

Digitization Guidelines Survey

Response ID:202 Data

1. Identification page

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

CA - Canada

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".

Canadian 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
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:

XML
Image-based
Paper

For patent collections:

XML
Image-based
Paper
Other (please specify): For national CIPO only accepts PDF and TIFF (not XML). For international, CIPO accepts XML and PDF via ePCT. For reference:
<https://ised-isde.canada.ca/site/canadian-intellectual-property-office/en/patents/mycipo-patents-user-guide>

For trademark collections:

XML
PDF
Image-based
Media-based formats (e.g., MP3)
3D formats
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.

CIPO publishes correspondence as submitted by clients. CIPO correspondences in French or English and publish in the language sent or received based on clients preferred language.

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

For industrial design collections:

WIPO Standard ST.86
Image-based

For patent collections:

WIPO Standard ST.96
PDF

For trademark collections:

Image-based
WIPO Standard ST.96
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:

XML

For patent collections:

XML

For trademark collections:

XML

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

Trademark: XML is validated upon publication

Patent: Paper is scanned as images, plus some subject matter is OCR'd so that it can be published and searched.

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 OCR'd or DOCX converted to XML.

Patent: Paper is scanned as images, plus some subject matter is OCR'd so that it can be published and searched. OCR using Datacap.

Trademark Industrial Designs and Copyrights: Paper is manually scanned.

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

Previously the Trademarks Office had a Service Level Agreement with another government department to digitize historical documents. The agreement ended, however, the digitization is not fully complete, so now we continue digitizing in-house.

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?

Patent digitization is currently a manual process within Patent Branch, using various free software like IBM Data Camp.

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 PCT minimum documentation the current manual process is very demanding. However CIPO-Patents does have license to use a robust technology though implementing this will require project planning.

For Trademark, there evidence on historical files that are considered "unscannables". The Office photographed them.

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.

CIPO is evaluating a solutions to automate the text extract of Abstract, Claims and Disclosure documents. One possible solution involves using a Cloud bases solution that can leverage AI modules (subject to security approval and costing) and other AWS services. AWS Textract may be used for simple text only documents and AI models to detect mathematical and chemical formulas. These will be extract as images and an external link to view the formula as an image. One of the biggest challenges is the security posture of our cloud environment does not currently allow use of AI models that are available in the AWS Market place.

CIPO-Patents has license for IBM Datacap which comes with form recognitions, OCR and also AI capabilities and may begin using this in the future.

No AI tools are being considered at this time for CIPO Trademark

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

N/A