

Survey Regarding Challenges Experienced by Offices in Digitizing IP Documentation

Collated responses

Identification page

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

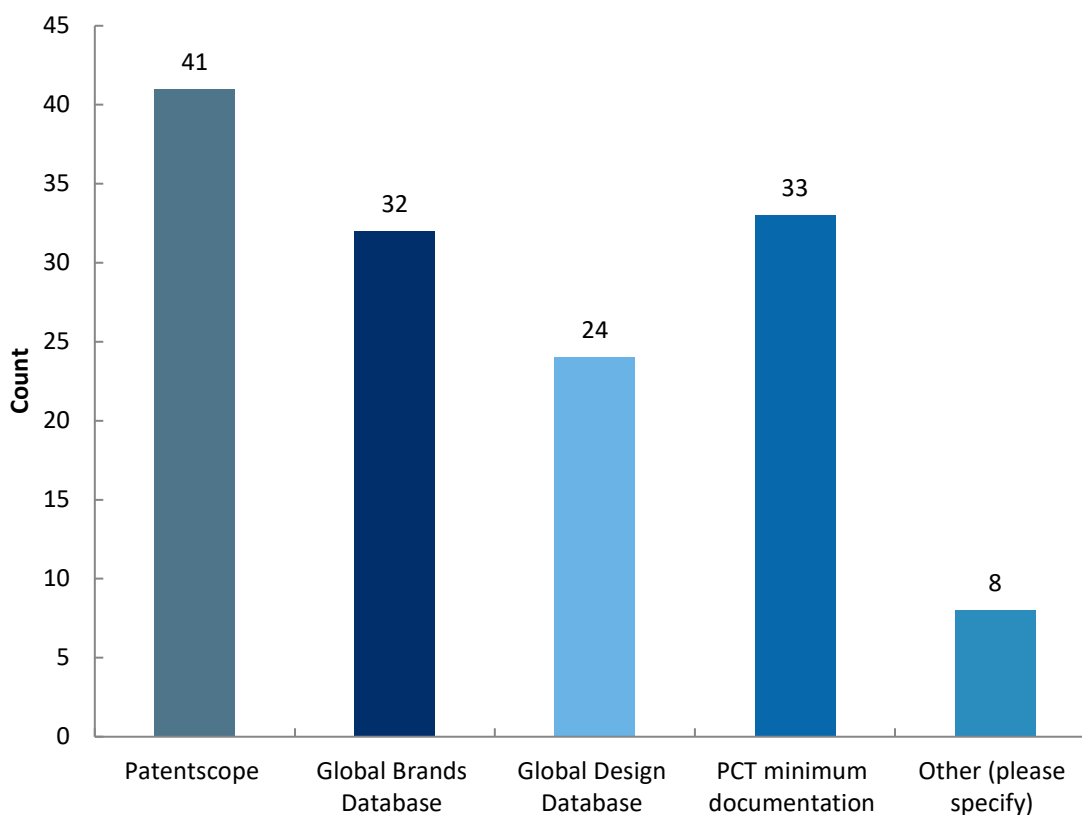
ST.3 Code	Country or Office
AP	African Regional Intellectual Property Organization (ARIPO)
AU	Australia, IP Australia
AZ	Azerbaijan, Intellectual Property Agency of the Republic of Azerbaijan (COPAT)
BA	Bosnia and Herzegovina, Institute for Intellectual Property of Bosnia and Herzegovina
BE	Belgium, The Belgian Intellectual Property Office
BG	Bulgaria, Patent Office of Republic of Bulgaria
BH	Bahrain, National Patent Office Directorate of Foreign Trade & Industrial Property Ministry of Industry and Commerce
BR	Brazil, National Institute of Industrial Property (INPI)
CA	Canada, Canadian Intellectual Property Office (CIPO)
CL	Chile, National Institute of Industrial Property (INAPI)
CN	China, China National Intellectual Property Administration (CNIPA)
CR	Costa Rica, Registro Nacional
CZ	Czechia, Industrial Property Office of the Czech Republic

DE	Germany, German Patent and Trade Mark Office (DPMA)
DK	Denmark, Danish Patent and Trademark Office Ministry of Industry (DKPTO)
DZ	Algeria, Algerian National Institute of Industrial Property Ministry of Industry (INAPI)
EA	Eurasian Patent Organization (EAPO)
EE	Estonia, The Estonian Patent Office
EM	European Union Intellectual Property Office (EUIPO)
EP	European Patent Office (EPO)
ES	Spain, Spanish Patent and Trademark Office (OEPM)
FR	France, French Patent and Trademark Office (INPI)
GB	United Kingdom, United Kingdom Intellectual Property Office (UK IPO)
GE	Georgia, National Intellectual Property Center (SAKPATENTI)
GQ	Equatorial Guinea, Scientific and Technological Research Council (CICTE)
HR	Croatia, State Intellectual Property Office of the Republic of Croatia (SIPO)
HU	Hungary, Hungarian Intellectual Property Office (HIPO)
IE	Ireland, Intellectual Property Office of Ireland (IPOI)
IL	Israel, Israel Patent Office (ILPO)
IN	India, Office of the Controller General of Patents, Designs and Trade Marks Department for Promotion of Industry and Internal Trade Ministry of Commerce and Industry
JM	Jamaica, Jamaica Intellectual Property Office (JIPO)
JP	Japan, Japan Patent Office (JPO)
KR	Republic of Korea, Ministry of Intellectual Property (MOIP)
LK	Sri Lanka, National Intellectual Property Office of Sri Lanka
LT	Lithuania, The State Patent Bureau of the Republic of Lithuania

LV	Latvia, Patent Office of the Republic of Latvia
MD	Republic of Moldova, State Agency on Intellectual Property (AGEPI)
MX	Mexico, Mexican Institute of Industrial Property (IMPI)
MZ	Mozambique, Industrial Property Institute Ministry of Industry and Commerce (IPI)
NO	Norway, Norwegian Industrial Property Office (NIPO)
NZ	New Zealand, Intellectual Property Office of New Zealand (IPONZ)
OM	Oman, National Intellectual Property Office Ministry of Commerce, Industry and Investment Promotion (MoCIIP)
PE	Peru, National Institute for the Defense of Competition and Intellectual Property (INDECOPi)
PH	Philippines, Intellectual Property Office of the Philippines (IPOPHL)
PL	Poland, Polish Patent Office
PY	Paraguay, National Directorate of Intellectual Property (DINAPI)
RO	Romania, State Office for Inventions and Trademarks (OSIM)
RS	Serbia, Intellectual Property Office of the Republic of Serbia
RU	Russian Federation, The Federal Institute of Industrial Property (FIPS) subordinate to the Federal Service for Intellectual Property (ROSPATENT)
SA	Saudi Arabia, Saudi Authority for Intellectual Property (SAIP)
SE	Sweden, Swedish Intellectual Property Office (PRV)
SG	Singapore, Intellectual Property Office of Singapore (IPOS)
SK	Slovakia, Industrial Property Office of the Slovak Republic (INDPROP)
TR	Türkiye, Turkish Patent and Trademark Office (TURKPATENT)
TZ	United Republic of Tanzania, Business Registrations and Licensing Agency (BRELA)
US	United States of America, US Patent and Trademark Office (USPTO)

General questions

2. 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.



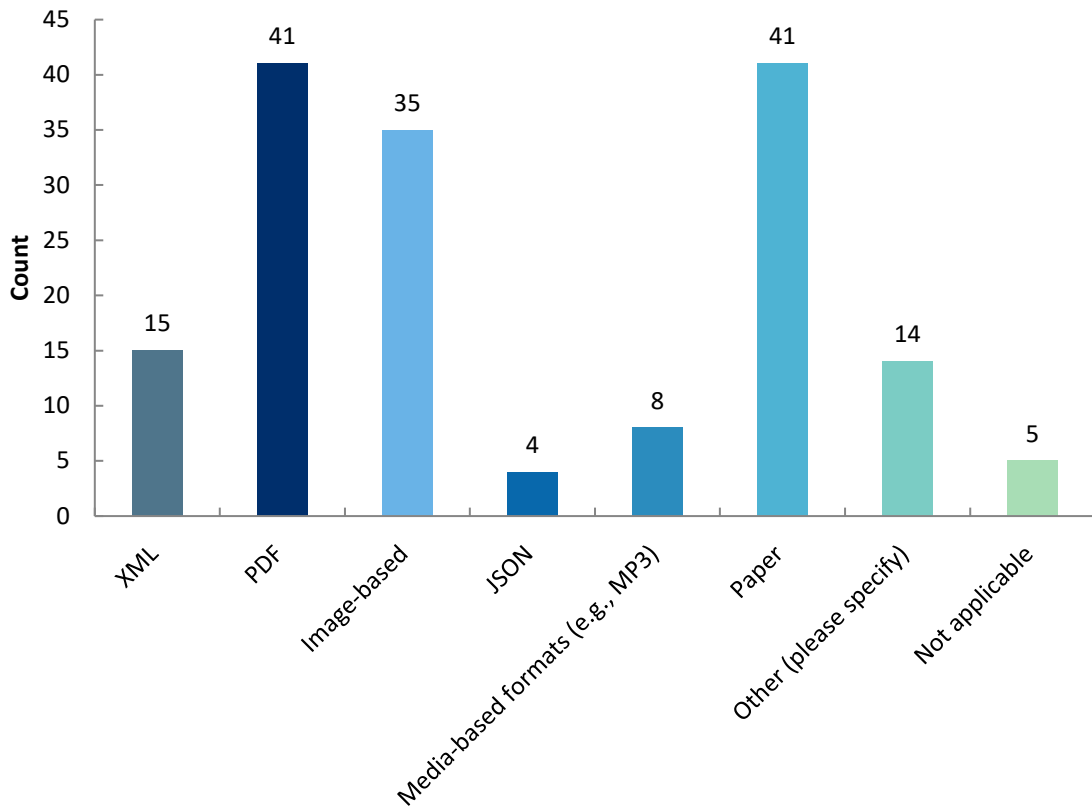
Value	Percent	Count
Patentscope	74.5%	41
Global Brands Database	58.2%	32
Global Design Database	43.6%	24
PCT minimum documentation	60.0%	33
Other (please specify)	14.5%	8

Other (please specify)

ST.3 Code	Response
BA	Our Office doesn't contribute neither stated above
CR	LATIPAT
GQ	Direct registration through OAPI
MX	LATIPAT
NO	TMView, Designview
PH	WIPO Publish, ASEAN IP Register
PY	TMview (EUIPO) and LATIPAT (OEPM/EPO; currently, bibliographic data)
TZ	Tanzania is a Member State of the African Regional Intellectual Property Organization (ARIPO). While national filings are handled locally by BRELA, patent applications filed regionally through ARIPO that designate Tanzania are included in WIPO's PATENTSCOPE via ARIPO's regional data stream.

3. What file or document format(s) does your Office accept at filing for IP applications? Please select all that apply.

a) For industrial design collections:

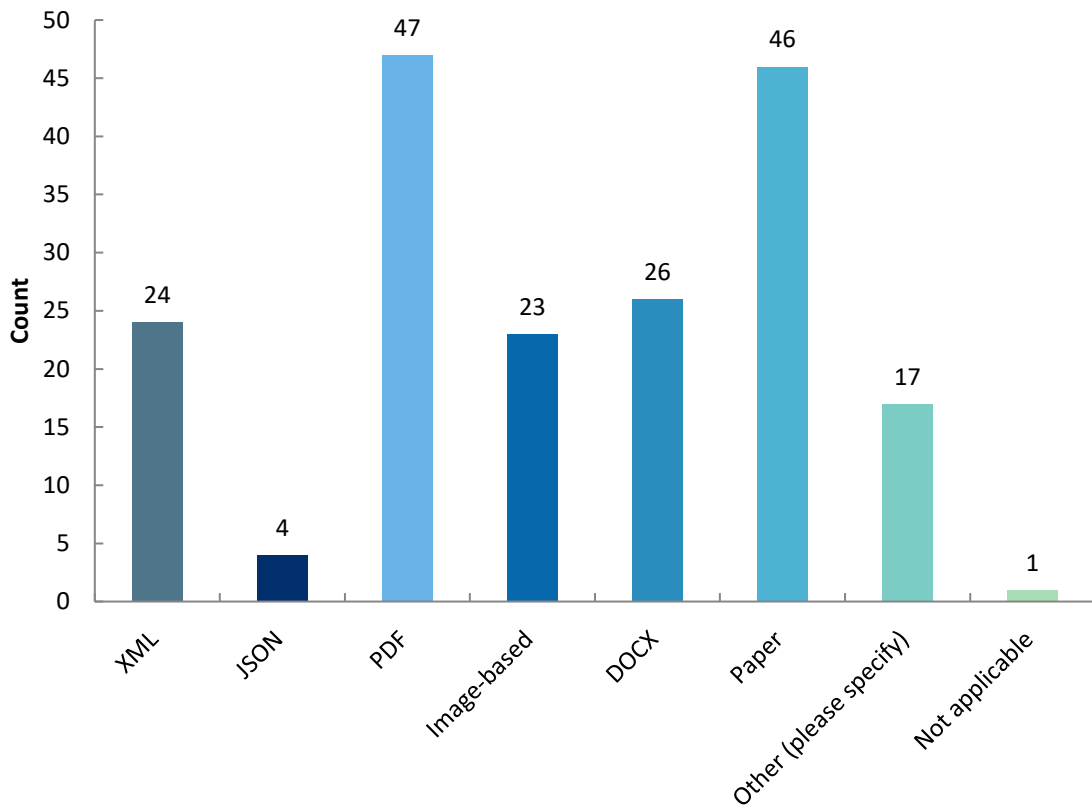


Value	Percent	Count
XML	26.8%	15
PDF	73.2%	41
Image-based	62.5%	35
JSON	7.1%	4
Media-based formats (e.g., MP3)	14.3%	8
Paper	73.2%	41
Other (please specify)	25.0%	14
Not applicable	8.9%	5

Other (please specify)

ST.3 Code	Response
BG	Specialized WEB interface for e-filing
BH	Word MS
CR	JPG for the inclusion of the design
DZ	Digital form
EM	EUIPO eFiling application generates a ZIP containing an XML file, PDF application forms and attachments and the representation(s) in JPEG, MP4, OBJ or STL format. As of 01/07/2026, X3D format and paper filing are decommissioned, MP4 format is added. JSON used for eFiling API.
IE	Word
JP	SGML, a product model
LV	DOCX, EDOC
NO	DOCX
NZ	Online form with image representations
RU	3D formats
SE	When filing via DSEfiling XML, images and pdf are used.
SG	DOC, DOCX
SK	DOCX

b) For patent collections:

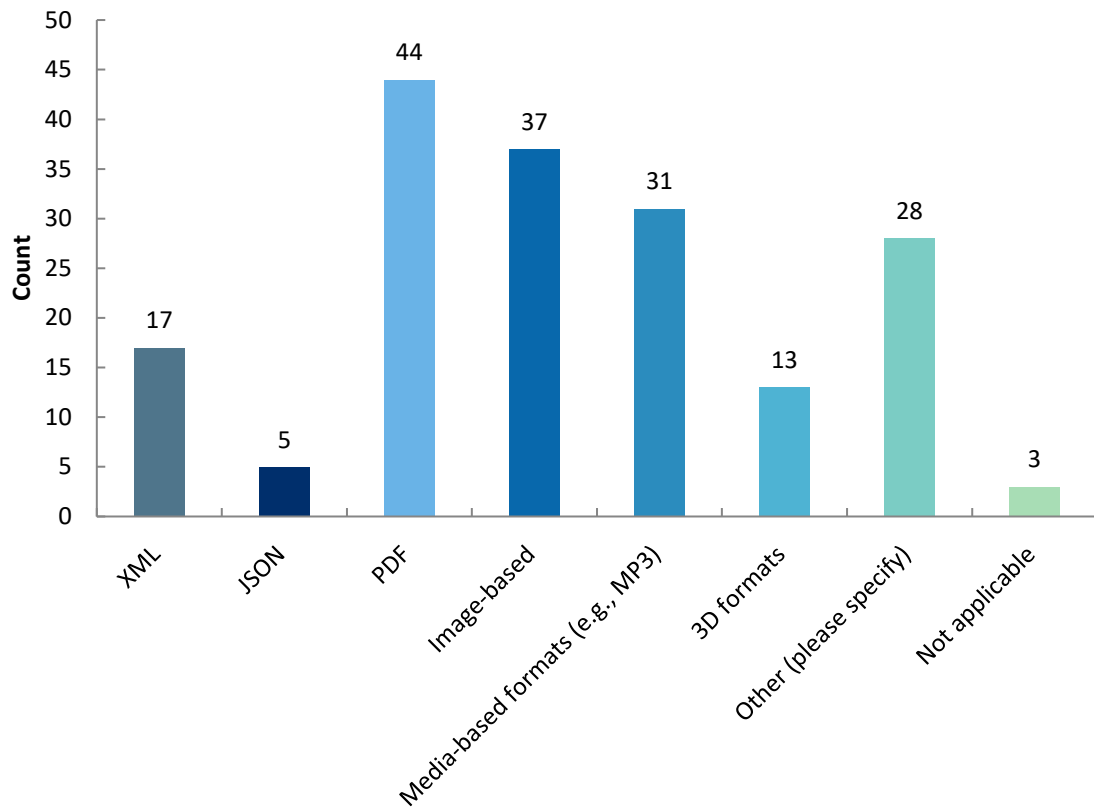


Value	Percent	Count
XML	42.9%	24
JSON	7.1%	4
PDF	83.9%	47
Image-based	41.1%	23
DOCX	46.4%	26
Paper	82.1%	46
Other (please specify)	30.4%	17
Not applicable	1.8%	1

Other (please specify)

ST.3 Code	Response
BG	Specialized WEB interface for e-filing.
BH	Word MS
BR	XML ST.26 for sequence listings.
CA	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
CR	JPG for the inclusion of figures where applicable.
DZ	Digital form
ES	XML for sequence files (WIPO Standard ST.26).
HU	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.
IN	ST.26 format XML for Sequence Listing only for PCT applications; Sequence listing is also accepted in .TXT format.
LT	All formats are suitable if zipped.
LV	EDOC, BMP
NZ	Online form or API submission with specifications in PDF.
PY	DOCX
RU	3D formats
SE	For sequence listings we allow XML (in accordance with to ST.26). We allow PCT filing via WIPO's filing tool ePCT, where it is possible to file DOCX.
SG	TXT
US	For indirect Hague filings, accept JPEG.

c) For trademark collections:



Value	Percent	Count
XML	30.4%	17
JSON	8.9%	5
PDF	78.6%	44
Image-based	66.1%	37
Media-based formats (e.g., MP3)	55.4%	31
3D formats	23.2%	13
Other (please specify)	50.0%	28
Not applicable	5.4%	3

Other (please specify)

ST.3 Code	Response
BA	Paper
BG	Specialized WEB interface for e-filing.
BH	Word MS
CA	Paper
CL	Paper
CR	Paper and JPG for the inclusion of the design.
DY	Currently, trademark application procedures at DINAPI are entirely in physical (paper) format until the conclusion of the procedure, with the exception of the Certificate of Registration, which is issued and sent in digital format (PDF) with a qualified electronic signature. For the filing of applications, the institution offers two approaches: In person: the applicant completes the electronic form, prints it, signs it, and physically submits it (paper) at DINAPI, upon payment of the corresponding fees. Receipt is formalized through the delivery of a proof of filing. Online: the applicant makes the electronic payment via Bankcard and submits the application through the institutional computer system, without the need for in-person assistance. Receipt is automatically confirmed by sending the application form to the email of the industrial property agent, together with the entry number, date and time of receipt.
DZ	Digital form
EE	DOCX
EM	EUIPO eFiling application generates a ZIP containing an XML file, a PDF application form and attachments and the representation in JPEG, MP3, MP4, OBJ, STL or X3D format. JSON used for eFiling API. XML used by WIPO transactions for IRs. PDF are also used For filing via communication backup.
HU	Paper
IE	Word, paper
IL	DOCX
JM	Paper, plain text

JP	SGML, Paper
KR	Paper
LK	Paper
LV	DOCX, EDOC
MD	Paper
MX	Paper
NO	DOCX
NZ	Online form or API submission with image marks in image formats.
PL	Paper
RO	Paper
RS	Paper
SE	Paper is also allowed. When filing via TMEfiling XML, images and pdf are used.
SG	DOC, DOCX
SK	MP4

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

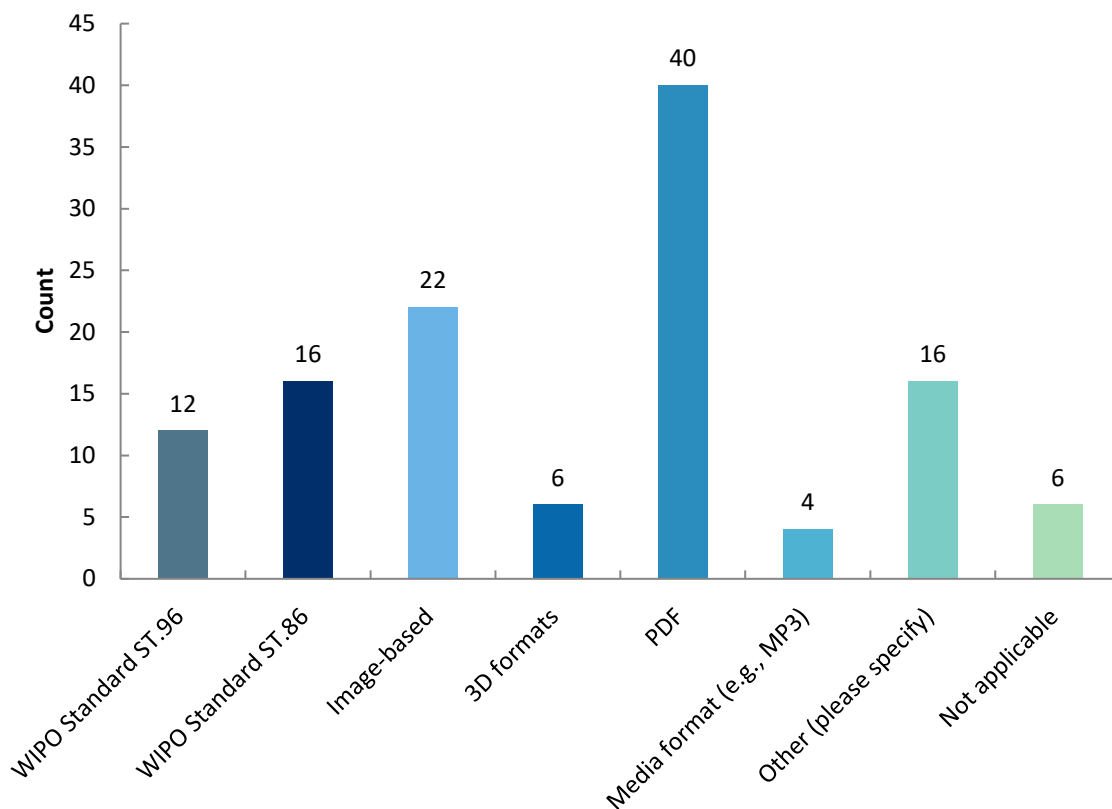
ST.3 Code	Response
AP	EN
AU	EN
AZ	AZ, RU
BA	BS, HR, SR
BE	NL, FR, DE
BG	BG
BH	EN, AR
BR	PT
CA	CIPO publishes correspondence as submitted by clients. CIPO correspondences in French or English and publish in the language sent or received based on clients preferred language (EN, FR).
CL	ES
CN	ZH
CR	ES
CZ	CS, EN
DE	DE
DK	DK, EN
DZ	FR
EA	RU
EE	ET
EM	Up to 23 EU languages, depending on the type of document or procedure: BG,CS,DA,DE,EL,EN,ES,ET,FI,FR,HR, HU,IT,LT,LV,MT,NL,PL,PT,RO,SK,SL,SV.

EP	EN, FR, DE
ES	ES
FR	FR
GB	EN
GE	KA
GQ	ES, FR
HR	HR
HU	HU
IE	EN
IL	EN, HE, AR
IN	EN
JM	EN
JP	JP
KR	EN, KO
LK	SI, EN, TA
LT	LT
LV	LV, EN (patent abstracts may also be published in English).
MD	RO
MX	ES
MZ	PT
NO	NO, EN
NZ	EN
OM	AR, EN

PE	ES
PH	EN
PL	PL
PY	ES
RO	RO
RS	RS (Latin or Cyrillic)
RU	RU, EN
SA	AR
SE	SV, EN
SG	EN
SK	SK
US	EN

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

a) For industrial design collections:

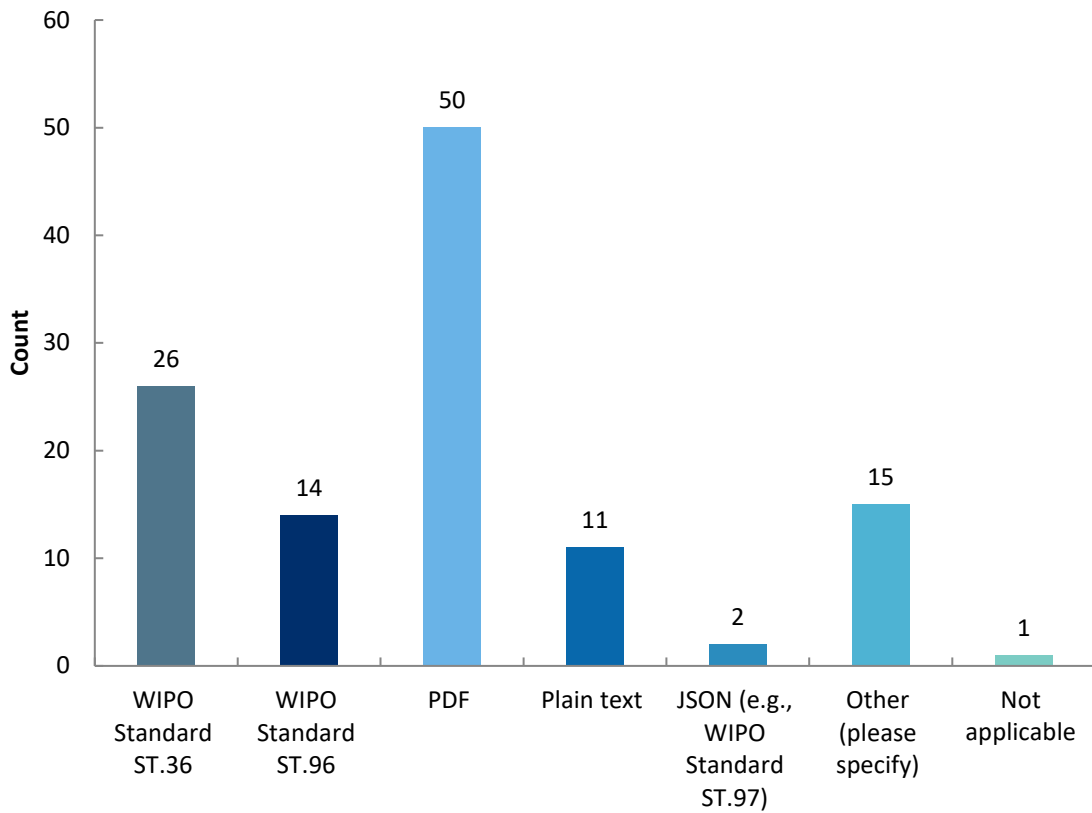


Value	Percent	Count
WIPO Standard ST.96	21.4%	12
WIPO Standard ST.86	28.6%	16
Image-based	39.3%	22
3D formats	10.7%	6
PDF	71.4%	40
Media format (e.g., MP3)	7.1%	4
Other (please specify)	28.6%	16
Not applicable	10.7%	6

Other (please specify)

ST.3 Code	Response
AU	API/JSON (Australian Design Search API)
BG	Specialized WEB interface for e-filing.
BR	Open data CSV files.
CR	JPG for the inclusion of the design.
DZ	Publish on WIPO Publish/BOPI.
EE	Web-based public database.
EM	EUIPO Download – the XML Design package follows a customized version of ST.86. The Design package also contains the design representation (static views: JPEG, 3D models: OBJ, STL, videos: MP4). EUIPO also provides web site links to download the animated/dynamic representations which cannot be embedded in a PDF document. Users can access the design details by using our search tools (eSearch plus and DESIGNview). EUIPO also disseminates design bibliographic data and representations via Design Search API.
ES	The ST.86 format is used by EUIPO and is XML compatible with ST.86.
IL	HTML (online)
IN	IPO submits its Industrial Designs data to WIPO Publish in ST.86 format.
LV	ZIP packages containing RTF publication files custom XML sent to the Public Database (PDB).
NO	Proprietary JSON format.
NZ	Available via API in XML with representations in image format documents.
PH	Plain text
PY	Publication is done in specialized external journals and magazines. We post the bibliographic data and image (figure) on a public inquiry web platform.
SG	In HTML format in our online IP filing system and in XML to Global Design Database.

b) For patent collections:

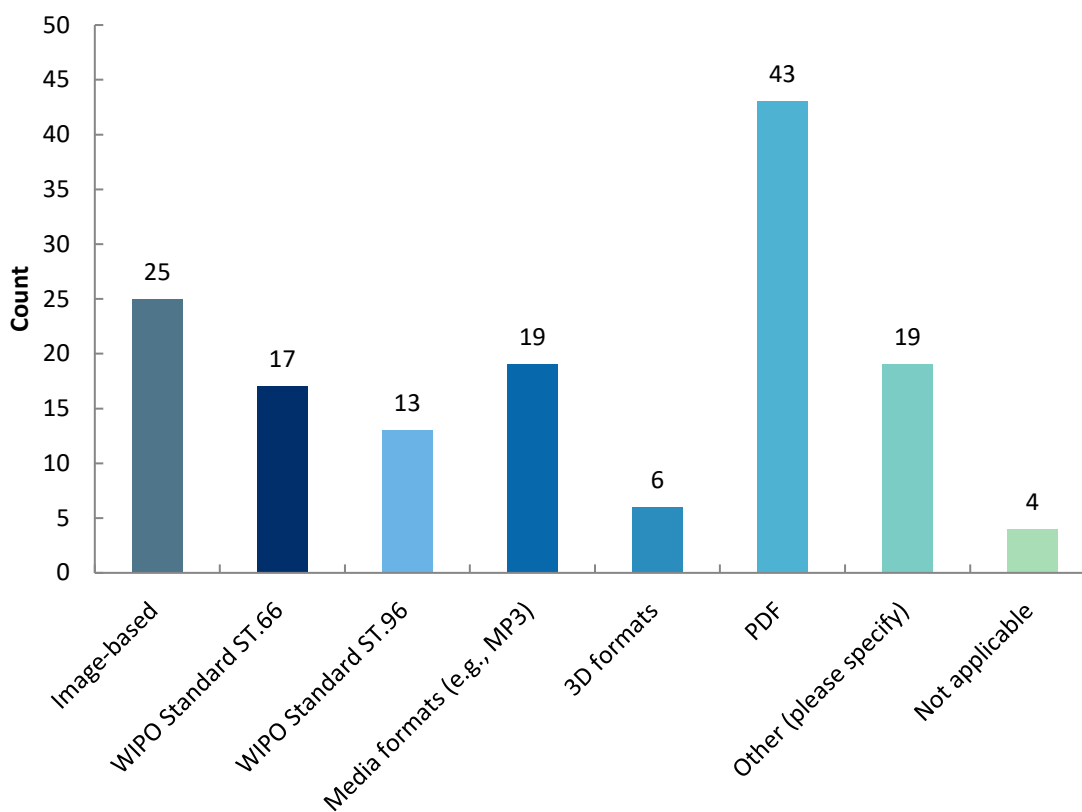


Value	Percent	Count
WIPO Standard ST.36	46.4%	26
WIPO Standard ST.96	25.0%	14
PDF	89.3%	50
Plain text	19.6%	11
JSON (e.g., WIPO Standard ST.97)	3.6%	2
Other (please specify)	26.8%	15
Not applicable	1.8%	1

Other (please specify)

ST.3 Code	Response
BG	Specialized WEB interface for e-filing.
BR	XML ST.26 for sequence listings.
CN	XML data based on our own standard.
CR	JPG for the inclusion of figures, where applicable.
CZ	WIPO Standard ST.26
DZ	Publish on WIPO Publish/BOPI.
EE	Web-based public database.
ES	Standard ST.26 for sequences, Standard ST.27 for legal data, Standard ST.32 for abstracts, Standard ST.37 for the Authority File.
IN	IPO submits its Patents data to WIPO Publish in ST.36 format.
LV	ZIP packages containing RTF publication files custom XML sent to the Public Database (PDB).
NZ	Available via API in XML with specifications in PDF format documents.
PY	Publication is done in specialized external journals and magazines. We post the bibliographic data and image (figure) on a public inquiry web platform (only published patents are displayed). In addition, application publication notices are uploaded to a dedicated section on DINAPI's main website.
SG	In HTML format in our online IP filing system.
SK	XML
US	TIFF

c) For trademark collections:



Value	Percent	Count
Image-based	44.6%	25
WIPO Standard ST.66	30.4%	17
WIPO Standard ST.96	23.2%	13
Media formats (e.g., MP3)	33.9%	19
3D formats	10.7%	6
PDF	76.8%	43
Other (please specify)	33.9%	19
Not applicable	7.1%	4

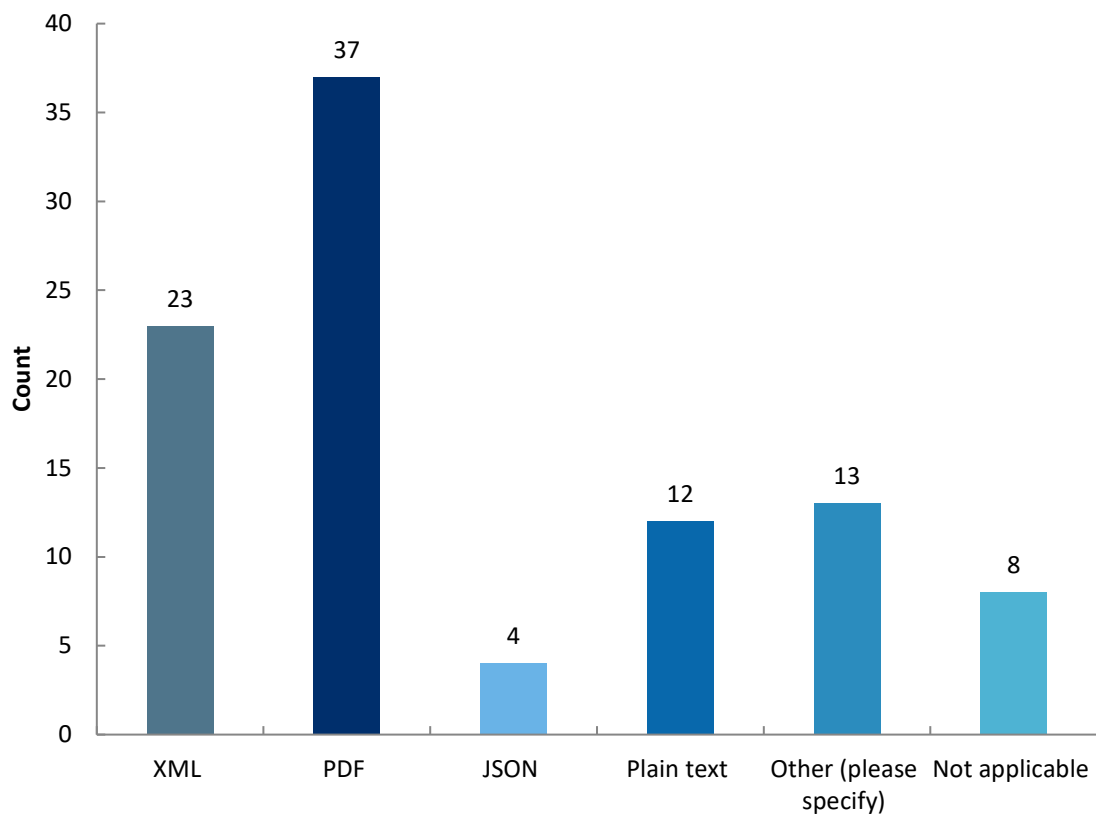
Other (please specify)

ST.3 Code	Response
AU	MECA; API/JSON (Australian Trade Mark Search API)
BG	Specialized WEB interface for e-filing.
BR	XML (proprietary format); Open data CSV files
CR	JPG for the inclusion of the design.
DZ	Publish on WIPO Publish/BOPI.
EE	Web-based public database.
EM	EUIPO Download – the XML Trademark package follows a customized version of ST.66. The Trademark package also contains the trademark representation (JPEG, OBJ, STL, X3D, MP3 and MP4). EUIPO also provides web site links to download the representations which cannot be embedded in a PDF document. Users can access the trademark details by using our search tools (eSearch plus and TMview). EUIPO also disseminates trademark bibliographic data and representations via Trademark Search API. Trademarks' Gazette available in PDF format.
ES	The ST.66 format is used by EUIPO and is XML compatible with ST.66.
FR	.tiff .jpg .gif .mp4
GB	XML - MECA format
IL	HTML (online)
IN	IPO submits its Trademarks data to WIPO Publish in ST.66 format.
LV	ZIP packages containing RTF publication files custom XML sent to the Public Database (PDB).
NO	Proprietary JSON format
NZ	Available via API in XML with image marks in image format documents.
PL	XML

PY	Publication is done in specialized external journals and magazines. We disseminate the bibliographic data and label (logotype) on a public inquiry web platform. In addition, we offer a specialized journal called RedPI, where trademarks are published in file format (PDF) and accessed through a public inquiry web platform.
SG	In HTML format in our online IP filing system and in XML to Global Brands Database.
SK	MP4, XML

6. 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.

a) For industrial design collections:

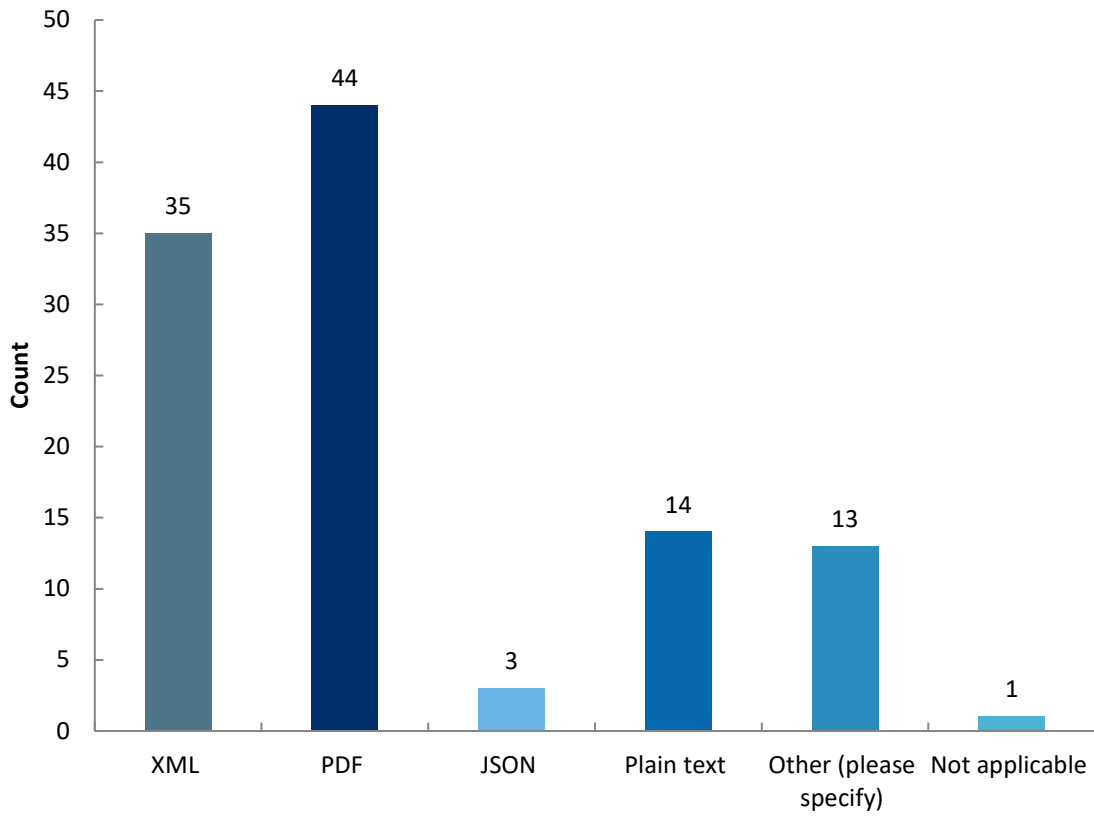


Value	Percent	Count
XML	41.1%	23
PDF	66.1%	37
JSON	7.1%	4
Plain text	21.4%	12
Other (please specify)	23.2%	13
Not applicable	14.3%	8

Other (please specify)

ST.3 Code	Response
AU	HTML (web-based search system)
BH	Word MS
CR	HTML generated by the WIPO OCR tool.
DZ	WIPO publish
EA	HTML (Publication Server at Web Site)
EE	Web-based public database.
EM	eSearch plus allows to export as Excel file: DESIGNview allows to export as PDF, MS Word and MS Excel.
JP	TSV
LV	RTF files included in ZIP publication packages.
PY	We post the bibliographic data and image (figure) on a public inquiry web platform.
SE	The bibliographic data is also presented as text in "Search PRV's Databases".
SG	In HTML format in our online IP filing system.
SK	HTML

b) For patent collections:

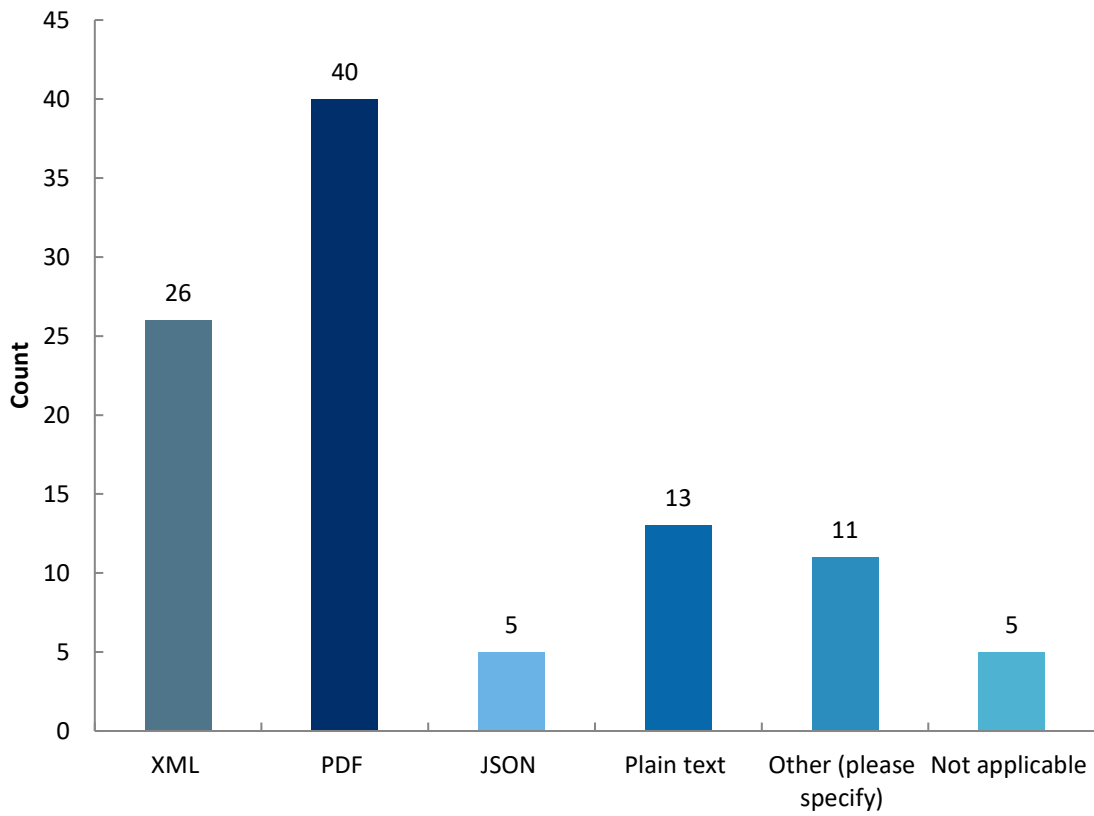


Value	Percent	Count
XML	62.5%	35
PDF	78.6%	44
JSON	5.4%	3
Plain text	25.0%	14
Other (please specify)	23.2%	13
Not applicable	1.8%	1

Other (please specify)

ST.3 Code	Response
AU	HTML (web-based search system)
BH	Word MS
CR	HTML generated by the WIPO OCR tool.
EA	HTML (Publication Server at Web Site)
EE	Web-based public database.
JP	TSV
LT	ZIP files
LV	RTF files included in ZIP publication packages.
PY	We post the bibliographic data and image (figure) on a public inquiry web platform (only published patents are displayed). In addition, application publication notices are uploaded to a dedicated section on DINAPI's main website.
SE	The bibliographic data is also presented as text in "Search PRV's Databases".
SG	In HTML format in our online IP filing system.
SK	HTML
US	XML- 2002-PRESENT; PDF- 1790 TO PRESENT; PLAIN TEXT- 1976 TO 2001

c) For trademark collections:



Value	Percent	Count
XML	46.4%	26
PDF	71.4%	40
JSON	8.9%	5
Plain text	23.2%	13
Other (please specify)	19.6%	11
Not applicable	8.9%	5

Other (please specify)

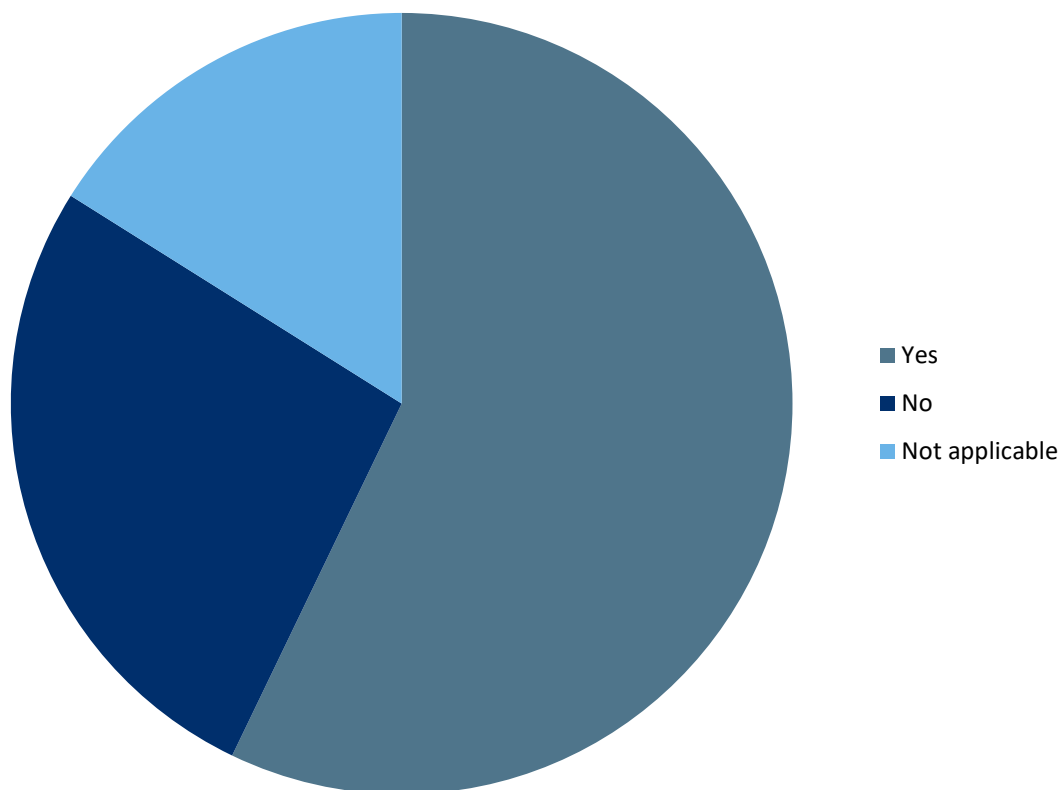
ST.3 Code	Response
AU	HTML (web-based search system)
BH	Word MS
CR	JPG for the inclusion of the design.
EE	Web-based public database.
EM	eSearch plus allows to export as Excel file: TMview allows to export as PDF, MS Word and MS Excel.
JP	TSV
LV	RTF files included in ZIP publication packages.
PY	We post the bibliographic data and image (figure) on a public inquiry web platform.
SE	The bibliographic data is also presented as text in "Search PRV's Databases".
SG	In HTML format in our online IP filing system.
SK	HTML

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

ST.3 Code	Response
AP	All published data is validated before publication.
AU	Testing was conducted on the process to build the XML for PCT Minimum Documentation requirements. Validation of every instance is not performed due to resource constraints.
AZ	The publication is not automated.
BG	Not applicable
CA	• 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.
CL	The data is validated before posting.
CN	Not applicable
CR	No validation processes are currently implemented, since the WIPO Publish extraction tool, provided by WIPO, is used.
DE	Does not apply
EM	The Office implemented some quality controls to ensure the validity of these instances. Although the validation of XML is implemented, we did not activate it by default due to efficiency matters.
GB	Not applicable
GQ	We do not produce XML or JSON instances, as we do not have a digitized database from WIPO.
HR	Not applicable
HU	Not applicable
IE	The XML file is cross-referenced and validated using an authority file.
JM	Not applicable

KR	Not applicable
LK	XML data used for dissemination and exchange is generated and managed through WIPO-supported systems. Validation controls are integrated into the system environment, and the Office does not currently undertake separate manual validation processes before publication.
LV	The Office produces XML instances for dissemination and data exchange purposes, including exports to the Public Database (PDB) and participation in EUIPO cooperation services. XML files are generated automatically by the Office systems and packaged as part of established publication workflows. For national publication processes, XML instances are generated from internal systems and transferred to downstream systems without a separate formal validation against an external XSD schema. The publication workflow relies on controlled generation processes and successful processing by receiving systems. For data exchanged with external cooperation services, such as EUIPO platforms (e.g., TMview and DesignView), validation requirements are governed by the respective data exchange specifications and supported WIPO standards.
MX	The IP collection of the Mexican Office is currently being processed using the ST.36 and ST.37 standards for text extraction and XML generation.
MZ	Not applicable
NO	We use internal JSON for gazette production, which is available for download. Plan to move to validated ST.97.
OM	Not applicable. The Office does not currently publish IP data in XML or JSON formats. Bibliographic and publication data are disseminated primarily via the online IP portal and in PDF/image-based formats. Therefore, XML/JSON validation prior to publication is not part of the current workflow.
PH	Not applicable, we do validation of produced XML.
PL	Our XMLs are generated in the correct XML structure directly from the central repository to the API, so they do not need any additional validation.
PY	Not applicable
RO	Not applicable
RU	The Office validates all XML and JSON instances prior to publication.
SG	Not applicable
TZ	We are not producing XML

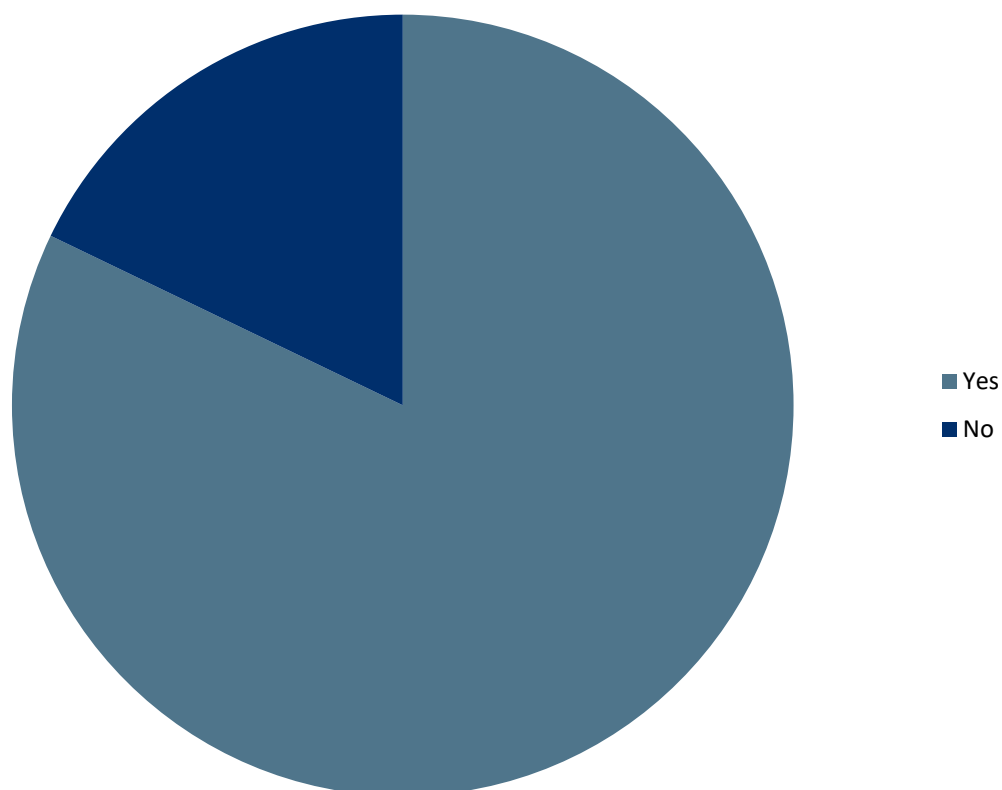
8. Does your Office publish the color drawings they receive, if they are permitted during the filing of a patent application?



Value	Percent
Yes	57.1%
No	26.8%
Not applicable	16.1%

Digitization activities

9. 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?



Value	Percent
Yes	82.1%
No	17.9%

9.1. 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.

ST.3 Code	Response
AP	They are scanned and OCRRed.
AU	To align with PCT Minimum Documentation requirements, OCRRed PDFs are converted to XML. Any paper documentation received is digitised by scanning to PDF. Not all documents are OCRRed, but anything forming part of a patent's specification is OCRRed. All IP rights documentation stored as PDF; data is converted to formats indicated above as needed.
BA	Paper document manually scanned and OCRRed.
BE	Paper document manually scanned (for about 5% of the Belgian patent applications that are still filed on paper).
BG	OCRRed
BH	DOCX and pdf files.
BR	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.
CA	Patent: Paper is scanned as images, plus some subject matter is OCRRed so that it can be published and searched. OCR using Datacap. Trademark Industrial Designs and Copyrights: Paper is manually scanned.
CL	INAPI has its own scanners with OCR software, which generate PDF/A files. As regards patents, starting in 2026, once a patent is granted it is converted to XML using WIPO-OCR and published on the institutional website.
CN	CNIPA mainly accepts patent application documents in XML format. As regards patent application documents submitted on paper, they will be scanned and processed into XML format. For trademark business, paper documents will be manually scanned into PDF format or digitalized into structural data.

CR	Following institutional policy, documents submitted in paper format are converted to PDF format. Likewise, documents received digitally are only accepted in that format. Exceptionally, drawings and designs are stored in the database in JPG format.
CZ	All paper documents associated with IP files are scanned, and OCR is applied where the quality of the source document allows. For national patent applications specifically, selected application documents are additionally converted into structured formats: claims are converted to HTML/XML, while descriptions are converted to DOCX with embedded drawings.
DE	• Paper documents: Physical documents are scanned at 300 DPI and converted to PDF/A-2U using the Foxit Rendition Server, with Optical Character Recognition (OCR) applied as an integral part of the process. Patent documents are predominantly scanned in black and white, reflecting the technical nature of the content. Greyscale or colour scanning is applied where document characteristics require it. Selected form fields are subsequently extracted and reviewed by qualified staff. • FAX: Documents received via fax are processed as TIFF files and converted to PDF/A-2U, following the same workflow as paper-based scanned documents. • Electronic submission channels: The DPMA operates a proprietary submission platform through which applicants are required to submit documents in PDF format. These are converted to PDF/A-2U using the Foxit Rendition Server. In cases where discrepancies are identified between the converted and original files, the original submission is retained for reference. Metadata accompanying electronic submissions is extracted and processed.
DK	Paper filings are scanned and OCRed and converted into PDF.
DZ	Documents received in paper format are scanned for archiving and electronic management. However, this concerns only the creation of digital copies of the documents and does not include any OCR processing or conversion to structured formats such as XML. Scanned documents are therefore kept in their original electronic form on WIPO Scan only.
EA	• Paper document manually scanned and OCRed -> normalized DOCX (DOCX with correct formatting, having only technically allowed content and markup of parts, claims, main drawing etc.) • Image-only pdf manually OCRed -> normalized DOCX • Text pdf is processed to extract the text layer (operator controlled) -> normalized DOCX • Docx is processed and normalized (operator controlled) -> normalized DOCX • Normalized DOCX is then converted to publication quality XML
EE	Paper document manually scanned and OCRed

EM	Paper documents manually scanned and OCRRed; PDF documents manually processed or OCRRed; Manual key-in and revision of OCRRed outcome.
EP	For an OCR text quality of >99.5% (internal use) - Specifications documents in paper are scanned and OCRRed internally - Specifications documents in image or PDF are OCRRed internally - Specification documents in DOCX are converted to ST36 internally - Forms are OCRRed by an external contractor and converted to XML - Citations in application documents and reports are extracted by an external contractor and converted to XML For an OCR text quality of 99.995% (Publication) - Specifications documents are OCRRed by an external contractor. Images of pages/pages of PDFs are OCRRed to be structured into ST.36. DocX are transformed into ST.36.
ES	Digitization at the OEPM is residual, since most data entries are digital and are made through electronic services (https://sede.oepm.gob.es/eSede/es/index.html). In some cases, OCR is necessary.
GB	Paper documents manually scanned and OCRRed to create XML.
GE	Paper documents are manually scanned but are not transformed , we keep them as scanned images.
GQ	For the most part, we handle application documents manually or physically, submitted by subscribers or other partners.
HR	Paper documents - manually scanned and OCR patent document - DOCX converted to XML (internal tool).
HU	Paper scanned and OCRRed
IE	Applicable to patent documents only: - Paper documents are manually scanned and converted to PDF without OCR processing. - Electronically filed applications submitted in Word format are converted to PDF for publication purposes.
IL	Regarding industrial design collections: almost all design application documents and communications with applicants are submitted through the e-filing system. Regarding patent collections: almost all patent application documents and communications with applicants are submitted through the e-filing system. Any documents received on paper are scanned immediately. To ensure accessibility, unsearchable PDF files automatically undergo Optical Character Recognition (OCR) processing before being published online. The conversion of these PDF files into HTML format is then performed on the fly. Regarding trademark collections: the Trademarks Regulations mandate digital filing. All documents are submitted electronically in a searchable PDF format, and in special cases we also accept DOCX format.
IN	Paper documents are manually scanned and OCRRed; The applications which are filed online, they are received in PDFs and these PDFs are OCRRed later.

JM	Paper document manually scanned into PDF documents and OCRed using Adobe 2017 Professional.
JP	An external organization performs OCR processing on paper documents and converts them into XML or SGML. Non-textual elements such as drawings are converted into image data by the same external organization.
KR	Paper document → Scanning → OCR processing → E-filing software → XML conversion DOCX file → E-filing software → XML conversion HWP/HWPX file → E-filing software → XML conversion International application XML format → Domestic application XML format conversion
LK	NIPO receives applications in paper and electronic formats. Paper records are scanned and converted into digital image files and PDF documents. Bibliographic data is captured and maintained in the IP administration system. Historical records are digitized through scanning, indexing, quality checking, and electronic storage for retrieval and dissemination.
LT	Manually scanned and OCRed.
LV	Historical paper documents were scanned into PDF and, as part of the EUIPO CSHF project, prepared with related metadata for bulk upload into the Back Office system. Current paper submissions are scanned and stored as PDF in PAIS/TMDS, while electronic documents are stored in their received formats and linked to the relevant dossier. Bibliographic and procedural data are entered or imported into structured fields for further processing, publication and data exchange.
MD	In cases where patent documents are not initially provided in machine-readable text formats, we formally request and subsequently receive them in such formats (DOCX or PDF). Any remaining paper-based documents are digitized through manual scanning followed by OCR processing.
NO	• Paper, scanned and OCRed to PDF • XML to PDF for forms • DOCX to PDF/A • PDF to PDF/A • TIFF to PDF • PDF to XML via WIPO Scan (from 2019)
OM	Paper documents, where received, are scanned into digital format (PDF). Any textual data is then manually captured or entered into the system.

PH	1. Paper documents were scanned, OCRRed and uploaded to the EDMS of WIPO-IPAS. 2. OCRRed patent documents being extracted, indexed and uploaded in WIPO-IPAS and once passed the validation, it will be readily available at the WIPO Publish for publication. 3. Readily available indexed patent documents from WIPO Publish were converted to DOCX and uploaded to WIPO-IPAS for conversion to XML for validation and shared as ST. 96 to PCT Min Doc.
PL	Manually scanned paper documents are not OCRRed or saved as DOCX files or converted to XML. Scanned documents remain in PDF format.
RS	Paper document manually scanned and OCRRed.
RU	Paper documents are manually scanned and processed using OCR technology.
SE	We use the tool WIPO OCR for national patent application.
SG	Documents are submitted to us in either PDF or DOCX format. Patent specifications submitted in PDF format are OCRRed.
SK	Both paper documents manually scanned and OCRRed or DOCX converted to XML.
TZ	Paper document manually scanned and OCRRed or DOCX converted to XML.
US	OCR for image files DOCX to XML paper is scanned to image which is then OCRRed.

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

ST.3 Code	Response
AP	Because of volume of the documents and expertise required.
AU	Not applicable
BA	No
BE	The BE Office does not rely on a third party to perform the digitization of its documents.
BG	Not applicable
BH	Easy access and affordable.
BR	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.
CA	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.
CL	Previously, digitization was done by an external provider, owing to lack of capacity given the volume. Internal resources are currently used to perform this function.
CN	Taking into consideration both regulations and work efficiency, CNIPA realizes the digitization of patent application documents via buying technical services from third parties.
CR	Not applicable
CZ	All digitization is performed internally.
DE	The DPMA utilises the Foxit Rendition Server as its primary solution for document conversion and digitization. This product was selected on the basis of its technical performance, in particular its highly effective compression capabilities for colour documents, as well as the reliability and quality of the output it produces. The vendor's technical support has consistently met the Office's requirements.
DZ	No
EM	Repetitive task with no added value.

EP	It is difficult to reach high quality, and the third party had the expertise. But we are getting there with AI. In some digitization steps, third parties are used. This is for flexibility reasons as well as to harness existing expertise.
ES	For the OEPM Historical Archive, we worked with the <i>Universidad Autónoma de Madrid</i> , given the large volume of the archive and its importance for academic research.
GB	Lack of capability at the time and more effective to use a third party.
GQ	We have not entrusted anyone with the digitization of our documents.
HR	Not applicable
HU	Not applicable
IE	The Office recently decided to rely on a third party for document digitisation by sharing data with WIPO, which currently provides OCR services. This approach was taken to improve the accuracy and efficiency of digitising documents.
IL	The digitization operations are performed in-house.
IN	The backlog collection of documents was too big to be managed by IPO; hence third party was engaged for the digitization.
JM	No
JP	The digitization of documents received by the JPO is a large-scale and routine operation that requires a certain level of specialized skills, knowledge, and substantial equipment. Accordingly, it is not performed in-house but is instead outsourced to external organizations. In utilizing such external organizations, the JPO implements necessary legal measures, including obligations of confidentiality.
KR	Entrusted to an institution equipped with the facilities and personnel prescribed by law to ensure efficient digitization operations.
LK	NIPO has engaged external service providers and technical assistance projects where necessary to accelerate the digitization of historical records, improve data quality, and benefit from specialized expertise, equipment, and resources that were not available internally.
LT	Because of the high volume of documents and high equipment quality.

LV	Historical digitization activities were carried out in cooperation with EUIPO under the Capture and Store Historic Files (CSHF) project. As part of that initiative, legacy paper documents were digitized and imported into the Office's systems. Current document-processing activities are mainly performed through internal workflows and electronic filing systems.
NO	In general, no, but backfile patent converted and qualified up to 2018 by Jouve.
OM	The National Office did not rely on any third party for the digitization and archiving of applications. Instead, the process was carried out internally by the existing staff members, in order to benefit from their experience and familiarity with the nature of the documents and internal procedures. This approach also helped ensure the confidentiality of information, reduce operational costs, and enhance the employees' skills and capabilities in document management and digital archiving.
PH	Not applicable, we are performing our digitization in-house.
PL	We seek assistance from another department in our Office in the following cases: • Non-standard formats, such as maps or posters. • If the colors of the document are dark (navy blue, dark green etc.) and appear black on our scanners. • Multi-page notarial deeds and contracts that are stapled together.
RS	In the past we used Jouve for digitalization of patent documentation backlog (EPO support). We used it because of the lack of financial and HR resources.
RU	We did not rely on any third parties. All digitization has always been done in-house.
SE	We use WIPO-OCR for front file production (regarding patent documents). For backfile before 2018 we used a third party, to be able to manage the volume.
SG	No, we have not relied on a third party for digitization.
TZ	We are not relied on third party.
US	We have both in-house and third party to perform digitization. Third party was used to maximize efficient use of resources.

9.3. Does your Office use or did you use any IT tools, in particular AI tools, in the digitization process that you find particularly useful?

ST.3 Code	Response
AP	No AI tools yet. We used WIPOscan.
AU	We run custom image AI models to detect mathematical formulas and recognise chemical structures. We run them on the AWS SageMaker platform.
BA	We don't use AI, mostly is used Abbyy Fine Reader with support for our languages.
BG	Abby Finereader, no AI tools.
BH	No
BR	No
CA	Patent digitization is currently a manual process within Patent Branch, using various free software like IBM Data Camp.
CL	For now, no AI tools are used. Traditional computer tools are used, such as Kodak Capture Pro and ABBYY.
CN	During the digitization of patent application documents, CNIPA uses the data processing production line which is developed by third-party service providers.
CR	No specific AI-based tools based have been used. However, the OCR tool provided by WIPO is used to support information processing in patent procedures.
CZ	We currently do not employ AI-based digitization tools.
DE	The DPMA has not yet deployed AI-based tools within its productive digitization workflows. However, the Office is currently conducting evaluations of a locally hosted Large Language Model (LLM) with a view to enhancing automated document classification and structured data extraction from forms. Initial results are considered encouraging, and further assessment is ongoing.
DZ	So far, our office has not used any specific computer tools, including AI tools, as part of the document digitization process. Digitization operations are limited to converting paper documents into digital format for processing, archiving, and document management purposes, without the use of AI technologies or advanced automated data processing tools.
EA	Currently Paddle OCR shows most promising results.

EM	Manual processing of PDF documents by using Foxit PDF Editor. Comparison of documents / USBs by using WinMerge.
EP	OCR: - Abby Fine reader - Mistral OCR - DOCX: Aspose Yes, LLMs are starting to prove a viable option to other external or traditional digitization systems/services.
GB	None
GE	No
GQ	Not yet
HR	SIPO doesn't have AI tools for the digitization process.
HU	ABBYY Fine reader is used for OCRing PCT documentation.
IE	Not applicable
IL	Regarding industrial design collections: IT tools are implemented for the scanning of design documents submitted in paper. Regarding patent collections: IT tools are implemented for the scanning of patent documents submitted in paper, performing OCR to unsearchable PDF files, conversion of PDF files to HTML. In addition, integrated AI tools are also applied, for example, for the automated validation of XML files. Regarding trademark collections: IT tools are applied for the automatic conversion of DOCX files to PDF.
IN	Yes, documents collected from 1.1.2026 is OCRed using paddle OCR tool for text extraction and for backlog digitization and scanning of document ABBY Fine software tool is used.
JM	Adobe 2017. No AI tools to date.
KR	None at present.
LK	NIPO uses the WIPO IPAS platform and document management tools for managing digitized records and bibliographic data. The Office is also exploring the use of AI-assisted technologies for document processing, classification, indexing, and data extraction.
LT	No
LV	No

MD	We use ABBYY Recognition Server v 2.0 or MS Office built-in OCR. No AI tools.
NO	No
OM	The Office used scanners and the "Wiposcan" software in the digitization and archiving of applications, but we did not use AI tools.
PH	IPOPHL uses Mathpix to convert OCR'd patent documents per section (AB, DE, CL, DR, SL, whichever is applicable) to DOCX which is being uploaded in WIPO-IPAS and a module is being ran to transform DOCX to XML package with corresponding HTML for validation. Once validated, the XML package being rendered into XML ST. 96 compliant, otherwise, we re-do the conversion, uploading and validation process. This process was made operational through the assistance from WIPO-SG.
PL	The Polish Patent Office has been using EZD (Electronic Document Management) system since 2014. All scanned PDF documents are exported to this IT system. Our Office does not use Artificial Intelligence in the digitization of documents.
RS	We intend to use AI tools in future for OCR and digitalization.
RU	The Office does not currently use any AI tools in the digitization process, nor has it done so in the past.
SE	We use the tool WIPO-OCR.
SG	Yes, we utilise AI-enabled OCR capabilities (e.g., AWS Textract) to support document digitisation, including extraction of text, identification and removal of non-substantive elements (e.g., headers, footers, pagination, line numbers), and detection of figurative elements for separate image extraction.
TZ	Yes
US	AI tools are not being used at present.

9.4. 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.

ST.3 Code	Response
AP	We had challenges with old documents, which had to be photocopied first, as well as some document sizes.
AU	Main challenges for PCT Minimum Documentation conversion of documents to XML came from document formatting such as: • Chemical formulas within sentences (not on a separate line) • Heavily condensed text • Identification of math compared to inline text • Claim documents with paragraph markers in similar position to claim numbers • Claim documents with multiple numbered lists within a claim • Drawings that overlap (multiple to a page) • Drawings with figure reference not in alignment with the image • Descriptions with sequence lists included in the document • Special formatting characters used by some jurisdictions • Superfluous text provided with documents, e.g., cover pages • Tables over hundreds of pages • Tables with misaligned columns These were identified by testing. Some but not all were overcome or improved by OCR coding enhancements. These general document formatting issues and lack of standardisation diminish output quality, given all the possible edge case variations that cannot be accounted for in the build. The most challenging conversions involved math and chem formula detection, as there was no initial guarantee that a feasible solution could be found. We overcame these challenges by training our own image recognition and classification AI models for the purpose. Tables have also been a challenge. Textract handled them well initially, but performed worse after an update was done by AWS. We haven't found a solution for that and our table quality has decreased.
BA	Extracting drawings and chemical formulae from a patent application, OCR handwritten applications, tables, etc.
BG	No challenges

BH	Set plans for few years to shift from documents to digital documents through scanning and uploading procedures.
BR	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 OCRed 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 OCRed 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.
CA	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.
CL	We have faced both technical and operational problems. Operational problems: We have a contract for external storage of documents and files. This contract is limited in its capacity to provide documents or files on a weekly basis, which hinders their continuous and timely processing. Technical problems: Managing the size of digital documents (mainly in patents) and character recognition (chemical formulas and others) for the processing of digital documents (mainly in patents). The way to overcome these difficulties is through the allocation of resources, both human and financial, in the case of operational problems; for technical problems, the use of ad hoc software to improve the results obtained in digitization is recommended.
CN	During automatic recognition, errors may occur to those complex data elements, such as mathematic expressions, chemical formulae and complex tables, which need manual verification or workarounds.

CR	Interfaces have been developed to transfer documents received through the online filing system (WIPO File – WIPO) to the EDMS document management system, given that the Office uses a third-party system (ePower) digitization tool. Similarly, intermediate tables were implemented to enable integration between the EDMS system and physical intake managed through the IPAS system.
CZ	Patents: Producing ST.96-compliant patent documents required a significant upgrade of the IPO CZ information system, including the development of a custom claims editor that helps users prepare the underlying ST.96 XML. Trademarks, Industrial designs: The preparation of images, such as drawings and depictions, for publication was challenging, as it previously required manual standardization. This process is now largely handled by automatic conversion under user supervision. All IP rights: Processing paper filings requires OCR and/or manual entry. Processing electronically fillable PDF forms is also a challenge, particularly due to issues with the national character set and the continued use of old revisions of our forms by the public. This has been partially addressed by guiding users towards the IPO CZ electronic portal, where text can be entered directly into an HTML-based form.
DE	The DPMA has identified the following as the principal technical challenges encountered in the digitization process: • Quality requirements for patent documents: Patent documentation is subject to strict requirements regarding image fidelity, particularly with respect to black-and-white technical drawings, where minimal quality loss is essential. Meeting these standards consistently has presented challenges, particularly in the context of transitions between scanner hardware generations. • Non-conformant PDF submissions: A recurring technical issue involves PDF files received from external parties that do not render consistently across different viewer applications. Addressing such non-conformant files requires additional processing and quality assurance measures. • Legal requirements for the disposal of physical documents: Compliance with national legislation governing the conditions under which original paper documents may be discarded following digitization represents a procedural and organisational challenge, rather than a purely technical one. • Chemical formulae and OCR limitations: Chemical formulae are not currently subject to structured extraction. While OCR processing is applied, the accuracy of the resulting output for such content cannot be fully guaranteed. The potential impact of this limitation on AI-based document search and retrieval has not yet been comprehensively assessed and may warrant further evaluation in the future.

DZ	The main difficulties encountered in the digitization of documents concern the optimal use of digitized data and its integration into intellectual property management processes. In particular, the lack of advanced tools enabling the automation of certain technical tasks, such as the classification of trademarks according to the Vienna Classification using AI, is a significant challenge. Furthermore, the establishment of fully online filing procedures remains a major objective in order to reduce dependence on physical documents and improve the efficiency of file processing. The use of reliable and high-performing software enabling the viewing, analysis, and full management of files in digital format is also essential to avoid systematic reliance on paper archives.
EA	• Incorrect block markup (detection of images in text, including inline images) • Incorrect recognition of letters, digits, math and other special characters • Mixture of Latin and same looking Cyrillic letters • Incorrect word/sentence/paragraph breaks • Excessive merging of paragraphs • Unpredictable changes to document when using AI tools like PaddleOCR (hallucination effects): parts of text may be removed, some digits and characters (not present in the original document) may be added, replacement of Russian words with English ones.
EE	The biggest technical challenge is extracting drawings, mathematical formulas, and chemical structures from patent applications.
EM	Possible issues: Scanning quality for representations with specific colors or with insufficient quality. Mitigation: Periodical scanner and screen calibration; staff training; applicant to provide better quality representations.
EP	Applications documents: properly reflecting the original structure (paragraphs numbering, claims numbering, lists, ...) in the XML document is a challenge because many different formats/layouts of documents received. We are still working on improving this. For the extraction of citations, the non-patent literature is more difficult to extract as generally not following any standard (contrary to patent literature that comprises clear patent numbers with country code etc.). In both cases we are working on it in training an LLM. For patent applications, extracting information from complex structures such as tables, mathematical formulas, chemical compounds, images embedded in text is the most challenging part. They can either be kept as image or, with more recent developments, structured and extracted as rich information.
ES	For patents, extracting drawings or chemical formulas. In old documents, the condition of the paper. In general, considering data protected under Spanish and European data protection legislation before publication.

GB	None
GE	Not applicable
GQ	We have found it difficult to capture chemical formulas and diagrams in patent applications; we do not attempt to improve their rendering within the detailed description of the applications.
HR	Bad print, chemical formulas, tables, handwriting.
HU	Data quality over OCRing process from poor quality documentation. Also extracting drawings and chemical formulas. Poor OCRed data lead to incorrect xml format that does not meet the prescribed requirements. To overcome these issues, we carry out manual corrections.
IE	The main challenges in digitising documents are limited resources, as many documents require manual processing, and the need for OCR processing. At present, WIPO provides OCR services to IPOI.
IL	Migrating our patent bibliographic XML data to WIPO Standard ST.96 (Version 9) was highly resource-intensive. However, by integrating an AI tool to handle the transition, we successfully accelerated the workflow, completing the entire migration efficiently and ahead of schedule with minimal manual intervention.
IN	The patent documents being techno-legal in nature contain a lot of mathematical and chemical formulas along with tabulated data and drawings. Hence, IPO India uses paddleOCR v3 structure (an open-source OCR library) for identification of drawings & chemical formulas.
JM	Unable to readily extract Chemical formulas from a patent application. We are presently using Adobe 2017 as the main tool for our digitization.
KR	Patent specifications contain various objects such as drawings, chemical formulas, mathematical expressions, and component diagrams. However, current OCR technology can extract only text, requiring operators to manually digitize non-text objects by capturing and inserting them as images. To improve efficiency, document analysis software is planned to be introduced to convert documents into XML format, followed by post-editing, thereby reducing the time required for the digitization input process.
LK	Digitization of large volumes of historical paper records. Poor quality or deteriorated documents affecting scanning and OCR accuracy. Manual verification and correction of extracted bibliographic data. Ensuring consistency and completeness of metadata. Limited technical and human resources for large-scale digitization projects. Integration of legacy records into modern electronic systems. These challenges have been addressed through quality-control procedures, staff training, phased digitization activities, and technical support provided through WIPO digitization initiatives.

LT	Extracting chemical formulas or drawings. We are reaching the applicant or perform it manually.
LV	One of the main challenges was the migration and digitization of historical paper-based collections, including document scanning, metadata extraction, indexing and import into operational back-office systems. Ensuring data quality, completeness of metadata and successful integration of digitized records into existing IP management systems required significant effort.
MD	When patent attorneys provide translations of European patent documents, they sometimes "translate" the drawings by adding text boxes in front of the original image in MSWord documents. Addressing this issue requires additional manual processing, which involves converting the DOCX file to PDF, then to PNG format, and subsequently cropping the relevant drawings.
NO	• Image compression/Size • OCR-level where print quality is low on images • Formula detection (for image capture) • Sectioning vs page • Page numbering and removal of unwanted elements
OM	One of the main technical challenges experienced during the digitization process was ensuring the quality and accuracy of scanned documents, particularly older paper files and applications containing unclear text, handwritten notes, stamps, or complex formatting. In some cases, document sizes and layouts varied, which affected the consistency of scanning and indexing. These challenges were addressed by using high-quality scanners, reviewing scanned documents manually when necessary, and improving scanning settings. Staff members were also trained to verify and correct archived records to ensure accuracy and completeness.
PH	1. Extracting drawings, tables, images or chemical formulae from the OCREd to DOCX and XML. 2. Some of the patent documents were submitted as an image being paste to a word file and being saved as pdf, thus, there is a need to OCR them first before conversion. 3. Existence of numbers or bullets in the claim, which is being recognized as claim number, thus claim number is inconsistent; i.e. Claim 1....comprising....and selected from..... 1. xxxxx.... 2. yyyy.....

PL	When scanning a document consisting of more than 500 pages, there is a major problem with exporting it to the EZD Electronic Document Management system. In such cases, the scanned document must be split into several smaller files. • For drawings in formats other than A4, we seek assistance from another department (printing division). • For multi-page, stapled notarial deeds, certificates, and contracts, we seek assistance from another department (printing division). The above challenges occur with regard to all IP rights.
RS	In digitalization of the old patent documentation, the problem was bad paper source quality, as well as small scan resolution of scanned documents which prevents high OCR quality. Manual control needed, challenges still exist since we need to improve quality of backlog patent documentation.
RU	The main technical challenges have included accurate recognition of formulas and tables, as well as processing lengthy documents. To ensure data quality and accuracy, we employ manual review and editing as part of our quality control process.
SE	One of our challenges is e.g., that all patent files/dossiers are saved as an image in our internal system. We have to reconvert the images to textformat and XML before we disseminate the publications. We do manual corrections of the OCR text if necessary in the WIPO-OCR-tool.
SG	During the digitization of patent specifications, we have encountered several recurring challenges: • The system is unable to properly separate paragraphs or individual claims. • Data extracted from tables is not rendered correctly and instead appears as continuous text. • Formulae embedded in documents are not extracted accurately in text format. • Text appearing in the margins of documents (e.g., PCT publication number, page number) is sometimes not removed.
US	• Limited requirements for application submissions including for image quality, and for standardized formatting requirements. • In ref to chemical formula, image quality may require manual intervention, • For WordMath, converted to MathML - otherwise manual recreation of math formula. • For sequence listings- based on size may be include in the publication or published separately.

10. 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.

ST.3 Code	Response
AP	Not at the moment
AU	Not applicable at this stage
AZ	The Agency began using AI provided by the EAPO.
BA	Not such a plan in the near future.
BG	Currently, no changes to the digitization process are planned, as no need has been identified.
BH	Thinking of on using AI tools if possible, to assist with workloads.
BR	- 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.
CA	CIPO is evaluating a solution 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 extracted 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.

CL	There are no ongoing projects, but the best way to address the backfile is being evaluated, particularly regarding the PCT minimum documentation that has not yet been digitized.
CN	CNIPA will continue promoting the application of new technologies in data processing, so as to improve its efficiency and quality.
CR	Currently, the implementation of digitization systems, specifically those based on AI, is not being considered.
CZ	IPO CZ is assessing the potential use of AI tools developed within the framework of cooperation agreements with the European Union Intellectual Property Office and the European Patent Office. These tools are being considered for their potential to support further digitization, automate selected processing tasks, improve the handling of IP-related data, and enhance the efficiency and user-friendliness of Office services. In the coming year, the Office plans to implement a national system called the Registry of Representation, enabling individuals and businesses to digitally create, store and manage powers of attorney and other authorizations.
DE	The German Patent and Trade Mark Office (DPMA) is currently considering a number of possible future digitization measures, some of which may involve the use of artificial intelligence (AI) tools, subject to further evaluation. Among the measures under consideration is the development of a new customer portal based on modern web technologies, which could replace the existing DPMAdirektPro and DPMAdirektWeb systems. Such a portal would be intended to enable applicants to file legally valid applications for industrial property rights online and to submit supplementary electronic documents relating to such applications. At this stage, no AI components are foreseen in this context; however, given its potential relevance to the Office's digital service offering, it is mentioned here for completeness. With regard to AI, the DPMA is exploring the possible future integration of AI tools that could provide support for search activities and document drafting. Such tools, if implemented, would be aimed at increasing efficiency and improving data quality. As any related procurement and contractual arrangements remain at a preliminary stage, no specific provider or product can be named at this time. In this connection, the Office has recently acquired licenses for OpenShift AI, which is intended, in principle, to serve as a platform for evaluating possible AI applications; no operational use has commenced to date. The DPMA also makes use, under certain conditions, of the publicly available translation tool DeepL, including its writing assistant function, for official purposes. In addition, the DPMA is following developments relating to large language models (LLMs) with interest. Given the variety of available LLM services, each typically requiring its own application programming interface (API), consideration is being given to the possible definition of a unified interface that could connect to multiple LLM providers, with a view to avoiding separate integrations for each service. In this context, the tool LiteLLM is among the options currently being looked at, alongside other comparable products; no decision in this regard has been made. Beyond the Office's own considerations, it may also be noted that the German federal IT service center (ITZ-Bund) has introduced a central AI platform, known as KIPITZ, intended to enable federal authorities to make secure and sovereign use of generative AI in their daily

	work. This platform is envisaged to support tasks such as document summarization, document-based dialogue ("chat with documents"), translation, and knowledge search. According to current planning, this platform could potentially become available to the DPMA around October 2026.
EA	• Automatic (AI-based) detection of the «complexity» and "type" of document • Depending on the complexity and type we plan to apply a proper tool or several tools (classic OCR, AI-OCR (now PaddleOCR is most promising), text layer extraction from PDF, math formulae detection using machine learning methods etc.) • We plan to continue testing all emerging AI OCR Tools as the technology is progressing rapidly.
EE	We would like to use AI-assisted OCR for processing and extracting information from documents.
EM	EUIPO is following a paperless strategy and is abandoning paper-based filings.
EP	Further work with Mistral OCR are ongoing. Our goal is to reach publication quality (99.995%). We also have activities on internalising the Form OCR and the Citation extraction. Exploring the potential of additional AI solutions/models is on our roadmap of further developments. Those solutions will improve quality and timeliness during the patent grant process, also allowing for increased usage of fully digital data end-to-end.
ES	We currently use the WIPO-OCR tool, and we are evaluating the use of the latest version of this tool, which incorporates AI.
GB	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)
GE	Not applicable
GQ	We work with standard digitization using computers, without using AI.
HR	We are currently analyzing the possibility of introducing an EPO/WIPO AI tool for OCR.
HU	AI OCRing tool developed by EPO is on our list for future improvement once available and fulfils expectations.
IE	Tools for OCR processing of documents would be a key area of interest, particularly AI-based OCR solutions. These tools could help improve the accuracy, efficiency, and searchability of digitised documents, especially for patent applications containing technical text, drawings, symbols, and chemical formulae.

IL	The ILPO is considering publishing all patent collection documents in XML format utilizing advanced AI tools.
IN	Yes, exploring the possibilities of using enterprise level proprietary OCR software for digitization of the specification and its corresponding Database insertion.
JM	Optical Character Recognition (OCR) like Microsoft Azure OCR, Semantic Search & Classification such AI classifiers that assign IPC codes automatically and assist in organizing digitized files and linking them to global patent databases enhancing JIPO's search capabilities. We would also seek AI Tools that will enable us to organise our data in accordance with WIPO's Standards Compliance ST.36 and ST.96 schemas, use of these applications in the future are dependent on the governments policy on cloud storage.
KR	AI-powered OCR technology enables the extraction of key-value data from various document templates, facilitating data reuse. In addition, document analysis software can analyze document structures and convert them into XML format.
LK	NIPO plans to continue the digitization of historical IP records and improve the quality of digital collections. The Office is interested in AI-based OCR solutions, intelligent document classification, metadata extraction, automated indexing, and quality-control tools. These technologies are expected to improve efficiency, accuracy, searchability, and access to intellectual property information while reducing manual processing effort.
LV	No
MX	AI use has not been considered.
MZ	Not applicable
NO	Started project on patent publication production workflow, to enable native XML ST.96 publications. Will evaluate tools, solutions from others, and are open to use own AI infrastructure (own network).
NZ	It is mandatory to submit documents in digital form and digitisation is not needed.
OM	The Office is seeking to further develop its systems and will, in the future, consider adopting advanced technological tools that help carry out examination processes with greater accuracy and efficiency. The Office also identifies the need for secure electronic document exchange and communication technologies. This includes systems that enable the secure sending and receiving of documents and information between the Office and applicants in an efficient and reliable manner.

PE	Although the operational processes for digitizing incoming applications and documents are institutionally centralized through the documentation office (both physical and virtual), there is no dedicated unit for the digitization and indexing of patent documents from the point of application filing. This task is currently performed by staff from the Directorate of Inventions and New Technologies once the application is published. Plans are being considered to establish a specialized unit, along with tools such as AI-assisted advanced OCR, automated extraction of bibliographic metadata, natural language processing models and automatic classification.
PH	Yes, the Office is considering future digitization, especially the backfiles. As of now, we consider using the digitization process as cited at item 12.3, since we are using WIPO-IPAS and support from WIPO-SG. We may also consider accepting DOCX or XML format of patent documents upon filing. Further, we are considering additional personnel that would speed up the digitization of backfile records.
PL	The Polish Patent Office is working toward a transition to exclusively electronic correspondence. Currently, the Polish Patent Office is receiving fewer and fewer paper documents compared to previous years.
PY	There are plans to digitize historical documents (backfile) and to apply a process for incoming documents (frontfile). We are currently exploring the possibility of designing a project for the digitization of industrial property files at this level.
RO	Not applicable (for the moment)
RS	We are still discussing this matter, since we need support of specialized AI expert in this field. Possible support of EPO on this matter (OCR customized Mistral AI model).
RU	No, our Office is not considering any future digitization activities at this time.
SE	We are in the patent process looking into the future possibilities to handle for example fulltext documents, DOCX and colour drawings. We are following the development of WIPO standards regarding all IP rights. It would be of interest to be able to use AI support where possible. A new upcoming Design Directive will in the future allow new features, for example the protection of dynamic designs, including animated and moving design representations.
SG	Building on current OCR-based digitisation and layout structuring capabilities, section identification (e.g., title, abstract, claims) is presently rule-based (e.g., wildcard matching). We are exploring GenAI to enhance document structuring, as rule-based approaches are less reliable for complex documents. This results in more accurate section and claim identification, which enables structured outputs to be integrated into downstream workflows (e.g., claim-level analysis).

TZ	If our Office is considering future digitization activities, we intend to explore the use of Artificial Intelligence (AI) and related digital tools to improve the efficiency, accuracy, and accessibility of intellectual property records and services. The tools under consideration include: • Optical Character Recognition (OCR) Tools – To convert scanned documents into searchable and editable digital text, thereby facilitating electronic records management and retrieval. • AI-Powered Document Classification and Indexing Tools – To automatically categorize and index documents, reducing manual effort and improving the organization of digital archives. • Natural Language Processing (NLP) Tools – To assist in extracting key information from intellectual property documents, enabling faster searches and analysis of records. • AI-Based Trademark and Patent Search Tools – To enhance prior rights searches by identifying similar trademarks, patents, and industrial designs more efficiently and accurately. • Machine Learning-Based Data Analytics Tools – To analyze filing trends, user behavior, and operational performance, supporting evidence-based decision-making and policy development. • AI Chatbots and Virtual Assistants – To provide automated responses to routine customer inquiries, improving service delivery and reducing response times.
US	Future activities are currently under discussion in the research and evaluation phase.

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

ST.3 Code	Response
AP	We require capacity building on AI tools that can be used for digitization
AU	Key capabilities involved in the digitisation for the PCT Minimum Documentation work: • Data science – AI/ML, image processing, NLP • Software engineering – Java in our case • Cloud infrastructure and DevOps – AWS in our case
BA	We have a lack of human resources in this field.
BG	At present, no additional feedback is available regarding capacity building for the digitization of IP document collections, as no specific needs or challenges have been identified.
CA	Not applicable
CL	INAPI has a team that works permanently on the digitization of IP documents, and since 2025 it has begun converting granted patent documents into XML format for PCT minimum documentation.
CR	It is considered necessary to improve the functioning of the OCR tool provided by WIPO, in order to maximize the processing, response, and information transmission capabilities of the Offices, thereby contributing to more efficient management.
EM	EUIPO is investigating on the use of AI tools to obtain efficiency gains in different areas.
GB	None
GQ	Difficulties in procuring resources for the digitization of the IP document collection and a lack of public awareness.
IN	Capacity building in developing ability for data capture conforming to WIPO standards in IPO's online filing systems, by using XML filing or docx filing implementation.
IN	Continued support and capacity building in OCR technology, AI-based digitisation tools, and staff training would help improve the efficiency and quality of digitising IP document collections. Collaboration with WIPO has also been beneficial in supporting these activities
JM	None

KR	Member Offices should share practical examples of adopting AI and other digital technologies in their digitalization efforts.
LK	NIPO would welcome additional capacity-building activities from WIPO, including training on digitization standards, WIPO Standards ST.96 and related XML technologies, OCR and AI-based document processing, metadata management, quality assurance, and best practices for large-scale digitization projects. Technical guidance, knowledge-sharing opportunities, and practical training would support the sustainable development of digitization capabilities within the Office.
LV	The Office would benefit from support in carrying out a structured gap analysis of its current digitization and dissemination processes against WIPO Standards (including ST.36, ST.66, ST.86 and ST.96) and recommended good practices.
MX	The IP collection of the Mexican Office is currently being processed using the ST.36 and ST.37 standards for text extraction and XML generation.
MZ	No feedback
OM	Capacity building for the digitization of IP document collections should continue to focus on strengthening staff technical skills and improving familiarity with digital tools and document management systems
PE	We promote staff participation in ongoing training programs (in coordination with national and international organizations) focused on digital public management, data analytics and the automation of administrative processes. We also highlight the importance of technical assistance from WIPO and other regional offices for the exchange of best practices on data standards (such as the transition to Standard ST.96), which enables office staff to understand the value of digitization oriented toward data structuring rather than mere image storage.
PH	More training for our in-house digitization team, especially on document validation and quality checking, would be beneficial. Additional personnel would also help speed up the digitization of backfile records and reduce backlogs. WIPO may also consider supporting a backfile digitization project, with the IP Office focusing on validating and checking the quality of the digitized documents. This would help make the digitization process faster and more efficient. Continued investment in training, technology, and knowledge sharing will help support the long-term success of IPO Philippines' digitization efforts.
PL	This question is unclear. Does it refer to existing IP document collections? Or to the creation of new IP document collections? Please explain in more detail with regard to capacity building for the digitization of IP document collections. Perhaps it would be useful if you could give some examples.
PY	Our institution has sufficient and robust technological infrastructure, and we have trained staff. Perhaps we do not have a sufficient number of staff, but we can guarantee a solid level of operations and immediate responsiveness to future implementation.

RO	Not applicable (for the moment)
RS	We lack financial resources as well as methodological approach for completion of patent documentation digitalization.
SG	None
TZ	Capacity building for the digitization of IP document collections should focus on enhancing staff skills in digital records management, document preservation, metadata creation, OCR and AI technologies, cybersecurity, and quality control, while also promoting knowledge sharing, technical cooperation, and access to modern digitization tools and best practices to ensure efficient, secure, and sustainable management of IP information.
US	Suggestion is for Offices to develop more structured requirements for filing, for example consider DOCX as a requirement for filing.

[End of document]