The calibrate
Prepress Coach

Radix Translate

Interface between sender and receiver

RADIX
Print 4.0 made easy

Contents

Management Summary Page 03

Introduction Page 04

Who runs the system? Page 04

When does data have a high quality? Page 05

Metadata Page 08

XML, JSON and structure Page 09

Readable for humans and machines Page 09

Benefit and Contact Page 10

As a service provider, calibrate Workflow-Consulting GmbH implements the automated process for the production of printed and electronic media worldwide. Radix is calibrate's modular product package consisting of services and software. The focus is on the automation of processes. Here, the focus is on security in the reconciliation of production data with the metadata for production. The goal is digitisation - Print 4.0 made easy.

The calibrate Prepress Coach consists of five parts. Part 4 „Radix Translate - Interface between sender and receiver" deals with communication in the workflow.

Management Summary

We often speak of the „lead system" in software system integration. The concept is useful when several systems in which the same information is used or in which information is derived (for example, metadata) are considered. The consideration arose from the dilemma of determining a source of information that can be assumed to be correct or original, for example.

Let us assume that a customer places an order with his address in a web store (system A). The information is stored in database A. In the order processing system (system B), this data is created as an order in database B. An employee in the order processing system detects an error in the address, checks the address and corrects it in system A. After the correction is made, there is a difference in the data in database A and database B. Another system can be used for dispatch (system C). At this point it must be clarified which system (A or B) is leading. Does C obtain its data from A or B?

In practice, one defines a main system for the distribution of metadata. One of the existing systems is defined as the lead system for a master data class (= address, product description and the like) and is thus the starting point for the distribution of the information. The initial creation of the master data is done in the lead system with the properties available there. Subsequently, the data records are enriched with further properties. From now on, changes in other systems must be reported to the lead system or may only be made in this system. The other systems networked with the lead system are then responsible for local reconciliation.

A complication arises from heterogeneous system landscapes, i.e., different systems from different manufacturers. It must be clear to the sender what data the receiver expects and what this data looks like. Often, even XML data is proprietary, i.e., manufacturer specific. This can mean that the use of the data is restricted by the unavailability of the source code. Proprietary means „owned (by the manufacturer)." This means that if someone offers an XML interface, it does not mean that we understand their language.

Introduction

In the printing industry, ERP systems claim to be the leading system. For classic print shops with a sales force and around 3,500 jobs per year, this may be true in most cases. But what about when 35,000 jobs or more have to be handled per year? The quantities are initial catalysts for the need to automate processes.

A major obstacle in communication are formats and structures – comparable to languages and dialects, which require translation between data sender and data receiver. **Radix Translate** can do more than a simple translation that follows defined rules. From the information, actions are derived that trigger subsequent processes.

Who runs the system?

Based on information regarding resources, ERP systems perform timely and demand-driven planning and control of order processing. In addition, master and transaction data enable the ERP system to create work orders transparently and efficiently.

The web store developers initially mapped very simple order structures, mostly products in a shopping cart from one customer at one delivery address. As Amazon has proven, this works very well for unchanging consumables. But for the printing industry with its possibilities for individual design, this was not enough. Design and more complex order structures are possible today. Like web stores, other systems such as PIM, MAM or DAM also supply data for production.

The providers of production workflows can be seen as the third in the group. Practically every major provider has at least one upload portal in its portfolio.

With the aim of creating greater dependency on their own system in the overall infrastructure, the software providers are constantly expanding the competencies of their own systems, creating overlaps in the handling of tasks. For example, the ERP providers launched a web store, while the web store developers offer order management, and the production workflow specialists have also entered into the field of all-in-one solutions. The situation can be said to be one of competition between the system worlds. Since all systems are necessary, it is important for the users to have a clean synchronization. The basic prerequisite for this is a loss-free communication of the system landscapes.

At this point, there are two pieces of good news:

1) There is no fundamental right or wrong. The suitability of the systems depends on the use case and always requires a proof of concept, i.e., a feasibility study that proves the basic capability of the networking.
2) Most systems are capable of exporting structured data. This data can usually be translated from the sender to the receiver. So which system is actually the „lead system" depends on a variety of factors. One important question is always: What should be done with all the collected data in the end? Do they have to be stored, or can we „forget" about the data set once the job is finished?

When does data have a high quality?

In principle, data is of high quality if it fulfils the purpose defined by the user. In this respect, data quality is also frequently described with the term „fitness for use". The more concrete determination of quality should take various criteria into account. These criteria are completeness, timeliness, consistency, accessibility and interpretability.

- Data is complete when all defined criteria are met, i.e., there are no more zero values. In production, for example, all product properties (e.g., format, colour, paper, etc.) must be available before production can begin.
- Timeliness describes how promptly data values are adjusted in the event of a change in the real state. Requirements for timeliness can vary: For example, the specification of delivery date and shipping address during production is sufficient, whereas the timeliness of product data must be retrievable in real time with Just-in-Time production.
- Consistent data sets are characterized by the fact that they are free of contradictions. Vendor master data is consistent, for example, if there is a unique identifier for each vendor, i.e., if there are no duplicates.
- The extent to which data is available or findable for the user is determined by the criteria of accessibility. In online sales channels, item information must be accessible to consumers in real time.
- Interpretability refers to the extent to which data is available in suitable formats, while adhering to clear rules and definitions. For example, data for payment (e.g., ZUGFeRD) must comply with certain syntactic specifications so that automatic processing can take place.

The criteria mentioned can render the quality of the data concrete and „measurable", since it is only through measurement that it is possible to derive actions to improve the quality of the data.

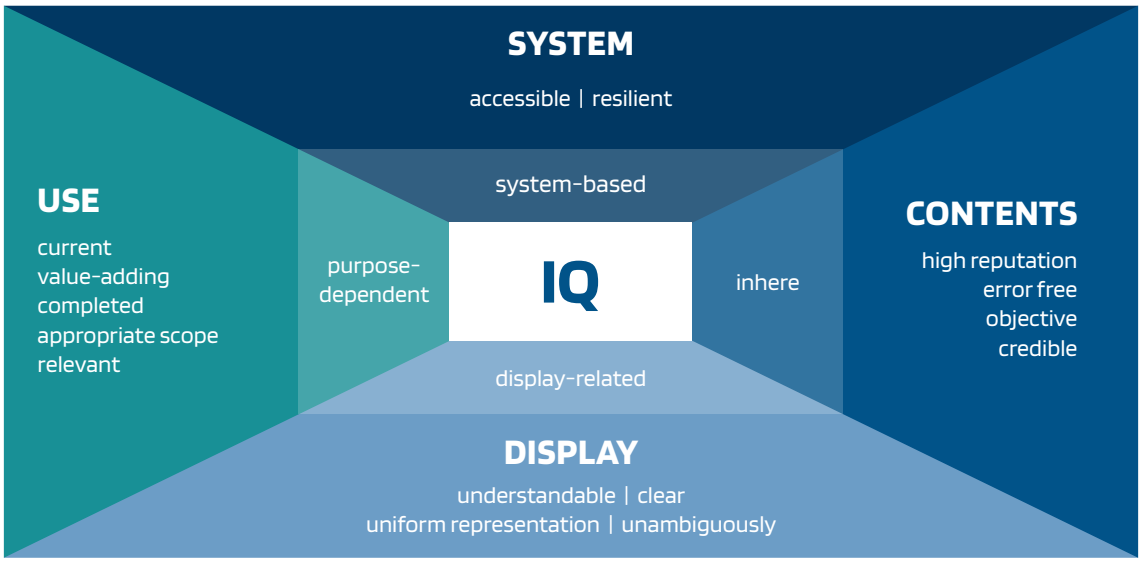

Fig. 1: The 15 dimensions of data quality: definition of the German Society for Information and Data Quality (Deutsche Gesellschaft für Informations- und Datenqualität, DGIQ) based on the research results of Richard Wang and Diane Strong (Source: DGIQ IQ-Definition - 2007).

Due to the increasing requirements, both the importance of master data quality in companies and the efforts to eliminate quality problems are growing. According to a study by Pricewaterhouse-Coopers, the causes of such problems are often the following factors:

- **Data entry error**
  Someone enters the data at the beginning of every process. Manual entries harbor the risk of errors. Typical errors include spelling mistakes, entering inappropriate information in mandatory fields that are not automatically checked for the use of allowed values, or incorrectly selecting values from predefined lists.

- **Contradictory definitions for data objects used across the board**
  Especially when a new system is added to an existing one, this problem can come into play. If the data management has historically been created independently in different systems, then contradictory definitions are hardly avoidable. This becomes clear when we ask several people about the definition for formats – open, closed, trimmed and so on.

- **Errors when migrating data to new systems**
  When introducing or merging business information systems (e.g. ERP systems), data must be transferred from legacy systems to the new system environment. If the data is not consistent and cannot be interpreted in a syntactically uniform manner, incorrect entries will be made during the – mostly automated – data transfer.

- **Errors when integrating data from multiple systems or external sources due to different data models**
  Data from different source systems often have to be merged and aggregated. If a company has several systems (ERP, web store, portal, etc.) but wants to perform an analysis on the orders, the goods master data, for example, must be available in a uniform model or at least be able to be transferred to a uniform model.

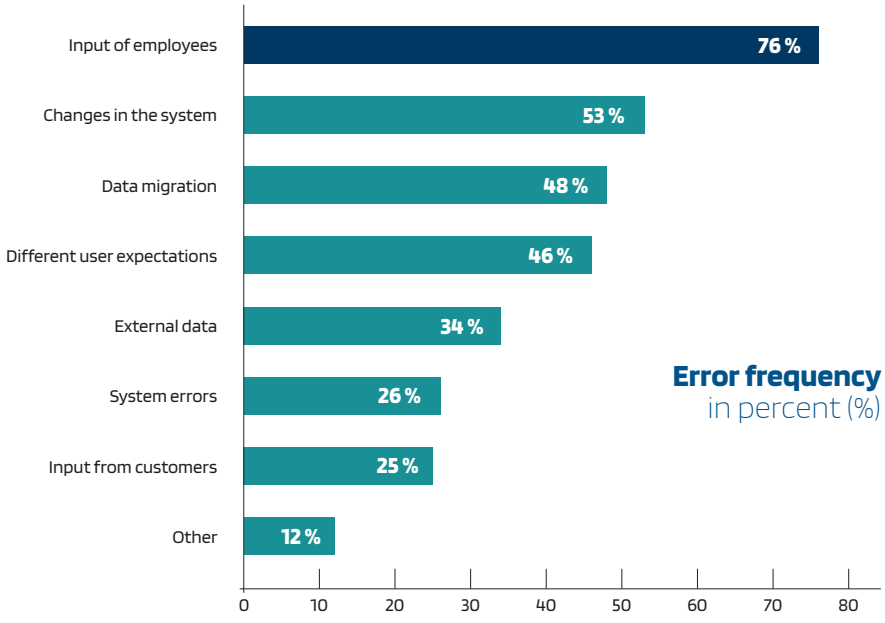

Fig. 2: Data quality depends on various factors. First and foremost, however, is human input. (Source: The Data Warehousing Institute).

To better understand the different requirements for translating data, it is worth taking a look at the different master data architectures. Four ideal-typical architectural approaches are available for implementing master data management in a distributed system landscape: Lead master data system, central master data system, connection via a directory (registry) and standards. The approaches can be differentiated on the basis of the dimensions „master data model" and „master data maintenance and storage". In a harmonized master data model, the master data attributes (e.g., name and field length) and their contents (e.g., unique identifiers or value ranges) are defined globally, i.e., company-wide or cross-company, depending on the scope. In master data maintenance and storage, a distinction is made between the extent to which master data is created and stored centrally. On the other hand, the central recording and maintenance of master data ensures the consistency of master data, since a „single point of truth" exists with the central system. However, the more centralized the approach, the less flexibility there is.

We speak of a central master data system when the management of master data takes place in a separate master data system that distributes it to the local systems. The collection and maintenance basically takes place in a central system, but on the basis of a harmonized master data model.

„Standards" do not lead to central storage and distribution of master data; rather only uniform structures are defined throughout the company. A harmonized master data model ensures that the structure and entry of a master data record is the same across different systems.

A „lead system" is the most commonly used method for master data distribution: one of the existing systems is defined as the lead system for a master data class and is thus the starting point for distribution to the other systems.

„Connection via a directory" (registry) is the way to implement an overarching directory that contains mappings of the various master data records to the various source systems. If, for example, a system needs data on a specific customer, it starts a query to the registry and receives an answer about which system contains the data on the customer in question. In a further step, the data is then retrieved directly from the corresponding system.

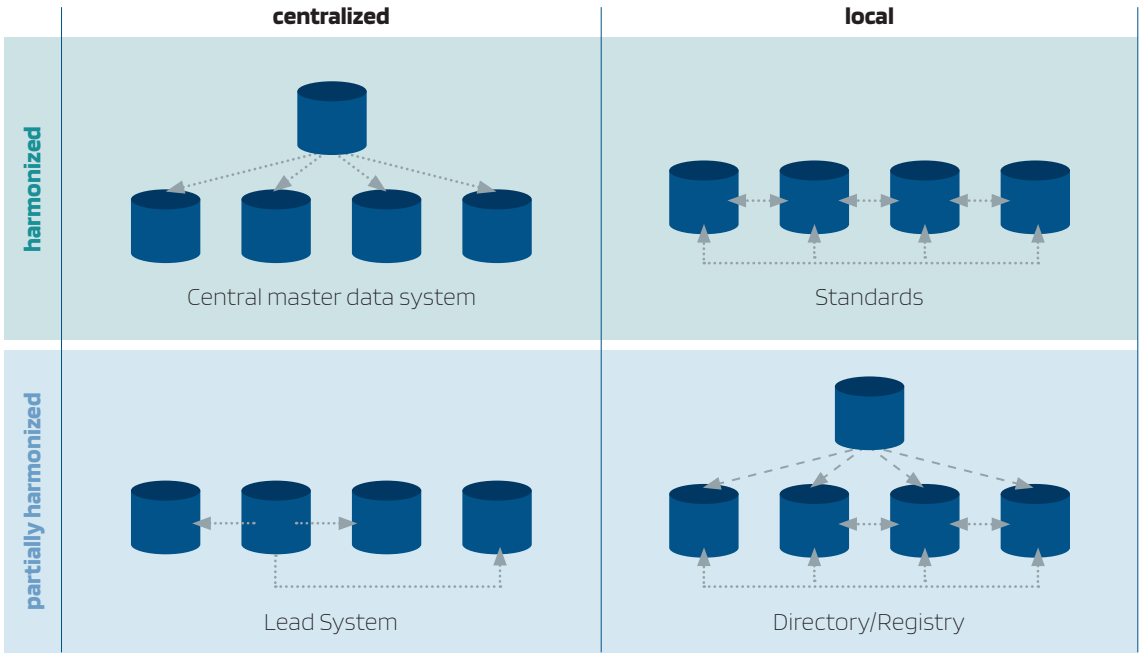

Fig. 3: Master data maintenance and management

Metadata

What data can the sender provide? A first step on the journey through PDF properties and metadata can start in a PDF opened in Adobe Acrobat. There, under „File/Properties", you will already find quite a lot of information about the properties of the document.

The description displays basic information regarding the document. Title, author, subject and keywords may have been set by the creator of the document in the source application. The Security parameter specifies which changes and features are allowed in the PDF document. If a password, certificate or security policy has been set up for the PDF document, the method is listed here. A big topic is the fonts. Here you can again find the information about whether the fonts are included or not. Things get really exciting under the „Custom" tab: Here the insertion of document properties has been made possible. This means that additional metadata that has nothing to do with the actual print image can be saved in the PDF. Here, under „Advanced", you will find a list of PDF settings, print defaults and reading options for the document.

PDF documents that were created with Acrobat 5.0 or higher contain document metadata in XML format. Basically, XML can be used to describe, store, and exchange data. The main advantages of XML are that it is widely used and requires minimal learning. In addition, XML can be easily interpreted by humans and machines. Metadata contains information about the document and its content. Using the Extensible Metadate Platform (XMP), a common basic XML structure is created that helps standardize the creation, processing, and exchange of document metadata between publishing workflows.

The metadata we need for a conversation with a print shop clerk is the format, size, color, substrate, and for the quote, the run length. It's the same with our digital colleague. If we order from a print shop in the USA, we have to do it in English. If we don't know English, we need a translator. This analogy can also be transferred to the automated process - whereby the digital processes also take into account whether the data can be processed at all.

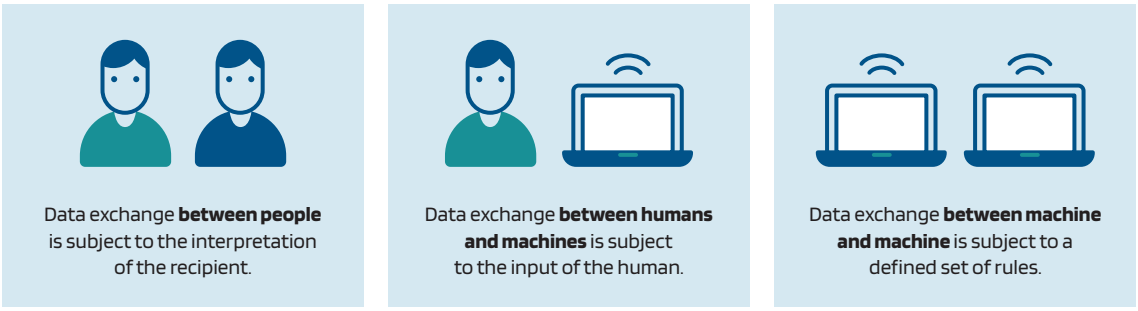

Fig. 4: Data is exchanged via communication between systems. Automation in the process foresees machines as senders and receivers here. Interpretations and input errors can be avoided by definition. However, every definition presupposes a decision.

XML, JSON and structure

Essentially, an XML document has the components layout, data and structure. Adobe added XML functions to PDF around 20 years ago, creating the basis for including forms in PDF files. However, XML is not type-specific. This means that if height="210.00" is passed, it is not defined whether the number 210 is involved here or the text 210.00. In the JSON data format, on the other hand, height: 210.0 is used - this means that height is clearly defined as a number which can be used for calculation.

The JavaScript Object Notation (JSON) is a compact data format in an easily readable text form and serves the purpose of data exchange between applications. JSON is independent of the programming language. This is one reason why **Radix Translate** relies on this format. XML files are converted to JSON format. Validation of the data is done with JSON schema files.

Readable for humans and machines

Information has to be prepared differently - for example, a print product is described as „flyer A4, printed in color on white paper". This can be read by humans. For automation, however, we need structured data with a defined structure and meaning. The production workflow needs it to be machine-readable, as shown in the example below:

```
{
       "dimensions": {
       "width": 297,
       "height": 210,
       "unit": "mm",
       "bleed": 3
    },
    "media": {
       "surface":"Coated",
       "paperclass":"PS1"
    }
}
```

Fig. 5: Machines need more structure to „read" information.

Additional complexity is added to communication by modern web services. In the broadest sense, a web service is a communication method between two applications over a network, usually the World Wide Web. Web services come in two different types: Simple Object Access Protocol (SOAP) and Representational State Transfer (REST). There are significant differences between SOAP and REST web services, which we will not discuss in detail here. However, proponents of REST or SOAP can be passionate about their web service architecture. Both SOAP and REST architectures have proven to be reliable, successful, and scalable. Therefore, the decision to use REST or SOAP has less to do with their efficiency and more to do with how both approaches fit into an organization's software development culture and project requirements. **Radix Translate** handles translation at many stages of the process, enabling error-free communication regardless of protocols and file formats. **Radix Translate** makes information readable for people and machines.

Benefit

Radix Translate ensures secure communication across the various interfaces. This can take place directly between systems, i.e. from the web store to the ERP system or directly to the digital printing machine or another digital front end, such as a Plotter. The same applies to the exchange between the ERP system and the systems for production.

Another use case is the translation of data from the ERP system into JDF or xJDF. xJDF, like JDF, is based on XML. While with JDF an attempt was made to describe a completely electronic job pocket with all the associated work steps, with xJDF the main focus is on describing a single work step or a print product. Not all ERP systems are capable of creating a CIP4-compliant JDF/xJDF export. This is also where **Radix Translate's** „language talent" comes into play.

Contact ▸

calibrate Workflow-Consulting GmbH
St. Pöltnerstraße 26
3130 Herzogenburg / Austria
Phone +43 / 732 / 68 26 60
office@calibrate.at
www.calibrate.at

Managing Director: Peter Kleinheider