

EPO-PatCom meeting
Tuesday, 20 September 2011, 10.00 hrs
Room 348, EPO Vienna

Summary of the discussions

1. MINUTES OF THE LAST MEETING

The minutes of the meeting of 22 March 2011 were approved.

2. FOLLOW-UP FROM PREVIOUS MEETINGS

2.1. WIPO - FRONTFILE DATA

For historical reasons, the EPO has for many years distributed the full text of PCT documents English French and German. It has announced that it will stop this by end of 2011 (product 14.18 will thus disappear from the price-list). Subscribers have been informed and their attention drawn to the fact that there is a corresponding product from WIPO.

Replying to a question from PatCom, the EPO confirmed that WIPO is aware that subscribers require electronic access to the data and that distribution on DVD is not an acceptable alternative. The EPO invited PatCom to raise the issue direct with WIPO.

2.2. EUROPEAN MACHINE-TRANSLATION INITIATIVE

The EPO described its on-going discussions with Google related to machine translation (MT):

- the Office signed a memorandum of understanding with Google in October 2010 and an agreement on 24 March 2011
- the agreement foresees co-operation over a longer period of time on providing machine translation for patent documents - the EPO will supply patent corpora in the languages at its disposal, and Google will supply a free-of-charge machine translation service.
- the co-operation is based on the principle of non-exclusivity.
- the plan will be to progress language by language, grouping the languages in batches according to how easy it is to implement them. Batch 1, covering EN, FR, DE, PT, IT, and possibly SE, should be up and running by end 2011. The steps for adding further countries are more complex. In some cases, where there is not a sufficient collection of patent texts in a particular language, the Office will work with

other organisations, eg scientific institutes. Work must start immediately in order to achieve goal of getting everything ready by 2014.

- the Office has created a data repository, to which it pushes the corpora and from which Google retrieves them. Google has indicated that they will "train" their translation engine twice a year. This "training" seems sometimes to take the form of a complete rebuild.
- the EPO is not the only partner for Google in the IP area - Google is, for example, also working with WIPO and OHIM.
- The Administrative Council has agreed to a budget of EUR 10 m over next four years. In addition, the European Commission has committed to providing an additional EUR 4 m.

PatCom asked what a provider should do in order to request access to the repository. The EPO explained that all requests of this nature are being fed to a central point (Paul Schwander at the EPO in The Hague).

Responding to concerns about confidentiality, the Office confirmed that the agreement obliged Google not to save the translations performed by their system. Moreover, requests for translations will go via an EPO server, so Google will only "see" the EPO as its customer, not the end user.

2.3. USER SURVEY RESULTS

The EPO informed PatCom that it would hold a special meeting exclusively for commercial providers on the results of its recent survey during the Patent Information Conference in Kilkenny.

2.4. VISUALISATION TOOL DEVELOPMENT

Progress on a new visualisation tool was good, reported the EPO, and it was now in a consolidation phase, stabilising the system, updating technology definitions, etc. The Office was investigating an extraction tool for PATSTAT to set up small databases. Currently, it was working on how to deal with spelling variants of applicants' and assignees' name and was investigating solutions proposed by the University of Leuven in Belgium and OECD. It was looking at Google Refine as a tool for cleaning the data.

2.5. BNS EXTRACTION

The EPO confirmed its announcement at last meeting that it wanted to offer data extraction from BNS. Some issues remained due to huge data volume, but it had started offering BNS extractions on a small scale. In the meantime it was working on a better solution that would allow larger

extractions in future. Currently the service was an ideal way to fill gaps in data collections, but not appropriate for major data acquisition.

2.6. STATUS OF XML ST 36 SCHEME

The Office remained fully committed to ST.36, it said, noting that WIPO's XML-for-IP initiative seems to be less in focus at the moment. The EPO was still fully involved, however, and participated in a task force coordinated by WIPO.

3. IP5 - PROJECT DEVELOPMENT (CPC)

CPC (Co-operative Patent Classification) is currently one of the most important developments in the IP5 area. The transition from ECLA to CPC will see a merge of ICO and ECLA codes, the Office expecting to have agreed an algorithm for applying the conversion from ECLA/ICO to CPC by the end of October.

According to planning, ECLA will be frozen on 1 July 2012 to allow a smooth transition. The expected day of entry into force of the CPC is 1 January 2013. The CPC scheme itself should be available to the public by the end of 2012. There will be a replicated storage master of the scheme at the USPTO and EPO, with both organisations having equal rights. This will enable a third partner to join the co-operation on equal terms, which is important considering that there is a growing expectation that Asian countries will also adopt CPC for classifying their documents.

The transition will require substantial training for examiners; the EPO will also train US examiners.

PatCom asked if there was any intention to maintain ECLA in parallel. The Office explained that this was very unlikely. PatCom replied that in that case it would be useful to have a flag showing whether a CPC symbol had been auto-generated during the conversion, or manually added later on. The EPO said there would be an audit trail showing the date of a classification symbol being allocated to a document and the source. It would only know at a later date to what extent the information will be publicly available.

With regard to the backfile, the Office said it would probably create a special CPC backfile (hopefully before the end of 2012) to facilitate the loading into systems - it would be a bit smaller than a complete reissue of DOCDB. It would also publish a concordance table for the conversion from ECLA to CPC. There would be no attempt to establish a concordance between CPC and US classes.

4. STATUS REPORT ON THE UNITARY PATENT

The Office reported that it was on target with preparations to fulfil its duties with the unitary patent if it entered into force, as planned, in 2013. The idea is to build on existing products and services at the EPO to fulfil this important mission for Europe.

Assuming no changes will be made to the EPC, the Administrative Council will need to set up a Select Committee which would need to establish the Implementing Rules for the unitary patent. The Select Committee would address issues such as a Register for the unitary patent, renewal fees, etc.

Entry into force of the unitary patent on time depends on a successful conclusion to the work of the Select Committee and, significantly, on a solution to the question of jurisdiction for the unitary patent.

5. QUALITY OF DATA SUPPLIED BY EPO

The EPO reported on the following developments:

DOCDB/XML

- Q3/2011 - US (re)classifications, Japanese FI and Fterm and Patent Abstracts of Japan; distribution of complete DOCDB backfile including this new data
- In the pipeline - considerable enhancements to cited references in the course of 2012; these improvements will mean, however, that it is necessary to make changes to the XML schema; this is also an opportunity to render the DOCDB XML schema ST.36 compliant; details will be made available as early as possible.

Survey

- A recent survey amongst subscribers to raw data products has shown that subscribers are generally well equipped for handling large data quantities. Therefore the EPO will reduce its practice of split up data deliveries into smaller chunks.

INPADOC worldwide legal status

Achievements since March 2011

- New Latipat countries: Cuba (done) and Argentina (coming soon)
- Extension of the legal status for Poland with additional events
- Issues with unexpected format change at DPMA resolved - here, the EPO thanked PatCom members for the support and understanding while it got to grips with the format changes. PatCom thanked the EPO in return for the good communication on the issues it faced

- Japanese legal status - the Office has received data from the JPO and started a detailed analysis

Modernisation of the legal status master database

There are plans to modernise the legal status database, the EPO announced. This will be an exercise that will stretch over many years and will be divided into three phases:

Phase 1	Infrastructure and architecture, improve connection between PRS legal status and DOCDB bibliographic data	<u>Objective</u> : Improved data quality within the constraints of the current product
Phase II	Inventory of user requirements - within both user communities, the end-users as well as the commercial providers	<u>Objective</u> To involve the user community at the very early stages
Phase III	Migration to new exchange format	<u>Objective</u> : Offer a richer and more versatile product, better suited to meet the business needs

With regards to Phase II, the Office will exploit the opportunity of the Patent Information Conference in Kilkenny to gather users' requirements:

- A. In a discussion round for end-users of legal status data
- B. In a separate meeting with the subscribers to the raw data product, focusing mainly on technical aspects and design features.

6. ESPACENET DEVELOPMENTS

The first half of the year saw the release of the newly designed GUI. Since then the Office has been collecting feedback on the new design, and plans to implement a set of changes based on that feedback in the next release. There will also be changes in functionality, including improvements in ECLA navigation and visualisation

Additionally, resources will be dedicated to improving the current level 1 full text search. For instance, new language specific indexes should provide better cross lingual and conceptual oriented searching.

The latest statistics showed 800 000 visits so far in September 2011, 350 000 unique IP addresses. 10 m PDF page downloads, and 5 m searches. In terms of geographical distribution, nearly 50% of users seem to be based in Asia, 30% in Europe and 20% in North America.

7. OPS DEVELOPMENTS

2011 was a year of consolidation, said the Office, following introduction of REST based architecture in 2010. OPS ver. 2.6.2, in both SOAP and REST was scheduled for October. Two products are available in REST only, all others in REST and SOAP. Future additions will only be in REST, so the message is that users should try to migrate to REST. As a result of the launch of the new version, there will be some small URL changes which the EPO will communicate to users in good time.

On the statistical side, the Office reported that the number of users seemed to be stable (2000-2500 unique IP addresses per week), as did usage (4 m requests per day, corresponding to about 100GB of data).

Finally the Office announced that it planned to introduce some automatic information to users on OPS traffic levels, so that they could self-regulate how they use OPS, avoiding periods of heavy loads on the system.

8. PUBLICATION SERVER - WEB SERVICES INTERFACE

The publication server is now available for machine access using site map protocol or REST. In response to a request from the EPO, PatCom agreed to ask its members to provide feedback as to whether they prefer site map or REST, and whether they experienced any problems with the limits on the system.

9. ANY OTHER BUSINESS

9.1. EPOQUE

PatCom asked about the EPO's plans to make EPOQUE publicly available. The EPO said that the subject was on the table within the IT roadmap project, but that there was no change of policy at the current time.

9.2. FREQUENCY OF MEETINGS

PatCom suggested holding the autumn meetings within the framework of the Patent Information Conference in future. The EPO agreed.

9.3. DATE OF THE NEXT MEETING

It was provisionally agreed to hold the next meeting on 27 March 2012, and to hold the annual XML raw data day on 28 March.

**EPO-PatCom meeting
Tuesday, 20 September 2011, 10.00 hrs
Room 348, EPO Vienna**

List of participants

PatCom

President of PatCom
Secretary of PatCom

Willem-Geert Lagemaat, Lighthouse IP
Georg Schultheiss, formerly FIZ-Karlsruhe

Pierre Buffet, Questel
Martine Massiera, Questel
Ann Chapman, Minesoft
Jurjen Dijkstra, LexisNexis Univentio
Barry Dunne, CAS
Armin Förderer, FIZ Karlsruhe

EPO

Principal Director PD 5.4
Director 5.4.1
Director 5.4.2
Director 5.4.3
Director 2.7.1.4

Richard Flammer
Pierre Avédikian
Heiko Wongel
Günther Vacek
Miguel Albrecht

Nigel Clarke
Valérie Gray
Davide Lingua
Johannes Schaaf
Daniel Shalloe