

## **Patent Cooperation Treaty (PCT) Working Group**

**Fifteenth Session**  
**Geneva, October 3 to 7, 2022**

### **PROCESSING INTERNATIONAL APPLICATIONS IN FULL TEXT FORMAT**

*Document prepared by the International Bureau*

#### **SUMMARY**

1. This document provides an update on the proposal to move towards full text processing of international applications presented at the fourteenth session of PCT Working Group (see document PCT/WG/14/8). In particular, the document discusses the principles and requirements for full text processing, covering the preparation of international applications in DOCX format, the submission, processing and tracking of subsequent changes to the international application, and the storage of original documents as a safeguard to any subsequent conversion errors.

#### **BACKGROUND**

##### **XML FILING**

2. The original plans for PCT electronic filing, embodied in Annex F of the PCT Administrative Instructions as first promulgated (see PCT Gazette Special Issue No. S 04/2001, dated December 27, 2001), envisaged that applications would normally be filed and processed in XML format. This was (and remains) the requirement of the “basic common standard”, which is intended to be accepted by all receiving Offices. Other formats, such as PDF were allowed for essentially as transitional arrangements on the way to the goal of well-structured, full text document processing.

3. However, despite offering fee incentives, XML filing has only represented a majority of international applications filed in the receiving Offices of Japan (RO/JP) and the Republic of Korea (RO/KR), where PDF filing is not permitted. At other receiving Offices, most applicants see PDF as a more convenient and reliable filing medium than XML. This is despite the necessity to perform an OCR of the PDF file for any application entering the national phase at

Offices that typeset their publications (all of the highest volume designated Offices), with the attendant risk of errors and loss of information that is dependent on layout or other typographical options not available in the typeset format.

4. XML filings at the receiving Office of China (RO/CN) have never risen above 20 per cent of the total at that Office, but the Office has recently decided to take advantage of new Section 705<sup>ter</sup> of the Administrative Instructions under the PCT to convert all PDF and paper filings into XML format for processing.

5. ePCT Filing supports XML filing both by directly uploading WIPO Standard ST.36 XML and referenced image files prepared using an external conversion process and by uploading a DOCX file, which is then converted to WIPO Standard ST.36 XML and the results shown to the applicant prior to filing. In all cases where a DOCX file is uploaded for conversion and in most cases where an external converter is used, the original DOCX file is provided as a pre-conversion file under Section 706 of the Administrative Instructions. However, excluding RO/KR (where PDF filing is not permitted), only 2 per cent of applications filed using ePCT Filing are XML filings.

6. Further to this, various Offices in their national capacity have moved towards promoting or requiring XML filing, typically by means of the applicant submitting the application body in Office Open XML (DOCX) format, which is then converted to XML format, either WIPO Standard ST.36 or ST.96.

#### XML PROCESSING

7. It has been possible to file international applications in XML format since 2003, but no standards have been agreed on for subsequent processing activities. If changes<sup>1</sup> are required, the application is generally published as if it had been a PDF filing, with replacement sheets created as needed according to the position of the changes.

8. Discussions on the subject of text processing have taken place in earlier PCT meetings. Notably in 2009, the Meeting of International Authorities under the PCT, at its sixteenth session, discussed a proposal by the Japan Patent Office (JPO) for paragraph amendment of PCT applications (document PCT/MIA/16/14). Paragraphs 98 to 101 of the Report of the session (document PCT/MIA/16/15) summarize these discussions. Paragraph 100 of the Report states that "Authorities recognized the need for a paragraph replacement system, they were not yet in a position to agree on the required details. In particular, one Authority was concerned by a number of details relating to the numbering system and arrangements on how changes other than one-to-one replacements should be made ...". In the absence of consensus, the proposal was not elaborated further.

9. In 2017, the European Patent Office submitted a Proposal for Change (PCT/EF/PFC 17/005), proposing to add DOCX as a possible filing format for application bodies specified in Annex F of the PCT Administrative Instructions and PNG as an accepted image file format. There was a good degree of support for this in principle, but no agreement on a consistent approach for the legal status and processing arrangements for the DOCX file itself and any views of the application body created from it, whether directly or via a conversion to ST.36 XML. Consequently, the proposal remains pending.

10. The issues have been discussed in general terms in a variety of other meetings. Most recently, the Working Group, at its fourteenth session in June 2021 discussed a document on the processing of international applications in full text format (see document PCT/WG/14/8). This document included an outline of the transition, over recent years, of the PCT System from

---

<sup>1</sup> In most administrative respects, corrections, incorporations by reference, rectifications and amendments are essentially the same process. This document refers to such actions collectively as "changes".

paper to electronic filing, provided an explanation of the benefits of full text processing and provided an overview of the filing of international applications in full text format at that time.

11. In that document, the International Bureau proposed to revise Annex F of the Administrative Instructions to allow the international phase processing of the application body in XML full text format as filed by the applicant or processed by Offices, which in turn would enable the transition to full text international publication. Furthermore, to encourage more filing of full text applications, the International Bureau proposed to make its DOCX converter and full text comparison tools available to all IP Offices through both demonstration web pages and web services, which would serve as a reference for implementation of processing applications filed in DOCX format and ensure consistency in the conversion from DOCX into full text XML.

12. In general, the electronic filing of international applications has largely enabled the transmission of the request form bibliographic information as part of the record copy in XML format. This document, where it mentions processing in XML format, is therefore concerned with the processing of the specification, the description, claims, abstract and drawings, in XML format. This will enable further automation in international publication and improvements in the quality of the full text content made available.

13. The JPO and the Korean Intellectual Property Office (KIPO) have shown that applicants generally have little difficulty with XML filing. However, no Office has persuaded applicants to file in XML format through choice or small fee incentives. Where permitted, PDF is always preferred. Among other benefits, PDF usually allows an exact, page-based representation of what is seen in the word processor before the PDF document is “printed”<sup>2</sup>. Moreover, the exact appearance of the filed application is highly likely to be visible in the event of litigation many years later. By contrast, DOCX has no standard rendering or WIPO Standard XML and application bodies are typically rendered to page views with no relationship to the views that had been seen in the original DOCX.

14. There are several different arrangements that Offices have in place for DOCX filing and conversion. These arrangements:

- (a) use different document conversion tools;
- (b) provide different levels of visibility of the conversion processes;
- (c) have different legal statuses for the DOCX files and converted XML files; and
- (d) have different conditions and time limits for identifying and correcting any conversion errors that exist.

15. Such differences risk introducing confusion and reducing confidence in the goals and arrangements for full text processing, especially in the context of the PCT where the application documents may be processed by several Offices in the international phase and then, where no translation is required, may form the basis of national phase processing in other Offices.

### **CURRENT STATUS OF FULL TEXT PROCESSING**

16. Up to now, applicants have been using a variety of tools and systems to create XML application bodies for filing international applications. While some national Offices have been processing these applications using WIPO Standard ST.36 format for a number of years, others, more recently, have started processing national applications using WIPO Standard ST.96. Furthermore, a number of DOCX to XML converters have been implemented and there have

---

<sup>2</sup> An exact page-based representation cannot be guaranteed – even though PDF is designed as a document presentation system, differences can occur based on viewer capabilities, even with text, but especially where the presentation is dependent on overlapping graphics or transparencies.

been efforts made to align the converters. However, alignment is an ongoing challenge for international phase processing as any incompatibilities will present problems to applicants and Offices. Furthermore, only limited provisions are in place for dealing with changes to international applications in full text format.

#### FILING INTERNATIONAL APPLICATION IN FULL TEXT FORMAT

17. In 2021, 98.7 per cent of international applications were filed electronically, compared to 93.6 per cent of full electronic filings in 2015. However, in percentage terms, XML filings over this period have remained at a similar level: 27.2 per cent in 2021, 28.2 per cent in 2015. Increasing the percentage of international applications filed in XML is therefore crucial to a transition to full text processing.

<i>Receiving Office</i>	<i>Full-text XML applications</i>	<i>PDF applications</i>	<i>Paper applications</i>	<i>Total applications</i>
JP	48,720	0	320	49,040
KR	20,399	0	126	20,525
CN	5,730	67,480	246	73,456
IB	283	13,122	101	13,506
EP	192	37,472	658	38,322
US	0	56,367	61	56,428
Other ROs	163	23,570	2,106	25,839
Total	75,487	198,011	3,618	277,116

Table 1: Filing Format of International Applications for 2021 by Receiving Office

18. Table 1 shows that approximately 91.6 per cent of international applications filed in XML are received at either the JPO or the KIPO; these two Offices receive 99.3 per cent and 99.4 per cent of their international applications in XML, respectively. By contrast, while the International Bureau as receiving Office provides the possibility of submission of a DOCX document and conversion to XML format in accordance with WIPO Standard ST.36, the take up of this option is low. It appears that most applicants remain concerned over the perceived risks of processing a full-text application and continue filing PDF despite the additional 100 Swiss franc reduction for filing in XML. The profile of filing format remains unchanged compared to 2020 (see Table 1 of document PCT/WG/14/8).

19. Several other IP Offices are continuing to work towards increasing XML filings. The China National Intellectual Property Administration (CNIPA) and the European Patent Office (EPO) are both running projects to facilitate the filing and processing of international applications in full text. Regarding national and regional systems, the EPO, the National Institute of Industrial Property (INPI) France, and the United States Patent and Trademark Office (USPTO) already have, or are in the process of, implementing systems to process full text applications as the intended primary processing route, instead of paper and PDF.

#### TRANSMISSION OF FULL TEXT FILINGS

20. The International Bureau receives international applications in page image format as part of record copies from some receiving Office systems that have been filed in XML format. This is currently necessary due to local systems having been developed in the era of page image processing. The IB offers a discount to filers for filing in XML format, but the PCT System as a whole is not currently benefitting from that XML. Similarly, there are some Offices that accept electronic filing, but continue only to accept image based filing. To enable the processing of these applications in full text format, the International Bureau requests that receiving Offices

that have not already transitioned to transmitting the record copy in XML format, when it has been received in that format, make that transition as early as possible.

#### CURRENT PROCESSING OF INTERNATIONAL APPLICATIONS IN FULL TEXT

21. Where applications are filed in full text format, much of the subsequent processing is based on image files that have been rendered from the XML, including transmission of image files between Offices. Amendments, corrections and rectifications are handled as replacement sheets. At the International Bureau, “double work” processing of the image application body and the XML application body in parallel occurs to maintain the full text while also producing a traditional publication, incorporating replacement sheets.

22. ePCT already accepts Article 19 amendments filed through the submission of a DOCX file or a text-based PDF file. However, this is the most simple case of changes, since the amended claims are shown in addition to those originally filed, rather than replacing them. Furthermore, these submissions rely on the user providing separate information describing the change, such as which claims have been amended.

#### INTERIM PROCESS FOR FULL TEXT BASED INTERNATIONAL PUBLICATION

23. Up to now, the International Bureau has processed corrections and rectifications in image format and in XML format (for XML applications) as a parallel operational activity; this duplication of effort risks introduction of errors and inconsistencies. As a first step to removing this duplication, the International Bureau is preparing to accept application bodies from CNIPA that contain corrections and rectifications in XML format, which the International Bureau intends to use as input in the processing of the XML to generate the page image files that will be published for these international applications. In this interim processing arrangement, the International Bureau will not generate true “replacement sheets” for changed content, but will generate new XML using “branch numbering” (“1.”, “1.1”, “1.1.1”) and empty paragraphs to ensure consistency of paragraph and figure numbers. The changed version will have the changed paragraphs marked and “replacement sheet” footers will appear at the bottom of any page containing paragraphs or figures that differ from the original. However, the page breaks will typically not match those seen in the international application as filed.

#### PRINCIPLES FOR FULL TEXT PROCESSING

24. The International Bureau continues to support the move towards full text processing of applications. The aim is that:

- (a) applicants should be able to submit full text formats of their application bodies in the international application as filed, and in the case of any changes see immediately, when using an online filing system, how this will be presented in the published application, or as amendments annexed to an international preliminary report on patentability (Chapter II);
- (b) the views generated from the full text filing be used equally at all stages of international phase processing; and
- (c) designated Offices and third parties should be able to view full text versions of the application showing all accepted corrections, rectifications and amendments clearly marked up to distinguish between original and modified content, similar to the “substitute sheet” markings of the current system, but at the level of headings, paragraphs, claims and figures, rather than pages.

25. The International Bureau is seeking to devise a simple and consistent process that meets the requirements of applicants, Offices and third parties alike. Some of the issues to be considered are set out below, noting the importance of the need to identify versions of the international application accurately in a distributed processing environment, where it may be

difficult to see all of the documents at a later stage in a way that allows the original context and timing to be understood.

#### FORMALITIES OF DOCX PREPARATION

26. At present, Rule 11 is drafted in terms of preparation of documents on paper. Many of the requirements are not relevant to XML documents, which have no concept of margins, font size or page layout. On the other hand, to ensure that applicants know how to prepare DOCX applications that will convert without difficulty to the XML formats necessary for processing and preparation, other regulation will be needed. Rule 11 should specify (either directly or by reference to the PCT Administrative Instructions) clear drafting requirements. Applicants should not be forced to use particular templates, but should be aware of the features that they may and may not use, as well as how particular types of special content will be imported and used. Examples of such requirements are as follows:

- (a) Paragraph text must be presented in a single font size of black text, decorated only with supported effects (bold, underline, italic, superscript, subscript, small caps). Any other formatting will be ignored (with a warning generated by the conversion software). Any case where other formatting features are required for the purpose of effective disclosure should be represented as images (except to the extent that markup for mathematical and chemical formulae may be supported).
- (b) All text must be presented in a Unicode font, except for certain characters typically inserted by word processor functions using the Symbol font, which will be mapped to the appropriate Unicode equivalent within the XML (with a warning generated by the conversion software).
- (c) Tables, chemical and mathematical formulae may be included, but may have limited formatting options (to be identified) and/or be converted to bitmap images referenced by the XML.
- (d) Drawings should be included as simple bitmap images or explicitly supported complex embedded content types, which may also be converted to bitmap images referenced by the XML.
- (e) Line numbering should not be used, and will be ignored if included.
- (f) Track changes and comments must not be used (ePCT at present will reject upload of a DOCX file using either feature).

#### GRANULARITY OF TRACKING CHANGES

27. While changes in full text processed documents may be tracked at several levels, for example at the paragraph or character level, the need to mark up the changes with the relevant authority, regulation and date makes their tracking difficult to display unless done at the paragraph, claim or figure level. Titles and headings may also be modified. All of these elements are specified in the application body DTD. Consequently, this document proposes the principle of tracking changes at the level of paragraph, claim or figure, title and heading. In this case, "heading" includes any general heading and any IP5 Common Application Format (CAF) heading (see paragraph 29 below).

28. In the tracking of changes at this level, it is furthermore proposed to follow the principle of retaining all original content in the document. By moving original content into a "changes" section with original identifiers, and at the same time, inserting the new content into the document with new unique identifiers, this model aims to enable the recipient of the document to be able both to render the published content using existing stylesheet logic and to view the superseded content. In cases where there is a change to the image content of a figure,

chemical formulae, mathematical equation etc. it is proposed that the original image content is retained in the XML document package and referred to by the content in the “changes” section.

#### COMMON APPLICATION FORMAT HEADINGS

29. The current WIPO DOCX converter maintained by the International Bureau when processing headings that it identifies as Common Application Format (CAF) headings, creates container elements defined by the CAF specification, grouping the paragraphs associated with each section. The International Bureau has observed that these heading elements add to the complexity of the rendering and tracking of changes in full text international applications. Unless any Offices are taking advantage of this additional structure in the XML, the International Bureau proposes to update its DOCX converter to process all headings identifying them simply as headings followed by paragraphs, without producing container elements for the sections. This will not require a change in the stylesheet.

#### REPLACEMENT SHEETS

30. The current system of page-based publication relies on changes in the application body being recorded on sheets that are inserted and removed from the application body. At present, the International Bureau processes replacement sheets in the same way as for PDF and paper filings. To make the full text available for these applications, the International Bureau is currently doing the “double work” of making replacements in both the full text and image copies of such international applications. As discussed in paragraph 23, above, the International Bureau is putting in place an interim solution to avoid the double work, but this is not sufficient to support full automation of the processing and delivery of high quality text to designated Offices and patent information users.

31. In the longer term, the aim is that the concept of “sheets” should disappear. Where a change is required, the applicant should supply the appropriate replacement content in a manner that is independent of the pagination. This should then be stored as XML providing a record of the original document, the submitted modification and the modified document such that views can be generated showing where the changes lie at the levels of paragraphs, figures, etc. Current prototyping envisages that the changes would normally be made by supplying a complete new element (description, claims, drawings or abstract) and the system detecting where differences from the previous version are to be found. It is likely that alternative mechanisms will also be needed to allow change of individual paragraphs or figures where the need for change results in difficulties with the conversion from the original DOCX format. However, this should be rare if the system enables the use of full color images, since conversion from color to black and white is the largest conversion difficulty currently encountered.

32. To support this, Rules 26.4, 46.5, 48.2, 66.8, 70.2, 70.16, 74.1 and related provisions would need to be amended to make clear provision for the submission and processing of replacement content in a form other than sheets.

#### NUMBERING PRACTICES

33. As part of the discussions on paragraph amendment at the sixteenth session of the Meeting of International Authorities in 2009 (see paragraph 8, above), Authorities considered the numbering of paragraphs. At that time, it was noted that system capabilities and practices were a consideration in the numbering system (see document PCT/MIA/16/14 and paragraphs 98 to 101 of document PCT/MIA/16/15). This issue should be considered again, taking into account legal, technical and administrative considerations to identify a clear and simple solution.

34. Paragraph numbers should not be considered part of the substantive content and it should be permitted for Offices processing the application to add, renumber or reformat them as an administrative matter. Paragraphs in the description should be numbered, but may, for

example, be reformatted so that “1.” becomes “[0001]” or renumbered to provide a continuous numbering, both within the application as filed and following changes<sup>3</sup>. Such renumbering should not be considered a change and not, of itself, result in a paragraph being annotated as having been changed. The same should, in principle, apply to claims and drawings, though further consideration is needed in view of their special characteristics, including the common use of sub-numbering of drawings (Fig. 1a, Fig. 1b) and the fact that claims are routinely cross-referenced, but under the present rules are replaced entirely if amended under Article 34 (Rule 66.8(c)).

35. Consequently, cross-referencing paragraphs within the description – already used infrequently – should be avoided. Further consideration will nevertheless need to be given to how paragraph references should be handled by the software and marked up in the case where a reference change is solely consequential to a change in the number of another paragraph.

36. This has the potential to result in confusion between applicants and examiners since the numbering in the application as being processed may be different from that seen in the DOCX files originally submitted (ePCT detects this and provides a warning when it occurs). To minimize this risk, the numbering arrangements must be clear so that attorneys reliably create applications where the numbering arrangements match the expectations of the conversion tool.

#### STATUS OF ORIGINAL DOCX DOCUMENTS AND CONVERSION SAFEGUARDS

37. To date, one of the key concerns that has been holding back the transition to full text processing has been the need for confidence in full text processing. When converting a DOCX file into XML, the applicant should be confident that the substance of the application is preserved, but in the rare cases of a conversion error that changes the substance of the disclosure, the problem can be corrected at any stage, including during the national phase. Consequently, where a DOCX document is filed, it should be regarded as the definitive original filing, even though all subsequent processing takes place on the basis of a simplified XML format.

38. From a technical storage point of view, the existing arrangements where a DOCX file is automatically stored as a “pre-conversion file” under Section 706 of the Administrative Instructions alongside a conversion to ST.36 XML appears to meet the requirements. However, further consideration is needed of the legal issues (where currently the ST.36 XML is considered the “original” despite the right to make corrections) and the issues of reliably determining the true original content, given that DOCX is a format that can result in different presentations depending on the word processor in which it is opened.

#### ANTICIPATED ePCT PROJECT WORK

39. The International Bureau proposes to continue to develop a pilot implementation that would offer to applicants filing full text applications to a receiving Office accepting ePCT Filing, and selecting an International Searching Authority that conducts the international search in ePCT, the possibility to submit Rule 26 and Rule 91.1 requests through the upload of a revised DOCX. This would directly create the desired XML for the proposed modification to the application body that would be ready for processing by the International Bureau and International Authorities in XML format. This would prepare the way for exposing the same technical functions to other receiving Offices and International Authorities as web services and/or packaging the relevant components as systems that could be deployed locally for use in services hosted by national Offices.

---

<sup>3</sup> This is different from the interim arrangements referred to in paragraph 23, which use branch numbering to avoid complete renumbering of subsequent paragraphs.



*40. The Working Group is invited to comment on the principles and requirements for full text processing in the PCT, with particular regard to the issues set out in paragraphs 24 to 39 of document PCT/WG/15/14.*

[End of document]