

## **Meeting of International Authorities under the Patent Cooperation Treaty (PCT)**

**Twenty-Eighth Session**  
**Geneva, March 24 to 26, 2021**

### **XML SEARCH REPORTS AND WRITTEN OPINIONS**

*Document prepared by the International Bureau*

#### **SUMMARY**

1. The International Bureau is now receiving XML data for search reports and written opinions from nine International Searching Authorities, representing about 85 per cent of total search reports and written opinions. This represents a great opportunity to improve the services available to applicants and designated Offices, but further work is needed to ensure that the data is consistent, accurate and useful.

#### **BACKGROUND**

2. Standards were developed for XML search reports and written opinions in the early stages of WIPO Standard ST.36 and Annex F of the PCT Administrative Instructions, with version 1.0 of the standards being released in 2005. However, practical implementations of XML systems proved difficult. It was not until 2011 that the European Patent Office began to send test XML search report data and only at the end of 2014 that data from ISA/EP and ISA/CN began to be used in production. Even now, a variety of problems remain with the quality and consistency of data.

3. Since 2016, ePCT has offered functions for preparing XML international search reports and written opinions and this is currently used by the examiners in five Authorities. The XML from this source is, of course, very high quality and consistent. However, the interface uses components that were optimized for different types of data entry and improvements are needed for efficiency and good user experience for examiners, as discussed below.

4. Production data is currently received in XML format for at least some international applications from nine International Authorities (AT, CN, EG, EP, JP, KR, PH, TR, XN). Additional Authorities are expected to start providing data in the course of the year.

#### **ADVANTAGES OF XML REPORTS**

5. XML search reports and written opinions offer a number of advantages if implemented well. The data from the reports can be extracted and checked automatically, allowing the receipt of most such reports to be fully automated, increasing efficiency of processing and eliminating the risk of transcription errors. Moreover, the data can be used to provide improved services:

- (a) The full text data can be passed to translators, properly tagged to be clear which parts require translation and allowing effective use of translation memories and other professional tools to assist the work of the translators. This improves quality and reduces the time and cost in establishing the official English language translations of the reports.
- (b) The data can be passed efficiently to machine translation systems, allowing the reports to be viewed in a wide range of languages.
- (c) Links can be provided to relevant material, such as definitions of classifications, copies of cited documents or copies of patent family documents.
- (d) Information can be accurately extracted from the reports automatically, allowing analysis of the contents in greater detail and at an earlier stage than has previously been possible with the manual extraction of information that has been done in the past.

6. With respect to these points, PCT Translation Division is already using the XML from most of the search reports and written opinions (and also some international preliminary examination reports) received in XML format. PCT Translation Division is also working with PCT Operations Division and Authorities concerned to improve aspects of the XML that make the data for some applications difficult or impossible to use directly in the translation systems.

7. Furthermore, PATENTSCOPE provides a service where the international search reports and written opinions received in XML format can be viewed in any of the 10 PCT languages of publication. The “boilerplate” text comes from the official templates in the relevant language and the other text is taken from the original or official English language translation XML where available, or else shown directly or passed to a machine translation system, according to the nature of the field involved. Links are provided to patent citations and family members for easy reference.

8. Preparation of XML reports, both national and international, to consistent standards will promote the sharing of useful information and the merging of related data from different sources, whether for individual applications or databases covering large collections of applications. The PCT third party observation system, established in 2012, collects citation information in the format used in international search reports. This was intended to permit easy merging of citation information from different sources to provide wide-ranging information in a useful format for the designated Offices. The sharing of national search reports in compatible formats would assist the effective flow and collation of data between different stages of processing, improving both the efficiency and quality of search and examination, and the information available for applicants and for system analysis.

#### **IMPROVEMENTS REQUIRED**

9. The XML supplied for international search reports and written opinions is created in a variety of different ways. In some cases, the information is entered such that XML is the original format of data, with the human-readable view of data being created using a stylesheet to render

a PDF or other document from that data. In other cases, XML is not the original format of the data. Instead, the reports are initially created in other ways, such that the pages sent to the applicant are output from one system, with the XML then created separately by extracting the relevant data and exporting it into an XML format. However, the options available within a visual editing system may not always correspond well to the options available within the document type definitions (DTDs) defining the formatting and data structures expected within the XML.

10. Problems that have occurred, or that still occur within the XML received from Offices include:

(a) Line breaks (<br> tags) are often used by examiners in some Authorities either to simulate logical groupings of data or to make visually appealing arrangements of text in the page-based views of the original language output for the benefit of applicants. However, unless the line breaks form genuinely logical units of text, this interferes with translation systems (both machine translation and translation assistance), since the text is seen as smaller units that do not correspond to texts in translation memories. Furthermore, in many cases, these small units do not make logical sense in isolation from the remainder of the sentence. In addition, the fact that the tags may be used both for visual purposes and to separate logical consequences means that it is not possible to automate either the removal of the tags intended only for visual purposes or the identification of tags representing multiple logical statements.

(b) Examiners may use formatting options in some fields that are not available according to the DTDs defining the XML. Consequently, information that is intended to be conveyed by that formatting is lost.

(c) The XML is not always created correctly according to the standard. While the most common examples of this have been identified and addressed, the reports include a wide variety of less frequently used options and some reports still fail to validate correctly against the relevant DTDs.

(d) The definitions for the search report are extremely flexible in the possible relations between relevant passages, citation categories (X, Y, etc.) and relevant claims, permitting a very wide range of possible structures. Different Authorities create significantly different data structures for this purpose. While the stylesheets have been developed to handle this range as effectively as possible, it creates difficulties in designing common database structures to handle the relevant information if detailed analysis is required. While not yet fully analyzed, it seems likely that this will also cause difficulties in creating accurate data transformations between the ST.36 and ST.96 Standards if, as seems likely, this becomes necessary in the future.

11. Consequently, the International Bureau needs to correct some XML data between its receipt and processing. This requires the preparation, maintenance and careful monitoring of import systems, as well as a degree of manual work. Where the XML processed is not identical to what was used to generate the view of the report sent to the applicant, this needs to be tagged with an indication that it is not the official view of the data. The International Bureau hopes to eliminate or at least significantly reduce the cases where this is required.

12. When creating systems for generating XML reports, it is important to ensure that those systems are efficient for use by examiners, but are also well adapted to ensuring that the relevant data items are properly identified, do not permit the use of inappropriate formatting, and encourage the presentation of information so that it can be automatically parsed based on its meaning, rather than manually adapting its visual appearance on the page.

13. For examiners using the ePCT interface to create reports, a number of performance and usability improvements are required. The system offers options for import or lookup of citation data from certain other systems, as well as validations, such as those for classification terms. However, the components available were designed for different types of data entry and are not as efficient as might be desired. Work is under way to improve the speed and usability of the components used, especially for citation and classification data entry and editing. It is also aimed to improve the navigation and information visibility between related sections of information, such as the citation entry for the international search report and the information on novelty and inventive step in Box V of the written opinion, as well as the effective coordination and reuse of the citation data between other search and preliminary examination forms, including PCT/ISA/206, PCT/IPEA/408 and PCT/IPEA/409.

## **OTHER ISSUES**

14. The variations in the structure of XML are partly a result of the differences in the ways that data is entered by examiners at different Authorities, but may also represent different attempts to seek user-friendly representation of the data on the search reports. International Authorities may wish to highlight best practices or issues of usability that could be addressed to provide more useful and consistent representations of data, both within the XML and when rendered for viewing in PDF or HTML versions of the report.

15. At present, all XML data representing international search reports and written opinions is received in the format defined in Appendix I to Annex F of the PCT Administrative Instructions. This is based on WIPO Standard ST.36. Several Authorities have indicated that they wish to develop system for preparing XML reports, but wish to do so according to the newer WIPO Standard ST.96.

16. The International Bureau has not yet been able to analyze the requirements to develop a specific implementation of international search reports and written opinions according to ST.96. If alternative formats are used, it will probably also be necessary to create a bidirectional transformation between the different formats if the data is to flow directly into national phase processing. This may well be complicated in view of the wide variety of citation formats supported by ST.36, as noted above. It is not clear whether a simple transformation would be possible between all options supported by the two different standards. Similar issues may also arise with regard to local transformation requirements where the native format used by an Office for search and examination report data is a different standard again. Irrespective of the specific standards used to exchange the relevant data, further standardization of the underlying data being presented within search reports may be desirable if efficient reuse of information is to be achieved.

*17. International Authorities are invited:*

*(i) to provide information on the status and plans for the introduction and development of XML reports in their Offices; and*

*(ii) to comment on the issues set out in this document.*

[End of document]