

Digitization Guidelines Survey

Response ID:128 Data

1. Identification page

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

IN - India

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".

Intellectual Property Office , India

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:

PDF
Image-based
Paper

For patent collections:

PDF
Paper
Other (please specify): ST.26 format Xml for Sequence Listing only for PCT applications; Sequence listing is also accepted in .txt format

For trademark collections:

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

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

EN

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

For industrial design collections:

Image-based
PDF
Other (please specify): IPO submits its Industrial Designs data to WIPO Publish in ST.86 format

For patent collections:

PDF
Plain text
Other (please specify): IPO submits its Patents data to WIPO Publish in ST.36 format

For trademark collections:

Image-based
PDF
Other (please specify): IPO submits its Trademarks data to WIPO Publish in ST.66 format

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:

PDF

For patent collections:

PDF

Plain text

For trademark collections:

PDF

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

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

Yes

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 OCRred or DOCX converted to XML.

Paper documents are manually scanned and OCRred; The applications which are filed online, they are received in PDFs and these PDFs are OCRred later.

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 backlog collection of documents was too big to be managed by IPO, hence third party was engaged for the digitization.

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?

Yes, documents collected from 1.1.2026 is OCRred using paddle OCR tool for text extraction and for backlog digitization and scanning of document ABBY Fine software tool is used.

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 patent documents being techno-legal in nature contain a lot of mathematical and chemical formula along with tabulated data and drawings. Hence, IPO India uses paddleOCR v3 structure (an open source OCR library) for identification of drawings & chemical formulas.

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.

Yes, exploring the possibilities of using enterprise level proprietary OCR software for digitization of the specification and its corresponding Database insertion.

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

Capacity building in developing ability for data capture conforming to WIPO standards in IPO's online filing systems, by using XML filing or docx filing implementation.