

**37th EPO-PatCom meeting
Tuesday, 19 March 2019, 14.00 hrs
Room 348, EPO Vienna**

Summary of the discussions

1. INTRODUCTION

The chairman introduced Susanna Kernthaler as the EPO's new stakeholder relations manager, who would in future be responsible for relations between PatCom and the EPO.

PatCom introduced Linda Williams of Questel and Cinda Harrold of CAS, both having travelled from the USA to participate in the meeting.

2. APPROVAL OF THE AGENDA

The agenda for the 37th meeting was approved.

3. MINUTES OF THE LAST MEETING

The minutes of the meeting of 16 October 2018 were approved.

4. OPS AND FULL TEXT ISSUES

The Office reported that version 3.2 of OPS was now fully operational. In terms of data coverage, the Office had extended full-text data to include documents from WIPO, the EPO, Canada and 16 member states. Due to restrictions in the distribution of full-text document via OPS, this is a subset of the full-text documents available via Espacenet, which total to 87 million. Further extensions to the Czech Republic, Italy, Lithuania and Slovakia were expected later in the year.

Currently OPS usage levels were at two to three million requests per day, with several thousand registered users. The Office added that it filtered out non-active users, so these figures should be a realistic reflection of reality.

On a side note, the Office communicated its intention of offering (within a few weeks' time) a "stripped" bulk collection of the EP full-text data more suited to machine-learning (this product will be in addition to the current packaging of EP full-text data). There were also thoughts on putting this data on cloud platform, but further discussions were required first.

PatCom asked if the Office saw any possibility of acquiring the full-text collections currently available in PatentScope but not in EPO repositories, eg India, Latin American and African countries. The Office asked PatCom to provide a list of countries so that it could investigate. PatCom agreed.

5. ESPACENET OLD AND NEW

The Office said that no further releases of "old" Espacenet beyond bug-fixing were planned. The beta version of new Espacenet had been launched late in 2018 and the Office was inviting users to try it out and give feedback. Current planning was to launch the first full version of new Espacenet in June, but that was not confirmed yet. Work was starting on preparing the national interfaces with special adaptations to the needs and language of the particular country.

Old and new versions would run in parallel for a couple of months, then once all localisations were in place, the Office would shut down old Espacenet.

Responding to a question put at the previous meeting, the Office informed PatCom that currently the only patent register with direct links from the new Espacenet was the European Patent Register. It was, however, working on offering links to as many registers as possible.

So far, feedback on the new Espacenet had generally been very good. It was not the intention to add any major new functionalities. For example, it was planned to keep the existing limit on the number of keywords that could be used in a search. This restriction might, however, be loosened over time. There might also be some improvements on "my Espacenet" list handling or the viewing options. Some modifications to the filtering would come too.

6. LEGAL STATUS INFORMATION PRODUCTS AND REGISTERS

At the end of 2018, the INPADOC database included over 260 million legal event records. Recently loaded data, the Office reported, was from Belgium and Luxembourg, with new data on SPCs from Spain. The database had also been fed with PCT entry data from PatentScope, filling all the gaps that existed compared to PatentScope. This work was not entirely completely, but essentially no authorities should now be missing for PCT entry information (compared with PatentScope).

Upcoming changes would include trying to fill any gaps in SPC data from Spain.

Recent difficulties with data from France had been resolved and normal data supplies were resuming with a slight change of format.

A change in format for Japanese data would enable the Office to speed up the up-loading into INPADOC. Test data had been received from Taiwan for pre-grant events and the Office was optimistic that it would be able to include this in the INPADOC database soon.

For the US, the Office was looking at PAIR data and the possibility of extracting status information to import it into INPADOC.

The Office noted that it was slowly changing the vocabulary it used, away from "legal status" and towards "legal event" data to describe the contents of the INPADOC database, as this was a more accurate reflection of the data.

PatCom asked if the EPO could make European Register data available in bulk at more regular intervals than currently done (e.g. weekly rather than twice a year). The Office replied that it had not abandoned the idea but hadn't yet found a good solution to offer.

PatCom suggested implementing a weekly data feed for OPS and the provision of an indicator that new data was there for retrieval, as this would be a way of reducing the amount of data that flowed for commercial providers to keep up-to-date. The Office agreed to investigate the possibilities.

As discussed at previous meetings, the Office confirmed that its classification scheme for legal event data was complete for the category level, and fully implemented in INPADOC. It was now working on the more detailed level. It was also involved in discussion with WIPO on detailed codes for WIPO ST.27. There was also an IP5 group on same topic that was trying to follow the WIPO approach; all IP5 offices had classified their own codes by category but not by the detailed events, at as far as the Office was aware.

For the European Patent Register, the Office reported on recent successes adding further countries to its deep linking and Federated Register services. It also reported that it had added two new status indicators to the Federated Register:

- "Patent lapse – still in grace period"
- "Request for EP validation pending".

7. NPL DATA; DOI

The internal databases have digital object identifiers (DOIs) for about 61% of the non-patent literature (NPL) they contain, the Office reported. For the NPL citations in search reports, this figures drops to about 50%. If DOIs are not provided, the EPO goes to service providers and tries to obtain a DOI for the respective document.

The Office asked PatCom to express a view on whether it was worth trying to obtain DOIs for the NPL backfile referring to EP-A and EP-B patent publications.

In principle, PatCom liked the suggestion, but had some reservations about the usefulness of DOIs if they were not used by the original publisher. This was because the DOI would not allow retrieval of the original document in such cases. One PatCom member reported that it was well known that journal titles outside US and Europe simply didn't have DOIs. There was no point in the creating them if they don't lead back to the original publisher.

The Office added that there had been a discussion at IP5 level on adding DOIs for patent documents, but was reluctant to proceed as the patent publication number was already a unique identifier. PatCom largely supported the EPO view, but noted that patent publication numbers had a format unknown outside the patent world and were sometimes misquoted in sources not used to such numbers.

On the question of enriching citation data with later-assigned DOIs, the Office said that this was not yet done, but that it would look into the possibility. PatCom suggested using the accession numbers provided in proprietary databases, but agreed it was preferable to refer to the primary rather than the secondary source.

8. DOCDB

In terms of news on DOCDB, the Office said that the frontfile for Italy was now up to date and that it was hoping to add utility model and full-text data for patent documents soon. It had also resolved recent problems with the data feed for data from France. For US data, it could announce improved coverage for design documents (kind code, bibliography, citations and full text), and for South Africa the image backfile was now available (meaning improved bibliographic data in DOCDB).

For citation data, the Office had been receiving data in "rich" format from Korea since the end of 2018, and the Office was optimistic that it would be able to load US rich citation data soon.

The expected replacement of CPC and CPC NO data by CPC International data in April 2019 had now shifted to August 2019.

In response to a question from PatCom on the display of CPC International data in Espacenet, the Office said the current idea was to have the classification symbol followed by the assigning offices in brackets. This would avoid unnecessary repetition of classification symbols. However, redundancy would be more visible in DOCDB and OPS where formal format requirements would mean that classification symbols would be repeated every time in conjunction with the individual generating authorities. However, as these products were accessed by

machine-processing, this should not be a problem. The Office was also looking into a proposal to present to subscribers of these services in order to reduce the redundancy for the XML data.

9. PATENT TRANSLATE

Patent Translate, said the Office, was now a mature product and thus in "maintenance phase" with no spectacular developments to report or planned. Current usage levels were at 18 000 to 20 000 translation requests per day. In the past, there had been some problems with Finnish, Turkish, Hungarian and Japanese translations, but translations seemed to have improved and the number of complaints had dropped. The most frequently translated languages were Chinese, Japanese and German, into English, with the greatest usage being in India, Japan, China and the US.

The Office explained that there had been no change in the situation with respect to providing the language corpora to third parties, as explained at previous meetings.

10. CLASSIFICATIONS

10.1. CPC

The Office said that the updating of the CPC classification scheme to reflect the latest revision of the IPC scheme had been completed in February 2019. Further CPC scheme releases were planned for May and August 2019.

10.2. IPC

The most recent IPC revision had entered into force on 1.1.2019, said the Office, the main changes being the introduction of G16B (bioinformatics) and G16C (chemo-informatics) subclasses.

11. NAME HARMONISATION / NORMALISATION

In the area of name harmonisation, the Office said that no new project was planned. PATSTAT already included four ways of presenting applicants' names:

- the original name as on document
- the DOCDB-standardised name
- the PATSTAT-standardised name (using the methodology of the KU Leuven)
- the name according to the OECD HAN table

The Office stressed that it was always pleased to receive feedback on name harmonisation.

It added that there was also an IP5 initiative to harmonised names on a patent family level, led by KIPO. The next step would be a global mapping table, which would help avoid typographical errors. The Office could publish this on IP5 website for public use, and for companies interested in improving this table, hopefully leading to an improvement in the quality at source.

In response to a question from PatCom, the EPO confirmed that the IP5 initiative would focus at first on the biggest filers and then develop.

Changing topic, the Office reported on developments in data regionalisation (ie data indicating where the applicant/inventor comes from) which permitted econometric or sociometric analysis e.g. for analysing the mobility of inventors. For Europe, the Office used the NUTS table, available in PATSTAT, but was investigating switching to a geotagging system based on coordinates, which would avoid issues such as changes in the postcodes. It would welcome PatCom's feedback if PatCom members already had experience in this area.

12. ANY OTHER BUSINESS

None.

13. DATE OF NEXT MEETING

It was agreed to look for a date for the next meeting between mid-September and mid-October, avoiding the week of 23 September and 14 October.

37th EPO-PatCom meeting

19 March 2019, 14.00 hrs, EPO Vienna, room 348

Participants

PatCom:

President	Ann Chapman, Minesoft
Secretary	Jane List, Extract Information
	Jurjen Dijkstra, LexisNexis Univentio
	Cinda Harrold, CAS
	Paul Peters, CAS
	Armin Förderer, FIZ Karlsruhe
	Linda Williams, Questel
	Robert Fokkema, Lighthouse IP
	Margit Höhne, PatentGate
	Lee Smith, RWS

EPO:

Principal Director Patent Information and European Patent Academy	Richard Flammer
Director Publication	Pierre Avédikian Davide Lingua Roland Feinäugle Susanna Kernthaler Johannes Schaaf Daniel Shalloe