

# Digitization Guidelines Survey

Response ID:185 Data

## 1. Identification page

1. Please enter the ST.3 code of the member state or inter-governmental organization you represent.

DE - Germany

2. Please enter the name of the Office or Organization you represent. For instance, International Bureau of WIPO.

If your organization name is the same as your ST.3 code name, you may put "n/a".

DPMA

## 2. General questions

3. Is your national patent collection part of the PCT minimum documentation or does your Office contribute your IP documentation to one of WIPO's databases? Please select all that apply.

Patentscope

Global Brands Database

Global Design Database

PCT minimum documentation

4. What file or document format(s) does your Office accept at filing for IP applications? Please select all that apply.

### For industrial design collections:

XML

PDF

Image-based

Paper

### For patent collections:

XML

PDF

Image-based

Paper

### For trademark collections:

XML

PDF

Image-based

Media-based formats (e.g., MP3)

3D formats

5. In what language(s) do you publish your IP documents in? Please provide a full list of 2-letter language codes.

DE

6. In what format(s) do you publish or disseminate your IP collection? Please select all that apply.

### For industrial design collections:

WIPO Standard ST.86

Image-based

PDF

### For patent collections:

WIPO Standard ST.36

PDF

Plain text

JSON (e.g., WIPO Standard ST.97)

**For trademark collections:**

Image-based  
WIPO Standard ST.66  
Media formats (e.g., MP3)  
3D formats  
PDF

7. In what format(s) does your Office publish the bibliographic data associated with the format(s) indicated in question 8? Please select all that apply.

**For industrial design collections:**

XML  
PDF  
Plain text

**For patent collections:**

XML  
PDF  
Plain text

**For trademark collections:**

XML  
PDF  
JSON  
Plain text

8. If your Office produces XML or JSON instances and does not validate these instances before publication, please provide your reasoning below.

does not apply

9. Does your Office publish the color drawings they receive, if they are permitted during the filing of a patent application?

No

### 3. Digitization activities

10. Does your Office perform digitization of documents you receive before publication or data dissemination, for one or more IP rights, or has experience of digitization of documents associated with IP rights?

Yes

10.1. How are or were these documents or associated data transformed from the format you received? e.g., paper document manually scanned and OCR'd or DOCX converted to XML.

- Paper documents: Physical documents are scanned at 300 DPI and converted to PDF/A-2U using the Foxit Rendition Server, with Optical Character Recognition (OCR) applied as an integral part of the process. Patent documents are predominantly scanned in black and white, reflecting the technical nature of the content. Greyscale or colour scanning is applied where document characteristics require it. Selected form fields are subsequently extracted and reviewed by qualified staff.
- FAX: Documents received via fax are processed as TIFF files and converted to PDF/A-2U, following the same workflow as paper-based scanned documents.
- Electronic submission channels: The DPMA operates a proprietary submission platform through which applicants are required to submit documents in PDF format. These are converted to PDF/A-2U using the Foxit Rendition Server. In cases where discrepancies are identified between the converted and original files, the original submission is retained for reference. Metadata accompanying electronic submissions is extracted and processed.

10.2. If you rely or relied on a third party to perform the digitization of your documents, why did you reach this decision?

The DPMA utilises the Foxit Rendition Server as its primary solution for document conversion and digitization. This product was selected on the basis of its technical performance, in particular its highly effective compression capabilities for colour documents, as well as the reliability and quality of the output it produces. The vendor's technical support has consistently met the Office's requirements.

10.3. Does your Office use or did you use any IT tools, in particular AI tools, in the digitization process that you find particularly useful?

The DPMA has not yet deployed AI-based tools within its productive digitization workflows. However, the Office is currently conducting evaluations of a locally hosted Large Language Model (LLM) with a view to enhancing automated document classification and structured data extraction from forms. Initial results are considered encouraging, and further assessment is ongoing.

10.4. What are or were the biggest technical challenges you have experienced in digitizing your documents? e.g., extracting drawings or chemical formulae from a patent application. Please provide further details in the comments box below, including how you overcame these challenges and the relevant IP right type.

The DPMA has identified the following as the principal technical challenges encountered in the digitization process:

- Quality requirements for patent documents: Patent documentation is subject to strict requirements regarding image fidelity, particularly with respect to black-and-white technical drawings, where minimal quality loss is essential. Meeting these standards consistently has presented challenges, particularly in the context of transitions between scanner hardware generations.

- Non-conformant PDF submissions: A recurring technical issue involves PDF files received from external parties that do not render consistently across different viewer applications. Addressing such non-conformant files requires additional processing and quality assurance measures.
- Legal requirements for the disposal of physical documents: Compliance with national legislation governing the conditions under which original paper documents may be discarded following digitization represents a procedural and organisational challenge, rather than a purely technical one.
- Chemical formulae and OCR limitations: Chemical formulae are not currently subject to structured extraction. While OCR processing is applied, the accuracy of the resulting output for such content cannot be fully guaranteed. The potential impact of this limitation on AI-based document search and retrieval has not yet been comprehensively assessed and may warrant further evaluation in the future.

#### 4. Digitization activities

**11. If your Office is considering future digitization activities, please provide details regarding the tools you are considering, in particular AI tools, and include your reasons for selecting them for potential use in digitization.**

The German Patent and Trade Mark Office (DPMA) is currently considering a number of possible future digitization measures, some of which may involve the use of artificial intelligence (AI) tools, subject to further evaluation.

Among the measures under consideration is the development of a new customer portal based on modern web technologies, which could replace the existing DPMAdirektPro and DPMAdirektWeb systems. Such a portal would be intended to enable applicants to file legally valid applications for industrial property rights online and to submit supplementary electronic documents relating to such applications. At this stage, no AI components are foreseen in this context; however, given its potential relevance to the Office's digital service offering, it is mentioned here for completeness.

With regard to AI, the DPMA is exploring the possible future integration of AI tools that could provide support for search activities and document drafting. Such tools, if implemented, would be aimed at increasing efficiency and improving data quality. As any related procurement and contractual arrangements remain at a preliminary stage, no specific provider or product can be named at this time. In this connection, the Office has recently acquired licenses for OpenShift AI, which is intended, in principle, to serve as a platform for evaluating possible AI applications; no operational use has commenced to date.

The DPMA also makes use, under certain conditions, of the publicly available translation tool DeepL, including its writing assistant function, for official purposes. In addition, the DPMA is following developments relating to large language models (LLMs) with interest. Given the variety of available LLM services, each typically requiring its own application programming interface (API), consideration is being given to the possible definition of a unified interface that could connect to multiple LLM providers, with a view to avoiding separate integrations for each service. In this context, the tool LiteLLM is among the options currently being looked at, alongside other comparable products; no decision in this regard has been made.

Beyond the Office's own considerations, it may also be noted that the German federal IT service center (ITZ-Bund) has introduced a central AI platform, known as KIPITZ, intended to enable federal authorities to make secure and sovereign use of generative AI in their daily work. This platform is envisaged to support tasks such as document summarization, document-based dialogue ("chat with documents"), translation, and knowledge search. According to current planning, this platform could potentially become available to the DPMA around October 2026.

**12. Please provide any other feedback regarding capacity building for the digitization of IP document collections.**