

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

**Twenty-Third Session**  
**Santiago, January 20 to 22, 2016**

**PCT MINIMUM DOCUMENTATION:  
DEFINITION AND EXTENT OF PATENT LITERATURE**

*Document prepared by the International Bureau*

### **SUMMARY**

1. The process of documenting and modernizing the definition of the patent literature part of the PCT minimum documentation is progressing slowly. The Meeting is invited to offer comments on the general issues, or on ways of improving the process to reach a positive outcome more quickly.

### **BACKGROUND**

2. At the nineteenth session of the Meeting of International Authorities and the fifth session of the PCT Working Group, the International Bureau presented documents (PCT/MIA/19/13 and PCT/WG/5/16) suggesting the concept of amending Rule 34 to automatically include the national patent documentation of any PCT Contracting State as part of the PCT minimum documentation, provided that it was made available reliably in a suitable electronic format which would be easy for International Authorities to load into their databases. Certain technical limitations would continue to apply relating to avoiding duplicate documents and searching documents in languages which are inaccessible to the examiner.

3. The aim of this initiative was to improve the availability of technical information from patent documents, both in terms of the technical coverage and linguistic coverage and, as a result, to help improve the quality of international search.

4. The process was put on hold, pending the outcome of work by the IP5 Offices (European Patent Office, Japan Patent Office, Korean Intellectual Property Office, State Intellectual Property Office of the People's Republic of China and United States Patent and Trademark Office) properly documenting their own national or regional patent documents. That work was completed with the publication of "Authority Files" from each Office. The International Bureau then put a number of issues to the PCT minimum documentation task force, which were divided into three sections:

- Documenting Patent Collections
- Making National Patent Collections Available
- Minimum Documents and Data for Records to be Included

5. Responses to the questions concerning these issues have only been received from a small number of International Authorities. At this stage, it is not clear that there is consensus for, or sufficient information to recommend a specific way forward.

## ISSUES

### Documenting Patent Collections

6. This area of work has two initial goals:

(a) Accurately document the extent of each national patent collection to allow database providers to be certain that their information set is complete – the aim was to agree a definition based on the "Authority Files" developed by the IP5 Offices, which national Offices could then use to document their own collection, based on seeking quality at source.

(b) Identify the practical scope of the minimum documentation at present, noting that in addition to those national collections explicitly defined as being included, parts of national collections with certain characteristics may be included if they are made available to the International Searching Authorities by the Office concerned (Rule 34.1(c)(vi)). However, the last inventory of such documents was taken in 2001 and it was not clear whether these collections were still being made available to all International Authorities, or whether others had begun to be made available.

7. The Authority Files published by the IP5 Offices were all in different formats (two variants of XML and 3 variants of CSV) and contained different levels of detail about the publications to which they referred. The International Bureau suggested that a more consistent standard would be highly desirable if Offices and database providers were to receive and process such files (updated on a regular basis) from a large proportion of PCT Contracting States and regional Offices.

8. Three Offices replied to the queries in relation to Authority Files. None expressed a strong preference as to format, though one (an IP5 Office) had produced an Authority File in a CSV format and another (a non-IP5 Office) had produced an Authority File in XML format. All Offices indicated that it may be desirable for the files to be useful for more than simply verifying that a collection of documents was complete. The International Bureau commented that a CSV-based format was slightly preferred for a simple list, but XML was strongly preferred if detailed information (such as details of all priority claims) were to be included. Two Offices suggested that the files should be produced only occasionally (likely once per year); the other suggested that the list should be kept as close as possible to up to date. It was felt to be important that any detected errors should be handled effectively. However, experience from the

IP5 Authority files suggested that once an initial definitive list had been established, there may be few further errors and it may not be necessary to have a complex, formal procedure to follow.

9. There were no responses (other than “not applicable”) to the questions concerning which Offices, beyond those whose collections are explicitly indicated in Rule 34.1(c), currently provided regular updated documents and information concerning their published patent collections.

#### Making National Collections Available

10. This area of work seeks to study the formats in which Offices make their patent collections available (file formats, packaging means, frequency of updates, etc.), identify which formats are particularly easy or difficult to load, and make recommendations for common standards in which Offices should be requested to publish their applications and bibliographic data for ease of use by others.

11. Again, three Offices responded to the questions. Two Offices provided either samples of the formats in which they electronically publish data or else links to a site from which data could be downloaded. All of these Offices provided at least their front-file data for download using FTP. The back-files were also available in electronic format, either also for download or else on a NAS (Network-Attached Storage) device sent to customers on first subscription. The depth of data varied with age. In general, very complete bibliographic data and full text application bodies were available for recent publications. Older publications might be image format only with only very basic bibliographic data.

12. The Offices indicated that they saw few problems in principle with providing data according to any well-defined standard (though not all of the bibliographic data was currently made available in accordance with WIPO Standards ST.36 or ST.96). Nor was it difficult to import consistent, well-structured data from a variety of different formats. However, it was observed that not all data was compliant with the formats in which it was said to be stored. In addition, it was observed that it was easier to implement reliable systems based on standards defined in schema, rather than DTDs.

#### Minimum Documents and Data for Records to Be Included

13. This area of work seeks to identify what the minimum content requirements should be for publications to be included in the minimum documentation.

14. Up to now, it has only been necessary for national Offices to make the published documents available. This was essentially all that was needed for paper-based searching, where the documents would be placed in search files based on either the IPC on the front page of the application or a national classification applied by the Office where the search collection was to be housed. However, electronic searching uses full text and bibliographic data. While many Offices or database providers will perform their own processing to add value to their database, it is highly desirable that as much of the basic information is provided in machine-readable format once, at source, rather than expecting other Offices and database suppliers to refine it.

15. On the other hand, it is also desirable to have as many publications as possible made available in electronic format, even if the relevant Office is not able to provide complete full text and bibliographic data. As recognized in the responses to “making national collections available”, even Offices which act as International Searching Authorities may not have the resources to digitize the backfile data as completely and accurately as that for the applications which were originally processed and published electronically. To give the best incentive for Offices to provide electronic data concerning historical as well as current publications, it may be desirable to set high requirements on the content for current publications, but lower requirements concerning older publications. There may then be opportunities for refining the

historical data over time in a manner which is open to all interested parties, providing new full text data or better bibliographic information as processes improve across all the languages, character sets and typefaces concerned for OCR of full text and information extraction from printed bibliographic data.

16. A different set of three Offices provided responses to this section, but all agreed with the general principle that more relaxed conditions would be necessary in relation to older publications. In many cases, the bibliographic data for older patent publications was no longer available in any means practical to extract, save for that which is printed on the front of the publications or in the relevant Gazettes. Even that information is likely to contain more errors than the more recent publications, where the data could be taken from the original electronic records which had originally been used for preparing the publication. Moreover, it was observed that abstracts had never been established for some applications in the first place.

## NEXT STEPS

17. The International Bureau does not feel that it has sufficient information to infer any consensus or to make specific proposals at this stage. One possible way forward could thus be to open the issue up to a further round of comments on essentially the same questions, perhaps complemented by better information or examples on which to base discussions, in the hope that responses will be received from a larger number of International Authorities. Alternatively, it may be desirable to take a different approach, such as using a different means of consultation with Authorities or appointing a task force leader from an interested International Authority.

18. *The Meeting is invited to*

*(i) confirm that modernizing the definition of the patent literature part of the PCT minimum documentation is still considered useful;*

*(ii) comment on any of the general issues referred to in paragraphs 6 to 16, above; and*

*(iii) recommend any actions which should be taken to improve the process in order to come to a conclusion which will achieve the goals of improving the information available to International Authorities and database providers to enable effective patent searching across a broad range of Offices and languages.*

[End of document]