

**35th EPO-PatCom meeting
Tuesday, 20 March 2018, 14.00 hrs
Room 348, EPO Vienna**

Summary of the discussions

1. INTRODUCTION

The PatCom secretary informed the EPO that PatCom had elected Ann Chapman as its president. Ms Chapman said that PatCom very much appreciated the regular meetings with the EPO as they facilitated dialogue and helped avoid problems.

2. APPROVAL OF THE AGENDA

The agenda for the 35th meeting was approved without change.

3. MINUTES OF THE LAST MEETING

The minutes of the meeting of 13 October 2017 were approved.

4. BIG DATA PROJECT

The EPO remarked that increasing numbers of consumers of data were using standard techniques to process data from different domains, of which patent data was just one. In the past, EPO interacted with patent information specialists, who then grew into performing data analysis. A recent study the EPO had done suggested this was changing and that a new users group was emerging: data analysts, i.e. people who analyse data in general, patent data just being part of the overall view. It was with these people in mind that the EPO was thinking about how to make patent data easier to explore and easier to understand for the non-patent specialist.

PatCom said that many of its members were working in areas such as patent analytics. Their observation was that a solid patent search was a prerequisite for the analytics part to function. There was however a growing appetite for patent data.

The EPO explained that the above-mentioned study had been done by an external expert and would be presented the following day at the Raw Data event. The remit had been to look at EPO data and make suggestions on how to make it more useful for more people in the context of other types of datasets. The EPO added that it would soon be abolishing the term

"raw data" to describe its data products, as it seemed to be misleading for some potential users.

5. WORLDWIDE LEGAL STATUS DATA IN XML

The EPO reported that its legal status database now contained 235 million events. It was currently looking at US PAIR data, in the hope of adding some new events (e.g. withdrawal or refusal information) once it was clarified whether the API at the USPTO was going to be kept up to date. The Office had also recently loaded SPC data from ES, and was expecting to go into production soon with data from BE and LU.

No country was exchanging data yet in conformity with the new WIPO standard ST.27, but the EPO had taken the 22 categories and tried to categorise its legal status codes according to them (about 2700 in total). Implementation was expected in Q2/2018, after which the EPO would be able to exchange "category information" in XML.

Responding to a question from PatCom, the EPO explained that the solution of using US PAIR was specific to the US due to the fragmented data supply, and would not be a model for other authorities. China, for example, delivered data in a unified way, with a time-lag of just two to three weeks.

PatCom asked about Chinese reassignment data. The Office explained that it used to receive reassignment data in transliterated form, but that this had been discontinued in 2015, and so the data was only available as a commercial product from the IPPH.

6. OPS (WITH POINT 8.: QUALITY AT SOURCE)

The Office reported that it had taken a first step in decommissioning OPS version 3.1 the previous week. Version 3.2 would then be the only version available. It used a new image repository, which had given rise to some transitional issues, but the service was now operating reliably. A new pricing scheme for OPS had entered into force at the beginning of the year without incident or complaints from users.

With respect to the Quality at Source project, the Office was making good progress on obtaining full-text data and hoped to make it available via OPS and Espacenet soon for the following countries: BE, PT, PL, RO, HR, IE.

The long-term goal was to have the complete full-text data for all member states back to 1973, but this will take some time. The EPO agreed to send a brief notification to the PatCom secretary whenever new countries' full text became available.

PATCOM confirmed that access to EPO file wrapper information via a web service (OPS) would be desirable, as was already the case for certain patent authorities, e.g. USPTO

7. DOCDB QUALITY

Responding to remarks from PatCom at the previous meeting, the EPO said it had tried to reduce the quantity of changes to DOCDB and increase the advanced warning it gave of such changes.

PatCom said it had noticed problems with Chinese utility model data. The Office confirmed that this was the case and that it was already in contact with SIPO about the matter.

8. QUALITY AT SOURCE” (QAS) PROJECT UPDATE, IN PARTICULAR

Dealt with under agenda point 6 above.

9. LEGAL STATUS INFORMATION PRODUCTS

With respect to the Federated Register, the EPO said it was making good progress. 25 member states and one extension state participated already, and for the deep linking service, a total of 34 member states and extension states were on board.

It added that the European Patent Register now included more information in the file inspection area on PCT-International applications treated by the EPO. Global Dossier now included RSS alerting service for CN and EP files; it also included PCT and CA data, but there were some problems with availability of CA data. The EPO will work with WIPO and with the supplying offices to ensure that data is actually available when a "Global Dossier" link is visible.

10. NPL (NON-PATENT LITERATURE) DATA

PatCom explained that they had put this point on the agenda because ever more publishers were opening up APIs to their data, at least to bibliography and abstracts. The EPO said that bibliographies were in any case available in the EPO's data products as part of the citation information. Increasing amounts of DOI information were also making retrieval easier. It added that it would certainly be prepared to look at ways of improving accessibility to NPL, maybe via linked open data as it had observed a trend towards open access publishing.

11. PATENT TRANSLATE

The Office said it was seeing that the most recent generation of Google/Patent Translate was leading to improved translation quality. The EPO was monitoring the improvements as it was essential to balance fluency and accuracy. In the field of patent data, accuracy was clearly a higher priority than fluency. Currently Patent Translate was processing more than 20 000 translation requests per day. The most requested language pairs were Chinese, Japanese and German into English.

12. GDPR – ANY IMPLICATIONS FOR DATA SUPPLY?

This agenda point referred to the EU's General Data Protection Regulation set to enter into force on 25 May 2018. The Office explained that its current data protection rules had been drafted on the basis of the previous EU regulation, dating back to 1995. Neither the EPO nor the EU institutions themselves were member states of the EU, so the GDPR would not apply direct to them. The EU institutions internal regulation was from 2001, and was, for example, applicable for the EUIPO. The Office said it planned to conduct a review once the GDPR is in force and the once EU had adapted its own regulation.

One concrete implication was in any case that the EPO would have to ask users to comply with the GDPR.

PatCom reported on discussions they had had with WIPO, saying that the view there was that nothing needed to change in patent publications. They realised however that PatCom members would need to be cautious and see what feedback they got after entry into force. The EPO confirmed that, broadly, its analysis agreed with WIPO's in this matter.

A brief discussion followed on "the right to be forgotten", but there was general agreement that this was unlikely to be a frequent issue where patent data was concerned.

13. NEW ESPACENET (STATUS OF DEVELOPMENT AND IMPLEMENTATION PLAN)

The Office explained that the new Espacenet project had two parts:

- Harmonisation of national interfaces – largely complete;
- A new interface, which should appear late 2018, starting with a restricted release (beta version) in summer.

It stressed that the new interface would include a completely new way of accessing Espacenet. Command line search would stay, but there would also be a more intuitive way of building queries, navigating, and filtering and reviewing results. Responsive design would also be a feature.

The new interface would run in parallel for quite a while with the old one, and an intensive communication campaign was in planning. The EPO agreed to inform PatCom when the beta version goes live.

In reply to a question from PatCom, the EPO said it was currently in internal discussion about details such as search limitations (i.e. the no. of allowable search terms, etc). It expected the limits to increase, but there would be no enhancement specifically designed to include "professional" features such as combining search statements or simultaneous searching in different databases. It confirmed that it was not the aim to compete with professional services, but rather to make it easier for the man or woman in the street to conduct a reasonable search.

The EPO invited PatCom to a demonstration of the new Espacenet in its current form, stressing that it was still undergoing a lot of development.

14. UPDATE ON PLANS TO ADD CHEMICAL STRUCTURE SEARCHING / IMPROVING ACCESS TO CHEMICAL INFORMATION

With its current focus on getting the new Espacenet ready, chemical structure searching was only a far-off goal for the EPO, it said. It was, however, aware that WIPO's PATENTSCOPE had started offering some features that went in the direction of chemical structure searching. It added that it was considering introducing some metadata (like MathML and CML) to EP data, but this would not concern the many documents in the databases that come from other patent offices.

15. UPDATE ON CPC

The EPO briefed PatCom on the current situation with the CPC as follows:

- 26 authorities classify with CPC, 16 from EPO member states, 10 sending CPC data to the EPO

Overall, 50 million documents have been classified in the CPC (at family or document level), incl. 47 million patent documents

- 3.3 million documents bear CPCNO allocations (i.e. classifications assigned by national patent offices), of which 2.6 million documents bear CPCNO symbols only (i.e. without CPC allocations at family level)

The CPC website (www.cpcinfo.org) contains everything users need on CPC, e.g. the CPC scheme and definitions, guide to the CPC, combination sets authorised areas, list of ongoing CPC change projects, Notices of Changes and a pre-release area where the upcoming CPC scheme is pre-published one month before entry into force. A special bulk area provides CPC products in electronic format (XML).

Four revisions are planned for 2018:

- 1 January 2018 - CPC 2018.01

- 1 February 2018 - CPC 2018.02
- 1 May 2018 - CPC 2018.05
- 1 August 2018 - CPC 2018.08

Future developments planned for the CPC website, in machine readable format (XML):

- Compilation of changes
- Cross-reference list (CRL)

Introduction of CPC International:

- Promotion of allocations from national offices to simple patent family level (CPCNO disappears)
- Originating office always populated
- Planning: full production April 2019
- No changes in DOCDB XML schema, but population of "originating authority" tag.

On the question of the quality of classification assignments, the Office explained that there was a quality checking system in place between the USPTO and the EPO, but that it was not aware of any formal checks for CPCNO classification work. Many documents in the EPO's working languages would however ultimately come to the EPO examiners to classify, and this would be a chance to give additional classification symbols.

16. AOB

Name standardisation: The EPO agreed to give an update on the IP5 WG on name standardisation at the next meeting between the EPO and PatCom

EPO Patent Information Conference: The Office confirmed that this year's conference would take place from 12 to 14 November, i.e. in a break with previous practice, the event would start on a Monday.

Authority file: The Office informed PatCom that a new WIPO standard for authorities files was available, with further details on the IP5 homepage

Linked open data platform: the Office informed PatCom that its new linked open data platform was set for launch at the beginning April, offering EP data in RDF format, enabling better linking to other data sets.

17. DATE OF NEXT MEETING

It was agreed to hold the next meeting in Vienna in late September or early October. The EPO will co-ordinate with PatCom to find a suitable date.

35th EPO PatCom meeting

20 March 2018, 14.00 hrs, EPO Vienna, room 348

Participants

PatCom:

| | |
|-----------|---------------------------------------|
| President | Ann Chapman, Minesoft |
| Secretary | Jane List, Extract Information |
| | Aurélie Brunet, Questel |
| | Jurjen Dijkstra, LexisNexis Univentio |
| | Armin Foerderer, FIZ Karlsruhe |
| | Robert Fokkema, Lighthouse IP |
| | Margit Höhne, PatentGate |
| | Miriam Plana, CAS |
| | Lee Smith, RWS |
| | Anna Maria Villa, PatentSight GmbH |

EPO:

| | |
|---|--|
| Principal Director Patent Information and European Patent Academy | Richard Flammer |
| Director Publication Director Promotion | Pierre Avédikian Heiko Wongel Davide Lingua Keri Rowles Roland Feinäugle Daniel Shalloe |