

Committee on WIPO Standards (CWS)

Fourteenth Session
Geneva, November 23 to 27, 2026

ANALYSIS OF SURVEY RESULTS ON DIGITIZATION OF IP DOCUMENTATION

Document prepared by the Secretariat

SUMMARY

1. The International Bureau analyzed the responses to the survey regarding the challenges experienced by Offices digitizing their Industrial Property (IP) documents. Survey analysis reveals that the major challenges for Offices exist in extracting complex content from patent documents during this digitization process.

BACKGROUND

2. At the thirteenth session of the Committee on WIPO Standards (CWS), the International Bureau presented a document regarding the digitization challenges that it had noted in particular during the period where International Searching Authorities were attempting to meet the Patent Cooperation Treaty (PCT) minimum documentation requirements (see document [CWS/13/28](#)).

3. The CWS approved that a survey regarding the challenges experienced by Offices in digitizing their IP document collections be sent to both CWS Members and PCT Contracting Parties via joint CWS and PCT Circular (see paragraph 212 of document [CWS/13/33](#)).

SURVEY RESULTS

4. In May 2026, the Secretariat prepared and distributed the survey by joint CWS and PCT circular to both CWS Members and PCT receiving Offices, International Searching Authorities and International Preliminary Examination Authorities, and/or designated or elected Office (see [C.CWS 206/C.PCT 1701](#)). The survey invited responses from circular recipients during the period from May 15 to June 15, 2026.

5. Fifty-six Offices or Organizations responded to the survey, comprising the following Member States: Algeria (DZ), Australia (AU), Azerbaijan (AZ), Bahrain (BH), Belgium (BE), Bosnia and Herzegovina (BA), Brazil (BR), Bulgaria (BG), Canada (CA), Chile (CL), China (CN), Costa Rica (CR), Croatia (HR), Czechia (CZ), Denmark (DK), Equatorial Guinea (GQ), Estonia (EE), France (FR), Georgia (GE), Germany (DE), Hungary (HU), India (IN), Ireland (IE), Israel (IL), Jamaica (JM), Japan (JP), Latvia (LV), Lithuania (LT), Mexico (MX), Mozambique (MZ), New Zealand (NZ), Norway (NO), Oman (OM), Paraguay (PY), Peru (PE), Philippines (PH), Poland (PL), Republic of Korea (KR), Republic of Moldova (MD), Romania (RO), Russian Federation (RU), Saudi Arabia (SA), Serbia (RS), Singapore (SG), Slovakia (SK), Spain (ES), Sri Lanka (LK), Sweden (SE), Türkiye (TR), United Kingdom (GB), United Republic of Tanzania (TZ) and United States of America (US); and the following Organizations: African Regional Intellectual Property Office (AP), Eurasian Patent Organization (EA), European Patent Office (EP) and European Union Intellectual Property Office (EM).

6. Individual responses and collated results to the survey questionnaire have been published in Part 7 of the WIPO Handbook. A workshop was held in July 2026, to share the results and preliminary analysis undertaken by the Secretariat. Further details on the outcomes of the workshop can be found in document CWS/14/8.

SURVEY ANALYSIS

7. In undertaking this survey, the International Bureau hoped to understand more clearly the challenges experienced by Offices in digitizing their IP document collections in order to provide recommendations for how to address these challenges, and in doing so, support Offices in their digitization activities.

8. While this survey addressed challenges in digitizing IP documentation, the results show that these challenges were primarily regarding patent documents, with more than three quarters of survey respondents indicating an issue. This is in comparison with 7 per cent of Offices for trademarks and 4 per cent for industrial designs.

9. The first question posed to Offices in the survey was to determine whether contributing to global databases hosted by the International Bureau could be a motivating factor for these digitization activities. However, there was no correlation between the Offices who conducted digitization activities and those contributing to databases hosted by the International Bureau. The results are shown below in Figure 1.

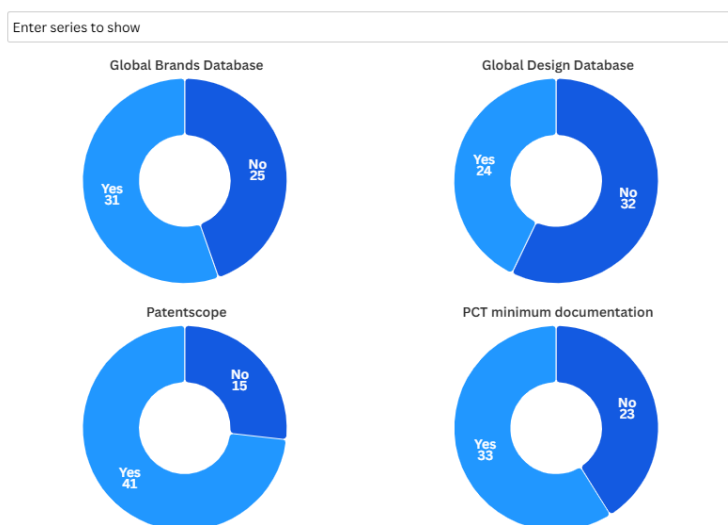


Figure 1: Proportion of survey recipients who contribute to WIPO databases.

10. The following questions related to the formats accepted at filing by Offices and the formats which Offices publish in. If these differ, then the Offices are undertaking a transformation process and, in certain instances, digitization. The top three formats accepted at filing for patents were PDF (84 per cent), paper (82 per cent), and DOCX (47 per cent). The top three formats for trademarks and designs were PDF (70-80 per cent), image-based (60-70 per cent) and paper for industrial designs (approximately 70 per cent). One hypothesis why there are a high number of Offices who accept paper filings could be that these are Patent Law Treaty (PLT) Contracting Parties. However, there are 45 Contracting Parties to this Treaty, and they do not correspond to those survey recipients who indicated they accept paper patent applications.

11. Regarding the formats that Offices publish in, for patents, this was overwhelmingly in PDF (over 90 per cent) with a large proportion also publishing compliant with either WIPO Standards ST.36 or ST.96 (over 70 per cent). For trademarks and industrial designs, PDF (over 80 per cent) and image-based (over 40 per cent) were the most popular publication formats, with over 50 per cent publishing in accordance with at least one WIPO Standard. The results indicate that part of the reason why there are less challenges for trademarks and industrial designs in digitization is because the format that is received at filing is the same format that is published.

12. For bibliographic data, the top three publication formats for each IP right type considered are shown in Table 1 below. These results indicate that Offices have a preference to publish their bibliographic data in PDF or XML, independent of the type of IP right.

IP right	Format	Number of Offices publishing in the format
Patents	PDF	43
	XML	34
	Plain text	14
Trademarks	PDF	40
	XML	26
	Plain text	13
Designs	PDF	37
	XML	23
	Other	14

Table 1: Top 3 bibliographic formats by IP right type.

13. The following question related to the language of publication. This was to determine whether the language of filing has an impact on how readily they are able to digitize an IP document. The top languages of filing are shown in Figure 2.

14. No challenges were reported by Offices, with the exception of three Offices who indicated some limited issues relating to Cyrillic script. This indicates that digitization challenges are independent of the language of filing.

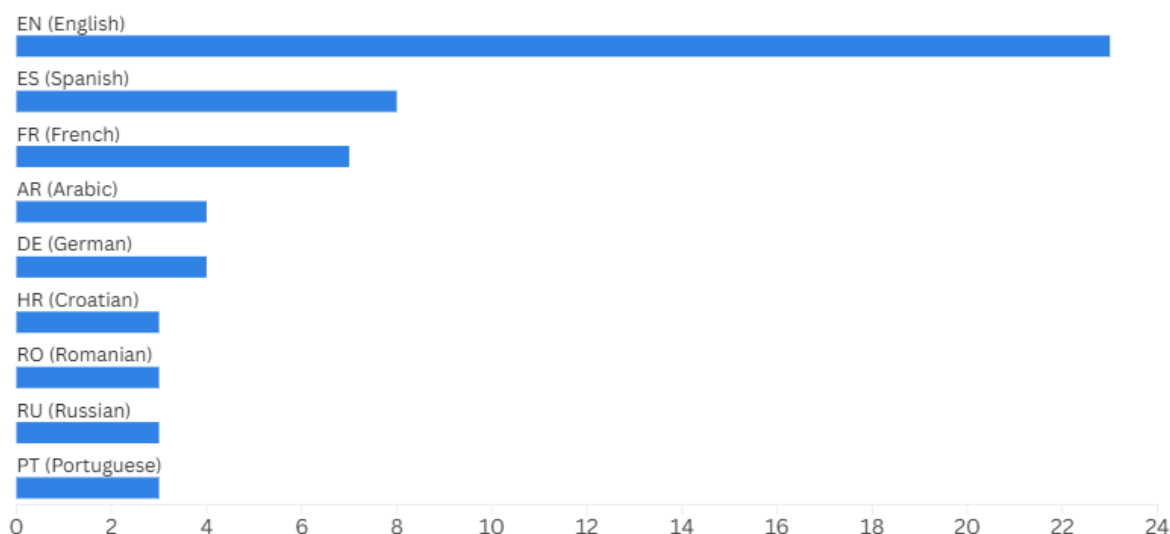


Figure 2: Top filing languages among survey respondents.

15. Of particular interest to the PCT system, in light of the work of the PCT Text Processing Task Force, was whether Offices publish color drawings they receive, or whether they publish only in black and white. The International Bureau is already capable of publishing drawings in color if the application is filed in XML, but wanted to better understand the needs of users before proposing the changes to Rule 11 and Annex F of the PCT Administrative Instructions. The results show that only approximately half of these Offices publish color drawings, if they allow them to be filed. The results are shown below in Figure 3.

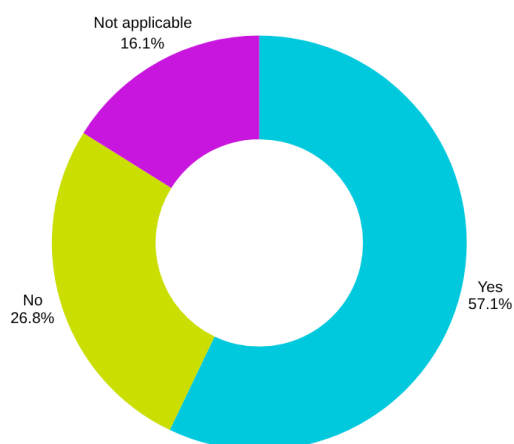


Figure 3: Do Offices publish color drawings?

16. The final part of the survey contained questions relating to digitization activities specifically conducted by Offices and the processes and platforms that they employed. Digitization in this context was defined as "the process of transcribing data into a digital form that can be directly processed by a computer"¹. Over 80 per cent of survey respondents indicated that they undertake digitization activities. Others that do not simply publish the formats they received during filing.

¹ [Collins English Dictionary](#)

17. Considering their digitization processes, Offices can be categorized into one of 5 tiers:

- Tier 1: Basic scanning (no Optical Character Recognition (OCR));
- Tier 2: Scanning then OCR to produce a PDF;
- Tier 3: Scanning, OCR to produce a structured output (XML for example);
- Tier 4: As above, but with an AI-enhanced pipeline; and
- Tier 5: Purely digital workflows from filing to publication.

18. If Offices outsource their digitization activities, the primary reasons typically include the large volume of documents requiring digitization, in particular for patent backfile collections, the specialized expertise that is required in performing this task, the cost efficiency of outsourcing and the speed at which this can be conducted externally.

19. In answering the original question and main focus of the survey, the responses indicate that the biggest challenges experienced by Offices relate to the extraction of complex content from patent documents. Even Tier 4 Offices indicated that they continue to have issues even making use of AI to extract the content, in particular the extraction of chemical formulae and mathematical expressions. Other major challenges include poor quality original paper documents, issues relating to the document structure (for example paragraph and claim numbering), and extracting drawings from the surrounding text.

NEXT STEPS

20. Following the survey responses and the analysis indicated in paragraphs 7 to 18 above, the International Bureau will consider the preparation of a set of digitization guidelines, providing recommendations where possible to address the challenges which have been highlighted by this survey, or recommend the use of a common tool such as the WIPO OCR Solution. Further details regarding these next steps can be found in document CWS/14/8.

21. Regarding future plans, 11 per cent of Offices indicated in their survey responses their intention to implement WIPO Standards, or expand the use of WIPO Standards in the future. Offices who are interested in implementing WIPO Standards are encouraged to seek technical assistance by emailing the International Bureau at cws.mail@wipo.int.

22. *The CWS is invited to:*

(a) note the content of this document;

(b) consider and approve the survey analysis as captured by paragraphs 7 to 18 and approve their publication in Part 7 of the WIPO Handbook; and

(c) encourage its Members to contact the International Bureau for technical assistance if interested in implementing structured output as part of their digitization activities.

[End of document]