

Patent Cooperation Treaty (PCT) Text Processing Task Force

First Session
Geneva, January 29 to 31, 2024

TEXT PROCESSING TASK FORCE INITIATIVES

Document prepared by the International Bureau

INTRODUCTION

1. This document provides the PCT Text Processing Task Force with an initial list of topics for discussion towards processing all international applications in full text format. The topics include seeking improvements to the submission, processing, and transmission of international applications to grow the proportion of applications filed in full text format by making it easier for applicants to file international applications in that format with legal certainty.
2. These discussions are aimed at finding a path forward to move away from sheet image-based processing of international applications. This should reduce formalities errors and provide self-service functions for applicants that improve efficiency for Offices and provide more accurate and accessible information to designated Offices and patent information users.
3. The International Bureau has grouped the discussions into three main initiatives:
 - (A) growing full text format filing volumes through customer experience improvements and receiving Office encouragement to move away from pdf;
 - (B) implementing common functionality across offices for the processing of international applications in full text format; and
 - (C) maintaining confidence in filing certainty through coordinated legal provisions.
4. The Text Processing Task Force is requested to review the proposed initiatives and advise the International Bureau on future work that can lead to processing all international applications using full text format.

(A) GROWING FULL TEXT FORMAT FILING VOLUMES THROUGH CUSTOMER EXPERIENCE IMPROVEMENTS

(A1) COMMON INFRASTRUCTURE FOR CONVERSION TO XML

5. The International Bureau considers that consistency is needed in conversion from DOCX to XML. Differences in drafting rules for applicants filing national, regional, or international applications cause confusion and errors, leading to a loss of confidence in the system. Ideally, all Offices should use a common and clearly documented piece of software for conversion. The International Bureau has developed and deployed a converter that is available for use in 10 languages and can easily be extended to others.

6. DOCX has been chosen for current development work as the *de facto* primary format in which patent applications are drafted and shared between inventors, applicants and attorneys. In principle, equivalent conversion tools could also be developed for other file formats if a need were established. However, conversion requirements from other file formats would need to be sufficiently clear and simple over which word processor features could be used. Moreover, the converted XML file would need to be identical, irrespective of the original file format.

7. It is proposed that the Task Force advise on a way forward in the development and deployment of a common DOCX converter software that will apply common requirements to the conversion of national, regional and international applications.

(A2) COMPLETION OF XML CONVERSION FEATURE SET

8. The International Bureau has provided a rich feature set in the DOCX converter embedded into ePCT. It is necessary to ensure that users find it clear and easy to use, and that simple arrangements exist for handling post-filing changes, including corrections, rectifications and amendments.

9. At the time of filing, it is necessary to consider whether the conversion tool provides easy access to all the functionality required. At present, the main risks are considered to lie in issues concerning alignment of text:

(a) WIPO Standard ST.36 does not generally allow for the use of tabs and spaces to align text – paragraphs usually need to be written as a simple flow of text, without any expectation of line breaks occurring at a particular point or text on one line aligning reliably with related text immediately above or below. The standard does allow an element <pre> to denote “pre-formatted text”, which is represented using a fixed width font, such as *Courier*, and maintains all spaces, which can be important for the representation of computer code or sections of nucleotide sequences. However, at present there is no way to trigger the use of this element from a DOCX file.

(b) The main form of text alignment is through tables. The XML table structure is fairly rich, including merged cells and column width information, but text alignment is lost. Further, the fact that all text is rendered in the same font at the same size means that in some cases it is difficult to fit the required information into the available space. Consequently, in some cases, it is more practical to provide large or complex tables in image form.

10. The Task Force is requested to identify and prioritize other significant limitations in conversion capabilities.

(A3) CUSTOMER EXPERIENCE IMPROVEMENTS ENCOURAGING POST FILING UPLOAD AND COMPARISON OF REPLACEMENT DOCUMENT(S) IN FULL TEXT FORMATS

11. The International Bureau aims to implement additional end user self-service features that benefit the applicant when filing in full text format. It should be possible to conduct all subsequent stages of processing in full text format and to allow effective comparison of different versions of the application. Ideally, applicants should be able to submit changes for themselves and immediately see the effects the changes will have if they are accepted. The automated comparison between versions should help to generate the letters required describing the changes and the basis for the new text in the application as filed when amending the claims before international publication, or amending the application during international preliminary examination (see PCT Rules 46.5(b) and 66.8, respectively).

12. The International Bureau expects that the normal method of making a change would be to submit either an entire replacement application body or a replacement element (description, claims or drawings) in the same format as the original application and for the system to detect changes, though alternative methods would also need to be supported, for example to allow correction of errors caused by the original conversion process.

13. For such a service to work in an environment where different roles are performed by the International Bureau and different national Offices, it is important that receiving Offices, International Authorities and the International Bureau all accept full text applications in a common format and perform the conversions consistently. Receiving Offices and International Authorities are requested to advise, working with the International Bureau, on strategies and implementation timelines to offer applicants full text processing improvements that improve the value proposition for applicants to file in full text formats.

(A4) ENABLING THE FILING AND PUBLICATION OF COLOR DRAWINGS

14. Support for color drawings is important. In some cases, effective disclosure is very difficult without the use of color drawings and photographs. Even where drawings are originally prepared in black and white, the conversions between different file formats causes degradation and is probably the largest single cause of formalities defects.

15. The International Bureau's publication systems are already capable of using original color drawings for the purpose of processing XML applications. As an interim measure, an unofficial full text and color version of the application body part of the international publication will soon be made available for XML applications in many cases. An aim of upgrading the publication processes is to allow official color publication at least for those applications filed in full text format and to ensure that systems are developed with a view to allowing further file formats to be added in the future, such as video or 3D diagrams, if the business need and technical standards are agreed.

16. The Task Force is requested to advise on these requirements and as necessary propose an update to Annex F and Rule 11 to enable color drawings in international applications.

(A5) REINFORCEMENT OF FILING AUDITABILITY

17. It is essential, for both applicants and designated Offices, that it is possible to prove exactly what was filed. Annex F of the PCT Administrative Instructions from the outset set out requirements for integrity and non-repudiation, which were fulfilled by issuing a filing receipt including a hash of the package submitted by the PCT-SAFE filing software. In the event of data loss by the receiving Office, the applicant would be able to prove that the package generated by PCT-SAFE, of which a copy was kept locally, was identical to what was filed. With the move to browser-based filing systems, the applicant no longer necessarily keeps a copy of the package filed, only of the individual files uploaded for inclusion within the package.

18. Although there has never yet been a need to use this particular safeguard, it would be desirable for the filing receipt to be updated to provide hashes of the individual files. This requirement is not directly related to the question of full text processing, but is closely associated with the questions of legal status of different versions of the application, availability of records and the ability to make corrections in the case of any conversion errors (see section C2, below). The Task Force is requested to consider the implementation of improvements to submission receipts and any corresponding records to enable browser-based submission systems to return a receipt that meets the needs of non-repudiation for each file uploaded.

(B) COMMON INFRASTRUCTURE ACROSS OFFICES FOR THE PROCESSING OF INTERNATIONAL APPLICATIONS IN FULL TEXT FORMAT

(B1) A FULL TEXT SUBMISSION PATH AND DOCUMENT EXCHANGE BETWEEN ALL RECEIVING OFFICES, INTERNATIONAL AUTHORITIES AND THE INTERNATIONAL BUREAU

19. For effective full text processing of corrections, amendments and rectifications, receiving Offices and International Authorities will all need to accept full text changes to the international application according to a common standard and to exchange these with the International Bureau. Otherwise, applicants will need to submit rectified sheets and, to avoid the application effectively reverting to being an image-based filing, the national Office or International Bureau would need to transcribe the changes into the international application.

20. National and regional Offices should consider how such arrangements could be made, for example through ePCT as either a browser-based function or else as a web service for integration into the systems of Offices. The International Bureau seeks advice on Office capabilities to update local transmission systems and advice on any necessary updates to the Administrative Instructions.

(B2) USE OF AN UPDATED STYLESHEET TO GENERATE IMAGE-BASED PUBLICATION FROM PROCESSED FULL TEXT APPLICATIONS

21. The International Bureau has developed a revised stylesheet that shows the changes made to an application through rectifications, corrections, and amendments in the international phase (see Circular C. PCT 1656). This has eliminated the need for sheet-based processing at the International Bureau for processing full text applications where previously the International Bureau had been doing “double work” modifying the full text and the sheet image-based documents in parallel.

22. The new stylesheet is already in use and the International Bureau would welcome any suggestions for further improvements.

(B3) EVOLUTION OF MARKUP REQUIREMENTS FOR CHANGES IN PATENT DOCUMENTS

23. Because the record of changes to the application body may be distributed across the files of several different Offices, it is particularly important that the international publication include clear markup showing which parts of the application are as originally filed and which have been the subject of corrections, amendments and rectifications. For applications filed in XML format, for the purpose of the arrangement referred to in section B2, above, the International Bureau marks up the changes at the level of paragraphs and drawings, rather than sheets. However, this is done by inserting text into the XML, which can be interpreted by a stylesheet to give suitable markup in HTML or PDF format, but does not give “clean” XML showing the proper text of the application.

24. The International Bureau has drafted an Annex F proposal for change (PFC) that would enable the markup to be transformed into a structured XML markup that would properly separate the markup from paragraph text. The resulting XML would include a clean view of the international application as updated, but also retain the original texts and images with information to indicate their original locations, allowing a reader to review and compare the versions of a document from the original to the published version. This would give a technically superior result but be more complicated than the present system to implement. A discussion of the benefits and the costs of the required functionality is desirable before other full text processing proposals are taken further. The International Bureau seeks advice as to whether it should proceed with this Annex F Proposal For Change (PFC).

(B4) SIMPLIFICATION OF THE PROCESSING OF DRAWING TEXTS

25. PCT Rule 11.11 only permits a limited use of text in drawings. However, the International Bureau observes that applicants often include significant text in drawings. Where such drawings are used for the front page of an international application, this text is sometimes difficult to read. Provision of this text to translation is done through reviewed OCR or re-typing. Furthermore, since the text is only shown in image format, neither the original nor translated text is available for indexation in search engines.

26. A new front page arrangement (referred to on page 4 of document PCT/MIA/30/6) will soon be used to address some of these problems, but further consideration should be given to the questions of when and how much text should be permitted in drawings, as well as how best to capture and process that text.

(C) DELIVERING CONFIDENCE IN FILING CERTAINTY THROUGH COORDINATED LEGAL PROVISIONS

(C1) AMENDMENTS TO RULE 11 TO SPECIFY FULL TEXT PROCESSING FORMALITY REQUIREMENTS

27. Rule 11 prescribes the physical requirements of an international application based on paper sheets and the layout of text matter, drawings, formulae, and tables on these sheets to enable page image reproduction. This has continued to be relevant in the transition from filing international applications on paper to electronic filing in PDF format. However, for full text filings, there is no need for a formalities examiner to check requirements related to page layout because the layout is delivered using appropriate stylesheets suited to the reader's display media, be that on paper or on screen.

28. A revision to Rule 11 should specify requirements for submitting the text of the international application in a manner that is independent of its layout. When using common desktop document editors, it is important that the authors do not use formatting and features that are not supported by the stylesheets used in the display and publication of applications. This is a necessary requirement where Offices are converting the original document editor file into a WIPO Standard based XML file. This ought not to be a significant change in practice for attorneys used to drafting patent application for submission to Offices that typeset their patent publications since it is important to ensure that the application as filed does not rely on particular fonts, colors, sizes or layout features that are not available in the format to be published. However, it is important that the rules on the features that are acceptable for use in filing the application are consistent across national, regional and international systems to avoid errors due to confusion between different requirements.

29. The International Bureau therefore seeks advice in drafting amendments to Rule 11 to focus primarily on the requirements of applications filed in full text format, with only the figures and certain special content being submitted in image formats.

(C2) STRENGTHENING PROVISIONS TO PROTECT APPLICANTS FROM FULL TEXT PROCESSING CONVERSION ERRORS

30. At present, the only formally recognized full text filing format for PCT filing is XML according to WIPO Standard ST.36. At some receiving Offices, the applicant is allowed to submit a pre-conversion format document, which can be used to correct conversion errors discovered during the international phase. In the case of applications filed using ePCT using a DOCX file, the conversion to ST.36 XML is performed before the application is filed; the applicant can see a PDF view of the converted application and the DOCX file is automatically submitted as a pre-conversion format document. However, to give greater confidence in full text filings, it may be necessary to strengthen the safeguards.

31. It is suggested that DOCX files should be accepted as a primary document of record, representing the official format in which the application is filed, even though it will go on to be processed in a less rich format. It would be desirable if corrections were possible in the case of errors in conversion irrespective of the receiving Office at which the application was filed and whether they were found in the international or national phase. Applicants have pointed out that DOCX does not guarantee the exact rendering of its content and this may vary over word processor type and version, and depending on the installed fonts of the system viewing the file. Consequently, it might be appropriate to permit a PDF file created from the DOCX file to be included to deal with situations where there may be doubt as to how a DOCX file would originally have been viewed.

32. The International Bureau seeks the opinion and advice of stakeholders regarding the existing provisions safeguarding applicants in case of document format conversions in the international phase. If needed, adjustments could be made to the legal framework to ensure that the documents submitted are captured as submitted and that all stakeholders can rely on their conversion into the international publication produced in a full text format.

33. The Task Force is invited to review the proposed initiatives and comment on the directions and priorities of work in the different areas.

[End of document]