tekom

magazine for international information management

tcworld

iiRDS SPECIAL
2025

Enter the age of digitization

Delivering digital information for use with the intelligent information Request and Delivery Standard

DITA meets iiRDS – a plugin for DITA-OT

The possibilities for DITA content

iiRDS for Trusted AI

How DITA and iiRDS enrich Large Language Models

Discover the **iiRDS** Consortium

Shaping

the Future of Intelligent Information

The iiRDS Consortium invites you to explore how we're driving the development and adoption of iiRDS – the **i**ntelligent **i**nformation **R**equest and **D**elivery **S**tandard.

With members from industrial companies, system providers, organizations, and consulting and service companies, we manage and advance iiRDS through collaboration and innovation.

Resources to Get You Started

- Free materials and tools like the iiRDS Open Toolkit, the iiRDS Plugin for DITA-OT, the iiRDS Request API, and the iiRDS Whitepaper.
- Funding support for pilot projects.
- Access to qualified iiRDS Consultants.
- Webinars and events to keep you informed and connected.

Join us!

Explore membership and learn how iiRDS can transform your approach to technical communication.

Members of the iiRDS Consortium

C&O · C-REX · DIN Media · DKE · doctima · DOCUFY · EMPOLIS · Endress+Hauser · etteplan · gds · HELLER · ICMS · kothes · NINEFEB · nexum · PANTOPIX · parson · plusmeta · Quanos · SEW EURODRIVE · SIEMENS · STAR · tekom · tcworld · text-it · T3

Ready to join the iiRDS Consortium?
Visit us at **www.iirds.org** and discover more!

tekom iiRDS

imprint

publisher

tcworld GmbH
Dr. Michael Fritz (CEO)
Claudius Mehne (CEO)
Heilbronner Str. 86
70191 Stuttgart
GERMANY

+49 711 65704 - 0

HRB 22804 Amtsgericht Stuttgart

www.tekom.de
info@tekom.de

advertising

tcworld GmbH
Sales Team
Natalia Zarychta
Tel. + 49 711 65704-58
sales@tekom.de

layout

Irmi Hobmaier
irmi@hobmaier.com

editor

Corinna Melville
www.tcworld.info
editor@tcworld.info

printing

Druckhaus Bonifatius GmbH
Karl-Schurz-Straße 26
33100 Paderborn
https://bonifatius.de

cover image

© sankai/istockphoto.com

tcworld magazine is published every quarter (4 issues per year).

Subscription price of a single issue: 9.00 Euro + VAT + shipping & handling.

Yearly subscription: 32.00 Euro + VAT + shipping & handling.

The minimum subscription period is one calendar year. This will automatically be renewed for a successive period of one calendar year, unless written notice to cancel the subscription is given to the publisher at least six weeks before the end of the calendar year. Printed in Germany

ISSN (Print): 1862-6386
ISSN (Online): 2942-6316

tcworld
magazine for international information management

Embracing the future with iiRDS

With this special edition, we have put together a collection of articles from the *tcworld magazine* about the intelligent information Request and Delivery Standard iiRDS. As we gather for this year's ConVEx conference, we are on the brink of a new era in technical communication, driven by advancements in Artificial Intelligence (AI) and the seamless integration of iiRDS with DITA.

iiRDS is not just a standard; it's a key to smarter, more efficient information management. In a world where data is king, iiRDS offers a structured approach to delivering the right information at the right time, tailored to the needs of users across industries. This magazine will provide you with a broad understanding of iiRDS, its versatility, and the many ways it can revolutionize technical documentation.

One of the most exciting developments in technical communication is the synergy between iiRDS and AI. AI algorithms are now capable to analyze user behavior and preferences, enabling them to dynamically generate and deliver content that meets the users' specific needs. This is facilitated by iiRDS, which structures content in a way that allows for its intelligent delivery, ensuring that information is not only accessible, but also relevant and timely.

Moreover, the integration of iiRDS with DITA opens up new possibilities for content creators. DITA's modular approach to documentation, combined with iiRDS's metadata-driven delivery, creates a powerful toolset for producing high-quality, user-centric content. This edition highlights case studies and expert insights on how iiRDS and DITA together can greatly enhance technical communication.

As tech writers, we shape the solutions and opportunities presented by these technological advancements, and it is therefore crucial for us to stay informed and engaged. This special edition of *tcworld magazine* aims to raise interest and inspire innovation. Whether you are a seasoned professional or new to the field, we hope to encourage you to explore the full potential of iiRDS and what it can do for your content.

For those who want to dive deeper, we also highlight the iiRDS online training, which will give you a more comprehensive understanding of the standard and its applications. And don't miss international conferences such as tcworld China, NORDIC TechKomm, Information Energy, and the tcworld conference, where iiRDS is a major focus theme and where you can network with experts and colleagues from around the world.

So, enjoy the content we have gathered for you in this special edition, and don't hesitate to get in touch. We hope to inspire you to envision how iiRDS can transform your approach to information delivery.

Ralf Robers
Vice Chairman of tekom Germany and President of the iiRDS Consortium

Empolis

Trustworthy AI for Content CREATION and Content DELIVERY

WWW.EMPOLIS.COM

Enter the age of digitization

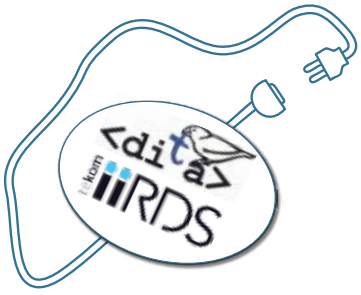
In 2016, a group of experts started to develop the intelligent information Request and Delivery Standard with the purpose of enabling the delivery and exchange of information for use irrespective of the manufacturer. Find out what it is, how it works, and how you can jump on board.

page 6

DITA meets iiRDS – a plugin for DITA-OT

The easy seperation of information into topic types and XML has made DITA a well-established authoring standard. The DITA-OT plugin enables the implementation of iiRDS for DITA content.

page 10



Ten questions about iiRDS

How exactly is iiRDS integrated in the creation of technical communication? In their interviews with *tcworld magazine*, two leading manufacturers explain why they chose the standard, how they put it to use, and what changes it has brought to their content.

page 20 and 24

intelligent information Request and Delivery Standard

» 6 Enter the age of digitization

» 10 DITA meets iiRDS – a plugin for DITA-OT

13 iiRDS for Trusted AI

16 AI-based iiRDS tagging of technical documents

Ten questions about iiRDS

» 20 Interview with the head of technical communication at Endress+Hauser and the iiRDS consultant

» 24 Interview with the technical writing team at Siemens Energy

28 From PDF to a collaborative Information Digital Twin

32 iiRDS for a better support experience

Enter the age of digitization

Delivering digital information for use with the intelligent information Request and Delivery Standard

Text by Susanne Lohmüller

Image: © gorodenkoff/istockphoto.com

The intelligent information Request and Delivery Standard (iiRDS) is a technical standard for the delivery of digital information for use or technical documentation such as electronic operating instructions and user manuals. iiRDS can be used free of charge and is published under a Creative Commons license.

A group of experts at tekom started to develop iiRDS in 2016. In 2018, the iiRDS Consortium was established to maintain and further develop the standard. In 2025, the consortium is focusing on raising awareness for iiRDS at industry events while also advancing international standardization efforts.

Benefits of iiRDS

iiRDS enables the manufacturer-independent delivery, exchange, and aggregation of information for use. This is particularly relevant in the context of Industry 4.0 and the Industrial Internet of Things (IIoT), where technical information from different manufacturers in plants and smart factories is combined.

iiRDS focuses solely on standardizing the exchange format and is not concerned with the way content is created and maintained. This focus on the delivery format for smart content enables iiRDS users to author the content using different programs, for example, content management systems.

The standard consists of a format that defines the file and folder structure of the package containing intelligent information as well as an ontology for enriching content with metadata.

Content enriched with iiRDS metadata supports functions for searching and filtering information and other content functionalities. This smart content can thus be found faster and more accurately in self-service portals, documentation portals, or apps.

A use case

Take the classic iiRDS use case of a service technician tasked with the maintenance or troubleshooting of a product: Here, iiRDS enables the display of the necessary information for the affected machine part that needs to be replaced and shows the relevant information – either directly on the product or via an associated app. The service technician doesn't have to browse through manuals or PDF documents. Even if the service technician is used to classic information for use, he would have to either search in the PDF manually or the search results would depend on the exact term he has entered into the search bar.

Metadata is key for intelligent information. iiRDS provides the appropriate tools for modeling such use cases, thus offering the option to present information for use in an individual, user-oriented, and context-related way. This is especially beneficial for manufacturers of more complex products that integrate various components and the respective information from external suppliers, who use differing terminology. iiRDS standardizes metadata and provides a packaging format for archiving and exchanging technical communication.

Find best practice examples of iiRDS implementations at iirds.org/iirds-consortium/best-practices.

Metadata model

iiRDS uses the Resource Description Framework Schema (RDFS) as the technical format for its metadata model. After registration, these RDF files can be downloaded from the iiRDS Consortium website for free.

iiRDS metadata can be assigned to text fragments, topics, or documents. The iiRDS metadata model is based on the PI classification by Prof. Dr. Wolfgang Ziegler (I4ICM) and defines the categories of metadata shown in Figure 1.

The metadata model can be extended by adding new subclasses or objects to existing iiRDS classes. iiRDS explicitly supports the extension of the metadata model and provides so-called "docking points" for linking external metadata, for example, product metadata or role and qualification definitions.

Delivery format

The iiRDS package is a ZIP file containing the content to be delivered and the associated metadata as RDF. iiRDS comes in two variants regarding the format of content:

- Unrestricted packages can contain any kind of format, for example, PDF, HTML, XML, MP4, SVG, or Office files.
- Restricted packages in the iiRDS/A format contain content only in predefined formats and are self-contained. The predefined formats are PDF/A, a restricted selection of media formats, and XHTML5 with element and attribute restrictions. iiRDS packages are self-contained when the content files only point to files within these packages. The restrictions of the iiRDS/A format aim to ensure that all iiRDS-capable applications process and present the content in similar ways without any additional implementation or customization effort.
- A new format, iiRDS/H, will be included in the next iiRDS version to be released in Q2, 2025. It will cover new developments, e.g., an optional JSON-LD serialization, thus coming closer to the Request API, offering more direct support for chatbots and many other new features.

RDF Schema (RDFS) provides a data-modeling vocabulary for RDF data. RDFS is an extension of the basic RDF vocabulary and belongs to the family of semantic web standards maintained by the World Wide Web Consortium (W3C). A common vocabulary – or ontology – is needed to interpret statements formulated in RDF. RDF is not a metadata vocabulary, but a data model for statements about resources. It can be used to develop vocabularies such as iiRDS, which contains a vocabulary for technical communication. There are also other ontology description languages such as Web Ontology Language (OWL). For further information, visit W3C website at w3.org.

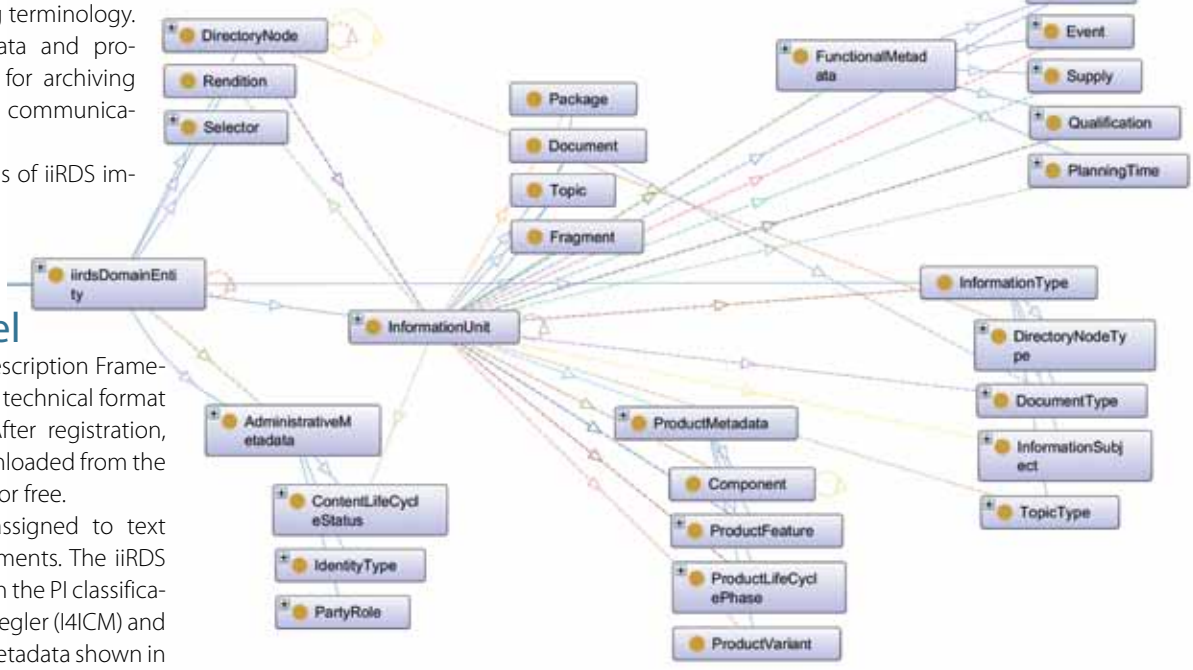


Figure 1: Excerpt of the iiRDS metadata in the Protégé tool

Ways of getting involved

- **Register on our website,** iirds.org, and subscribe to our newsletter to receive relevant updates on new releases, events, webinars, training sessions, and more.
- Upon registration, you can **access the iiRDS materials** such as the current iiRDS version, the sample content, and the iiRDS Whitepaper to get started.
- **Start an iiRDS pilot project** accompanied by a Consortium member from our pool of consultants. Insights into previous pilot projects can be found on our website under Best Practices. Funding is announced annually. The application phase opens in spring.
- **Take part in our iiRDS Online Training** to gain a solid foundation of theoretical and practical knowledge about iiRDS in six weeks of self-study supplemented by group work, online meetings, and quizzes. The next training will take place from April 28 to June 18, 2025.
- **Become a Consortium member.** Discuss your experience with iiRDS experts and get first-hand insights into various projects on intelligent information. In our Consortium, you get the chance to be part of the development of the standard.

Free iiRDS tools to get you started

iiRDS defines a standardized format for delivering smart content from an application that creates iiRDS packages from content (e.g., CMS, authoring or converter tools, iiRDS Open Toolkit) to an application that receives and processes iiRDS packages (e.g., content delivery portals or self-service portals).

To support the application, the iiRDS Consortium provides an **Open Toolkit** that can be used to enrich content with metadata. Users can upload the content in common formats such as PDF, Word, HTML, and XML files. It then generates a valid iiRDS package from content and metadata that the user can check, revise, or add custom metadata values to. The iiRDS Consortium also provides an **iiRDS Validation Tool** to validate iiRDS packages from other systems.

The **iiRDS Request API specification** enables access to data via iiRDS, thus meeting the growing requirements of the economy. It offers system providers and programmers a standardized interface, ensuring modern functionality to support seamless integration. It also offers extended capabilities by covering a broad range of use cases, and allows developers to choose the most suitable query method for their specific needs.

In November 2024, the iiRDS Consortium released the iiRDS Request API mockup. This tool is designed to help developers get hands-on experience with the iiRDS Request API, offering an entry point for implementing the full iiRDS specification.

Key Features of the iiRDS Request API mockup include:

- **Comprehensive route coverage:** This includes all routes defined by the iiRDS Request API specification.
- **Sample responses:** It generates responses that are syntactically correct per the iiRDS Request API specification, based on static sample data. While the mockup doesn't provide dynamic or parameter-sensitive responses, it does validate the syntax and structure of API call parameters.
- **Built for flexibility:** Implemented as a Node.js project, the mockup is easy to run in any environment with Node.js installed. A Dockerfile is also included, allowing you to run the mockup in a containerized environment effortlessly.

The API mockup leverages the OpenAPI 3 definition of the iiRDS Request API to generate its structure and integrates with the Node.js Express library. It operates as a standalone web server, making it a breeze to test and experiment with. Whether you're looking to prototype integrations or explore the iiRDS Request API, this mockup provides a robust starting point for your development journey.

These tools are available as a free web service on the iiRDS website (iirds.org).

Find more information about the iiRDS plugin for DITA-OT, released in March 2024, in our article starting on page 10.

Standardization and interoperability

The iiRDS Consortium has contributed significantly to developing the "Intelligent Information for Use" submodel of the Asset Administration Shell (AAS), published by the Industrial Digital Twin Association (IDTA). IDTA submodels are designed for standardization by the IEC, supporting automated communication between assets, and serve as the foundation for interoperability in digital twins of industrial systems. This submodel provides a standardized link between assets and software systems, addressing the challenges posed by traditional document-based formats like PDFs, which often lack adequate metadata. Based on the iiRDS standard, the submodel aims to bridge information gaps with a metadata model for

intelligent information, enabling efficient, cross-manufacturer exchange and aggregation of technical data.

The iiRDS Consortium is collaborating with the Digital Data Chain Consortium (DDCC) on a joint standardization effort to establish iiRDS and digital manufacturer information as international standards. As a result, a New Work Item Proposal (NP) is now undergoing a vote in the IEC Technical Committee 3. The voting results are expected in March 2025. This NP, based on IEC PAS 63485, incorporates iiRDS content, digital manufacturer information, and references to the digital product passport and relevant IDTA submodels like "Handover Documentation" and "Intelligent Information for Use." If approved, the IEC TC 3 Working Group 28 will work on the draft, further boosting iiRDS' relevance and impact as it becomes anchored in international standardization.

Why iiRDS?

Digitization is transforming industries. From Bluetooth and WLAN to smart factories and Industry 4.0, connectivity is everywhere. But how does this affect technical writing? Understand the framework, approach, and benefits of iiRDS. Discover how iiRDS enables structured, intelligent content delivery and how it supports future-proof documentation. Learn about the iiRDS Consortium and how you can actively contribute to shaping the standard.

ABOUT THE AUTHOR

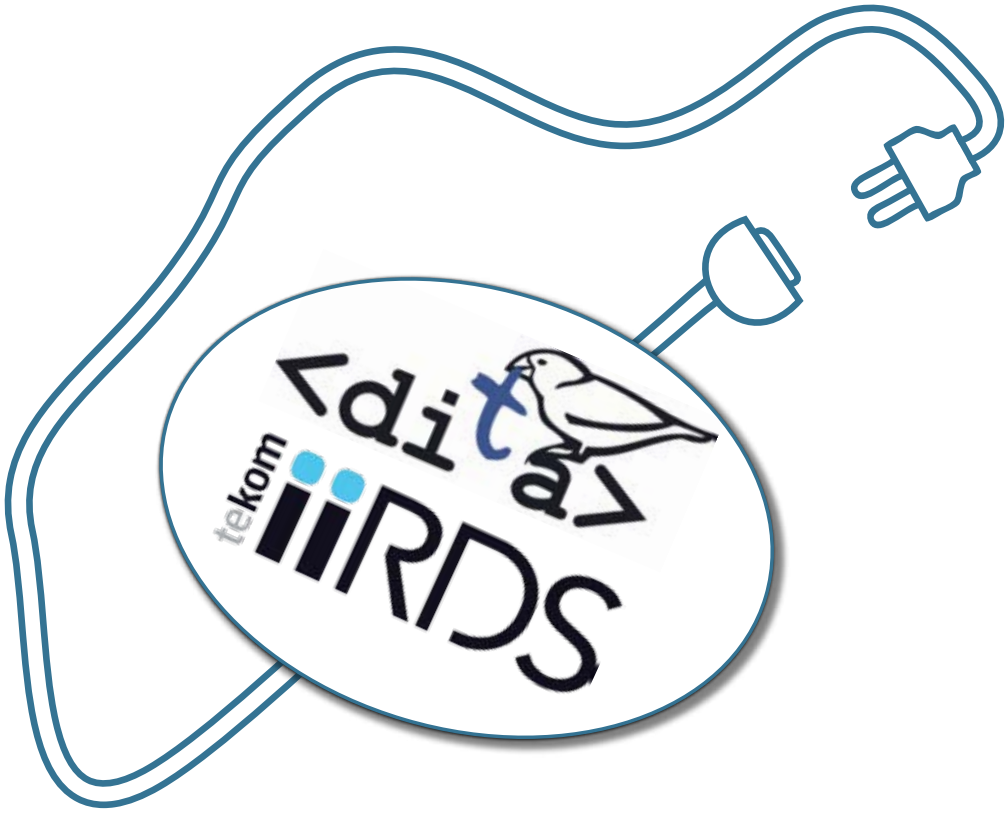
Susanne Lohmüller is the Community Manager and iiRDS Project Manager at tekom Europe. Contact Susanne to get more information and find out how you can get involved.

@ s.lohmueller@tekom.org
website www.iirds.org
@company/the-iirds-consortium

DITA meets iiRDS – a plugin for DITA-OT

The iiRDS Consortium has developed a plugin to support DITA content. The iiRDS package is easy to use and can be customized to meet your project's needs.

Text by Mark Schubert and Dr. Martin Kreutzer



DITA is a well-established authoring standard. It is well structured, customizable, and well supported by an open-source toolkit, the DITA Open Toolkit DITA-OT. You can use DITA-OT to create HTML, PDF, and many other output formats from your DITA files. Single-source publishing is one of the core ideas of DITA and DITA-OT.

The separation of information into topic types and XML as the foundation makes DITA well suited for today's technical communication needs. With microcontent as the new gold standard, DITA allows the authoring of granular information that feeds content delivery portals, chatbots, and AI-based scenarios such as retrieval-augmented content generation.

Introducing iiRDS

For microcontent to be useful, however, it requires metadata. And here's where the Intelligent Information Request and Delivery Standard (iiRDS) comes in. DITA and iiRDS both focus on topic orientation and allow users to get exactly the information they need. For that, iiRDS provides an exchange format that is commonly used to publish technical documentation from content management environments to content delivery scenarios.

iiRDS defines a packaging format and an extensible metadata schema. The packaging format is a ZIP container holding the content files and metadata. Metadata is not part of the content but rather provided in a separate file conforming to the W3C standard "RDF". The Resource Description Framework (RDF) allows you to make Subject-Predicate-Object statements about anything, for example, "Topic A - is about - product B". And that is the metadata we need for our microcontent.

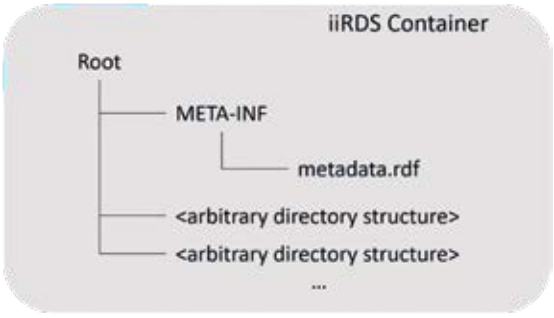


Figure 1: iiRDS package with content and metadata folders

iiRDS has been around since 2016 and is supported today by many content delivery portals. However, DITA-OT has so far only been supported by plugins that were meant as proof of concept. The plugins required specialized XML elements, were difficult to integrate into existing projects, and lacked documentation. In 2023, the iiRDS Consortium commissioned the development of a new plugin. Developed by Empolis and tested and documented by parson, the plugin was released in March 2024. It builds an iiRDS package, is easy to use, and can be customized to meet your project's needs.

iiRDS DITA-OT plugin

The iiRDS plugin follows the principles of DITA-OT. After installation, it provides a new transformation type, `iirds`. With this transformation type you get an iiRDS package that contains HTML files for each of your DITA topic files. The package also contains a `metadata.rdf file`. This file provides metadata for all the HTML files in the iiRDS package.

The metadata is derived from the XML elements and attributes in the DITA files. For example, the iiRDS metadata `Component` is derived from the XML element `component` while the iiRDS metadata `ProductVariant` is derived from the combination of the `prodname` element and the `product` attribute.

Even when there is no explicit metadata in the DITA file, the plugin generates useful iiRDS packages with documents, topics, topic types, and a navigation structure. However, as the main objective of the new plugin was to make it easier to use, not all iiRDS metadata is supported out of the box. What may seem like a shortcoming is, in fact, one of the biggest improvements of the new plugin: You can take your DITA content, even specialized DITA XML, and generate iiRDS packages. The only challenge left is to convince vendors to provide you with a content delivery portal to import your iiRDS packages.

Customization

While the new iiRDS DITA-OT plugin will get you up and running fast, you may end up wanting more. And the new plugin delivers. Basically, every aspect of iiRDS package generation can be customized, including:

- HTML look and feel
- Metadata extraction
- Metadata identifiers
- Processing steps

The iiRDS package contains the basic HTML files generated by DITA-OT, but the iiRDS plugin supports all DITA-OT transformation parameters to customize HTML. You can use your own CSS and XSLT. This will add your custom HTML to the iiRDS package.

But, as we learned, iiRDS is all about metadata. Limited to certain metadata by design, you may want to customize the Java-based metadata extraction. You can do this by registering your custom metadata extractors and supporting all your metadata elements and attributes. Don't want to use the `component` element, but want to use your own grouped attribute values in your specialized attribute? Implement the Java service provider interface of the `IirdsMetadataHandler` and you're ready to go!

Identifiers are a more technical aspect. To assign metadata to content, both need to be addressed in a consistent and stable manner. Make sure that the same information is always identified by the same identifier. The same goes for metadata concepts. If the identifier for the component `gasket` changes over time, a content delivery portal will not know that the two components are the same. But just like the metadata extractors, you can implement your own Java service provider and customize the default way of assigning identifiers.

Finally, the iiRDS plugin allows you to add your own processing steps. DITA-OT extension points allow you to add your own ANT processing.

All in all, the plugin is ready for production use. The code base is tested, the functionality is documented and, if there are project-specific needs, you can customize it with parameters or your own additional plugins.

Where to get it

You can find the iiRDS plugin in the DITA-OT plugin registry (www.dita-ot.org/plugins). It allows for easy installation. Open a terminal in the bin folder of your DITA-OT installation and type the following:

```
dita install org.
iirds.dita.package
```

If you don't want to install it right away, you can download the plugin from https://iirds.org.

The plugin is published under the Apache 2.0 license and is open for use and customization by the DITA community.

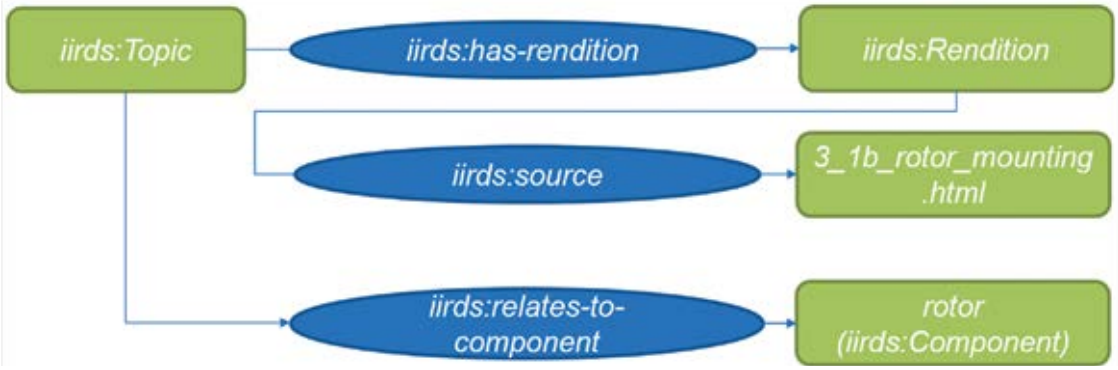


Figure 2: iiRDS metadata of a topic assigned to an HTML file

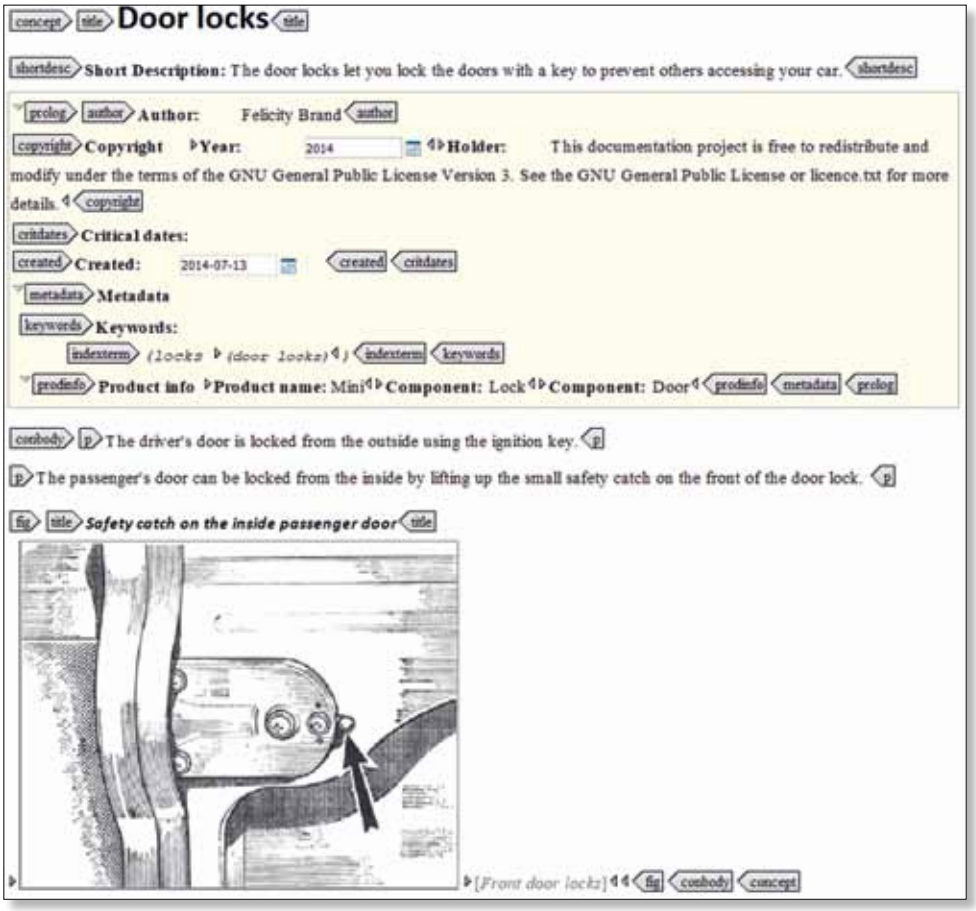


Figure 3: DITA content

ABOUT THE AUTHORS

As a Principal Solutions Engineer at Empolis Information Management GmbH, **Dr. Martin Kreutzner** supports customers in introducing iiRDS at the interface between content management and service portals. He is involved in the maintenance and further development of the standard in the iiRDS Consortium.

@martin-kreutzer
www.empolis.com

Mark Schubert is a technical consultant at parson AG and supports clients from the software industry with the use of authoring tools and publishing solutions. DITA, DocBook, and other document type definitions are part of his daily routine. He has been an acting member of the iiRDS working group "Development" from the start.

www.parson-europe.com

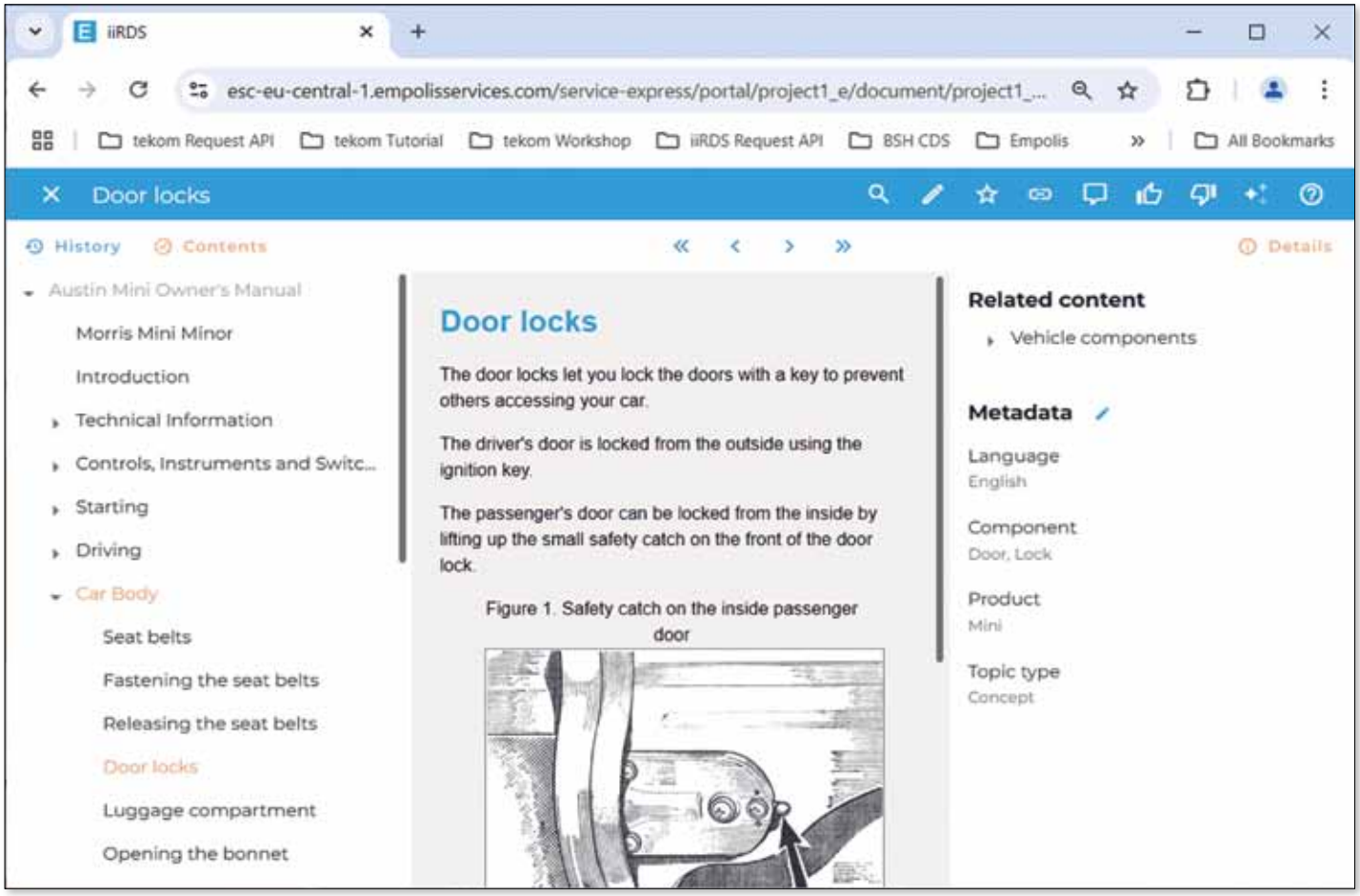


Figure 4: Topics resulting from DITA in a content delivery portal

iiRDS for Trusted AI

Can we trust AI to deliver reliable information in industries where accuracy is crucial, even vital? Large Language Models enriched through DITA and iiRDS deliver promising results.

Text by Markus Kronfellner, Helmut Nagy, and Harald Stadlbauer

Large Language Models (LLMs) and Generative Artificial Intelligence (GenAI) are the major focus of today, not only in technical communication. They promise to solve almost everything and cost almost nothing compared to metadata or master data management projects.

But how accurate is the information produced by LLMs? Accuracy is crucial, particularly for safety-critical information in technical documentation. In this article, we will compare the accuracy of information produced with three different systems: a standard LLM; a standard GenAI retrieval mechanism using structured data with DITA; and a trusted knowledge base using structured data with DITA enriched by iiRDS to retrieve the right context for optimized GenAI output.

We will reveal how a knowledge-based approach to enriching DITA in combination with iiRDS produces far more accurate results, which is vital for safety-related information.

Why do we need Intelligent Information?

Intelligent Information is information that is contextually retrievable, thus helping users to find the specific information they need to solve a problem.

Various developments have greatly increased the need for Intelligent Information: First of all, due to the heavy usage of digital devices, Generation Z has shorter attention spans than previous generations. Secondly, we are facing an upcoming skills shortage due to the retirement of the baby boomer generation, which will lead to a gap in practical knowledge. Moreover, current upskilling processes will no longer work.

Just consider the following case that is quite common in our industry: In former times, out of a service team of ten technicians, one retired per year and was replaced by a young one. The new member was trained for half a year, and the other nine technicians took over the additional work until the new recruit was fully trained. Nowadays, however, we are facing a scenario in which out of a team of ten service technicians, six to seven will retire over the next three to four years, causing drastic workforce shortages.

As the only solution to solving this brain drain, newcomers need the right information, in the right context, at the right moment, and in the right output format, to be productive right from the start. For that, a trusted retrieval and recommendation mechanism is essential.

Is AI a valid answer to everything?

In the hype of OpenAI's ChatGPT, GenAI/LLMs are seen as a panacea for everything. However, LLMs are showing numerous drawbacks:

- **They lack traceability and explainability of their results.** Traceability is a

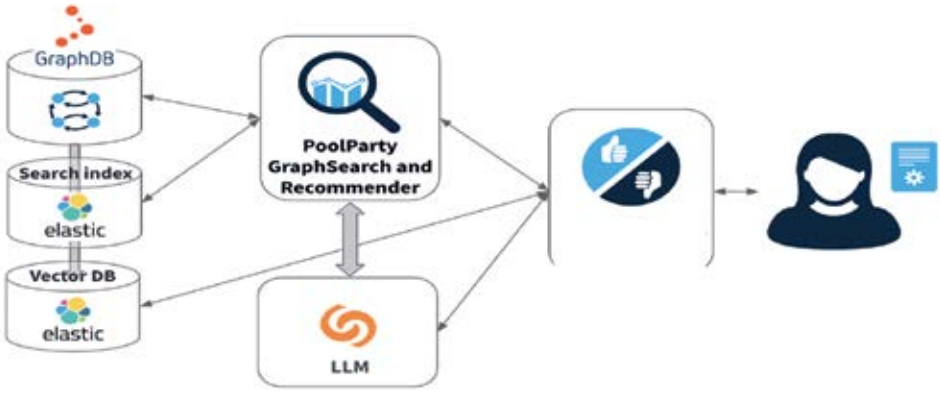


Figure 1: The iiRDS RAG pipeline

legal requirement in heavily regulated sectors, such as medical technology.

- **LLMs seem to struggle with sequences.** Sequences are necessary to retrieve the right order for task instructions. If mixed up, this might compromise safety.
- **LLMs have no sense of the varying magnitude of importance of different content.** In machinery, for example, it is a legal requirement to have hazard statements in the right place.

Solving these issues requires a lot of money and resources for additional training as well as for the search of suitable prompt chains.

Introducing Trusted AI

So, how can we counter the flaws of standard LLMs?

In order to find the best solution, we have put several LLM scenarios to the test. For our studies, we used a SANDVIK bolter miner and its documentation. In this scenario, hydraulics work is dangerous due to the high pressure of the hydraulics oil and the fact that you cannot see a pressurized oil stream coming through a pinched hose. Troubleshooting for such hydraulics systems is complex and a fault can have different causes. Also, it is highly important to retrieve the right warnings in hazardous situations.

For this study, we investigated the accuracy of the information produced by

1. An LLM alone
2. An LLM optimized with standard (vector-based) Retrieval-Augmented Generation (RAG)
3. An LLM optimized with an iiRDS-based RAG, where we enriched DITA topics and maps with the corporate taxonomy, mining-specific domain taxonomy, and the iiRDS core and machinery ontology to establish a knowledge graph that optimizes the retrieval of context for the RAG system.

By referencing a corporate knowledge base, the iiRDS RAG selects the specific relevant iiRDS-tagged content. Thus, only trusted knowledge is fed into the LLM, resulting in what we call **Trusted AI.**

Bringing iiRDS and GenAI together

We created iiRDS entities and relations from the topics in our technical documentation as well as additional e-learning nuggets. These were stored as a knowledge graph in a graph database.

A graph search algorithm is getting the right topics and documents out of the database as recommendations using the knowledge graph structure of iiRDS (for more information, see www.iirds.org). This is then used as input for the LLM. Results are again checked by the iiRDS-based RAG.

What are the results?

Using the content (DITA topics) of the operating manual of a bolter miner, we tested the three scenarios:

1. LLM alone
2. Vector-based RAG with an LLM
3. iiRDS-based RAG with an LLM (i.e., a Graph RAG based on iiRDS).

As you can see in Figure 2, the results were overwhelming.

Due to its trusted basis of input, iiRDS-based RAG scored far better than all the others. Adding more granularity to the hazard statements can raise the percentage of awareness even further.

RESULTS Question 2

	LLM LLM alone	**Vector search** Vector-based RAG with LLM	**Knowledge graph** iiRDS-based RAG with LLM
What causes this fault?	8.4 %	30.6 %	72 %
What needs to be checked?	8.2 %	27.8 %	72 %
What needs to be done?	6.7 %	23.3 %	90 %
What requires awareness?	17.0 %	0.0 %	32 %
Summary	**9.9 %**	**20.5 %**	**68 %**

Figure 2: Performance comparison (LLM only vs. vector-based RAG vs. iiRDS-based RAG)

Conclusion

Our study shows that iiRDS as the only existing standard schema for the retrieval of technical information secures the best LLM output taken from the basis of technical documentation. [1] Interfacing iiRDS with the language capabilities of LLMs promises accurate results in the right context, at the right time – and future-proof solutions.

Reference

[1] IEC PAS 63485:2023 – Intelligent Information Request and Delivery – A process model for the exchange of information for use; 2023. https://webstore.iec.ch

ABOUT THE AUTHORS

Helmut Nagy is VP Sales Enablement of Graphwise, the provider of the Graphwise Platform. He is a well-known expert in the fields of knowledge management and semantic AI.

@helmutnagy
https://graphwise.ai

Harald Stadlbauer holds a Master's Degree in Software Engineering and a Doctoral Degree in Electrical Engineering. He is the acting General Manager and shareholder of the NINEFEB Group of companies.

@harald-s-3074a595
www.ninefeb.com

Markus Kronfellner, M.Sc., holds a Master's degree in Technical Documentation. He is a consultant at NINEFEB Technical Documentation GmbH.

www.ninefeb.com

AI-based iiRDS tagging of technical documents

An IUNTC presentation in November 2024 explained the use and challenges of Large Language Models for the automated indexing of technical documents. Here is a summary.

Text by Gerald Adam Zwettler

Large Language Models, such as ChatGPT, are used by many people for various tasks, including translation, summarizing texts, improving texts or writing short letters. During my presentation at the IUNTC meeting, I introduced the Text-IT project as a case study of using AI for iiRDS tagging of technical documentation. Findings from this case indicate the challenges of using such tools in a professional, deterministic environment, because it is difficult to achieve consistent and reliable results with a stochastic system as is required for the indexing of technical documentation.

Image: © Valeriy_G/istockphoto.com

Training Large Language Models

Large Language Models (LLMs) are trained on a broad base of text data. Their unsupervised learning is based on recognizing patterns in the order of letters and words. Basically, they learn how letters follow each other in languages such as German or English. This approach makes it possible to identify linguistic patterns. However, there are also areas where Large Language Models show weaknesses. Their performance can be improved through additional methods such as reinforcement learning and other learning approaches. One particularly effective approach is self-supervised learning. Thanks to the vast amount of text on the internet – from books to articles to websites – language models can be trained efficiently. The concept of "cloze texts" is often used here: Parts of a text are removed and the model is trained to insert the missing words or phrases correctly. This method is extremely flexible, texts from almost any field – whether law, history or fiction – can be used. The principle is simple: A text is partially emptied and the model learns to fill the gaps with the best possible words. This enables an efficient learning process based on an almost unlimited amount of training data. The approach allows models to be continuously improved and further developed. It is the core of how Large Language Models are trained and why they are so powerful.

Improving prompt engineering and tuning

To achieve near-optimal or even deterministic results with LLMs, an important question is how to improve prompt engineering and prompt tuning in a standardized and professional way. The concept of so-called "mega prompts" involves dividing each task that you want the LLM to perform into clearly defined sections. These sections should be formulated as precisely as possible.

Assign a role

The first step is to assign a clear role to the language model. For example, we could say: "You are an expert in technical documentation. Please help us." The model should always know in which role it should act – be it as an expert, teacher, or advisor. This is particularly important in educational contexts: Students should learn to make specific requests, e.g.: "I am a student, I am having problems with this task, and I need specific support." Clearly assigning roles in this way significantly improves results.

Outline the task

The task should be described precisely, including the individual work steps. LLMs benefit when complex tasks are broken down into smaller steps that can be processed iteratively. An example:

1. Analyze the text.
2. Think about the content.
3. Structure the text.
4. Create a summary in bullet points according to certain specifications.

By specifying such steps and instructions, the model can work more efficiently and in a more targeted manner.

Provide context

A language model can only work with the information it is provided with. Without context, such as the geographical, temporal or situational background, the model will not be able to provide optimal answers. An example: "We are in Central Europe, it is winter. Should I wear a jacket today?" This type of context must be specified explicitly as the model has no situational awareness of its own.

Define the output

The desired output format should also be clearly defined, especially if there are specific requirements, such as compliance with certain standards or structural specifications. Structured formatted texts are much more useful than unstructured output.
An effective mega prompt could look like this:

- Role: "You are a research expert."
- Task: "Formulate a precise summary."
- Steps: "1. Analyze the source. 2. Structure the content. 3. Create a summary according to a list of points."
- Secondary conditions: "Follow specific guidelines and provide the response in XML format."

Such detailed prompts lead to much better results than simply entering a text and hoping for useful results.

Prompt optimization

Language models can also be enriched with specific commands, e.g., for multiple-choice tasks, XML exports or other automated processes. For example, in the education and e-learning sector, exams or content could be created efficiently in this way. By introducing parameters, models can even be controlled programmatically in order to meet specific requirements – such as legal issues or content-related topics.
A particularly interesting aspect of prompt optimization is that the LLM itself often knows best how to handle its functions. If a prompt needs to be improved, AI itself can help. For example, if you are unsure how an optimal prompt should be designed, you can ask the language model directly: "You are now a prompt generator. Please define the best prompt for this task." This is a new and effective approach where the model itself suggests how it can best be used. In an earlier project, the LLM was even able to act as a tokenizer: It suggested splitting texts into smaller units to accomplish the task more efficiently.

A step-by-step guide

The following procedure is recommended to achieve good output results:

1. **Divide tasks into smaller steps:** LLMs deliver better results when tasks are broken down step by step and into manageable units. In this way, the model can work more efficiently and avoid errors.
2. **Validate results:** There is always a certain risk of so-called "hallucinations" (false or invented information). Repeat the same task several times and compare the results to ensure consistency. Several runs can help to identify optimal solutions.
3. **Provide examples:** Show the model concrete examples of the expected results or solutions. The more examples provided, the better the model can learn to fulfill the task.

NINEFEB

Technical Documentation

compliant, consisent, interactive and structured

iiRDS & AI

iiRDS integration,
Trusted AI using GenAI and iiRDS based RAG

Graphical Illustrations

interactive spare parts catalogues,
2D & 3D animations

eLearning & Training

design for interactive courses,
XR experience spaces,
microlearnings,
personalised learn paths

Telephone: 0043 1 600 2000 390
Email：info@ninefeb.com
Website：www.ninefeb.com

4. **Iterative improvement:** If the results are not satisfactory, work iteratively and adjust prompts or framework conditions.

AI and LLMs can be used for almost any task – be it for optimizing prompts or structuring complex processes – making stochastic language models more deterministic and reliable in their use.

Automatic indexing of documents

The objective of the "Text-IT" project was to implement automatic or semi-automatic tagging of documents, whereby the tags had to comply with the iiRDS standard. There were restrictions regarding the input format. Image data was excluded as it was not part of the project scope, although good progress is now being made in image analysis. LLMs were used to achieve the project goals. Some instructions were extracted from the document text and made available to the model via prompts. This required not only importing the commands, but also the entire text, which slightly increased the workload. The model should respond with a standard-compliant list of texts, which was available in a textual form. A structured output, ideally in XML or JSON format, makes further processing much easier. A pure text output, on the other hand, would have been less useful from an IT perspective. An example of an expected output could look like this:

- Language: German
- Information topic: Generic collection
- Conformity: Generic conformity

Of course, it also required post-processing to put everything in the right context.

Different variants of interaction with the language models were also examined. A distinction was made between tagging (assignment of predefined tags) and labeling (user-defined assignment).

Tagging uses predefined options (e.g., subject areas) from which the language model selects the appropriate tags. This approach is easier to handle and enables consistent results.

In the second type of interaction, labeling, user-defined tags were used. Training was somewhat more difficult here. While automated tagging quickly reaches its limits, especially when adapting to new subject areas, labeling offers a little more flexibility, but is more complex.

Another aspect of keywording is the number of calls. A single call to complete the entire task and obtain a result would only work to a limited extent, as the language model would be overloaded. A better strategy was to split the document and the text and divide the task into several calls. With this so-called multi-call strategy, the results were much more precise as smaller tasks could be targeted.

The evaluation compared the different strategies and their impact on outcomes. We specified context, purpose, topics, rules, location, aspects and boundaries, among other things, in order to formulate the prompts precisely. But the most important factor is how prompts are defined. Strongly defined prompts are crucial to achieve the best results.

The model configuration was deliberately generic so that parameters such as the model, the input prompt, and the temperature could be easily adapted via configuration files. This means that when new or cheaper models become available, a simple line change can be made without modifying the code.

The "temperature" in this context influences how creatively or neutrally the model responds. For technical documentation, strictly neutral responses were to be achieved, so the temperature was set correspondingly low. The model also needed to be provided with the iiRDS statistics and associated commands.

The prototype of the user interface was very similar to the final goal: The AI automatically assigns texts to tags, with the user having the option to make subsequent adjustments if errors or inconsistencies occur. When the text is extracted from a PDF, the automatic assignment takes place and the user can correct it as required.

When evaluating the accuracy of the emulation, it was found that the results for closed tagging were very good, especially for predefined texts. However, problems occurred with the number of tags, which is why an emulation metric was developed to correctly evaluate missing tags in the error statistics. In the free text analysis, it was more difficult to make an exact assignment, as the similarity of the text had to be compared. A similarity matrix was used to help identify and correct errors, although this is still a challenge.

The results were also compared with those of human experts in the field of technical documentation. The agreement was good, but there were discrepancies, especially in the interpretation of tags such as "caution", "warning", "danger" or "note". These terms lie in a gray area where different interpretations are possible, even if their definitions are relatively clear. This was an important learning point in the project.

In terms of cost, the calculations for tokens and a good model were relatively inexpensive, minimizing the need for manual review by experts.

Findings from this case suggest that if the human effort required to check and correct the results is less than the entirely manual processing of the document, AI costs can become insignificant. This shows the potential of a business model in which AI provides support while manual control of the results is retained.

In summary, it can be said that Large Language Models can certainly be used for technical documentation, but some critical aspects require fine-tuning.

This is a summary of the IUNTC meeting on November 14, 2024, by Gerald Adam Zwettler.

ABOUT THE AUTHOR

Gerald Adam Zwettler is teaching and researching at the Hagenberg Campus of the University of Applied Sciences Upper Austria in the fields of signal and image processing, computer vision, project development, and machine learning. He is head of the Advanced Information Systems and Technology research group, which conducts research in the fields of data science, machine learning, and computer vision.

@gerald-zwettler-59b3a8104

Part 1

Ten questions about iiRDS

Two founding members of the iiRDS Consortium explain the importance and potential of iiRDS and its integration at Endress+Hauser. An interview with Thomas Ziesing from Endress+Hauser and iiRDS consultant Ulrike Parson from parson.

Interview: Susanne Lohmüller

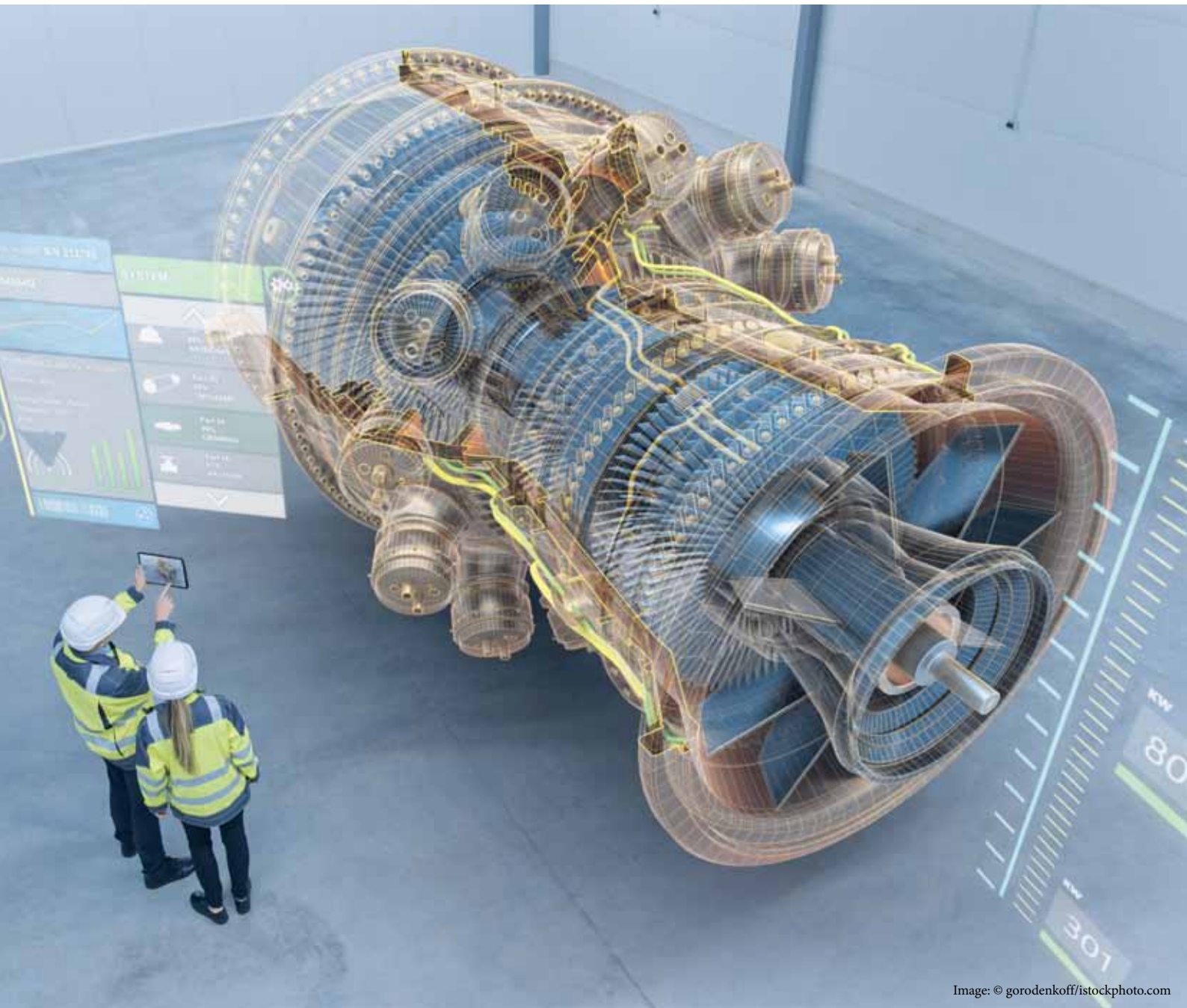


Image: © gorodenkoff/istockphoto.com

How did you come across iiRDS?

Thomas Ziesing: tekom brought the topic to our attention. We found the approach and possibilities of iiRDS so interesting , we joined the Consortium as a founding member in 2018.

Since then, we have been working in the Consortium on the further development of iiRDS.

Why did you choose iiRDS?

Thomas Ziesing: We see the greatest potential in the openness and expandability of iiRDS.

As one of the leading providers of measurement instrumentation, services, and solutions for industrial process engineering, Endress+Hauser has gained sufficient experience in the past with standards for describing and classifying our products, e.g., NE100, ECLASS. In doing so, we have often encountered the problem that certain aspects and features and, in some cases, entire device types cannot be mapped or cannot be mapped completely.

In addition, many of these standards are tailored to the requirements or needs of a specific industry. However, as Endress+Hauser is active in a wide range of industries, this often presents us with problems and generates increased effort.

We see iiRDS as an opportunity to map our information comprehensively and holistically. In addition, iiRDS is combinable with other industry standards such as ECLASS, and the consortium actively cooperates with such industry standards. This makes it possible to map all our metadata in a single model and thus optimize our processes. Efforts towards international standardization as a process model for the exchange of information also benefit our activities regarding digitalization and Industry 4.0.

Ulrike Parson: iiRDS offers a standardized information model for technical documentation that is ideally suited as a blueprint for company-specific metadata models, regardless of the technical implementation. And this knowledge can be used by any company free of charge, as iiRDS is an open-source standard.

For which products do you use iiRDS?

Thomas Ziesing: We are currently working on a project in which we are completely redesigning the authoring, management, and delivery of technical communication information. This information is authored as topic-oriented information units and stored in a comprehensive information model. These information units are then linked to the contexts for which they are valid via a graph-based product model. The aim is to be able to generate order code-specific documentation in the future and to make these information units available to other applications in a context-oriented manner.

iiRDS serves both, as an ontology for structuring and classifying the content and as the delivery format for making the content available.

How do you use iiRDS in technical writing?

Thomas Ziesing: As part of the project mentioned earlier, the content of the technical documentation is created in information units based on topics. These information units are DITA-based and are automatically enriched with the metadata that they receive through their mapping to the information model and product model. As iiRDS is the underlying ontology, the metadata is based on it.

In the future, the information units will not only be made available in information products such as operating instructions but will also be addressed and delivered individually

ABOUT THE INTERVIEW PARTNERS

Thomas Ziesing worked as a media documentalist in the chemical-pharmaceutical industry until joining Endress+Hauser more than 25 years ago. He was involved in several projects in the online sector before taking over responsibility for the translation management process and the coordination of the technical editing departments of the Product Centers at Endress+Hauser Consult AG.

Ulrike Parson is the founder and CEO of parson AG, a service provider for technical communication. Her areas of expertise are consulting for content strategy, digitalization, and intelligent information. She is a member of the iiRDS Steering Committee.

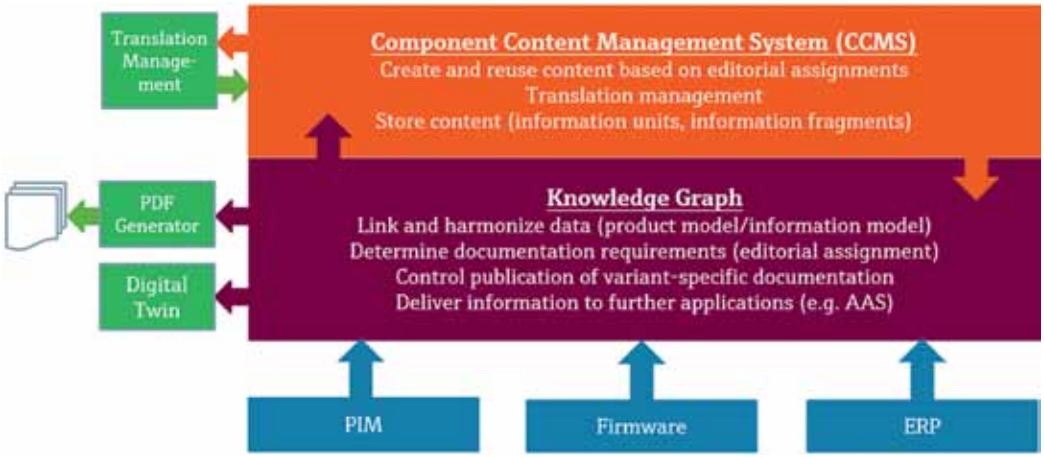


Figure 1: Technical writers work with various sources of product information.

and specifically via the metadata in order to be used, for example, in an app or website. iiRDS will serve as the transfer format: The information units are transferred to a content delivery portal in the form of iiRDS packages. End users and apps can find the information from the portal via metadata and retrieve it for further use.

Have you brought in external support?

Thomas Ziesing: We called in parson AG for technical and information architecture consulting in this project. Like Endress+Hauser, parson is a member of the iiRDS Consortium and brings a great deal of expertise to this area as well as to DITA, which helps us a lot.
We are also working together with Empolis, another iiRDS Consortium member, as a technical partner. Here, we use their Knowledge Graph component for modeling product information and documentation metadata as well as linking CCMS content with product data.

Which other departments are involved in the implementation of iiRDS?

Thomas Ziesing: The project in technical communication is a pilot with which we want to test and demonstrate the possibilities of iiRDS. In principle, it is feasible to extend this concept to other types of information, and we are planning to do so.

Ulrike Parson: In our iiRDS project, information from different groups and sources is already merged: The product data comes from different systems such as a PIM and a software configuration system, the information model is maintained by the technical writing groups, and product management is responsible for ordering options for the products. Thus, the project also has the major task of harmonizing data from different sources.

Thomas Ziesing: There is great interest in this harmonized product data and documentation content, which can be compiled to match the order code, especially in the area of digitalization activities. On the one hand, we are planning to integrate this content into our own applications and, on the other, to deliver the technical documentation in an administration shell for the digital twin. The corresponding submodel "Intelligent Information for Use" is also based on the iiRDS concept. [1]

Do you involve your suppliers?

Thomas Ziesing: We are already talking to one of our customers, a plant manufacturer, who is very interested in our concept and would like to integrate iiRDS packages into its engineering processes. They see considerable potential for simplifying and optimizing their processes.
I expect that such collaborations will increase as soon as the project is completed and we can make the content of the technical documentation available in the form of topic-oriented information units.

How have processes in technical writing changed with the introduction of iiRDS?

Thomas Ziesing: The implementation of the project will fundamentally change the processes in technical writing. The document-oriented authoring and management of content will be replaced by topic-oriented authoring and management based on an information model and product models.
Technical writers will no longer have to work in the context of documents but based on editorial assignments. In an editorial assignment, the content for a specific subject area is compiled according to the product vari-

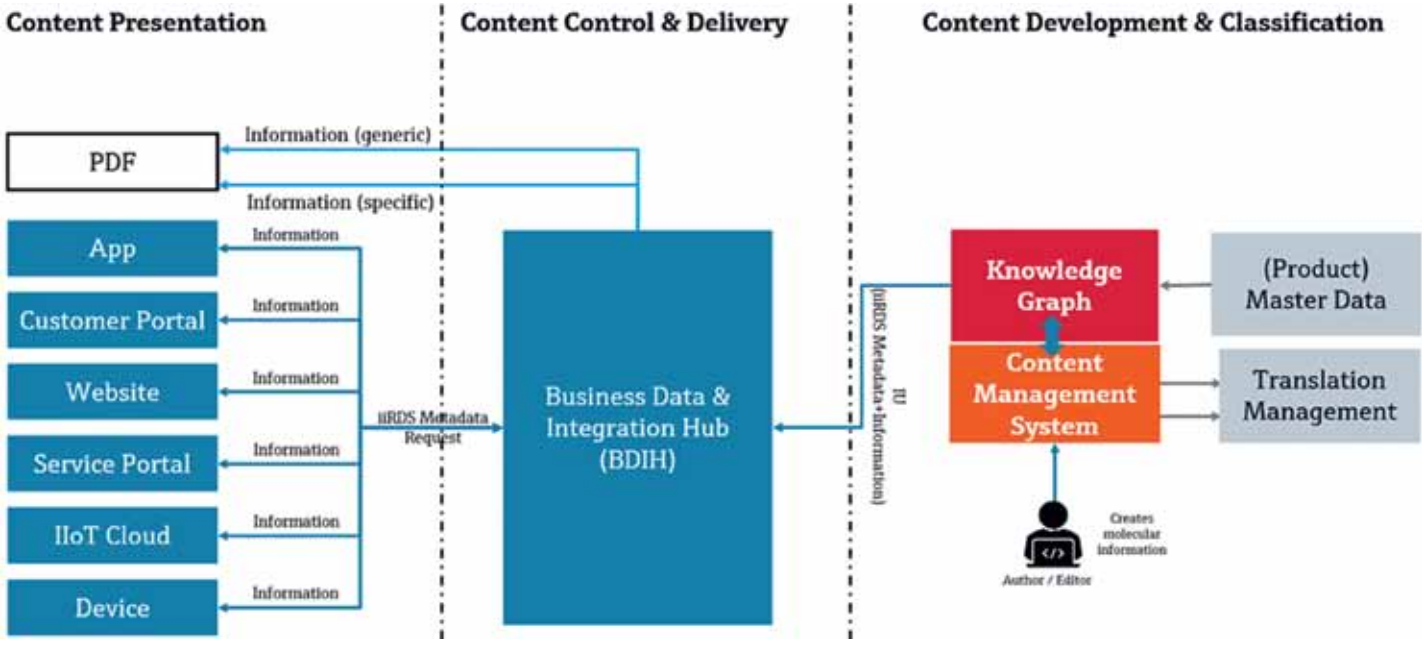


Figure 2: Overview of the targeted information process at Endress+Hauser from content development to presentation

ant, e.g., for the commissioning of the product. The content is compiled based on the data and relations in the Knowledge Graph, which manages the product and information models. Existing information units are automatically determined based on the assigned metadata, and missing information units are created. The structures of these information units are DITA-based.
Once the technical writer completes the editorial assignment and the content is approved, the information units are translated and published. As a result, they are available in all necessary languages for use in a content delivery portal.
Compiling and publishing information products, such as operating instructions or data sheets, are independent of the authoring processes. Information products are generated on demand. Mechanisms such as an automatic completeness check ensure that all the necessary information for an information product is available. The information products are equipped with all the required metadata and can be published in various output formats. We hope that this will significantly speed up the authoring process, as editorial assigments can be started earlier and in a more focused manner. The higher granularity of the information units will give us greater flexibility in terms of use, allowing us to use content in different information products and contexts. This has taken us to the point where we will be able to deliver order code-specific documentation and information in this way. It also prepares the technical writing team for the growing requirements associated with digitalization and enables them to contribute to formats such as the administration shell or the digital twin.

What future benefits do you expect from iiRDS?

Thomas Ziesing: As already mentioned, iiRDS serves as an ontology for structuring and classifying content and will be the delivery format for making it available. The openness and individual expandability allow us to classify the information on our devices and products without major restrictions.
The search and delivery options also meet our requirements. This makes it easier for us to provide technical documentation information for different applications.
We also see the cooperation and coordination with other committees such as ECLASS or VDI 2770 as an advantage, as this will considerably simplify the exchange between different systems.

Ulrike Parson: Using standards such as iiRDS and DITA makes the new documentation concept less dependent on the technical implementation in a specific content management system. At the same time, we are using the knowledge contained in these standards and do not have to develop all concepts from scratch. We can use future additions to the standards for our project.

Where do you see potential for improvement in iiRDS?

Ulrike Parson: For a long time, one area for improvement was the lack of a request interface for iiRDS content. However, the Consortium is currently working on this, and we are hoping to be able to use this API for future delivery of Endress+Hauser content[2].
Also, iiRDS deliberately does not map the product dimension, e.g., product functions, product features, and product variants. iiRDS is an information model, not a product data standard. However, this must be taken into account in a project such as Endress+Hauser's – the product model must be developed specifically for the company and linked to iiRDS.

References:

[1] https://landkarte.interopera.de/wp-content/uploads/2023/08/Submodel-Specification-Document_Intelligent-Information-for-Use.pdf
[2] The API for querying and retrieving iiRDS RDF resources is now available at iirds.org/tools/request-api.

iiRDS Request API

The specification for the request interface has been published by the iiRDS Consortium on their website (iirds.org), along with a mockup. iiRDS tool developments, standardization efforts, other advancements, and use cases were presented at tcworld conference 2024. The tools and presentation recordings are available on the iiRDS website and YouTube.

iirds.org
tcworldconference.tekom.de
youtube.com/@tekom-iiRDS

ABOUT ENDRESS+HAUSER

Endress+Hauser is a global provider of measuring instruments, services, and solutions for industrial process engineering. The company offers process solutions for flow, level, pressure and temperature measurement, analytical measurement, data logging and digital communication, and optimizing processes in terms of economic efficiency, safety, and environmental impact. Customers operate in a wide range of industries, such as chemicals, energy and power plants, basic materials, metals & mining, food, life sciences, oil and gas, and water/wastewater.

ABOUT THE TECHNICAL WRITING TEAM

More than 40 technical writers work across technical editorial offices in five Product Centers. They currently use COSIMA Enterprise to create operating instructions, short instructions, safety instructions, installation instructions, and other documents, primarily in the form of print documents, which are made available in up to 30 languages.

Part 2

Ten questions about iiRDS

How is iiRDS proving itself in businesses, for example, in power plant construction? What potential does the standard have, and where is there room for improvement? An interview with Michael Straeter and Tobias Köffer from Siemens Energy provides answers.

Interview: Susanne Lohmüller

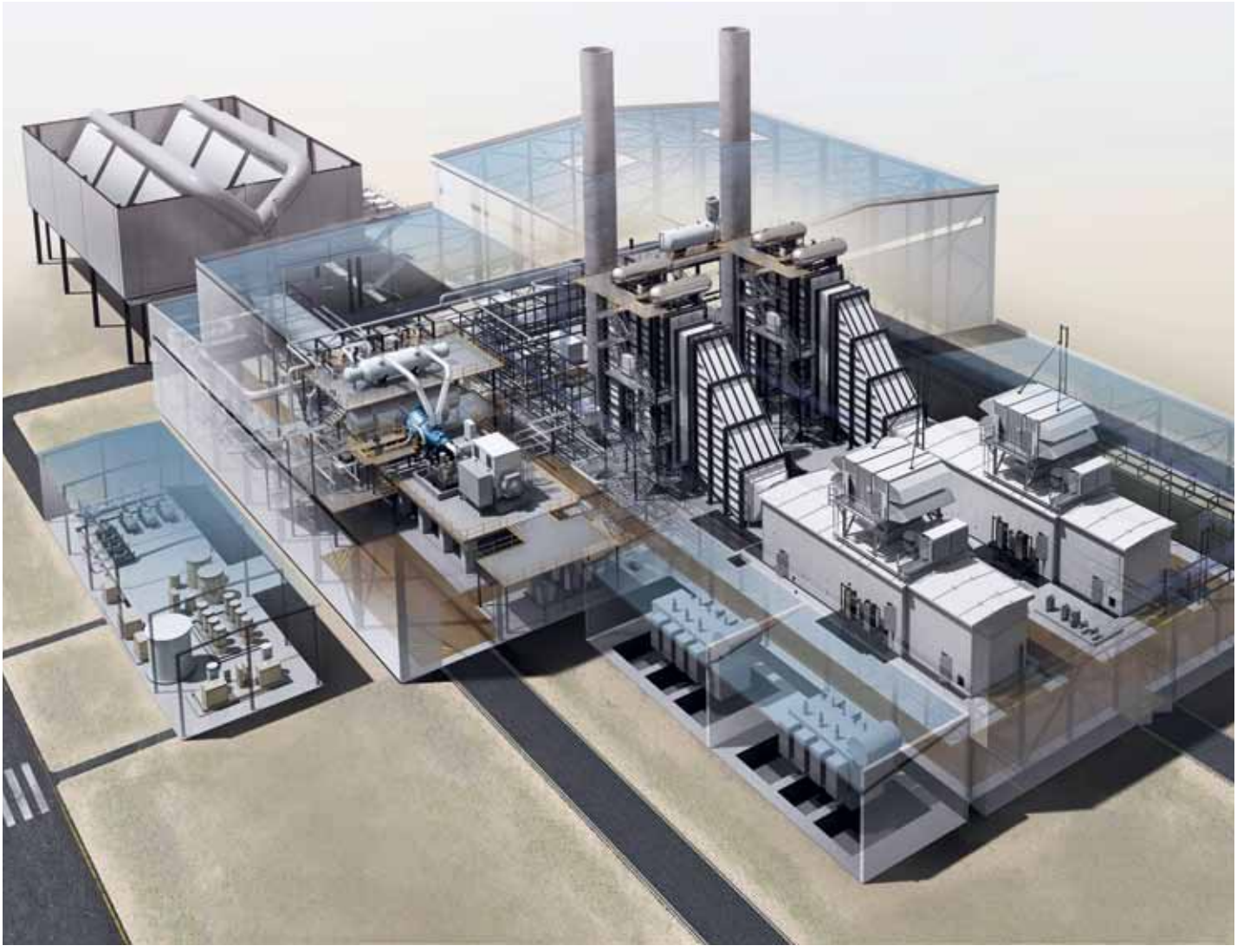


Figure 1: Model of a gas and steam power plant; the technical editorial department now also produces documentation for the large components supplied with the aid of iiRDS. *© Siemens Energy*

The intelligent information Request and Delivery Standard (iiRDS) defines uniform metadata for all information products. This simplifies the exchange of information, for example, between a manufacturer and its suppliers. But iiRDS can also improve information processes within the company, as it does at Siemens Energy.

How did you come across iiRDS?

Michael Straeter: We first became aware of the topic through tekom publications and events. We have been discussing the topic internally since 2018 and have been talking to service providers and suppliers. We then set up our first pilot project in 2019. The question was if we could display our central information products in iiRDS. And that worked.

Tobias Köffer: Independently of the changes to our own information products, we also looked into the possibilities of (semi-)automatically assigning metadata. The aim was to efficiently process content created outside the CCMS [Component Content Management System] to make it usable in a targeted manner. It very quickly became clear that this could not continue in the long term without a standard, and iiRDS was the obvious choice.

Why did you choose iiRDS?

Michael Straeter: In a large company like Siemens, later Siemens Energy, there are many competing requirements, developments, and standards side by side. When we introduced our CCMS, we decided in favor of PI-Mod and against DITA. But we soon realized that without a guiding standard, we were making our taxonomy immeasurable and increasingly incomprehensible – precisely because in the power plant sector we are dealing with huge product ranges, a vast amount of required information, and a great many "standards" in the departments involved. When we started to look at expanding our editorial portfolio – in case the era of fossil fuel power plants does come to an end – it was clear to us that it would not work without a radical pruning of the taxonomies and a standardization and minimization (keyword IEC/IEEE 82079-1). On the information side, iiRDS seemed to us to be a promising approach. On the product side, after some analysis and trial and error, we also found a more generic approach. However, both together also had an impact on our information products and editorial processes.

Tobias Köffer: We are trying to create a tool-independent and standardized way for information recipients to access and work with our content. We are also making our work a little easier for the future and laying the foundations for increasingly digital and granular use, even in our fairly traditional environment of power plant construction.

Which products do you use iiRDS for?

Michael Straeter: At the moment, mainly components of turbine generator sets. Specifically, this involves new construction and service documentation for gas turbines, condensers, generators, and steam turbines for power generation. New additions include hydrogen production plants, power transmission plants (high-voltage converters), and tool and fixture documentation for power plant construction. Not all areas are at the same level in terms of the implementation of iiRDS; some are still being converted or developed.

How do you use iiRDS in technical writing?

Michael Straeter: So far, we have mainly used the standard for classification, structuring, and building a cross-system ontology (we have two different CCMSs in use), but also in the area of automatic metadata recognition and assignment. We do not yet use it as an exchange format. But that is an important goal.

Michael Straeter has been active in technical documentation and information development for more than 20 years and has accompanied a large number of documentation and content management projects. He has been working as a Content Management Project Manager for Siemens Energy since 2014.

Tobias Köffer has been working at Siemens Energy since 2009 and has been involved in various areas of technical writing since 2014. His areas of responsibility focus on the (further) development of information products, support for existing generator systems, and digitization projects in the field of technical editing.

Have you brought in external support?

Michael Straeter: We had expertise right from the start. It wouldn't have worked without it. After the initial internal deliberations, we started a PoC [Proof of Concept]. We then investigated how we could generate iiRDS packages from heterogeneous sources, largely without converting the CCMS. Now that this also seems achievable, we are focusing on automated interfaces and process automation. CCMS manufacturers are now also supporting us, and more could be added in the areas of CDP [Content Delivery Portal] and customer clouds (which we would like to operate with iiRDS). The topics of language analysis and terminology also play a role. We have been working closely with partner companies in this area for some time.

Which other departments are involved in the implementation of iiRDS?

Michael Straeter: We plan and pursue the implementation within the Technical

Figure 2:
Large power plant gas turbines in production; construction time several months; the Technical Writing department delivers the documentation before the gas turbine arrives at its destination.
© Siemens Energy

Editing department in small agile project teams. In the direct environment of our department, we had to deal with our mostly internal customers at an early stage. After all, the standard also has an impact on our information products. Once we had made our main goals clear – standardization, a leaner and more future-proof information model, and additional automation options – we were readily supported. It certainly helped that a whole series of processes were and still are undergoing change as a result of the spin-off from Siemens AG to Siemens Energy.

Beyond our department, we are now networked with teams that deal with issues of enterprise data, data quality, data governance, and ontology at a higher organizational level. We are trying to get involved there and make the topic of "usage information" a strong one – after all, this is part of every product and every service. iiRDS is met with great interest here. In the future, we will try to incorporate our expertise in this field into higher-level data and information models. This promises to be all the more successful, the more prominent iiRDS becomes as a standard, and the better our information products and processes reflect this.

Tobias Köffer: In the short term, we are initially implementing iiRDS in the technical writing environment. In the medium and long term, we also want to use the ontologies mentioned, especially in those areas that cannot be specified by iiRDS. These include, in particular, the structure of our products and services as well as the detailed stages of the life cycle that a power plant goes through.

Do you also involve your suppliers?

Michael Straeter: Not now. At the moment, it is us and some of our network partners who are promoting the standard within the company. Given the vast number of internal and external information suppliers – we are on the move with our products worldwide – we cannot assume that iiRDS will establish itself as a standard across the board in our product area. We are also dealing with a traditionally conservative, long-term, and highly specialized sector in energy generation. This also applies to technical documentation. Thus, we are working more on making the supplied information "iiRDS-ready" and integrating it into the standard.

Tobias Köffer: The integration of information suppliers is traditionally somewhat difficult. This applies to both internal and external suppliers. We can't assume that suppliers will take on the issue of their own accord. We therefore tend to try to lower the entry barrier and make it as easy as possible for us to process the information, for example, through automated metadata recognition and assignment.

How have processes in technical writing changed with the introduction of iiRDS?

Michael Straeter: So far, hardly at all. Our current assessment is that the editorial process will not necessarily have to change after the introduction of the adapted classification structure, at least not as a result of iiRDS. As soon as the creation of iiRDS packages – whether inside or outside the CCMS – is largely automated, employees will basically only have to quality assure a different format.

Of course, we had to bring all employees on board, especially in the area of taxonomy and classification, which is change management. Some information products are already iiRDS-compliant, others have to be converted retrospectively. This is an expensive and lengthy process that is repeatedly slowed down by delivery deadlines and the resulting workarounds. But we hope that in one to two years we will have reached the point where old taxonomy structures have reached the end of their life cycle.

Tobias Köffer: The fact that the editorial process remains largely unchanged is definitely a big plus. In practice, harmonization often goes hand-in-hand with a certain amount of disruption to day-to-day business or established processes. In view of our order situation, it is not advisable to intervene too deeply in the process.

Michael Straeter: But something else has changed as a result of iiRDS, namely, the desire to standardize and minimize other areas in the same way and make them more comparable. From our point of view, DIN EN IEC/IEEE 82079-1 provides a kind of framework for subjecting information products ("information for use"), creation and provision processes, necessary qualifications, and required resources to a fundamental revision. This is because, as the standard clearly describes, ultimately these areas cannot be separated from each other but, on the contrary, should build on each other in a meaningful way.
As already mentioned, some of the content structures have changed. Since the introduction of iiRDS, we have adopted a much more generic approach and no longer try to map project and customer-specific tables of contents at all costs, for example.

What future benefits do you expect from iiRDS?

Michael Straeter: We are pleased that iiRDS practice has picked up speed in recent years. The dissemination and practical experience support the efforts to spread and standardize iiRDS internationally. We have also seen that iiRDS can now be integrated into other standards, such as Industry 4.0/AAS, ECLASS, or VDI 2770. In this way, information and products can be brought closer together. Things that are comparable to each other, such as information for use, can now be presented in an information space and applied to a wide variety of product areas. This is where we hope to see real progress in the digital exchange of information. Above all, however, we hope that the iiRDS standard will make information provision much more dynamic and target group-oriented. On the content management side, we have been able to provide more than just PDFs for some time now.

Tobias Köffer: If we look at the big issues of recent years, such as ChatGPT, these are also making waves in the traditional power plant construction environment. This has led to an increased interest in metadata and the various ways of providing information, especially outside of technical editing. iiRDS seemed to be the natural choice to meet this demand and provide solutions across the entire spectrum of knowledge graphs, Artificial Intelligence (AI), Large Language Models (LLMs), and more.

Where do you see potential for improvement in iiRDS?

Michael Straeter: Frequent criticism concerns deficits in the "request", and we can agree with that. What could a query language based on a standard look like that focuses primarily on the user's perspective? We believe that this is less a technical question and more a question of agreement. Target groups, individual product variants, and their product life cycles require a complex information structure. On the one hand, this requires quick and comprehensive orientation, but also the shortest possible path from question to answer. That remains a challenge.

iiRDS Request API
The iiRDS Request API specification is now available for comment on the iiRDS Website.
iirds.org

Tobias Köffer: By its very nature, the standard unfolds at the meta-level, and this is where most of the discussions and arguments in favor of this standard arise. This debate is reserved for a relatively small specialist audience and requires a considerable amount of familiarization with a complex topic. There is currently a lack of suitable tools to better emphasize the advantages, especially at the semantic level, outside the specialist audience. This applies not only to the request part but also to the entire life cycle of the information. The easier it becomes to create and interact with iiRDS content, the greater the overall interest will be.

ABOUT SIEMENS ENERGY

As a global energy technology company, Siemens Energy supports its customers along the entire energy value chain: low-emission or emission-free generation, transportation and storage, reduction of greenhouse gas emissions, and energy consumption in industrial processes.

ABOUT THE TECHNICAL WRITING TEAM

More than 40 employees produce technical documentation in the "Field Data Manuals" department. They are supported by around ten partner companies. Every year, 200 technical documents are produced for global projects. The Technical Writing department works on over 40 information products, which are used in around 40 product series and for services.

From PDF to a collaborative Information Digital Twin

Standardized asset information is crucial to improving operations and maintenance practices.
However, most content is available in unstructured formats only.
Information Digital Twins, created automatically from PDF documentation, offer a solution.

Text by Dr. Simone Turrin and Dr. Jan Oevermann

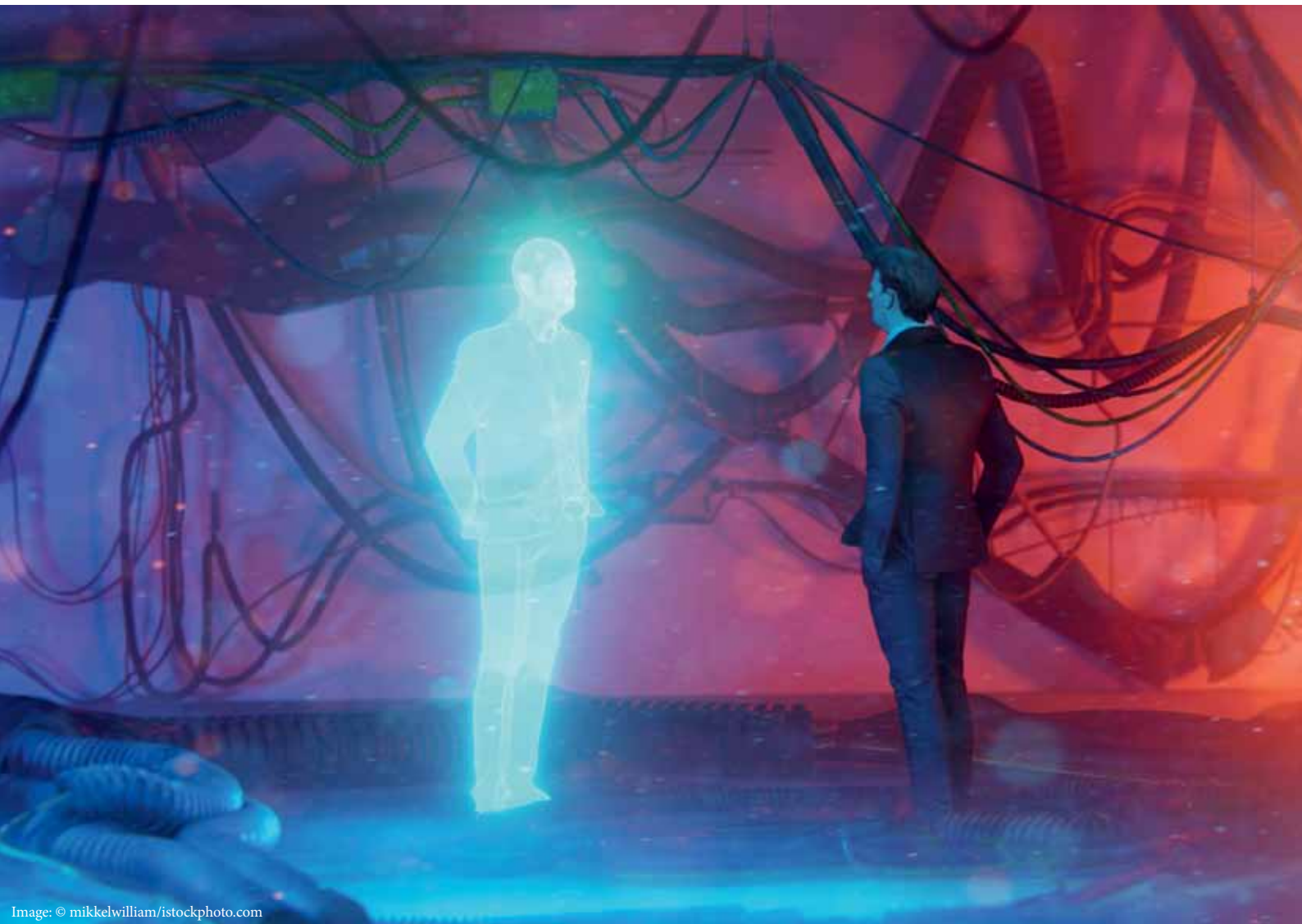


Image: © mikkelwilliam/istockphoto.com

Most component manufacturers and original equipment manufacturers (OEMs) must deal with unstructured documentation in PDF or Word format. Artificial Intelligence (AI) can enable the automated processing of unstructured documentation (out of PDF and/or Word files) in order to create Information Digital Twins that are then efficiently onboarded, enriched, and shared with the help of an Information Exchange Platform (IEP).
The digital data chain combines a standardized documentation exchange format (VDI 2770) and unique asset identification (IEC61406) with an IEP. In an IEP, Information Digital Twins are created by component manufacturers and OEMs. Information Digital Twins are then enriched and shared with asset owners/operators, dealers, engineering partners, and external service providers (as part of what is called a digital asset handover).

Digital Data Chain

The Digital Data Chain Consortium (DDCC) was founded in October 2021 to establish the complete digital data chain from OEMs to asset operators (eventually also including additional trading partners like component suppliers, dealers, engineering partners, and external service providers). The main objectives of DDCC include:

- Implementing the international standard IEC 61406 for unique asset identification across different trading partners and along the complete asset lifecycle.
- The further development of the national standard VDI Guideline 2770 into international ISO/IEC standards as the standardized documentation exchange format.
- Defining the requirements for cloud-based IEPs to provide digital manufacturer information and collaboration around Digital Twins along the complete asset lifecycle.

Currently, the DDCC consists of members from 60 leading companies worldwide (component manufacturers, OEMs, asset operators, and service providers) across different regions. For more information on DDCC, please refer to www.digitaldatachain.com.

From PDF to a collaborative Information Digital Twin with AI

This article describes the end-to-end process for the creation of Information Digital Twins out of PDF documentation. The key target segment for this process is component manufacturers and OEMs adopting the VDI 2770 standard as an exchange format for technical documentation (TD) and an IEP for sharing the Information Digital Twin with asset owners/operators or other trading partners. The process can then be easily adapted to support different documentation exchange formats, such as iiRDS or CHIFOS. A suitable workflow as shown in Figure 1 consists of the following steps:

1. Uploading existing TD in unstructured formats such as PDF or Word files
2. Performing an automated knowledge extraction from TD with Artificial Intelligence
3. Checking and validating extracted document metadata
4. Generating VDI 2770-compliant documentation
5. Creating the Information Digital Twin in an Information Exchange Platform
6. Enriching the Information Digital Twin in the Information Exchange Platform
7. Handover of the Information Digital Twin to the asset owner/operator

The main benefits for component manufacturers and OEMs adopting this process are:

- Intelligent and fast extraction of document metadata from existing technical

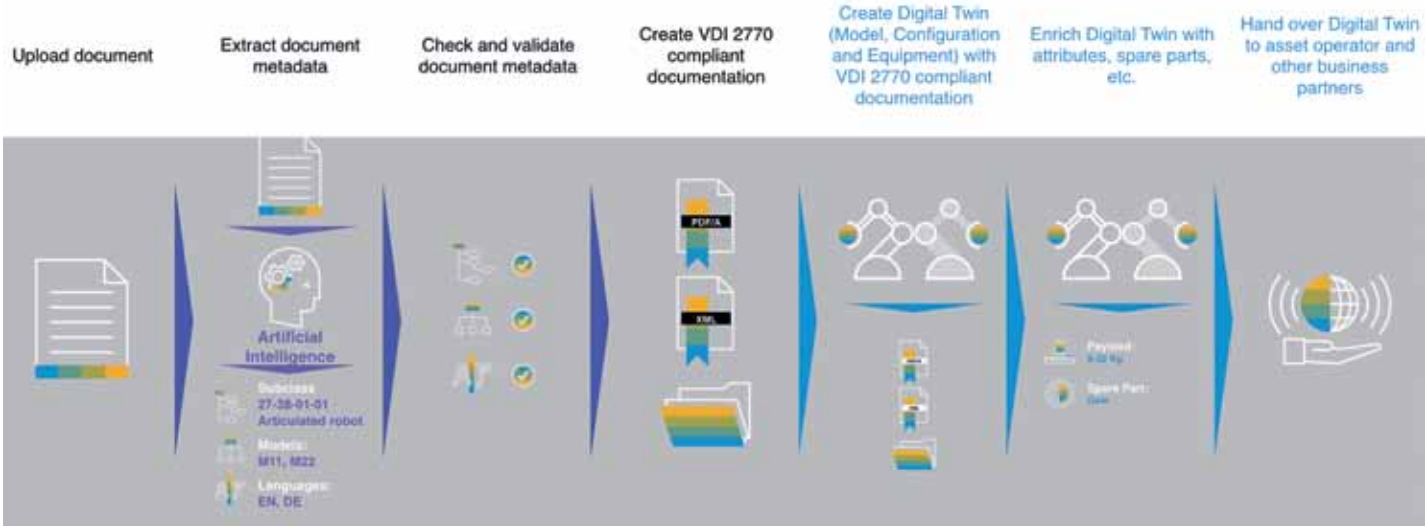


Figure 1: Process implementation: From PDF to Digital Twin with AI

documentation (such as PDF or Word files)
- Easy creation of VDI 2770-compliant documentation out of existing documents
- Linking VDI 2770-compliant documentation to a collaborative Information Digital Twin (supporting the asset structuring in terms of model, configuration, and equipment)
- Standardized handover of the Information Digital Twin with the related documentation through a central cloud-based Information Exchange Platform
- Leveraging the Information Exchange Platform to support any collaboration across different trading partners along the complete asset lifecycle

In the following sections, the process steps are described in more detail.

Figure 2: Product-related information for Digital Twins

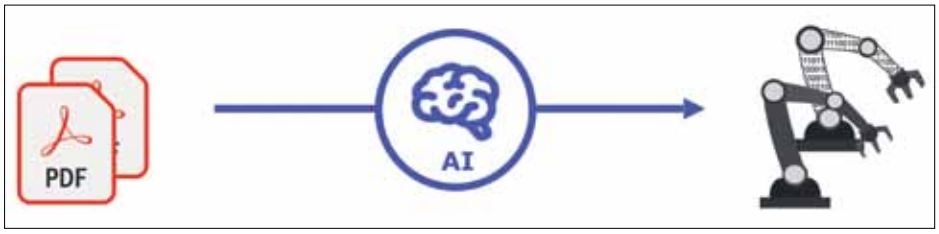


Figure 3: Schematic overview of process

1. Unstructured data as the basis for Digital Twins

Digital Twins are digital representations of assets and contain different types of information that relate to the asset in one way or another. One of these information types is technical content, which is usually human-centered as opposed to more granular data for machine-to-machine communication. Most content is available in unstructured formats, such as PDF or Word files, and contains valuable knowledge for Digital Twins. However, it is unsuitable for automated consumption by an IEP or similar software systems. AI can fill this gap and enable a semi-automated process for knowledge extraction out of unstructured technical content.

2. Automated knowledge extraction with AI

AI-based knowledge extraction is a combination of several automatable methods that can be based on rule-based systems or machine learning. The most common methods include:

- **Rule-based systems** act on a list of predefined values and predict the most relevant values for analyzed content.
  Examples: Product type, manufacturer
- **Pattern matching** looks for predefined patterns, extracts pattern matches, and normalizes their values.
  Examples: Serial number, technical data
- **Knowledge graphs** formalize product knowledge in semantic networks to make them accessible for AI processes and allow for new insights based on relations between entities.
  Examples: Product-manufacturer relation
- **Machine learning** can learn text characteristics based on examples and use this knowledge for classification tasks that require a high confidence.
  Examples: Target group, product lifecycle
- **Deep Learning** can be used for image classification or some tasks that require a deeper understanding of language.
  Examples: Image type
- **Large Language Models** can be used to classify documents based on definitions from the specification document or to generate text for administrative metadata. (Examples: summary, keywords.)

Knowledge extracted from asset-related technical content can then be used to set up the data for the Digital Twin. Examples include technical data, information about identification and location, or the relation between the instance and model of a product.

3. Standardized documentation with VDI 2770

The VDI 2770 standard specifies a minimum set of requirements for the digital handover of technical documentation. The core concepts can be boiled down to:

- Format of documents
- Structure of the documentation package
- Classification of documents via metadata

VDI 2770 requires every part of the documentation to be available as a PDF/A document with additional (optional) free file formats. The structure of the documentation within a VDI 2770 container is aligned with the physical hierarchy of the product that is described in the documentation. All documents within a VDI 2770 container must be classified as one of twelve document categories.

4. Creating Digital Twins in an IEP

After extracting knowledge from unstructured content and structuring the documentation according to the VDI 2770 specification, a first Digital Twin can be built for use in an Information Exchange Platform. In this process, information about existing instances and models stored in the IEP are considered for the creation process and synchronized with newly gained data points from the AI analysis.

5. Enriching Digital Twins

The Digital Twins in the IEP can then be directly enriched with information and data that might not be part of the product documentation. Unique asset identifiers (ideally according to IEC 61406), additional technical specifications, spare parts, trading partners, asset location, etc., are typical information that component manufacturers and OEMs can add to the digital twin created out of unstructured data as described in the previous steps.

6. Handover to asset owner/operator

Once the Information Digital Twin has been created by component manufacturers or OEMs in the IEP, this can then be published and shared with asset owners/operators and other business partners (i.e., dealers, engineering partners, and service providers) as part of the digital asset handover. The digital asset handover can be proactively embedded in the standard sales processes (i.e., linked to sales orders) or directly triggered by asset owners/operators in the IEP on demand as a digital asset request (this is the typical trigger for brownfield scenarios).

After the digital asset handover, component manufacturers, OEMs, asset owners/operators, etc., are connected and can collaborate around the shared Digital Twin along the complete asset lifecycle from delivery to disposal. During the asset lifecycle, for example, additional or up-to-date documentation (i.e., new firmware) will be available, and the Digital Twin might be enriched with data and information from operations and maintenance practices. Information about the usage and location of the asset, live data (e.g., from sensors or simulations), maintenance and service history, changes in asset components out of maintenance or upgrades, etc., may be added for further insights and use cases. This information can be shared with business partners involved in asset management to support business models and best practices for the reduction of the asset's total cost of ownership (TCO) and the increase in asset reliability and operational safety.

Conclusions

Asset owners/operators today expect to receive operations- and maintenance-relevant information directly from component manufacturers and OEMs in a digital and standardized format as part of the procurement of new assets and, on demand, for assets currently operated and maintained. Standardized asset information is of critical importance to improve operations and maintenance practices with the scope of increasing asset efficiency and availability while reducing operation costs and risks. The IEC 61406 and the VDI 2770 standards are considered the main standards for both, the unique asset identification across the complete asset lifecycle and the standardized manufacturer information for more and more asset owners/operators across different regions and in multiple industry sectors. Consequently, component manufacturers and OEMs are currently experiencing some market pressure in providing the required information in a digital format and according to the required standards.

In this article, we have shown a very convenient way for component manufacturers to create Digital Twins out of existing unstructured documentation by leveraging Artificial Intelligence to extract asset metadata from relevant documents. Adopting an Information Exchange Platform as a central cloud-based platform to create, publish, and share the Digital Twins with multiple trading partners brings further benefits. In this way, the digital asset handover with standardized technical documentation is supported on one single platform for multiple trading partners (i.e., OEMs and asset operators). Finally, with the use of an IEP, all trading partners involved in asset operations and maintenance management

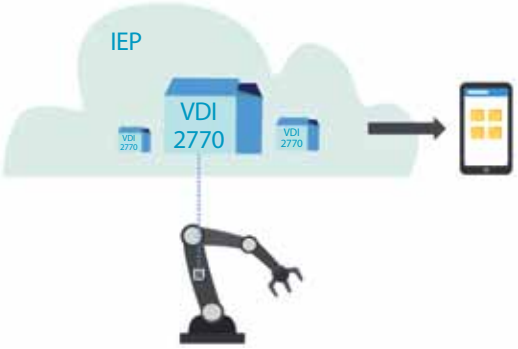


Figure 4: Digital Data Chain with Auto ID, VDI 2770 & IEP

can collaborate throughout the asset lifecycle around a Digital Twin that is continuously up to date with current documentation, maintenance and service history, sensor data, etc. This will help component manufacturers and OEMs to enhance their customer support and satisfaction as well as to improve their service business. Asset owners/operators will benefit from getting digital asset information in a standardized format from multiple suppliers in a single cloud-based, vendor-neutral platform.

ABOUT THE AUTHORS

Simone Turrin is Chief Product Owner for the SAP Business Network Asset Collaboration at SAP SE. As part of the product management team, he is responsible for the definition of strategy, implementation of roadmaps and development of the SAP Business Network Asset Collaboration.

@ simone.turrin@sap.com
www.sap.com

Jan Oevermann studied Technical Writing, Communication and Media Management at the Karlsruhe University of Applied Sciences and holds a doctorate in Computer Science from the University of Bremen. He is CEO and founder of plusmeta GmbH, a software company focused on AI-based information automation. He specializes in the application of AI and semantic technologies to the field of technical documentation.

@ jan@plusmeta.de
plusmeta.de
@dr-jan-oevermann-11845417a

iiRDS for a better support experience

A dynamic intelligent content graph based on iiRDS has the potential to deliver targeted support for service technicians. A use case.

Text by Helmut Nagy and Harald Stadlbauer

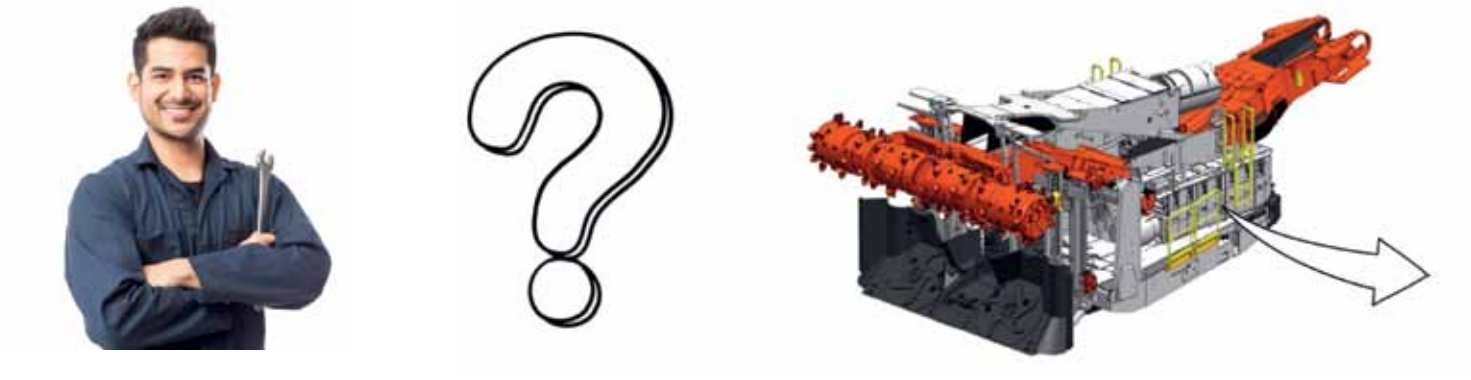


Figure 1: Use case of a service technician

In this article, we present a proof of concept based on a use case from a multinational corporation. The aim was to enable intelligent information retrieval for detailed questions and problems. This information is based on intelligent content created from structured content enriched by a knowledge graph. The proof of concept looks quite promising and has already led to further developments.

A use case

The Machinery Directive 2006/42/EC as the basis for technical documentation was originally aimed at technical documentation in a handbook structure. Even when written in DITA, documentation that is available only as a handbook or PDF makes search and retrieval difficult for service technicians. As an example, let us take a look at Karl, a support engineer. The hydraulic system of a bolter miner is not working properly, so Karl is called on site. He is new to the product. The manual has about 1200 pages.

Karl now needs to disconnect the level sensor, find the root cause using the troubleshooting information, and fix the problem, all the while staying alert for safety issues. To do this, he needs to quickly find:

Information about the problem

- Where to find the level indicator, page 260
- Where to find the level display on the machine, page 1084
- Where to find the sensor to be checked, page 261

What to do

- Short description of how to fill the machine's fluid, page 536
- What kind of fluid to use, page 759

What to watch out for

- Hazard statements for working on the hydraulic systems, page 186

In an ideal scenario, Karl is presented with a tailored information package including the right answer to his problem (ideally only one answer) to enable him to solve the issue fast and directly in front of the machine. Our proof of concept addressed the hydraulics part, a system with a high risk of injury due to its high-pressurized hydraulic oil. In this scenario, it is vital to point out dangerous situations to the service technician and raise awareness through warnings.

Step 1: From structured content to a first content graph

So, how can we achieve our ideal scenario? The basis is structured content. This is not always a given. In our case, Tridion Docs had been used to create the documentation, providing us with structured content based on DITA. (We consider DITA to be the ideal standard, but our approach is not tied to any specific standard). We also integrated e-learning data (slides and links to videos), which were unstructured.

Based on the DITA ontology, the provided DITA data for our proof of concept was transformed into a graph, see Figure 2. The granularity of the graph and what is represented in it can be tailored to the use case. This means not all data from the CCMS has to be included in the content graph, and the content graph structure can be adapted to the use case. We followed an iterative approach, starting small and growing the graph based on our needs and use cases. From our experience, this approach allows for early results, which can then be built upon. Graphs are suitable for agile use case development, as they are easy to extend.

After this, we transformed the XML schema into a graph structure that opens up new ways to interact with content.

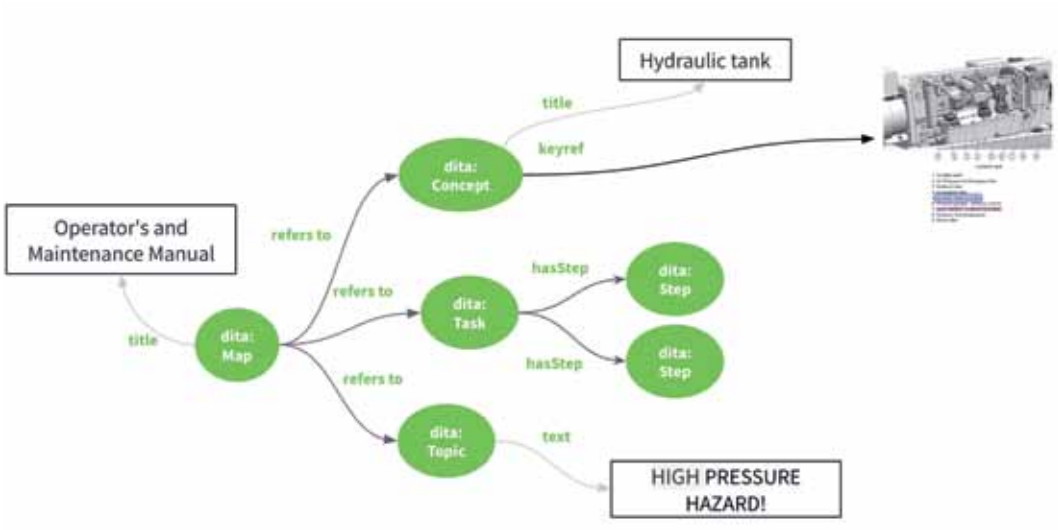


Figure 2: DITA topics as RDF graph

Step 2: From terminology to smart tagging

Step 2 was to enrich the content graph based on a developed taxonomy. In addition to the existing terminology database of the company, we used industry taxonomies from WAND (a provider of industry vertical taxonomies) and the taxonomy included in the iiRDS standard. These taxonomies were used to automatically tag the content and thus provide high-quality metadata to each DITA element. Semantic tagging goes beyond the traditional metadata approach. Just like the ontologies used, the taxonomies are part of the intelligent content graph we created. Providing a semantic footprint for each content element allows recommending similar content based on it. Ideally, this tagging is already done when content is created in the CCMS. If this is not the case, this step can be added downstream when content is delivered.

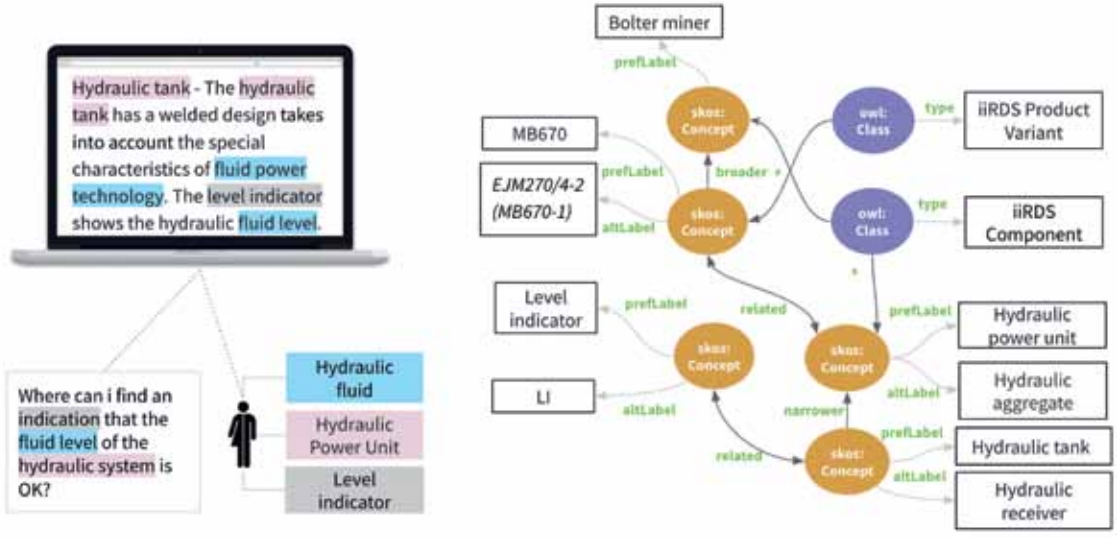


Figure 3: Enrich the content graph via tagging

We now have the enriched content graph as a basis for what we wanted to achieve for our use case (see Figure 3).

Step 3: iiRDS for targeted intelligent search

The Intelligent Information Request and Delivery Standard (iiRDS) is a great match for DITA. Topic types for DITA and iiRDS

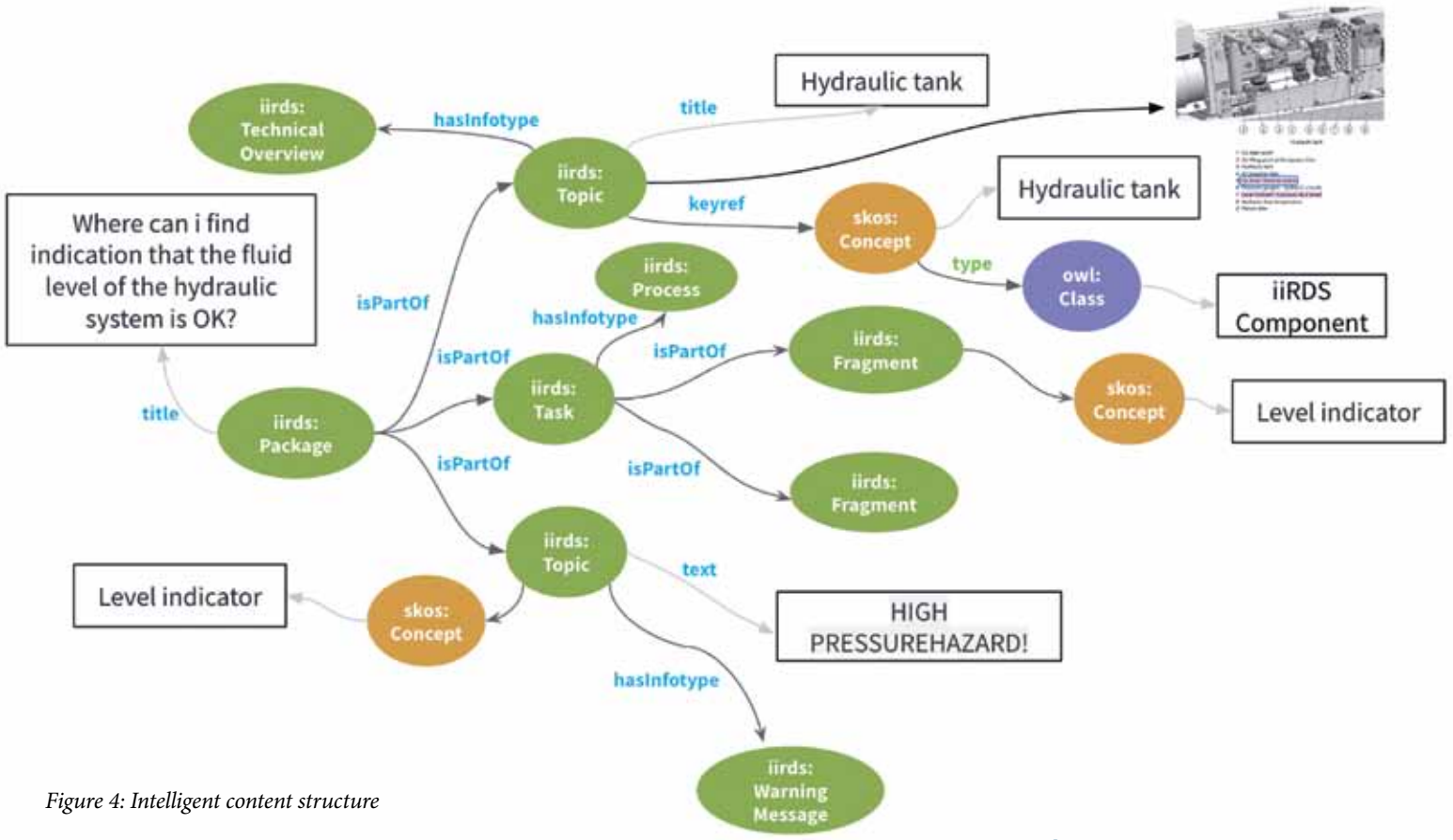


Figure 4: Intelligent content structure

are identical, making it easy to extend our enriched content graph using iiRDS ontology. This is the last step in our journey to intelligent content.

Using iiRDS, we can connect to product data and bring context into the content structure, allowing us to identify exactly the information needed for our use case. Troubleshooting information is created automatically: For each troubleshooting topic, the appropriate technical input, the respective process information, and the related warning messages are dynamically generated. This enables customized results, switching from a document-centric approach to a user-centric search process, providing support engineers with the right information to get the job done (see Figure 4).

Conclusion

Our proof of concept shows how the combination of an RDF view of DITA metadata, a company-specific taxonomy, and iiRDS can create a dynamic intelligent content graph to support the focused retrieval of information for service technicians. Figure 5 shows the proof of concept setup and high-level process with all systems involved.

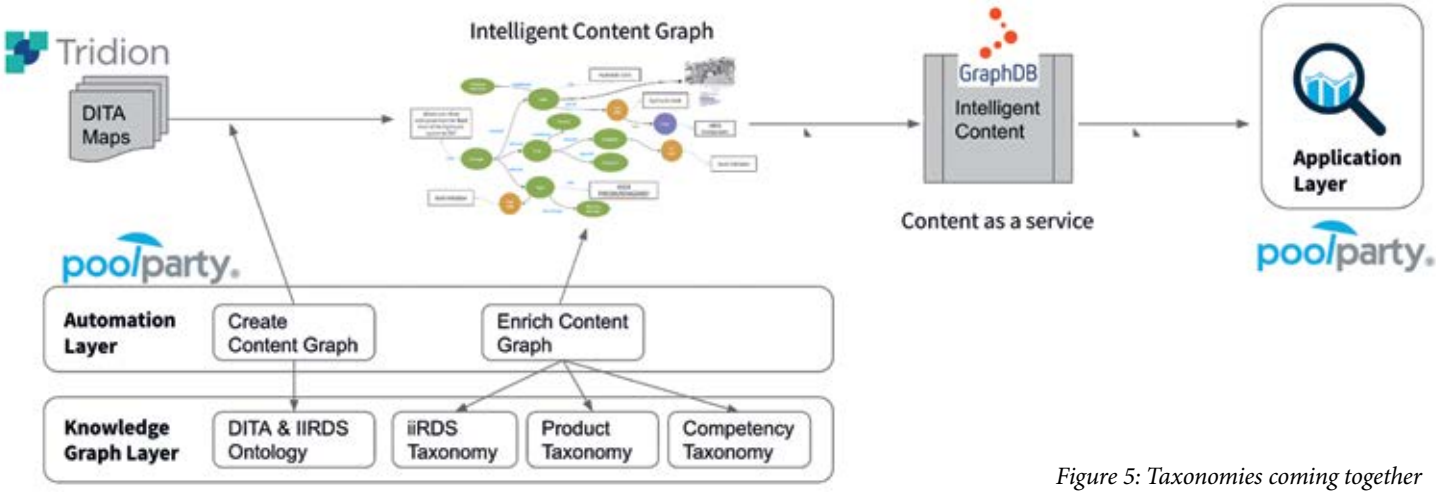


Figure 5: Taxonomies coming together

Delivering smart content is what we do best

Struggling to pronounce iiRDS?

We can't help you with that, but we can help you to understand the standard and use it to create and deliver smart content for your products.

At parson, we are iiRDS experts and also specialize in implementing CCMS and delivery solutions.

www.parson-europe.com

parson
we create knowledge

tekom iiRDS
CONSORTIUM MEMBER

iiRDS Online Training

Get started

with the Standard for Intelligent Information

Join our comprehensive two-month online training to gain a solid foundation in iiRDS. Through live web sessions, self-study, group work, quizzes, and hands-on assignments, you'll explore the basics of iiRDS and how to apply it effectively.

Key Topics and Learning Objectives:

- **Introduction to smart content and iiRDS**
  What is iiRDS? Where can I use iiRDS?
- **The iiRDS metadata model**
  Which types of metadata are in iiRDS? What do they mean? How do I define metadata for specific data?
- **The iiRDS package format**
  How is iiRDS transferred? What is an iiRDS package? What are iiRDS variants?
- **iiRDS and other standards**
  How does iiRDS relate to classification systems? How is iiRDS connectable to other standards?
- **Toolchains and systems for iiRDS**
  What tools can I use to generate and validate iiRDS packages? How do I generate iiRDS metadata from DITA? How do I connect search applications via the iiRDS Request API?
- **Practical applications of iiRDS in companies**
  How do I use iiRDS in real-life scenarios? How do I map external ontologies to iiRDS?

Ready to get started with iiRDS?
Visit us at **www.iirds.org**

tekom iiRDS