

Digitization Guidelines Survey

Response ID:138 Data

1. Identification page

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

GB - United Kingdom

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".

UK IPO

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
Image-based
DOCX
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.

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

For patent collections:

WIPO Standard ST.36
PDF

For trademark collections:

Image-based
Other (please specify): XML - MECA 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:

Plain text

For patent collections:

XML

PDF

For trademark collections:

PDF

Plain text

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

Not applicable

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 OCR'd or DOCX converted to XML.

Paper documents manually scanned and OCR'd to create XML

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

Lack of capability at the time and more effective to use a third party

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?

-

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.

-

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.

We are currently investigating the use of docling (with tesseract as the OCR engine) to digitise our patent documents. This is a lightweight and free solution but requires pre-and post-processing to get the most from it. Depending on how the trial goes we will either move to full production or look into other solutions (such as LLMs or Miner)

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

-