Digital Transformation of Manufacturing Record Books – An Ontology-Based Case Study

Bjørn Jæger[1[0000-0002-4661-5102]] and Beni Ruef[2[0000-0002-2436-2087]]

[1] Molde University College, Specialized University in Logistics, Molde, Norway
[2] Swiss Law Sources Foundation, Zurich, Switzerland
bjorn.jager@himolde.no, bernhard.ruef@ssrq-sds-fds.ch

Abstract. Manufacturers in the oil and gas industry face increasing documentation requirements of assets for compliance, quality assurance and contractual purposes. The documentation includes hundreds of documents for each asset that are compiled into a Manufacturer Record Book (MRB) to be delivered with the asset. The MRB is created by collecting all necessary documents and storing them in the MRB as scanned copies in PDF format. Creating an MRB is a cumbersome, error-prone and time-consuming activity. Using an MRB in PDF format is equally hard as it requires humans to search and extract data. Several initiatives have been undertaken by joint industrial projects and standardization agencies to represent the documents by their data in a language understandable to computers and humans. It has been proven that ontologies are effective tools in digitalization by making data from diverse sources interoperable. This paper presents a case study of MRB digitalization for CodeIT, a company developing software solutions for industrial applications. The focus is on creating an ontology for a welding certificate, following an Extract-Transform-Load (ETL) procedure that extracts data from a welding certificate in PDF format and uses the ontology for transforming the data into subject-predicate-object triples in a Resource Description Framework (RDF) model. The results demonstrate the applicability of ontologies for digital transformations.

Keywords: Manufacturing Record Book, Ontology, Interoperability, FAIR Principles, Digitalization.

1 Introduction

The oil and gas industry requires proper certification and authorization of any product to ensure the purchased product is qualified according to legislative and safety requirements. Manufacturers and engineering companies must document the systems delivered including documentation of product designs, materials specifications, certificates, operating procedures and inspection reports. Each system delivered is accompanied by a Manufacturing Record Book (MRB) with the documentation related to the system. Historically, an MRB consisted of a compilation of paper documents. The increase in system complexity, as well as increasing safety requirements, cause a typical MRB to consist of a large number of documents, often in the number of several hundred. As an example, an MRB to support an offshore accommodation module may be 170 pages, containing documents such as technical drawings, bill-of-materials, certificates, electrical equipment manuals, component certifications, external vendor information, quality plan, and acceptance test [1, 2].

The MRB contains documentation verifying that a given product or system meets its requirements. Contemporary MRBs are PDF documents, often in the form of scanned paper documents. This allows computers to handle the documents, but not their contents. This is not sufficient for business processes that require access to the contents, especially since many of the business processes are inter-organizational processes. For example, a company A approving the welding quality of a product by another company B requires an employee in A to fetch, read, compare and verify several documents fetched from B with A's own requirements, resulting in final approval by signing other documents. To automate such inter-organizational processes, a basic requirement is that the data of each document are in a machine-readable form that can be understood by each actor. A digital transformation of processes requires interoperability between the processes and the related data contained in the MRB. The transformation includes the transition from a document-centric to a data-centric documentation approach that allows for automatic processing across organizations. Interoperability based on semantic models and standards are required since ontology-based semantic approaches capture more of the meaning than traditional conceptual modelling like Entity-Relationship, Unified Modelling Language class diagrams or Object-Role Modelling [3]. An ontology supports a richer information interchange, not only exchange of data, especially since each class and relation (property) in the ontology must have a unique URI identifier. Thus, having an information model according to an ontology means it is ready-made for the web allowing interoperability across organizational boundaries.

Each document represents a context giving meaning to the data in the document. This context must be expressed formally in addition to the data to allow automated processing. The typical approach to add meaning to data is by using semantics expressed by formal ontologies. Ontologies allow data to be understood by a computer in the same way across various computer systems [4, 5, 6, 7, 8]. Ontologies are explicit formal specifications of the terms in the domain and relations among them [9]. In practice, the Resource Description Framework (RDF) can formalize existing data according to an ontology. RDF was initially a language for encoding knowledge on Web pages to make them understandable to electronic agents, developed by the WWW Consortium (W3C) [10], but is used now mostly for data interchange in general.

RDF expresses data in the form of subject-predicate-object triples formatted as URIs (Uniform Resource Identifiers). URIs identify a specific resource and point to its attributes or other resources.

Each document in an MRB has unique content and structure; for example, drawing documents and weld-test-certificate documents have different formats. Generally, these formats are difficult to convert into graph data automatically because each document has a specific application domain, a unique content, a non-standardized structure, is a PDF file, and there might be no ontology defined. For each file, one first needs to find out whether an ontology for this type of document exists. If not, an ontology is created, and an Extract-Transform-Load (ETL) procedure must be executed for each document:

1. Extract the data from the PDF file
2. Transform the data to key-value pairs (subject-object pair in RDF terminology) supported by the ontology
3. Load the key-value pairs into an RDF file identifying subject-predicate-object triples

In step 2, the data must be analyzed in order to transform them into key-value pairs. Our approach analyzes the data and proposes likely key-value pairs based on the ontology. In the beginning, the key-value pair generation must be supported by humans. After a few times, it is envisioned to be automated by machine-learning methods based on previous transformations. Additionally, proper nouns—which are the essence of certificates—can be identified by Named-entity recognition [11].

In step 3, the key-value pairs must be converted to subject-predicate-object triples that capture the semantics by expressing relations among subject-object pairs by predicates.

1.1 Manufacturing Record Book Solutions

There exists some Manufacturing Record Book (MRB) IT applications in the market. To our best knowledge, all of these are document-centric systems managing the documents and not their contents. Some of these are:

- CodeIT eMRB. This is the case company's solution described below.
- Conoco Phillips, Technical Information Requirement Catalog (TIRC) [12]
- TIRC by the NORSOK Z-TI Joint Industry Project (JIP) [13]
- Global Cents' MRB solution [14]

2 Case Study

2.1 The case company CodeIT

CodeIT AS is a Norwegian SME company founded in 2011 to develop and market a software solution, CodeIT Enterprise, for implementing automatic identification (AutoID), labelling, marking and traceability of manufactured components and products. [15]. During the development of the CodeIT Enterprise software solution for industrial applications, CodeIT became aware that manufacturers needed to process and store an increasing number of documents to comply with national and international regulations. Manufacturers were documenting their product and manufacturing process using electronic document versions of paper documents. This invariably led to manual operations to use the information in the documents in business processes. For example, the process of verifying that a product's welding procedures adhere to international quality standards require manually searching, fetching, reading and comparing PDF documents from the manufacturer and the standardization agency. CodeIT identified an unfulfilled requirement for an automated software solution that can capture all the data from a manufacturing process as digital data elements. This would allow intelligent processing of information simplifying current manual tasks and providing a foundation for new service-based business models, especially within Life Cycle Maintenance.

2.2 Ontology definition for a Welding Certificate

For this case study, we selected a Welding Certificate document since certificates are a common type of document in MRBs [2].

In defining the ontology, we have followed the steps in the seminal paper by Noy and McGuiness [16]. The ontology describes the entities relevant for a welding certificate, their attributes and their relations to other entities. In our case study, we have six types of entities, i.e. six classes (names of classes and predicates in bold, classes start with a capital letter):

- A **Certificate** has a **certificateID**, is **issued** at a date, **validFrom** a date and **validUntil** a date (these are the attributes). It **coversStandard WeldingCode**, is **validFor Agent** and is **issuedBy CertificateIssuer** (these are the relations).
- A **StandardIssuer issuesStandard WeldingCode**.
- Both **CertificateIssuer** and **Agent hasAddress Address**.
- An **Address** consists of a **street**, a **city**, a **zipCode** and a **country**.

For the sake of clarity our ontology is slightly simplified: In reality a welding certificate does not cover a welding code [17] in its entirety but is valid for a specific range of certification, specifying one or more welding processes (arc welding, gas welding, soldering etc.) with their material groups, i.e. type of metal involved (steel, aluminium and its alloys, cast irons etc.), cf. table in Figure 2. Furthermore, a welding certificate also specifies a responsible welding coordinator (i.e. a specifically nominated person), not only an agent (i.e. a company).

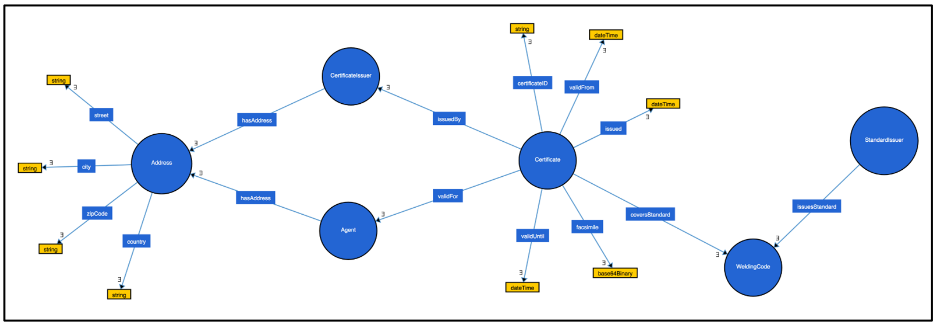

Fig. 1. An ontology for welding certificates

2.3 Document selected for the Case Study

The document selected is the welding certificate shown in Figure 2.

SVS ASS

CERTIFICATE
Welding of railway vehicles and components according to
EN 15085-2

This is to certify that **HAKAMA AG**

Hauptstrasse 50

4112 Bättwil
SWITZERLAND

is qualified to perform welding work within the range of certification of:

Certification level CL1 according to EN 15085-2

Field of application: • New construction of components for rail vehicles specially interior, without design and purchase of welded components.

Range of certification

Welding process according to EN ISO 4063	Material group according to CEN ISO/TR 15608	Dimensions	Comments
141	8.1 8.1	t = 1 - 2 mm t = 1.05 - 3 mm	BW FW
212	8.1	t = 1.5 mm	-
786	8.1	t = 1.5 mm	M3, M8

Responsible welding coordinator: Dipl.-Ing. (TU) Christian Plötner (IWE) [external]
born: 17.05.1980

Deputy with equal rights: -

Deputy: Nicolas Schneider (IWS) born: 15.01.1972

Certificate no.: SVS/15085/CL1/215/0A1/17

Valid: from 21.11.2017 to 20.11.2020

Issued on: 15.11.2018

Auditor: WILKE

General regulations (see reverse)

Schweizerischer Verein für Schweisstechnik

Grütter
head of certification body

Swiss Accreditation sas.admin.ch SCESp 0123

1 of 2

Fig. 2. Sample certificate by Schweizerischer Verein für Schweisstechnik [18]

2.4 The Extract-Transform-Load (ETL) Procedure

- Step 1: Extract the data from the PDF file. This is done by OCR processing of the image-based PDF file. The OCR process used in the case study for the certificate in

Figure 2 creates a two-layer PDF file with the recognized text as a layer under the graphic. This text layer can easily be extracted.

- Step 2: Transform the data to key-value pairs (subject-object pairs in RDF terminology). This is a manual process in the case study.
- Step 3: Load key-value pairs into an RDF-file with subject-predicate-object triples. The ontology guides the transformation.

```
<http://www.ontologies.org/WC/Certificate/#cert001>
   a              Certificate ;
   rdfs:comment   "Welding certificate for HAKAMA AG"@en ;
   certificateID  "SVS/15085/CL1/215/0A1/17"^^xsd:string ;
   validFor       [ rdfs:label    "HAKAMA AG"@de ;
                    rdfs:comment  "A Swiss solution provider in the field of thin sheet metal"@en ;
                    rdfs:seeAlso  <https://www.hakama.ch/> ;
                    Address       [ country   "Switzerland"^^xsd:string ;
                                    city      "Bättwil"^^xsd:string ;
                                    zipCode   "4112"^^xsd:string ;
                                    street    "Hauptstrasse 50"^^xsd:string
                                  ]
                  ] ;
   coversStandard [ rdfs:label    "EN 15085"^^xsd:string ;
                    rdfs:comment  "Welding of metallic materials for the manufacture as well as
                                   maintenance and repair of rail vehicles and parts thereof"@en
                  ] ;
   issuedBy       [ rdfs:label    "Schweizerischer Verein für Schweisstechnik"@de ;
                    rdfs:label    "Association suisse pour la technique du soudage"@fr ;
                    rdfs:seeAlso  <https://www.svs.ch/> ;
                    Address       [ country   "Switzerland"^^xsd:string ;
                                    city      "Basel"^^xsd:string ;
                                    zipCode   "4052"^^xsd:string ;
                                    street    "St. Alban-Rheinweg 222"^^xsd:string
                                  ]
                  ] ;
   issued         "2018-11-15T00:00:00"^^xsd:dateTime ;
   validFrom      "2017-11-21T00:00:00"^^xsd:dateTime ;
   validUntil     "2020-11-20T00:00:00"^^xsd:dateTime .
```

Fig. 3. RDF representation of the certificate in Figure 2 coded as TTL (Terse RDF Triple Language) [19]

3 Discussion

Interoperability is crucial in digital transformation. In the case of Manufacturing Record Books, interoperability is important because of the diversity of documents and the number of actors involved. Our case study showed that even when regarding only welding certificates, these can look very different because of the many existing issuers of standards (ISO, EU standards, British Standards, several [sic!] American professional associations etc., cf. [17]). Thus, a structured approach involving ontology-based modelling is called for as a foundation for semantic interoperability across standards. Interoperability is closely related to the concepts of findability, accessibility and reusability. These four terms are also known as the FAIR Principles (Findable, Accessible, Interoperable, Reusable). The term FAIR has its origin in scientific data management, first coined at a workshop in 2014. In 2016, the 'FAIR Guiding Principles for scientific data management and stewardship' were published in Scientific Data [20]. The authors provided guidelines to improve the findability, accessibility, interoperability, and reuse of digital assets. We are convinced that the FAIR principles are highly relevant for business as well, and that they can be applied more or less unchanged following e.g. the FAIRification Process [21].

As we have discussed in the introduction, contemporary Manufacturing Record Books are still a long way from being FAIR compliant. In the worst (and unfortunately quite common) case they are a collection of digital images originating from scanned paper documents, disguised as PDF files. In the best case, they are digitally born (i.e. the PDF derives from a program like Microsoft Word or a CAD/CAE tool like Auto-CAD) but even then, there exists no standard structure. Additionally, PDF is an inapt format for our goals as it only describes the look of a document and its pages, respectively. However, for automatic processing of a document one needs a format which describes the structure and the semantics of a document. A much better choice than PDF would be an XML-based format.

4 Conclusion

This paper presented a case study of MRB digitalization for industrial applications. An ontology enables standardized digitalization of documents supporting interoperability across actors. The case study demonstrated the feasibility of using an ontology for a welding certificate that is part of an MRB.

References

1. HB Rentals, Manufacturer Record Book, Accommodation Module. https://www.rmbids.com/wp-content/uploads/2018/06/443-HB-Rentals-MRB-Manual-8-2016-1.pdf
2. TechnipFMC Master Document MRB Index template. https://www.technipfmc.com/-/media/technipfmc/Downloads/suppliers/master-document-mrb-index-rev-t.docx
3. Henricksen, K., Indulska, J., & McFadden, T. (2005, October). Modelling context information with ORM. In *OTM Confederated International Conferences" On the Move to Meaningful Internet Systems"* (pp. 626-635). Springer, Berlin, Heidelberg.
4. Verscheure, O., Kiritsis, D.: The Meaning of Data. WMF Report (2018)
5. Martinez Lastra, J.L., Delamer, I.M.: Semantic web services in factory automation: fundamental insights and research roadmap (2006)
6. Iarovyi, S., Mohammed, W.M., Lobov, A., Ferrer, B.R., Martinez Lastra, J.L.: Cyber–physical systems for open-knowledge-driven manufacturing execution systems (2016)
7. Kádár, B., Terkaj, W., Sacco, M.: Semantic Virtual Factory supporting interoperable modelling and evaluation of production systems (2013)
8. Hildebrand, M., Tourkogiorgis, I., Psarommatis, F., Arena, D., & Kiritsis, D. (2019, September). A Method for Converting Current Data to RDF in the Era of Industry 4.0. In IFIP International Conference on Advances in Production Management Systems (pp. 307-314). Springer, Cham.
9. Gruber, T.R. (1993). A Translation Approach to Portable Ontology Specification. Knowledge Acquisition 5: 199-220.
10. Klyne, G., Carroll, J.J., McBride, B., Cyganiak, R., Wood, D., Lanthaler, M.: RDF 1.1

concepts and abstract syntax. World Wide Web Consortium (2004–2014). https://www.w3.org/TR/rdf11-concepts/
11. Nadeau, D., & Sekine, S. (2007). A survey of named entity recognition and classification. *Lingvisticae Investigationes*, *30*(1), 3-26.
12. Conoco Phillips, TIRC. Presentation of Technical Information Requirement Catalog, by B. R. Ydstebø, ConocoPhillips. The NORSOK Z-TI Joint Industry Project (JIP) Launch, 23 May 2018, Norway. https://www.dnvgl.com/software/campaigns-2018/synergi-norsok-jip-project-event.html
13. A Joint Industry Program (JIP). Presentation including Technical Information Requirement Catalog by the NORSOK Z-TI JIP. by P. Rylandsholm, DNV-GL. The NORSOK Z-TI Joint Industry Project (JIP) Launch, 23 May 2018, Norway. https://www.dnvgl.com/software/campaigns-2018/synergi-norsok-jip-project-event.html
14. Global Cents: Manufacturing Records Books: A Guide to Automation with SAP, Whitepaper. https://globalcents.com/resources/manufacturing-records-books-a-guide-to-automation-with-sap
15. CodeIT AS. https://codeit.no/
16. Noy, N. F., & McGuinness, D. L. (2001). Ontology development 101: A guide to creating your first ontology.
17. List of welding codes. https://en.wikipedia.org/wiki/List_of_welding_codes
18. Schweizerischer Verein für Schweisstechnik, https://www.svs.ch/, Certificate: https://www.hakama.ch/fileadmin/HAKAMA/Downloads/Schweiss_Zertifikate/2018-11-15_Cert-Original_EN_15085-2_CL1_-_HAKAMA_AG_EN.pdf
19. Terse RDF Triple Language (Turtle). https://en.wikipedia.org/wiki/Turtle_(syntax)
20. Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., ... & Bouwman, J. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data*, *3*.
21. *FAIRification Process.* https://www.go-fair.org/fair-principles/fairification-process/