

Digitization Guidelines Survey

Response ID:225 Data

1. Identification page

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

BR - Brazil

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".

INPI Brazil

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

Global Brands Database

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

For patent collections:

PDF

Other (please specify): XML ST.26 for sequence listings

For trademark collections:

PDF

Image-based

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

PT

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): Open data CSV files

For patent collections:

WIPO Standard ST.36

PDF

Other (please specify): XML ST.26 for sequence listings

For trademark collections:

WIPO Standard ST.66

PDF

Other (please specify): XML (proprietary format); Open data CSV files

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
Plain text

For patent collections:

XML
PDF

For trademark collections:

XML
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.

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.

For patents: Scanned, OCRred

For Trademarks and Industrial Designs:

Yes, for legacy documents. The digitization effort concerns exclusively the Office's historical paper archive of trademarks and industrial design files. These files are scanned, OCRred and converted into searchable PDF files, which are then made available to the public through the Office's web search system (BuscaWeb), where they can be consulted alongside the corresponding application data.

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

For patents: not enough internal resources. For Trademarks and Industrial Designs: We outsourced the digitization of IP documents due to lack of human resources in the Office.

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?

No

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.

For patents: The challenges encountered in the digitization of patent documents included:

- digitizing certified paper copies of priority documents;
- extracting information from chemical formulas and tables;
- digitizing figures, including photographs and other graphical elements;
- managing the large volume of documents to be digitized by the outsourced company; and
- carrying out quality verification of the digitized documents.

For patents, digitization was considered to consist of the generation of OCRred PDF files. The management of the large volume of documents outsourced for digitization, together with quality verification, represented the biggest operational challenges. To address these challenges, we had to establish a dedicated team to oversee the digitization process and perform the necessary quality checks.

For trademarks and industrial designs: Digitization considered only generation of OCRred PDF files. Managing the huge amount of documents to be digitized by the outsourced company and dealing with quality verification were the biggest challenges. We had to set up a dedicated team to be able to tackle these challenges at the time.

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.

For Patents: Yes. We need to digitize around 70,000 publication documents to comply with PCT MinDoc requirements. We will outsource the process. New company uses modern techniques and provides an integrated quality-verification interface.

For trademarks and industrial designs: Yes. The Office is currently carrying out a project to engage a new outsourced digitization company, and the greatest challenge

we need to face is the great amount of industrial design and trademark documents to be handled, which exceeds one million documents. Beyond the physical transport of these files, the quality control of the digitized output and the return of the collection to document custody are the most demanding aspects of the entire activity. The Office has previous experience with digitization, which places us in a more favourable position regarding the precautions that need to be taken. The procurement was conducted with great care, and we are currently in the preliminary testing phase; only once every aspect is properly settled we will move on to large-scale digitization.

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