

# Digitization Guidelines Survey

Response ID:129 Data

## 1. Identification page

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

HU - Hungary

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".

Hungarian Intellectual Property Office

## 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

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): In order to comply with the relevant WIPO standard, filing sequence listings in xml format is mandatory and hence, xml format is accepted in this case.

**For trademark collections:**

PDF

Image-based

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

Other (please specify): Paper

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

HU

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:**

PDF

**For trademark collections:**

Image-based

Media formats (e.g., MP3)

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:**

PDF

For patent collections:

PDF

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.

N/A

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 scanned and 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?

N/A

- 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?

ABBYY Fine reader is used for OCRing PCT documentation

- 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.

Data quality over OCRing process from poor quality documentation. Also extracting drawings and chemical formulas. Poor OCRRed data lead to incorrect xml format that does not meet the prescribed requirements. To overcome these issues, we carry out manual corrections.

### 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.

AI OCRing tool developed by EPO is on our list for future improvement once available and fulfils expectations.

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