

Committee on WIPO Standards (CWS)

Fourteenth Session
Geneva, November 23 to 27, 2026

CONVERSION OF IP DOCUMENTATION INTO XML FORMAT

Document prepared by the Digital Transformation Task Force Leader and the Secretariat

SUMMARY

1. Under Task No. 62, the Digital Transformation Task Force has been working to establish common functional requirements and a harmonized output profile for Microsoft Word Open XML Document (DOCX) to eXtensible Markup Language (XML) conversion solutions (DOCX2XML converter) of intellectual property (IP) documentation. Although various IP offices currently employ differing conversion tools, treatment rules for complex features (such as tables and equations), and legal determinations of the authoritative file, developing a shared baseline will ensure XML interoperability, consistent renditions, and reliable data exchange without harmonizing Offices' individual legal or operational frameworks. To support this effort and explore the development or reuse of a common conversion tool, the Task Force Leader and the Secretariat proposes conducting a structured consultation with interested IP offices to collect use cases, technical requirements, and test data, with the objective of presenting a revised DOCX2XML Converter Functional Specification Document for consideration at the fifteenth session of the Committee on WIPO Standards (CWS).

BACKGROUND

2. At its eleventh session, the CWS approved a revision to the description of Task No. 62 as follows:

“Review paper or image-based WIPO Standards in view of electronic filing and publication and exchange of IP documentation and propose revisions of those Standards or new recommendations if needed; and prepare a proposal for the recommendation on a common requirements specification for a DOCX to XML (DOCX2XML) converter.”

(See paragraph 59 of document CWS/11/28.)

3. At its thirteenth session, the CWS noted a working draft of DOCX2XML converter functional specification document which was reproduced in the Annex to the document [CWS/13/12](#), presented by the Digital Transformation Task Force. In addition to the development of the common functional specification, several delegations expressed interest in the development of a common DOCX2XML converter, implementing the agreed functional specification. The CWS noted that several Offices, particularly those with limited technical resources, would benefit from this common solution. (See paragraphs 74 and 75 of document CWS/13/33.)

DOCX2XML PRACTICES AND REQUIREMENTS FOR COMMON OUTPUT

DOCX2XML converter functional specification document

4. Following the agreement reached at the tenth session of the CWS, the Digital Transformation Task Force began preparing a document for a DOCX2XML converter functional specification. In response to subsequent requests from several Offices for a common DOCX2XML converter, the Task Force broadened its discussion to consider both the possible development or reuse of such a tool, and the alignment of output produced by converters used by several Offices.

5. This broader approach is intended to ensure that, even when Offices continue to use different conversion tools, the resulting XML is harmonized and interoperable. The functional specification document is therefore not being submitted for final consideration at this session of the CWS, in order to give the Task Force additional time to incorporate requirements for a common tool and a harmonized output requirement, taking into account existing solutions and Office practices. The Digital Transformation Task Force invites the CWS to encourage its Members to review and provide feedback to the working draft of the *DOCX2XML Converter Functional Specification Document*, which was reproduced in the Annex to document [CWS/13/12](#).

6. The consideration of a common DOCX2XML converter does not replace the need for common functional and output requirements. Rather, the Task Force intends to establish a shared baseline that could be implemented by a common solution or by converters already operated by individual Offices. This would allow Offices to retain DOCX2XML converter suited to their legal, technical and operational environments while producing XML that follows an agreed structure and can be exchanged, validated and reused consistently across Office systems.

7. Collaboration among Offices will be essential to this work. Offices that have already developed or implemented DOCX2XML converters could contribute technical knowledge, conversion rules, representative test cases and lessons learned, while Offices considering new solutions could provide their operational requirements and expected use cases. These collaborative efforts would help the Task Force identify reusable components, compare outputs against a common set of examples and progressively align conversion results, while avoiding duplication of effort and taking account of different levels of technical capacity.

Existing DOCX2XML conversion practices

8. The Digital Transformation Task Force identified DOCX2XML converters used by the China National Intellectual Property Administration (CNIPA), the Ministry of Intellectual Property of the Republic of Korea (MOIP), the United States Patent and Trademark Office (USPTO), the European Patent Office (EPO), and the International Bureau for Patent Cooperation Treaty (PCT) applications. The review of the information collected by the Task Force showed that converters may treat the same DOCX features differently and produce different outputs. For example, equations may be converted into MathML or images; complex tables may be represented in XML, rejected or converted into images; and unsupported features may be removed, reported as warnings or treated as errors.

9. In addition to the work done by the Digital Transformation Task Force and following requests at the thirteenth session of the CWS for a common DOCX2XML conversion, the Secretariat held an online meeting for interested Offices in September 2026. The meeting presented solutions and Offices practices for converting DOCX patent applications into XML documents and explored opportunities for cooperation. The International Bureau reported that the PCT DOCX Converter has operated for more than ten years. Integrated into PCT filing and available as a standalone tool and web service, it generates XML documents compliant with WIPO Standard ST.36, referenced images and a Portable Document Format (PDF) preview, supports multiple languages and scripts, and provides numbering and validation messages. The National Institute of Industrial Property of France (INPI France) reported using the PCT converter as a web service for national patent applications. Implementation required piloting, guidance, training and specialized support, while differences in underlying DOCX structures continue to create new conversion cases. The USPTO presented that its in-house DOCX2XML converter used for converting patent applications into XML documents and other downstream formats, performs pre-submission validation, and identifies errors and warnings. It processes common document objects and flags content that cannot be mapped reliably.

10. Offices reported differing approaches to the status and use of the converted XML output. Under the PCT, the XML output constitutes the legal copy. INPI France retains the DOCX file for examination purposes and uses the XML and PDF outputs for dissemination and publication. The USPTO regards a corrected DOCX file as the legal copy and generates XML and TIFF files for downstream processing. Despite these differences, a common technical profile could still define interoperable XML structures, validation results and renditions.

11. At the meeting, several elements that could form the basis of harmonized conversion outputs were identified: consistent conversion and rendering across systems; clear validation and preview mechanisms; preservation of submitted content; identification of the legal or reference copy; and compliance with WIPO Standard ST.36 or ST.96. Agreement on these elements would facilitate cooperation while accommodating different national implementations and practice.

Rationale, scope and requirements for common DOCX2XML outputs

12. The findings set out above demonstrate the value of closer cooperation among Offices to define common, harmonized DOCX2XML outputs and practices. Although Offices may continue to use different conversion tools and assign different legal or operational roles to the source and converted documents, agreement on a common output profile would improve the consistency, interoperability and reuse of the resulting XML documents across the systems and workflows of Offices.

13. As an initial scope, cooperation should focus on the patent applications, including descriptions, claims, abstracts and drawings. Consistent structured XML documents for these contents would support filing and examination processes, the preparation of PCT minimum documentation, digitization, publication, dissemination, data exchange and reuse in subsequent workflows. The scope could later be extended to other text-based IP documents where Offices identify a clear operational benefit.

14. Within this scope, the common approach should define agreed use cases and a minimum structure for patent application bodies, including the sections and content that converters must recognize and any optional or Office-specific elements. It should also specify the expected treatment of complex content, such as tables, equations, chemical structures, images and embedded objects. For other DOCX features, including hidden text, comments, headers, footers, content controls, Object Linking and Embedding (OLE object), tracked changes, shapes, hyperlinks, bookmarks, watermarks and charts, the functional specification should indicate whether each feature is to be retained, removed, converted, reported as a warning or treated as an error. Predictable treatment of such content is essential for applicants, examination processes and data dissemination and exchange.

15. The requirements should further address variations in DOCX structures and drafting practices; recognition of document sections; generation of valid XML documents that meet operational needs; handling of corrections and amendments; processing of large files and peak filing volumes; and arrangements for maintenance, guidance and support. As artificial intelligence (AI)-assisted drafting may introduce unsupported or unexpected structures, this deserves careful consideration as well.

16. The requirements should also accommodate the different roles that Offices assign to source documents and converted outputs. Accordingly, the common requirements should identify the authoritative or reference copy, the files to be retained, the applicant preview and discrepancy-resolution process, and the formats to be used for examination, publication, dissemination and exchange, without prescribing the legal status assigned by an Office.

PROPOSED FOLLOW-UP ACTIONS

Consultation on common requirements

17. Building on the rationale, scope and requirements outlined above, the Secretariat and the Digital Transformation Task Force Leader propose a structured consultation with interested Offices. The consultation would enable the Digital Transformation Task Force to validate the proposed scope, identify common use cases and define common requirements for a harmonized DOCX2XML output, while taking into account existing conversion solutions, national practices and different levels of technical capacity.

18. It is proposed that the Digital Transformation Task Force address, but not limited to, the following questions:

- (a) Why does an Office need a DOCX2XML converter, and for which processes will the output XML be used, including filing, examination, PCT minimum documentation, digitization, exchange, publication or dissemination?
- (b) Which IP rights and document types are in scope: patent application bodies only, or identified trademark and industrial design documents as well?
- (c) What minimum patent application-body structure and section headings should be recognized, and what Office-specific adaptations are required?
- (d) Which DOCX-producing applications, languages, scripts, fonts, images, tables, formulae, chemical structures and other objects must be supported?
- (e) How should unsupported content, validation failures, warnings, corrections and amendments be handled?

- (f) Which version of document is official or authoritative: the original DOCX document, corrected DOCX document, converted XML document, rendered PDF document or another representation?
- (g) What limits apply to file size, image size, number of pages, processing time and peak transaction volume?
- (h) What lessons have been learned regarding testing, training, applicant support, examiner review, maintenance and AI-prepared applications?
- (i) Can an existing DOCX2XML converters be reused, and if so, what deployment, confidentiality, governance and Office-specific adaptation are required?
- (j) What common XML output and PDF or HTML rendition requirements would enable interoperability even where Offices continue to use different converters?

19. To support this consultation, it is proposed that Offices be invited to share information on their requirements for the converter, conversion workflows, functional requirements, implementation experience, test cases and expected outputs, as well as any existing tools or components that could be reused or adapted. The Task Force would consolidate these contributions to identify areas of using conversion tools, justified Office-specific differences and the minimum requirements needed to achieve interoperable XML output and consistent renditions across different DOCX2XML converters.

Revision of the DOCX2XML converter functional specification document

20. On the basis of the consultation and the consolidated contributions from Offices, the Digital Transformation Task Force will revise the working draft of the *DOCX2XML Converter Functional Specification Document* to incorporate the agreed use cases and common requirements, including a minimum interoperable output profile.

21. The revised document will be presented to the CWS for consideration at its fifteenth session.

22. *The CWS is invited to:*

(a) *note the content, including existing DOCX2XML converters, Office conversion practices and the rationale for harmonizing conversion outputs, as set out in the present document;*

(b) *review and comment on the working draft of the DOCX2XML Converter Functional Specification Document reproduced in the Annex to the document [CWS/13/12](#) as indicated in paragraph 5 above;*

(c) *request the Digital Transformation Task Force to consult interested Offices on common use cases, scope and requirements for a harmonized and interoperable DOCX2XML output as indicated in paragraphs 17 and 18 above;*

(d) invite its Members to support the consultation by sharing their conversion workflows, functional requirements, implementation experience, test cases and expected outputs and, where feasible, information on existing tools or components that could be reused or adapted as indicated in paragraph 19 above;

(e) request the Digital Transformation Task Force to update the working draft of DOCX2XML Converter Functional Specification Document on the basis of comments received and the results of the consultation as indicated in paragraph 20 above; and

(f) request the Digital Transformation Task Force to present the revised DOCX2XML Converter Functional Specification Document, together with the outcome of the consultation and cooperation among Offices, for consideration by the CWS at its fifteenth session as indicated in paragraph 21 above.

[End of document]