

**EPO-PatCom meeting**  
**Friday, 13 October 2017, 11.00 hrs**  
**Room 348, EPO Vienna**

## **Summary of the discussions**

### **1. INTRODUCTION**

Richard Flammer, Principal Director for Patent Information and the European Patent Academy, welcomed the participants and opened the meeting. He informed the participants that two items (New Espacenet and neural machine translation) had been added to the agenda under “any other business” and a new item, namely “data policy issues” had been introduced under point 4 of the agenda.

The PatCom secretary informed the EPO that Martine Massiera had left Questel and terminated her function as PatCom president, too. Until a new president would be appointed at the next AGM meeting in March, Ann Chapman had been nominated as acting president.

### **2. APPROVAL OF THE AGENDA**

With the addition of the topic “currency of PDFs in OPS” under “any other business”, the agenda for the 34th meeting was approved.

### **3. MINUTES OF THE LAST MEETING**

The minutes of the meeting of 21 March 2017 were approved.

### **4. DATA POLICY ISSUES**

The EPO reported on the progress made in the **Linked Open Data** project. Explaining that this was strictly limited to EP publications, the EPO said it currently had a REST service on the Publication Server to fetch data and planned to implement a new access route based on semantic web concepts. At the beginning of 2018 a SPARQL endpoint would be available, allowing sophisticated linking between EP data and other data sources on the web. The SOAP interface would be abandoned, but the REST service would continue to be available. The EPO was interested in the feedback on the Linked Open Data initiative from the users.

Turning to the topic **Big Data**, the Office reported that it had launched a study with the Fraunhofer Institute looking at ways to use big data techniques under the 3-v principle: velocity, variety and volume. Following

a question from the PatCom group, the Office confirmed that it would make available the results and share them with the user community, possibly during the Raw Data Day next March or the EPO Patent Information Conference 2018.

The EPO informed PatCom that the **prices for patent information products** as of January 2018 had been published. EP data products were now free, but a service fee of EUR 150 applied, to access the data repository. There were some revisions to the prices for non-EP-data products but no big changes. For OPS there was an annual subscription fee of EUR 2,800 as of January 2018 for users who wished to download more than a weekly limit of 3.5 GB. This limit corresponded to the volume necessary to download all newly published EP publications. The Office also explained that it was moving away from written contracts to online terms and conditions, which needed to be accepted before downloading. All data customers would be contacted as part of the implementation of this transition.

PatCom members addressed briefly the issue of timeliness of PDFs in OPS. Some of their clients had reported differences to Espacenet. The Office invited PatCom to provide examples, as the data came from the same repository.

## 5. **WORLDWIDE LEGAL STATUS DATA IN XML**

Concerning the issue of timeliness of INPADOC legal status data compared to Espacenet, the EPO provided the following explanation: over the course of a week, the EPO accumulated new legal status data added to the database, and processed the data over the weekend so that the complete update to the database was ready on Monday, and could be distributed to subscribers on Tuesday. This meant that the time lag between adding the data to the database and provision to the subscribers was between three and seven calendar days (one and five working days). Espacenet on the other hand would provide access to legal status events one day after the data were loaded in the database. This fact meant that time-lag for legal status data added to **existing publications** compared to Espacenet was seven days. However the time-lag for legal status data added to **new publications** compared to Espacenet was only one day.

This was due to the fact that new publications were added to the families in processes run over the weekend, so these became visible in Espacenet & OPS on Monday morning and available to raw data customers on Tuesday. This was the natural consequence of providing a weekly service. The alternative would be a daily service, which would be complicated for all parties.

A PatCom member remarked that the EPO had introduced a publication identifier in DOCDB and that seemed to cause a delay, compared to the earlier practice before 2014. The EPO explained that in the past DOCDB and INPADOC were not linked by the identifier. The introduction of the

identifier provides a major advantage for users of both products. The benefit of the identifier was that the two products matched each other when delivered without a necessity for the user to do the matching work. Linking the two products required the EPO to extract data from both services at the same time, creating this delay.

Answering to a question from PatCom the Office confirmed that the delay for non-EP documents was maximum seven days.

PatCom enquired whether the EPO planned to further enhance the coverage of INPADOC legal status, e.g. through the addition of US pre-grant data.

The EPO reported that it was currently working on US, BR, EP and DE data, namely:

- BR: addition of opposition data
- US: investigate PAIR data on status of applications and granted patents (“abandoned”, “expired”), status not exactly a legal event,
- US fee payments: complete events currently not taken, current data might need to be deleted and reloaded completely,
- EP, DE, UA: review of data loading or introduce new data loading (UA)
- more SPC data from participating states (in the context of UPP)

For EP and JP the EPO had recently improved the event coverage. For EP there was now information on the request for entry into the regional phase for PCT applications, which was available earlier than the entry in the regional phase itself.

## **6. OPS**

The Office informed the PatCom group that a migration from version 3.1 to 3.2 was coming soon, due to the change of the platform for the repository for the PDFs to a new MOSES system (Multimedia Object StoragE Service). Users needed to migrate before end of November (if using image services) and definitely before end of December 2017. There should be no problem if users followed the EPO's recommended route for picking up documents, but those who retrieved data direct might experience some issues. It was recommended that users check their own practice. Version 3.1 would only be updated to the end of November and shut down at the end of December.

Anonymous access would no longer be possible and registration would be mandatory. All necessary information would be published via RSS feeds for OPS, announcements on the Forum and on the website.

The new release would also offer improvements such as full text for FR publications, a png format for images and display of colours and character coded PDFs.

## **7. DOCDB QUALITY**

Answering PatCom members' reports of various short-notice changes to formats in DOCDB, in particular regarding number formats and kind codes, the Office confirmed that there had been a number of such short-notice changes recently. These were largely due to an effort to harmonise the formats used by the EPO with those used at national level. The EPO had carried out many checks with the NPO authority files to ensure conformity with original data. In the future it was planned to be more cautious in number format changes relating (in particular) to application numbers and the EPO would be able to provide a longer period of notice as well.

Concerning number format changes for publication numbers (including kind code), the EPO said it did its best to inform subscribers as early as possible.

The EPO tried to establish consistency with original formats used by the respective national offices according to the principle: what is on the printed document should be the same as in the databases.

Answering a PatCom member's concern that richness of the original data could be lost (e.g. for BE data), the Office explained that it was not familiar with the particular case, but aware that BE had started delivering data regularly again and the EPO was trying to be as faithful to the data as possible, based on today's practice. The EPO was also continuously working on the addition of missing backfile data (e.g. MX).

PatCom members wanted to know whether the limitation with US numbers (10 digits for 11-digit numbers) still persisted. The EPO confirmed that it was still bound by the length of the respective data field.

The Office wanted to know if four weeks' advance notice were sufficient. The PatCom member stated that two months would be better. The Office said it could not promise this, but it would keep in mind the users' requirements.

Concerning IT data, the Office expressed that it was optimistic of making progress in 2018 with resumed data delivery, including full text and some backfile.

## **8. "QUALITY AT SOURCE" (QAS) PROJECT UPDATE**

The Office reported that lately a large number of offices had started regular deliveries of full-text data of national applications, thanks to the EPO's efforts in the context of the Quality at Source project. The offices that now delivered front file full-text were: CH, HR, EE, ES, LT, PT, AT, BE, SK, PL, CZ, IE, GR, RO, BG. Some more were nearly there (IT, NO, DK) and were expected to join soon.

This data was not in Espacenet or OPS yet, as the EPO only planned to load it when full-text search for these offices was activated. This would be done when the backfile was loaded as a critical mass of data was needed for the full-text search to make sense.

According to the project description, the backfile collection would only be able to start after the frontfile was up and running. Tender procedures had been started for the backfile with some few offices but this would still take some time.

Concerning the provision of full text for the commercial providers, the EPO had received the approval of the member states to make the data collected in the context of QaS available to the OPS users with the general terms and conditions applying in OPS.

The PatCom member wanted to know why the front file could not be made available now. The Office replied that it would confuse occasional users and it was difficult to explain why so much data was missing and what was actually searched. Following a suggestion from the PatCom group to provide clear information on the starting point and an indication that data were complete from that point onwards, the Office committed to discuss this internally.

## **9. LEGAL STATUS INFORMATION PRODUCTS**

The Office reported that the Federated Register was being progressively rolled out with 22 offices (AT, BG, CH, CZ, ES, FI, GR, HR, IE, LT, LU, MC, MK, NL, PL, PT, RO, RS, SI, SK, SM, TR) currently in production. A number of member state offices had announced their intention to join by the end of 2017 or early 2018. (DE, FR probably in the next six months). Deep linking was operational with 32 countries.

A new version of the European Patent Register would be rolled out this year, mainly finalising the technical preparations for including Unitary Patent information. The European Patent Register was fully prepared for the Unitary Patent.

Referring to an agenda point of the previous meeting, a PatCom member requested a recap on ECLI. The EPO explained that this was an EU website, where the public could search centrally for decisions from national courts of ten countries, plus ECJ and EPO BoA decisions. This way, one system was bringing twelve sources together for searching. Each decision had a standardised ECLI identifier.

## **10. UNITARY PATENT (UPP)**

As an update to the UPP, the Office explained that, for DOCDB XML, a new kind code "C0" would be introduced with the EP publication number.

This would also be visible in Espacenet as a family member, and in the Register in the dedicated unitary patent register section.

In INPADOC XML all major events relating to UPP would be taken over from the EBD/UPP into INPADOC (about 45 events planned so far). The codes would start with a “U”, for example “Filing of request for unitary effect” (UFI), “Renewal Fees” (UPY) and “Patent expired” (UEX).

Unitary patent would have a bibliographic record but no real corresponding publication.

## 11. FILE HISTORIES

Referring to an item from the previous meeting, a PatCom member explained that procedure-related US patent data was available from a server as a zip file. A US company had built a product that allowed users, for example, to work out the success rate of a particular patent attorney. He asked whether such data could be made available for PatCom to download.

The Office enquired whether the service was based on a web service to retrieve individual files inspections or the bulk data. In the case of bulk data provision there could be a data protection issue in Europe. PATSTAT might be a partial solution to this issue.

The Office promised to investigate internally whether it was possible to create this kind of metadata.

## 12. AOB

A PatCom member addressed cases of **difference in the currency of OPS PDFs and Espacenet PDFs**. The Office proposed that PatCom send concrete examples. In principle, all PDFs were located in one single repository, so there should be no differences in the documents available between the systems.

The EPO reported that it was currently overhauling the interface of Espacenet (**New Espacenet**). A prototype would be presented at the EPO Patent Information Conference in Sofia. The new interface would look very different and change the user experience considerably. More complicated queries would be possible via a query builder and it would have filtering features for refining searches. Fulltext collections would also be added.

PatCom wanted to know if certain limitations of the current Espacenet would be removed (e.g. the number of keywords in one query). This was not set at this stage, the EPO said, but some issues annoying the users, like the “approx” in the number of hits would be deleted. The smart search functionality would be maintained.

The Office explained that focus was on the change of the interface without reducing the functionality, compared to the current system. In parallel, programmers were working on improvements to the back end. Other functionalities like semantic searching or chemical formula searching would definitely not be in the first release but might come later.

Replying to a remark from PatCom that a distinction between the EPO's (free) products and commercial products should be retained, the Office confirmed its intention to stay within its mandate; it was not aiming to enter into competition with others on the market. New Espacenet was rather a facelift and meant to facilitate the use by wider user groups, in particular less experienced ones.

On **neural machine translation** the EPO underlined that three factors were important: quality, productivity and price. An important aspect of Patent Translate was that it was quick and integrated smoothly into the products. The big change recently had been the switch to neural MT for all 32 languages in Patent Translate. The EPO went through an API to use Google – the status now was that all languages in Patent Translate had been moved to neural machine translation to/from EN.

A PatCom member wanted to know if commercial providers could get the aligned corpora in order to do bulk searching in a particular language. Commercial providers have their own systems, but the quality depended on having a corpus to train the system. The Office explained that Google had used patent families with fulltext for training the engines. The EPO had no insight into the technical systems doing the translation. It would, however note the wish and look into it. The corpus delivery to Google had been linked with many caveats, including its use to provide a free service, and not to use it for any other purpose.

Following a question from PatCom as to whether the training of the translation machine with patent families was continuing, the Office explained that the new machine did not distinguish between patents and other texts anymore. Training was a permanent process.

### **13. DATE OF NEXT MEETING**

The date for the next meeting was agreed: 20 March 2018, afternoon.

## **34th EPO PatCom Meeting**

**13 October 2017, 11:00 hrs, EPO Vienna, room 348**

### **Participants**

#### **PatCom:**

Acting President  
Secretary

Ann Chapman, Minesoft  
Jane List, Extract Information  
Jurjen Dijkstra, LNU  
Christiane Emmerich, FIZ Karlsruhe  
Robert Fokkema, Lighthouse IP  
Renaud Garat, Questel  
Miriam Plana, CAS  
Anna Maria Villa, PatentSight GmbH

#### **EPO:**

Principal Director  
Patent Information and  
European Patent Academy  
Director Promotion

Richard Flammer  
  
Heiko Wongel  
Davide Lingua  
Keri Rowles  
Johannes Schaaf  
Daniel Shalloe  
Heidrun Krestel