is timely and in adherence with the retention periods required by law
- Using stored information to create reports for tax auditing and product liability
- Retrieving information for legal cases and defining legal holds
- Destroying stored information after the retention period has expired, assuming there is no legal hold

More details on these capabilities will be provided in Chapter 13.

4.5.2 System Decommissioning

System decommissioning enables you to consolidate multiple large legacy systems into a small single instance and retain on-demand access to data from legacy systems as well as respond to tax audits and create reports. The legacy data is stored in the SAP NetWeaver ILM retention warehouse, which serves as a single repository for multiple systems. When the data is required for legal, audit, or tax purposes, it is exposed to a data warehouse such as SAP NetWeaver BW (and, in the future, SAP HANA) for flexible and modern reporting capability. The business knowledge and context is preserved during the decommissioning even as retention policies are being enforced. After moving retention-relevant data from the legacy system to the SAP NetWeaver ILM retention warehouse, you can shut down the legacy system for good.

In a typical decommissioning scenario, you have one or more systems that will be shut down and a dedicated retention warehouse system that will receive the data you want to retain from the original systems. The retention warehouse consists of three pieces:

- A dedicated SAP ERP application running only SAP NetWeaver ILM
- A dedicated implementation of SAP NetWeaver BW
- Third-party SAP NetWeaver ILM-certified storage hardware

In the retention warehouse, data from decommissioned systems is organized according to audit areas and covered by the retention policies you define and manage using the software. This means retention times and legal holds apply to the data maintained in the retention warehouse, just as they do for non-decommissioned (or live) systems.

The system that is to be shut down is first emptied of its data using standard and SAP NetWeaver ILM-enhanced archiving objects and programs. You can archive all the business-complete data and then use snapshot functionality to archive data from objects that are not yet business-complete, as well as any related master and context data. Figure 4.15 shows the decommissioning of several systems to a single SAP NetWeaver ILM system and related storage. Next, we discuss a typical system decommissioning scenario and some of the required steps.

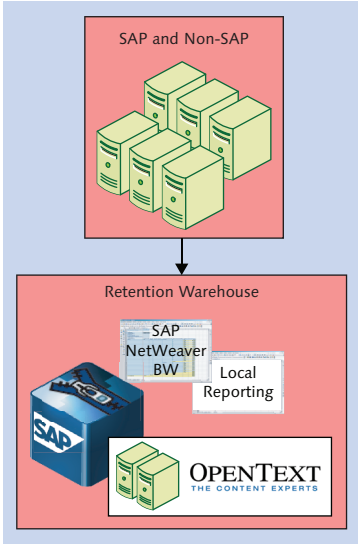

Figure 4.15 Decommissioning Systems Using SAP NetWeaver ILM

A Typical System Decommissioning Scenario

For a decommissioning project, there are multiple steps that can be roughly grouped into two categories: steps for the legacy system and steps for the retention warehouse in SAP NetWeaver ILM.

- **Steps for the legacy system**
  - Analyze the system
  - Enable the legacy system for archiving
  - Prepare the legacy system for shutdown
  - Extract data from the legacy system
  - Conduct testing and quality assurance
- **Steps for the retention warehouse**
  - Install an SAP ERP application with only SAP NetWeaver ILM and SAP NetWeaver BW, and connect to an SAP NetWeaver ILM-aware storage system
  - Set up the retention warehouse
  - Set up audit areas, policies, and rules
  - Move archived data to the retention warehouse system

- Convert and store the archive files and apply rules
- Generate business warehouse queries using archived data or run a local reporting option

The following sections discuss several of the most important steps in more detail. You can find more details on SAP NetWeaver ILM in Chapter 13.

Analyze the Legacy System

It's not uncommon for IT departments to have only limited knowledge of their legacy systems, particularly with respect to the amount and types of data they contain. If you don't know what your legacy system "looks" like, you have the option of either moving everything over to the retention warehouse system or, once you run an analysis of your system, moving only the specific data you will need later for auditing and reporting.

Enable the Legacy System for Archiving

Once you have identified the data you need to retain and the tools you need to retain it, you can enable the legacy system for archiving. Since SAP NetWeaver ILM is delivered with SAP ERP, your older, legacy system has to receive the necessary SAP NetWeaver ILM functions. (SAP NetWeaver ILM natively supports legacy systems back to R/3 4.6c. For older releases additional consulting is required.) This includes enhancements for archiving objects, snapshot functionality, and the context data extractor. The context data extractor functionality is preconfigured for tax reporting and product liability reporting.

Extract Data from Legacy Systems

Once the legacy system and retention warehouse are prepared, you can extract the data from the legacy system. This includes enhanced data archiving functionality, such as the snapshot option (one of the archiving options), and the context data extractor functionality. After you have "emptied" the legacy system, moved data to the retention warehouse, and performed testing, you can shut the legacy system down.

SAP NetWeaver ILM provides a streamlined IT infrastructure and lowered costs through the safe and efficient decommissioning of redundant systems. It provides reduced risk by delivering full auditing and reporting functionality and rules-based enforcement of retention policies and preventing destruction of important data.

This enables greater legal compliance by automatically managing the retention of data from your SAP applications based on retention policies and rules you define.

Set Up the Retention Warehouse

After preparing your legacy system to be shut down, you can set up the retention warehouse to receive the data from the system you want to decommission. Your dedicated SAP ERP application should be set up and connected to a third-party storage system, which must be an SAP NetWeaver ILM-certified product. If you are using SAP Archiving by OpenText for your image and content storage, you can automatically use it for SAP NetWeaver ILM as well. (To find an SAP NetWeaver ILM-certified partner, visit *http:/www.sdn.sap.com/irj/sdn/interface-certifications*.) In addition, you need to connect SAP NetWeaver BW. SAP recommends that you use a dedicated business warehouse system for your data retention warehouse.

This section introduced SAP NetWeaver ILM. More details can be found in Chapter 13.

4.6 SAP Solutions for Data Migration

Migration projects are relevant for all types of companies, including those with SAP applications, non-SAP applications, and those who are upgrading—any company that needs to migrate data from one application to another. This includes companies that are:

- Acquiring or merging with other companies
- Expanding into new lines of business
- Initiating green IT projects to reduce hardware and environmental and monetary costs
- Harmonizing data and processes across organizational boundaries
- Conducting a global SAP rollout

Because of this, data migration is often a first entry point with EIM, and it serves as a compelling use case for starting data management and information governance initiatives across the enterprise.

Planning for a data migration project is a major task. Getting data cleansed, transformed, and loaded into a new application can seem daunting. Add to that task the following common questions:

- How much of my data and related content do I take?
- Is my data ready to be migrated?
- What do I do with the legacy application after the migration?
- How long do I need to keep the data in the new application?
- How do I handle contracts and documents related to my data?

Dealing with the source system, content, and how long to keep data are often discussion items that go to the back burner because of time constraints, resource constraints, political constraints, and the tactical pressure to get the base migration completed. By extending the conversation from data migration to systems migration, the questions of the source system, information retention, and content management can be addressed, feeding into an overall coherent information management strategy.

Next, data migration will be discussed in more detail, explaining SAP Best Practices and Rapid Deployment Solutions for rapid data migration, and then systems migration.

4.6.1 SAP Data Migration Software

SAP Data Migration uses many of the software components already discussed. This means the data migration solution uses Data Services and includes some SAP BusinessObjects Web Intelligence reports.

As discussed earlier, Data Services delivers data integration and data quality as part of its data foundation. With its native connectivity to many applications, it is the perfect software capability to drive data migration projects. Information Steward lets you get to know your source data very well, and has metadata capabilities to reconcile what was loaded in a migration cycle with the source data. The SAP BusinessObjects BI platform provides reports on migration progress. Using business intelligence capabilities, such as SAP BusinessObjects Web Intelligence, enables a business user to easily distinguish the good and bad data in an easy-to-read report. The six major steps of SAP Data Migration are shown in Figure 4.16.

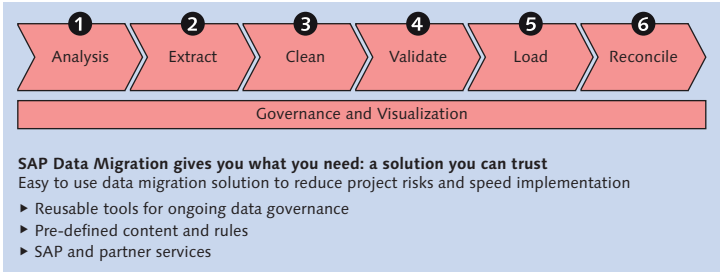

Figure 4.16 SAP Data Migration

The first step in SAP Data Migration is analysis, which includes data profiling and getting to know your data prior to migrating it to an SAP application. This includes the profiling capabilities delivered with Data Services that were discussed in Section 4.1.1, but also the more advanced data profiling capabilities provided with Information Steward. (Information Steward data profiling capabilities were mentioned in Section 4.2 and will be discussed in more detail in Chapter 6.)

Once you understand more about your source data, you can extract source data and place it in a staging area in Data Services. The extraction is based on the Data Services connectivity options mentioned above. After the data extraction occurs, data cleansing starts. This can be simple data cleansing such as checking for null values and correctly formatted zip codes, to very complex cleansing such as address cleansing, duplicate checking, and so on. A good example of data cleansing is provided in Chapter 8.

Once data has been cleansed, the validation phase begins with value mapping. *Value mapping* is when you map how specific values need to convert to other values for the new application. For example, consider a situation where the country in the source system is *United States*, and the country code in the new system is *US*. This would mean that the process of value mapping would map all records with the country code of *United States* to *US*. Value mapping is done with the Migration Services tool provided with the content for data migration. More details on the validation and value mapping are discussed in Chapter 11.

Once you know the data is good, then it is loaded into the SAP system. The load will happen based on the application's requirements for loading. For example, many core objects are loaded via IDocs, which are a proprietary format for loading data into SAP systems. Other options include loading via an industry-specific technique, for example, when loading utilities for banking data or loading via the

Legacy System Migration Workbench (LSMW) or other SAP NetWeaver loading mechanisms. Examples of SAP NetWeaver loading mechanisms include BAPIs, ALE (application link enabling), etc. Once the data has been loaded, a reconciliation is performed to check what was loaded versus what was planned to be loaded.

In Figure 4.16, notice that the data migration solution is based on reusable tools for ongoing data governance, predefined content and rules, and SAP and partner services. The reusable tools for ongoing data governance refer to Data Services, Information Steward, and the SAP BusinessObjects platform. The predefined content and rules will be discussed next, but before moving on to the migration content, we'll explain the SAP and partner services. Normally with a first migration project most companies choose to use an experienced consulting organization. The SAP consulting services organization includes a global team trained in EIM, including data migration. As well as SAP consulting services, SAP has several consulting partners who are also experts in Data Services and the predefined content and rules delivered for data migration. It is recommended that your first data migration project be done with SAP or a partner who has been trained using the migration content provided by SAP (discussed next).

4.6.2 Predefined Content and Rules for Data Migration

When migrating to SAP applications such as SAP ERP, SAP CRM, and SAP ERP Human Capital Management (HCM), as well as industry solutions such as utilities, banking, and so on, it is important to understand what data is expected in the SAP system. This includes understanding the required fields, how the data must be formatted, relationships between data objects, and so on. This knowledge is well known to SAP but unknown to the source system. To expedite the understanding of the SAP data structures, SAP has provided migration content that is available in Data Services. This means SAP has looked at the objects most commonly migrated (business partner, material, cost center, plant, etc.) and has created all required elements in Data Services, mapping to the SAP objects, and reports, in a bundle for data migration.

For data migration, over 100 objects are supported for migration to SAP applications and industry solutions related to SAP ERP, SAP CRM, SAP ERP HCM, utilities, and banking industries. More migration content for additional objects is released each year. The predefined migration content includes:

- **Data validation content**
  This enables mapping of legacy source data to SAP data, including validation of key values. Specific examples of the migration content include:
  - Business partner content such as customer and vendor masters.
  - Logistics content such as material masters, bills of material, purchasing documents, and sales documents.
  - Financial content such as receivables, payables, and cost elements.
  - Sales content such as activities, leads, opportunities, and contracts.
- **Documentation**
  There is a Quick Guide on how to install the content, order to load the data, and how to extend the content provided by SAP.
  - A business process document for data migration provides a process flow and the sequence in which the objects need to be loaded into Data Services.
  - An extension guide provides the information you need to extend and modify the content according to your specific needs.

In addition to migration content for many core objects, there is also content for data quality capabilities including address validation, duplication detection, and data remediation. Also included in the migration content is a structured approach for governance for data migration projects.

All the data migration content is provided by the SAP Best Practices and Rapid Deployment Solutions organization at SAP. This group creates preconfigured solutions that ease implementation and provide scenario-specific solutions out-of-the-box. This content includes everything from specific Data Services content to provided reports, to migration tools to map source data, to validation of SAP fields (for example, U.S.A. and United States must all migrate to the country code US in the target SAP system).

In addition to creating content for data migration, the Rapid Deployment Solutions organization creates other solutions for EIM as well as other components in the SAP system. For example, there is a Rapid Deployment Solution for integrating customer data on SAP NetWeaver MDM. Because new solutions are released quarterly, the best place to look for new solution packages is at *http://service.sap.com/solutionpackages.*

For more details on data migration, go to *http://service.sap.com/bp-datamigration*.

4.6.3 Systems Migration

Systems migration is the migration of data and associated document content from one or more existing stores of information to a new application deployment, including the retirement and decommissioning of legacy data stores. Information includes both data and rich content (text and binary documents, PDF, email, and pictures).

Systems migration includes the data, metadata, unstructured content, and information required to support application migration projects. Also included is the decommissioning of source systems that are no longer required, as well as the archiving of data no longer required in the existing system.

Systems migration combines the capabilities of Data Services, SAP NetWeaver ILM, and SAP Document Access to provide a holistic offering for migrating to new systems and decommissioning the systems no longer required. For more information on systems migration, see *http://wiki.sdn.sap.com/wiki/x/sAWcDQ*.

4.6.4 SAP Rapid Marts

SAP Rapid Marts (hereafter Rapid Marts) provide data extraction from SAP systems, put the data in a data mart (currently supported by various databases and SAP HANA and will be supported soon by Sybase IQ), and provide reporting content. Rapid Marts deliver departmental business intelligence solutions for key subject areas. In most cases, Rapid Marts serve as jump-start blueprints but require customization to meet the customer's requirements.

SAP's array of Rapid Marts is expanding, but it currently provides the following:

- **Finance**
  General Ledger, Accounts Receivable, Accounts Payable, Fixed Assets, and Cost Center
- **Manufacturing**
  Plant Maintenance, Production Planning, and Project Systems
- **Operations**
  Sales, Inventory, and Purchasing
- **Human capital management**
  Human Resources

New Rapid Marts are released each year. For more information on Rapid Marts see *http://www.sap.com/solutions/sapbusinessobjects/data-warehousing/rapid-marts/index.epx.*

4.7 Information Governance in SAP

We introduced the discipline of information governance in Chapter 2 of this book. In this section, we review the concept and then explain how SAP solutions can assist with information governance. It is important to understand that all information management must include governance from both product capabilities and across all processes that update critical information in your enterprise. Information governance is a discipline that provides oversight to ensure successful execution of your information initiatives. It involves the management of people, technology, and processes, as well as policies, standards, and metrics, including the ongoing oversight necessary to manage information risks.

Information governance can be applied to one project, can be started with one data domain, and should be extended to the enterprise. Information governance is a sustained activity.

How do you know if problems you are seeing in your company can be attributed to a lack of information governance? The first step is always to listen. Words such as *mergers, risk, compliance, migration, acquisition, shared services, security,* and *Sarbanes-Oxley* can all point to information governance.

Case Study: New Driver's License in Indiana

In Indiana, 58,600 residents lost confidence in state IT practices when, as Chris Sikich noted in a *USA Today* article, the Bureau of Motor Vehicles (BMV) sent out a notice for drivers to renew their driver's licenses. The notices were intended as a proactive measure, since the state is expecting a glut of renewals in 2012. Because 2.4 million people will need to renew their license, the state spent $247,450 to mail postcards to about 700,000 people with birthdays from January through August.

Of those 700,000 people, 58,600 (approximately 8.3%) received incorrect information. Though, for many recipients, addresses and last names were correct, first names were not. For example, Rod Humphrey received a notice for Brandon Humphrey. Rod just threw out the renewal notice, effectively subverting the proactive measure taken by the BMV. The consequences of this are twofold:

- The renewal notices are more likely to be ignored, and therefore the proactive renewal-avoidance process is less effective.
- The mailing cost of those postcards was wasted. At a cost of 35 cents per mail piece, the postal cost of the error was approximately $20,510.

The next section discusses scenarios for how to get started with information governance.

4.7.1 Information Governance Use Scenario Phasing

The Indiana driver's license case study indicates a data quality problem. However, many other kinds of use scenarios can spur you to start an information governance project. Figure 4.17 highlights common use scenarios that your company may be approaching. The size of the bubbles indicates the relative complexity of each of the use scenarios, paying particular attention to the organizational change management and business process changes required to succeed in the individual use scenarios.

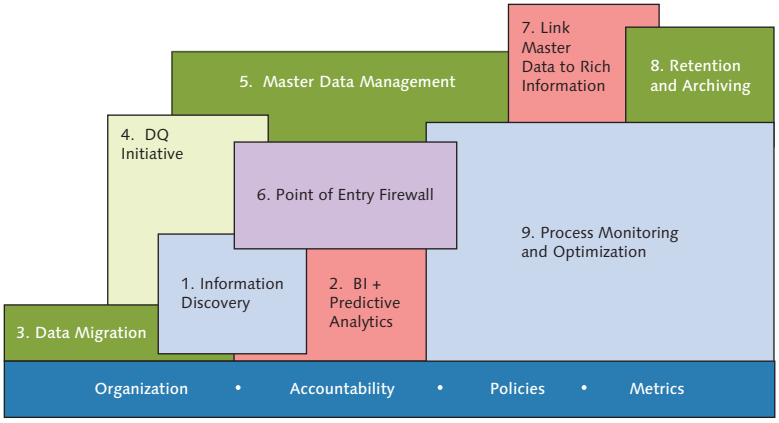

Figure 4.17 Information Governance Use Scenario Phasing

Notice that the use scenarios overlap. For example, if you do the work for an information discovery project, you can reuse some of that work for your business intelligence and predictive analytics project. For a business intelligence and predictive analytics project, you need to know the location of the best information (and the quality of that information) from which to drive the decision making. Information discovery includes activities such as knowing where the information is kept as well as quality level base-lining.

The following list describes each of the use scenarios.

- **Information discovery**
  Finding all of the critical information sources in your company, including discovering (and documenting) quality levels, freshness, and timeliness of those information sources.

- **Business intelligence and predictive analytics**
  Using high-quality, fresh, and timely information to drive decision making throughout the enterprise.
- **Data migration**
  Moving and integrating data as you move from one system to another (for example, moving data from a retiring system to a new version of SAP ERP), including aggregations and quality corrections.
- **Data quality initiative**
  Establishing policies to answer questions such as "What is a customer? What is a material? How can we create a single view of a material? Who needs to create and approve these master records?"
- **Manage master data**
  Creating single points of reference for master data records, including setting up review and approval workflows for creating, changing, and flagging records for deletion.
- **Point-of-entry firewall**
  Proactively fixing and verifying information before it enters productive systems in a live state, including enrichment and verifying against third-party truth sources.
- **Link master data to rich information**
  Correctly linking rich information (documents, PDFs, videos, etc.) to the master data that needs that information, requiring accurate identification and tagging of both rich information and master data.
- **Retention and archiving**
  Establishing retention and archiving policies in compliance with both corporate policies and enforced compliance policies, including both enforcement and manifestation of the policies.
- **Process monitoring and optimization**
  Actively monitoring how well both business processes and information governance processes are performing, including where key bottlenecks are forming and recommendations for optimizing both business processes and information governance processes.

Refer back to Chapter 2 for more on getting started with governance initiatives.

4.7.2 Technology Enablers for Information Governance

You can implement information governance without any new technology purchases. Many customers use standard software to facilitate their information governance project: email, departmental Microsoft SharePoint, and spreadsheets. You can do this. In fact, you *should* start information governance this way because it helps you gain experience and perspective, and grow your information network.

However, keep in mind human nature. You can establish all of the good policies and standards in the world, and people can agree on the inherent goodness of these policies and standards. But if these policies and standards are inordinately difficult to comply with, people will find a way to get around them.

How do you make your information governance program easy to comply with? Technology is a great force multiplier here. Figure 4.18 ties in the EIM technologies mentioned in this chapter with information governance.

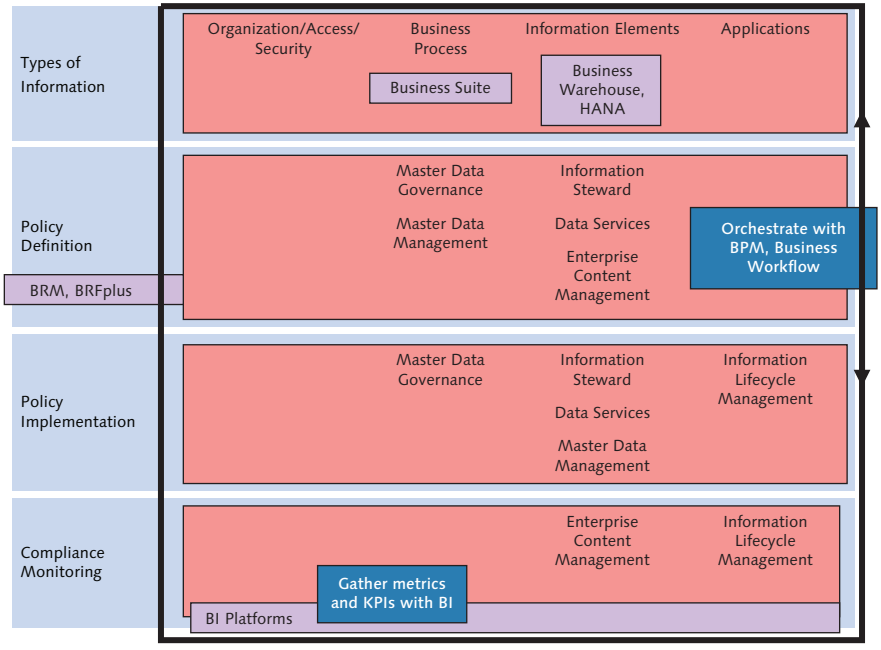

Figure 4.18 Enabling Information Governance Technologies

Figure 4.18 demonstrates that information governance does not only apply to master data. It applies to the organization, including governing user access and monitoring security. Information governance must be concerned with how the business processes use your enterprise information and where all of the data is stored. Finally, which systems access, use, and update the business processes and data elements that truly are enterprise assets?

For all of the above kinds of information (organization, business process, data elements, and systems), you have the following kinds of activities:

- **Policy definition**
  What does "good" look like? How should this information behave? These policies are defined in plain language by the key business stakeholders. The following list mentions the major products where information policy definitions are created.
  - *Information Steward:* This is the place for data stewards to define policies for master data.
  - *SAP Extended ECM:* SAP Extended ECM supports the definition of retention policies determining how long documents should be retained and how access rights should be changed over the lifespan of content. SAP Extended ECM is also able to use SAP data from the SAP Business Suite to define retention periods.
  - *SAP NetWeaver ILM:* SAP NetWeaver ILM supports the definition of retention policies determining how long data and associated documents should be retained. It is possible to synchronize (to create, update, and delete) SAP NetWeaver ILM retention policies via APIs with SAP Extended ECM retention policies.
- **Policy implementation**
  How can you seamlessly implement and enforce the policies you defined—with as much collaboration and as little rewriting as possible? The following software and components can help:
  - *Data Services:* This is the place for IT developers to design, move, aggregate, and understand master data, transactional data, and text data. Also, data stewards and IT developers can collaborate on core questions such as how to define a unique customer at the global level.

- *SAP NetWeaver MDM:* This is an environment for data stewards and IT developers to design and enforce master data storage, provisioning, and quality rules in analytical master data management scenarios.
- *SAP MDG:* Much like SAP NetWeaver MDM, this component takes advantage of the SAP ERP environment as a master data repository, business rule repository, and business workflow processing.
- *SAP Extended ECM:* Records management manages the retention and disposition of paper documents and digital content according to compliance requirements supporting international records management standards.
- *SAP NetWeaver ILM:* Establishes retention and policies for transactional data, master data, and rich content (such as documents), and enforces those policies.

- **Compliance monitoring**
  How can you prove that you are complying with the information policies defined above? For a legal hold, can you prove that you are following industry standards when the audit department comes calling? The following software and components can help:
  - *SAP BusinessObjects BI platform:* Here you can organize information and package it in a way that allows the business users to make decisions based on that information.
  - *Information Steward and SAP MDG:* These enable you to monitor over time how well the data is complying with the business rules.

In addition to the software capabilities previously mentioned for policy definition, policy implementation, and compliance monitoring, SAP NetWeaver Business Process Management (BPM), SAP Business Workflow, Business Rules Framework Plus (BRFplus), and SAP NetWeaver Business Rules Management (BRM) are software components that have a cross-functional role in information governance. They can be used to support policy definition, implementation, and compliance monitoring.

As you can see, information governance requires a holistic look at policies, processes, and procedures to ensure that you are provisioning trusted, relevant, and reliable information for business processes.

4.8 NeedsEIM, Inc. and SAP's Solutions for EIM

Now that you understand the EIM solutions offered by SAP, let's take a high-level look at how these solutions may help NeedsEIM, Inc. (We will go into more detail about this in subsequent chapters of the book.)

Chapter 1 introduced NeedsEIM, Inc. and discussed its business challanges. SAP technologies that can help address them are shown in Figure 4.19. (Note that this is only one possible solution. Depending on your specific landscape, there are other alternate options that can provide similar resolutions for your business challenges.) Notice that the EIM components cover the cleansing and consolidation of information for both operational use in the applications and analytical use in the SAP BusinessObjects Business Intelligence (BI) suite. Using SAP NetWeaver Portal, SAP NetWeaver Business Process Management, and SAP Business Workflow, a master data creation and updating process can be designed that provides clear audit trails of approval. The record is stored in the MDM system, prior to which it goes through standardization, cleansing, and any enrichment processes using SAP Data Services. The data can now be syndicated to subscribing systems such as SAP ERP, SAP Customer Relationship Management (CRM), and so on. If there is any unstructured content that has to be managed and attached to a master data record (such as a W-4, company financials, or a vendor record), the attachment can be triggered as part of the approval workflow, and the documents can be stored in a content repository along with requisite metadata. A one-time migration of existing master data from various systems includes cleansing, matching, de-duplicating, and enriching, which can all be done using SAP Data Services and SAP Master Data Governance (MDG).

The issue of contracts may require a complex contract lifecycle management (CLM) solution or a combination of workflow and ECM solutions along with integration between systems. The analysis required to identify supplier spending may need a combination of information from a vendor master data governance process and an application that tracks spending. This data can be housed in an analytic database such as SAP NetWeaver Business Warehouse (BW) and/or SAP HANA for further analysis. A similar solution can be applied for customer-level profit and loss that would need all types of expense data collected from various applications such as SAP ERP Human Capital Management (HCM), SAP CRM, and SAP ERP along with revenue data from SAP ERP to arrive at the comprehensive profit and loss figure by customer.

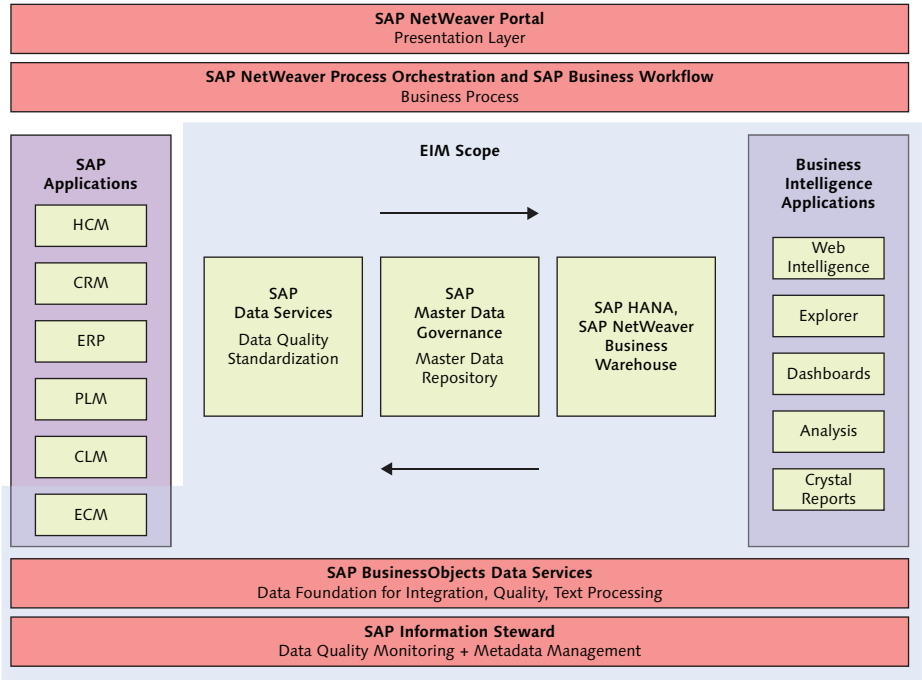

Figure 4.19 Sample Solution Architecture

As you move through Part II of the book, you will see how different capabilities within SAP's EIM portfolio can be used to address the major issues at NeedsEIM, Inc.

4.9 Summary

This chapter overviewed of all of SAP's solutions for EIM. It introduced Data Services for data integration, data quality, and text data processing. More details on Data Services can be found in Chapters 7, 8, and 10. You have seen how managing master data includes both SAP NetWeaver MDM and SAP MDG as an embedded master data solution. SAP MDG is discussed in more detail in Chapter 12. You now understand the breadth of ECM solutions offered by SAP and OpenText. SAP Extended ECM is the primary offering from OpenText and is discussed in more detail in Chapter 14. You understand that information is a risk if kept too long, so data archiving, retention policies, and systems decommissioning must be included in an information management strategy. Chapter 13 provides more details on SAP NetWeaver ILM.

In addition to details on the specific software capabilities, in Chapter 5 you will learn about real-world applications of EIM that are being used today. Chapter 5 is provided by named and unnamed customers who describe how they are using EIM and SAP's solutions for EIM to solve specific business problems.

Index

A

ABAP Web Dynpro Floorplan Manager, 404
Access server, 258
Active area, 390
Address
  cleansing, 110, 112, 152, 290, 296, 299, 337
  correction, 313
  directories, 309
  information, 60
  parsing, 299, 300
  profiling, 237
  validation, 154
AIS, 463
All-world address directory, 309
Analytical EIM, 37, 39
Analytical use, 253
Application integration, 197
Application link enabling (ALE), 387, 397
Architecture, 253
  retention management, 435, 436
  system decommissioning, 435, 436
Archival data, 176, 179
Archive administration data
  transfer, 456
Archive Development Kit (ADK), 429
Archive hierarchy, 434
Archive Management, 458, 459
Archiving, 34, 142, 470
  policies, 158
  strategy, 216
Archiving object, 147, 149
  definition, 429
  ILM-enabled, 434, 453
Archiving-object specific customizing, 458
Assessment, 64, 239
Atomic data, 174
Audit area
  definition, 437
  demo, 438
Audit area (Cont.)
  product liability, 437
  set up, 453
  tax, 437
Auditing, 149
Audit package
  create, 460
  extract to BI, 461
Audit Package Handling, 461
Automated electronic discovery, 145

B

BAdI, 387, 394, 395
BAPI, 265
Bar codes, 477
Batch data flows, 335
Best record strategy, 332
BI, 37, 122, 123, 155, 157, 315, 317, 357
Big data, 45, 46, 47, 48
Binding, 240, 244, 348, 355
Break group, 320, 323, 324
BRFplus, 161, 386, 392, 397, 409, 410, 411, 413
  single value decision table, 413, 415
  user agent decision table, 414
Business Address Services, 109, 334
Business-complete data, 452
Business efficiency, 131
Business glossary, 125
Business-incomplete data, 452
Business intelligence (see *BI*)
Business process descriptions, 125
Business process manager, 50
Business process owner, 50
Business rules, 188, 239
Business Rules Framework, 386
Business term glossary, 119, 223
Business term taxonomy, 224
Business user, 121

C

Capture, 470, 473
CDE, 443, 451
  archive data, 450
  extraction services, 453
Centers of excellence, 60
Central Management Console (see *CMC*), *225*
Central Management Server (see *CMS*)
cFolders, 136
Change request, 389, 390, 417, 421
  create, 421
  process, 386
  type, 411, 413, 418
Checksum function, 455
Cleansing package, 307, 324
Cleansing package builder, 123, 125, 224, 336
Cleansing process, 298
Cleansing rules, 119, 223, 224
CMC, 225, 226, 229, 230, 247, 257
CMIS, 477
CMS, 257
Collaboration, 470, 473
Compliance, 41, 65, *67*
Monitoring, 161
  requirements, 70
Condition alias, 414
Consolidate, 315
Content, 34
  access, 470, 472
  management, 151
Content Management Interoperability Services (see *CMIS*)
Content Server, 137
Context data, 452
Context data extractor, 149, 443
Context information, 450
Correction, 292
CRM, *67*, 108, 109, 136, 141, 153, 333, 335
CRM content management, 136
Culture dimension, *59*
Customer information, 64
Customer relationship management (see *CRM*)
Custom extraction rules, 362

D

DART Browser, 451
Data, 28
  cleansing, 73
  consolidation, 330
  correction, 113, 303, 309
  distribution, 180
  domain, 61, 156, 343, 351
  elements, 292
  enhancement, 313, 319
  foundation, 151
  integration, 56, 151, 163, 174
  lineage, 224, 228
  loading, 112
  management, 180, 181
  parsing, 303, 336
  policies, 57
  remediation, 154
  retention, 172
  sources, 76, 192, 348
  standardized, 113, 303, 306, 307
  standards, 59, 176
  synchronization, 112
  transformation, 272
  validation, 112, 113, 154, 189, 256
  warehouse, 111, 170, 173, 176, 252, 277
Data administrator, 177
Data analyst, 51, 90, 94, 228, 230, 231
Data architect, 188
Data archiving, 143, 144, 146, 149, 211, 216, 217, 218, 220, 428, 470
  basics, 429
  process, 431
Data assessment, 39
Database administrators, 217
Databases, 111
Data cleanse, 106, 112, 152, 187, 290, 292, 300, 302, 303, 313, 324, 337
  data correction, 309
  data standardization, 303
  data validation, 311
  standardization, 305
Data cleanse transform, 123
Data destruction, 439, 463
  in the live database, 440

Data destruction (Cont.)
in the retention warehouse, 463
security considerations, 465
Data enrichment, 113
Data extract browser, 451
Data extraction, 152
Data flow, 194, 255, 267, 268, 269, 270, 271, 279, 283, 290, 324, 332, 366, 367, 370
Data insight, 224, 227, 229
project, 229, 230, 245, 249, 340, 349, 355
Data integration, 201, 251, 256, 359, 373
Data integrator transforms, 277
Data lifecycle, 430
Data lineage, 40
Data load, 184
Data Mart, 111, 166
Data migration, 55, 71, 105, 106, 109, 112, 153, 158, 185, 187, 197, 252, 315, 373, 374
Data migration content, 373, 376, 382
Data model, 401, 405
Data modeling, 225, 226, 393, 400, 403
Data profiling, 108, 119, 120, 152, 188, 195, 223, 224, 228, 229, 232, 241, 245, 292, 319
basic, 233, 236
projects, 229
Data quality, 39, 43, 44, 62, 72, 112, 116, 118, 123, 151, 154, 157, 158, 163, 181, 182, 188, 195, 196, 200, 209, 229, 241, 249, 250, 251, 256, 324, 334, 335, 336, 339, 352, 353, 354, 359, 362, 374, 395, 396, 421
blueprints, 335
dashboard, 248
levels, 57
measurement, 122
metrics, 182
monitoring, 119, 120, 224, 353
process, 115
requirements, 181, 223
root cause analysis, 354
score, 245, 248, 250
scorecard, 120, 239, 340, 343, 344, 346, 347, 349, 352, 353, 355
Data quality (Cont.)
scorecard tile, 341
scores, 339, 349
Data quality dimension, 340, 344, 345, 346, 349, 351
accuracy, 344, 346
completeness, 345
conformity, 345, 347, 349
integrity, 345
timeliness, 345
uniqueness, 345
Data replication, 397
framework, 387
Data steward, 50, 62, 72, 77, 121, 124, 125, 160, 181, 200, 240, 255, 340, 344, 349, 414, 419
Data stewardship, 165
Datastore, 263, 264, 366
Data synchronization, 203
Data transfer, 399
Decision tables, 411
Decommission, 28
Decommissioning, 144, 155
De-duplication, 110, 152, 154, 187, 190, 195, 311, 317, 319
Demographic data, 315
Dependency profiling, 237
Derivations, 395
Digital asset management, 132
Direct linkage, 354
Direct marketing, 317
Direct Marketing Association, 335
Discovery, 64
Discrete format, 294
Document access, 104
Document archiving, 470
Document-centric workflow, 470, 472
Document management, 131, 470, 472
Document presentment, 105
Documents, 467
DRF, 387
DSO, 461
Dun and Bradstreet, 31, 68, 335
Duplicate check, 395

E

Easy Document Management, 136
ECC, 109
ECM, 44, 49, 56, 104, 130, 132, 138, 142, 206, 467
integration layer, 478, 479
workspace, 479, 480, 481
ECMLink, 479
Editions, 393
EIM, 27, 28, 29, 30, 40, 44, 48, 73
operational, 36, 37
strategy, 48
Email response system, 136
Emails, 467
Enterprise application integration, 173
Enterprise architecture, 203
Enterprise content management (see *ECM*)
Enterprise data warehouse, 192, 289
Enterprise information architect, 176
Enterprise information architecture, 165, 166, 167, 168, 184, 193
Enterprise information management (see *EIM*)
Enterprise search, 137
Enterprise services, 387
Enterprise workspace, 136
Entity, 361
extraction, 361, 367
Entity types, 393, 394, 400, 402
relationships, 394, 402
ERP, 78, 153, 161, 180, 333
ETL, 72, 100, 106, 111, 173, 190, 192, 202, 203, 205, 224, 225, 226, 339, 353, 447
Executive sponsor, 60
Extraction, transformation, and load (see *ETL*)
Extractors, 266

F

Failed records, 351
Family match, 320
File source, 264
Financial master data, 128

G

Generic object services, 485
Geocode, 195, 196, 314, 317
Geo directories, 315
Geolocation, 175
Geospatial, 313
Geospatial data, 313
Global address cleansing, 237, 239, 296, 300, 305
Global data manager, 50, 62
Global standards, 45
Governmental regulations, 44
Governmental standards, 60
Grammatical parsing, 360
GS1, 125

H

Hadoop, 46, 90, 93, 95, 96, 97, 98, 100, 102, 268
HBase, 94
HDFS, 92
Hive, 94, 98, 100
Mahout, 95, 97
map reduce, 92, 97
master node, 91
Pig, 94, 97, 98
worker node, 91

I

IDEA, 463
IDoc, 152, 261, 266, 387, 397, 399
ILM, 142, 144, 166
ILM-aware storage, 432
system, 433
ILM-BC 3.0
certification, 433
ILM object
definition, 429
Images, 35
Impact analysis, 184, 228, 354
Individual match, 320
Industry standards, 37

InfoCube, 353, 461
Information
access, 57
discovery, 56, 80, 157
policies, 58
strategy, 60
Information governance, 28, 30, 34, 36, 37, 40, 49, 55, 57, 58, 59, 61, 63, 65, 66, 68, 72, 73, 74, 75, 76, 80, 98, 100, 105, 118, 154, 156, 158, 159, 160, 165, 179, 180, 181, 188, 194, 200, 215, 377
committee, 72
council, 72
customized, 129
Information lifecycle management, 100
definition, 425
drivers and pain points, 425
external drivers, 426
for legacy data, 441
in live systems, 436
internal drivers, 426
Information lifecycle management (see *ILM*)
Information management, 168, 176
Information management strategy, 41, 130, 142, 143, 151
Information platform services, 257, 258
Information retention manager, 439
Information security, 165
In-memory computing, 47
Invoice management, 105
IPS, 257
IRM, 439
IT administrator, 229, 230

J

Java Message Service, 261
Job, 284

K

Key mapping, 397, 398
Knowledge workers, 77
KPI, 77

L

Legacy System Migration Workbench (see *LSMW*)
Legacy systems, 149
Legal compliance, 216
Legal hold, 38, 146, 161
management, 145
setting, 441
Legal hold management
overview, 440
Legal requirements, 57, 145
Lifecycle, 29, 40
Local reporting, 459, 463
LSMW, 153, 374

M

Maintenance notification, 139
Management reporting, 170
Managing content, 35
Master data, 31, 34, 38, 41, 49, 69, 73, 126, 127, 158, 160, 163, 171, 172, 180, 181, 194, 198, 199, 385, 393
consolidation, 127
customer, 128
harmonization, 127
management, 169, 252, 317
material, 129
Master data governance, 411
application framework, 387
Master record, 320
Match, 290, 315, 337
comparison options, 319
configuration, 321
criteria, 320, 327
group, 320, 321, 328, 330
level, 320
performance, 324
scenarios, 320
score, 328
set, 320
standards, 324
thresholds, 320
Match Criteria Editor, 328

Match Editor, 321, 327
Matching, 113, 114, 117
process, 315
score, 395
standards, 308
strategy, 318, 319
techniques, 319
Matching method
combination, 321, 323
rule-based, 321
weighted scoring, 321, 322
Matching routine, 195
Match Wizard, 321, 325, 327, 328
Mergers and acquisitions, 184, 426
Metadata, 122, 151, 169, 171, 177, 178, 208, 455
analysis, 119, 122, 224
integrator, 225, 227
management, 118, 123, 223, 224, 353, 354
Metapedia, 125, 224
Migration, 29, 34, 42
Migration content, 153
Missing data, 76
Monitor, 118
Multiline data, 295
Multiline format, 294
Multiline hybrid format, 294
Multi-source integrator, 225

N

No-match thresholds, 329
Nondiscrete data components, 295
Nonparty data, 336
Nonrelational data, 99
Non-SAP systems, 441
NoSQL, 89, 94

O

OLAP, 89, 111
OLTP, 89
Open hub, 266
OpenText, 104, 133, 469
OpenText Knowledge center, 484
Operational analytics, 170
Operational data, 171, 172
Operational efficiency, 41, 43, 201
Operational master data management, 126
Operational reporting, 170
Operational use, 253
Optimical character recognition (OCR), 209
Organizational change management, 157
Organizational ownership, 177
Output management, 131
Output schema, 370

P

Parsed data, 296, 299, 307
Parsed output, 299, 302
Parsing, 292
Party data, 335
Platform transformations, 275
PLM, 136, 139
Point-of-interest, 315
Policy
define, 160, 453
engine, 146
implementation, 160
set status to live, 454
Policy category
residence rules, 453
retention rules, 453
Portal Content Management, 134
Portal Site Management, 134
Power Builder, 105
Preconfigured solutions, 154
Predictive analytics, 58, 157
Pre-parsed data, 299
Print lists, 439
retrieve, 433
Process modeling, 396, 400
Process patterns, 415
Procurement, 139
Product data, 335
Product liability, 149

Product lifecycle management (see *PLM*)
Product quality, 364

Q

Quality, 42
standards, 47
Quality dimensions, 341, 421
Query transform, 270, 272

R

Rapid Deployment Solutions (see *RDS*)
Rapid Marts, 155, 353
RDS, 151, 154, 375, 376, 379
Real-time data flows, 335
Real-time service, 286
Records management, 132, 161, 216, 470, 471
Redundancy profiling, 237
Reference data, 169, 171
Regulatory compliance, 44
Replication Server, 105
Reporting
local, 463
Residence time
definition, 430
Retention, 158
limits, 176
Retention management, 144, 145, 146, 211, 428
capabilities, 436
unstructured data, 439
Retention Management cockpit
administrator, 445
line of Business, 445
Retention period
definition, 430
maximum, 455
minimum, 455
Retention policies, 30, 147, 149, 160, 216
Retention rules, 179
basics, 455
Retention time unit, 455
Retention warehouse, 147, 148, 150
Retirement, 36
Rule-based, 321
workflow, 407, 410
Rule bindings, 342, 349, 351, 355
Rule tasks, 246

S

SAP ArchiveLink, 141, 146, 424, 479
attachments, 455
documents, 433
SAP Archiving by OpenText, 104, 134, 140, 141, 145, 150, 424
SAP Audit Format, 463
SAP Best Practices, 151, 154, 375, 377
SAP BusinessObjects, 153
SAP BusinessObjects Business Intelligence, 79, 107, 110, 111, 151, 161, 225, 226, 257
SAP BusinessObjects Web Intelligence, 110, 151, 376
SAP Business Suite, 130, 139, 207, 208, 210, 213, 392, 474, 475
SAP Business Workflow, 71, 79, 127, 161, 214, 386, 406, 409, 410, 411, 417, 477
SAP CRM Customer Interaction Center, 214
SAP Customer Relationship Management (CRM), 107, 209, 289, 377
SAP Data Migration, 151, 375
SAP Data Quality Management, 104, 112
SAP Data Quality Management SDK, 112
SAP Data Services, 66, 69, 71, 72, 78, 97, 106, 107, 108, 109, 110, 111, 112, 114, 126, 127, 130, 144, 151, 152, 153, 155, 160, 163, 185, 189, 190, 193, 195, 197, 201, 203, 204, 225, 226, 227, 237, 239, 251, 252, 254, 263, 264, 267, 269, 272, 290, 296, 307, 309, 311, 313, 314, 315, 319, 334, 335, 336, 339, 353, 355, 359, 364, 365, 375, 376, 379, 380, 385, 396, 420, 424

SAP Data Services (Cont.)
administration, 256
batch jobs, 261, 288
breakpoints, 280
built-in functions, 272
central repository, 258
Data Integrator, 104
designer, 255, 259
ETL capabilities, 447
function categories, 272
functions, 272
history preservation, 277
job, 261, 271, 280
job server, 255, 258
lineage analysis, 256
Local Object Library, 260
local repository, 258
lookup function, 273
management console, 255
mappings, 255
metadata, 256
project area, 259
query transform, 269
real-time job, 261, 282
real-time service, 281, 284
tool palette, 260
SAP Digital Asset Management, 134
SAP Document Access by OpenText, 134, 140, 145, 155, 424, 433
SAP Document Presentment, 133
SAP Employee Management, 134
SAP Enterprise Asset Management, 139
SAP ERP, 107, 108, 128, 141, 146, 147, 148, 149, 150, 265, 288, 377, 388
SAP ERP Human Capital Management (HCM), 153
SAP Extended ECM, 69, 71, 104, 135, 137, 139, 140, 142, 160, 161, 206, 207, 208, 209, 211, 212, 214, 215, 383, 469, 472, 475, 478, 481, 484
metadata, 480
SAP GUI, 482
SAP HANA, 46, 49, 54, 64, 97, 98, 99, 108, 110, 111, 112, 147, 155, 217, 253, 263, 265, 268, 287, 288
SAP IMG, 392
SAP Information Steward, 66, 78, 104, 108, 118, 123, 125, 126, 130, 151, 152, 153, 160, 161, 188, 189, 191, 225, 226, 227, 228, 229, 230, 235, 240, 244, 250, 253, 255, 256, 336, 339, 340, 348, 374, 375, 378, 420, 421
SAP Invoice Management, 134, 212, 214
SAP Landscape Transformation, 374
SAP Master Data Governance (see *SAP MDG*)
SAP MDG, 69, 70, 71, 104, 108, 110, 126, 127, 128, 130, 161, 163, 190, 333, 335, 385, 386, 388, 391, 397, 399, 402, 403, 407, 408, 411, 420, 422
flex mode, 391
reuse mode, 391
SAP NetWeaver Application Server ABAP, 334, 374, 484
SAP NetWeaver Business Client, 387, 399, 445, 446, 453, 456, 479
ILM cockpits, 446
SAP NetWeaver Business Process Management, 66, 161
SAP NetWeaver Business Warehouse (BW), 72, 108, 110, 111, 112, 145, 147, 148, 150, 225, 266, 267, 353, 443
reporting, 460
SAP NetWeaver Folders Management, 136
SAP NetWeaver Identity Management, 107
SAP NetWeaver Information Lifecycle Management (ILM), 104, 136, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 155, 160, 161, 163, 211, 374, 471
architecture, 434
cockpit roles, 445
cornerstones, 428
data archiving, 428
object, 453
retention management, 428
retention rules, 439
Store Browser, 465
system decommissioning, 428
SAP NetWeaver Information Lifecycle Management (ILM) conversion, 445, 455
replace old sessions, 458

SAP NetWeaver Master Data Management (MDM), 68, 71, 104, 108, 110, 126, 127, 154, 161, 163, 385
SAP NetWeaver Portal, 134, 136, 138, 141, 210, 214, 387, 399, 479
SAP NetWeaver Process Integration (PI), 201, 203, 204, 205, 287
SAP NetWeaver Process Orchestration, 69, 70, 71, 79, 201
SAP Plant Maintenance, 139
SAP Portal Content Management by OpenText, 210
SAP solutions for information lifecycle management, 423, 466
overview, 423
SAP StreamWork, 136
SAP Travel Receipt Management, 134
Scanned invoices, 146
Schema, 282
Scorecards, 118
Semantic disambiguation, 360
Sentiment, 36
Sentiment analysis, 116
Service-level agreement, 77, 229
Similarity scoring, 319
Slowly changing dimensions, 277
Snapshot, 433, 449
SN_META
file, 455
Social media, 46, 47, 473
Solution packages, 154
SPRO, 392
SRM, 67
SRS, 436
Staging, 389
Staging area, 390
Standardization, 292
Standardization rules, 336
Step type, 418, 419
Storage, 145, 148
Storage and retention service, 436
Storage system
ILM-aware, 433
Structured data, 34, 36, 169
Subordinate records, 321
Supplier, 30
Supply chain, 47
Sybase, 263
Sybase IQ, 112, 155
System consolidation, 374
System decommissioning, 49, 145, 147, 148, 428
archive transactional data, 448
configure retention warehouse system, 443
convert data, 445
data analysis, 442
data transfer, 456
data transfer and conversion, 452
define audit areas, 445
detailed example, 447
enable system for ILM, 443
extract data, 448
non-SAP systems, 446
preliminary steps, 442
process overview, 442
reporting, 459
report on legacy data, 445
set up audit areas and rules, 453
transfer and convert files, 455
transfer archive administration data, 456
transfer data, 445
System decommissioning cockpit
administrator, 445
line of Business, 445
System landscape harmonization, 426
Systems decommissioning, 34
Systems migration, 151, 155, 383

T

Tax audit, 143
Tax reporting, 149
Technical requirements, 239
Text analytics, 357
Text data, 36, 364, 365
Text data processing, 104, 106, 115, 116, 117, 163, 251, 335, 359, 360, 363, 364, 369, 370
dictionary, 363
entity, 362
entity types, 369

Text data processing (Cont.)
extraction, 362
rule, 363
transform configuration, 366
Text mining, 40
Time reference, 455
TOAx
tables, 433
Transaction
ILM, 450
ILM_DESTRUCTION, 463
ILM_TRANS_ADMIN_ONLY, 456
IRM_CUST, 455
IRMPOL, 453
SARA, 448
TAANA, 443
Transactional application, 173
Transactional data, 169, 448
Transform, 262, 290, 367, 369
address cleanse, 310, 311, 313
case, 275
data cleanse, 189, 300, 307, 313, 336
entity extraction, 359, 363, 364, 366
geocoder, 314, 315
global address cleanse, 301
history preserving, 278
key generation, 279
Map_Operation, 276
match, 319, 320, 326, 332
merge, 276
query, 275
Row_Generation, 276
SQL, 276
table comparison, 278
transform configuration, 367
user defined, 276
validation, 276, 339

U

UI
configuration, 406
modeling, 400
Uniqueness profiling, 237
Universal data cleanse, 196
Universal Worklist, 409
UNSPSC, 125
Unstructured content, 168, 169, 474
Unstructured data, 95, 96, 99, 100, 145, 155, 168, 177, 357, 371
retention management, 439
text, 359
Unstructured information, 36, 47, 467
User interface, 474

V

Validated address, 334
Validation, 389, 395
Validation rules, 118, 121, 223, 224, 235, 239, 240, 242, 243, 245, 246, 339, 342, 344, 346, 347, 348, 349, 351, 352, 355, 421
Validation transform, 311
Value mapping, 397, 398

W

Web content management, 132
WebDAV
ILM-enhanced interface, 432
Web Dynpro, 406, 407
Weighting, 346, 347, 349
Weight scoring, 321
Workflow, 387, 389
Work order, 139
Workspaces, 138, 139, 140, 209, 211, 212, 260, 484
binder workspace, 481
business workspace, 481
case workspace, 481
Write program
log, 449

X

XML, 399
XML DAS, 436
XML data archiving service, 436
XML schema, 282, 287