

1. Identification page

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

SG - Singapore

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 of Singapore

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
Other (please specify): doc, docx

For patent collections:

XML
PDF
DOCX
Paper
Other (please specify): txt

For trademark collections:

PDF
Image-based
Media-based formats (e.g., MP3)
Other (please specify): doc, docx

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:

PDF
Other (please specify): In HTML format in our online IP filing system and in XML to Global Design Database

For patent collections:

WIPO Standard ST.36
PDF
Plain text
Other (please specify): In HTML format in our online IP filing system

For trademark collections:

PDF

Other (please specify): In HTML format in our online IP filing system and in XML to Global Brands Database

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

Other (please specify): In HTML format in our online IP filing system

For patent collections:

PDF

Plain text

Other (please specify): In HTML format in our online IP filing system

For trademark collections:

PDF

Other (please specify): In HTML format in our online IP filing system

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

NA

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

Not applicable

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.

Documents are submitted to us in either PDF or DOCX format.

Patent specifications submitted in PDF format are OCRred.

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

No, we have not relied on a third party for 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, we utilise AI-enabled OCR capabilities (e.g. AWS Textract) to support document digitisation, including extraction of text, identification and removal of non-substantive elements (e.g. headers, footers, pagination, line numbers), and detection of figurative elements for separate image extraction.

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.

During the digitization of patent specifications, we have encountered several recurring challenges:

- the system is unable to properly separate paragraphs or individual claims.
- data extracted from tables is not rendered correctly and instead appears as continuous text.
- formulae embedded in documents are not extracted accurately in text format.
- text appearing in the margins of documents (eg PCT publication number, page number) is sometimes not removed.

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.

Building on current OCR-based digitisation and layout structuring capabilities, section identification (e.g. title, abstract, claims) is presently rule-based (e.g. wildcard matching). We are exploring GenAI to enhance document structuring, as rule-based approaches are less reliable for complex document. This results in more accurate section and claim identification, which enables structured outputs to be integrated into downstream workflows (e.g. claim-level analysis).

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

NIL