SUBJECT: Report of the meeting between the EPO and the PatCom of 17.3.2009

SUBMITTED BY: President of the European Patent Office

ADDRESSEES: Working Party on Technical Information (for information)

SUMMARY

The PatCom Group was established towards the end of 1999 and is an association of commercial companies that provide patent information products and services that include EPO data. It has been agreed to hold regular meetings with PatCom on a six-monthly basis. The meeting on 17 March 2009 was the 18th such meeting.

This document has been issued in English only.

TABLE OF CONTENTS

———————————————

I.   STRATEGIC/OPERATIONAL

Operational

II.   RECOMMENDATION

For information.

The Working Party on Technical Information is requested to:

- take note of the following report of the meeting between the EPO and the PatCom of 17.03.2009.

III.   MAJORITY NEEDED

N/A

IV.   CONTEXT

The PatCom Group was established towards the end of 1999 and is an association of commercial companies that provide patent information products and services using EPO data. The EPO has agreed to hold regular meetings with PatCom on a six-monthly basis.

The 18th meeting between the PatCom Group and the EPO took place in Vienna on 17 March 2009. The present document is a report of that meeting.

The agenda for the meeting, chaired by Principal Director Patent Information, Richard Flammer, was based on a list of points submitted in advance by PatCom.

PatCom was represented by:

Lighthouse IP (PatCom Presidency), Incom, Lexis-Nexis Univentio, Questel, Thomson Reuters, FIZ Karlsruhe, CAS, Dialog, Matrixware, unycom and Scipat.

V.   ARGUMENTS

A.   REPORT OF THE LAST MEETING

PatCom confirmed their full agreement with the report of the previous meeting(CA/T 40/08).

B.   FOLLOW-UP FROM PREVIOUS MEETINGS:

The EPO submitted a written status report on issues addressed at the previous meeting (see Annex). The chairman stressed that this new practice of providing a written report, and of monitoring actively agreed actions was something that he personally welcomed, and that he hoped to be able to continue in future.

a)   Chinese services

PatCom remarked that China remained high on the agenda for commercial service providers and that PatCom members were always interested in hearing what the EPO and the Chinese Patent Office (SIPO) had to report, particularly with regard to translations from Chinese to English.

The EPO it had worked on a number of different solutions with regard to translations, some of which were beginning to bear fruit. Moreover, the work done by SIPO had led to a substantial improvements in automatic translation and the EPO was now able to offer a better service to its examiners as a result.

The EPO reported that it had also been able to agree with SIPO the procurement of manually translated abstracts in English of Chinese utility models. The EPO would, it said, be able to provide the backlog for this data, as well as frontfile information. The delay from publication date to loading into the EPO databases would probably be some three months, but this needed to be confirmed when the data began to arrive.

The Chinese utility model data will be available for purchase as part of the usual DOCDB raw data product.

b)   Automation Plan – latest status

It has been agreed that the EPO will provide PatCom with an excerpt of the Automation Plan to PatCom each year. The 2009 Plan is in preparation at the moment and will be ready for the autumn PatCom meeting.

The EPO informed PatCom that, in comparison to previous years, future plans would be smaller and contain less detail - this was in response to a request from the Administrative Council. Furthermore, a budget decrease (currently set at 5%) in automation projects was expected for 2010.

c)   Use of "ICO" for "spam" patents

At the previous meeting, PatCom presented the issue of patent applications being "abused" as a means to publish promotional information. The EPO's legal analysis has shown that there was very little room for manoeuvre, legally, and that these documents would certainly have to be published. PatCom had suggested using the EPO's internal "ICO" codes to identity these "spam" patents and filter out of any search. The EPO's investigations in the meantime had, however, shown that this would involve a major change in the EPO's classification tools and would thus be considered a low priority for the moment, especially considering the low number of cases.

d)   Availability of dictionaries for machine translation

With regard to the machine translation dictionaries that it was developing in co-operation with various national patent offices, the EPO explained that the underlying policy, agreed in the Council in 2004, was that the dictionaries would in principle be made available as a raw data products to interested users. Some member state delegations had taken the line that a co-decision with the member states might be required in order to release the dictionaries to the public.

At the time of the meeting, the EPO was able to report that the "closure reports" on the development of dictionaries for German/English and Spanish/English were under discussion and included a sentence stating that parties agree to the distribution of the dictionary as a raw data product, according to the general policy line. Thus, PatCom could expect to see these two dictionaries very soon in the EPO price list as raw data product, for purchase at marginal price, in line with EPO policy.

The EPO added that, in the context of the London Protocol, it intended, through its Publication Server, to provide translation for European patent documents into the three official languages of the EPO. It was planned to make the language pairs English/German and English/French available in a first step.

C.   "BIG FIVE" (=IP5) PROJECT DEVELOPMENT

The Chairman explained that an IP5 meeting had taken place on 9 and 10 March 2009 in Tokyo. The draft minutes that had been produced following that meeting showed that considerable discussions had taken place on the ten "foundation projections" and that a detailed list of points had emerged.

It was nevertheless clear that these projects were all in a very early stage. There would be, he said, a meeting of the deputy heads of the IP5 offices in Munich on 10/11 June, at which point more detailed plans should be available. The next meeting of the heads of offices was likely to be in September.

PatCom asked the EPO to explain some of the thinking behind the two projects it was leading:

a)   **Hybrid Common Classification**

A main goal, reported the EPO, of the foundation projects was to facilitate the development of trust across the Offices. A common classification would be an important step forward in that direction, on the understanding that if all examiners in the IP5 offices had similar classification schemes, they could carry out similar searches.

The hybrid scheme would seek to bring in the best in classification field from around the IP5. The vision was that publicly accessible, very rich classification schemes would come into existence that represent a very powerful tool for search for anyone concerned with IP. All parties understood very well that this project, if it went ahead, would mean substantial investments by the IP5 offices to reclassify all their documents (2000-4000 man-years, perhaps), and there was no horizon or time-planning for the project available yet at this stage.

b)   **Common Documentation Database**

The situation with this project was that there was a common understanding that substantial efficiency benefits could be achieved quickly. For instance, if all offices were to publish bibliographic data that entirely corresponded to WIPO standards, there would already be major savings. Discussion among IP5 members had shown that it was important to identify benefits both at a global and a local level.

Some offices still exchange paper, so getting into fully encoded electronic exchange is an important first step.

The EPO said that in its view, of all the projects in IP5, this one promised the best results in the shorter term.

PatCom asked if there was a role in this project for players outside the IP5. The EPO responded that was not currently the case because it would be premature to complicate the situation with additional players.

The EPO promised to included a status report on IP5 projects in time for next meeting (ACTION for status report!).

D.   FUTURE OF THE EPO PATENT INFORMATION CONFERENCE

In late 2008, the EPO carried out a professional survey among participants of its Patent Information Conference and other similar events. The results demonstrated that there was a strong support for the conference among users, and that it was seen as a major event in the patent information community. An article on the survey was published in Patent Information News 1/2009, which appeared in March 2009.

After having seen the survey results, the EPO's President had given the green light to continue the conference in 2010 and beyond.

PatCom welcome the news that the conference was not going to be abandoned by the EPO and said that it had been worried by rumours that this might happen. It would be very unfortunate, they stressed, if the conference were to be stopped, because it would mean there was no "objective" event (i.e. not driven by commercial interests) of this kind in Europe. If necessary, PatCom would be ready to help in organising the conference, or providing support. It was a very important event, also from the commercial side, they said.

The EPO remarked that the survey results confirmed that people came to the conference to discuss technical issues with all the players. The exhibition - and the participation of representatives of the commercial patent information sector - was very important in this context.

E.   USER SURVEY

The large-scale survey on the use of patent information in Europe that the EPO had wanted to carry out in 2008 had ultimately to be abandoned for various technical reasons. The EPO still intended to carry out the survey, and this would happen in 2009, it said. The survey would be performed in batches, which would make it easier to manage and to budget, and possibly also make the analysis easier. The basic format and philosophy would be the same as the survey it performed in 2003.

F.   QUALITY OF DATA SUPPLIED BY EPO

a)   Data delivery

PatCom reported that they had observed a more stable delivery of data and of throughput since the last meeting, and they thanked the EPO for the steps it had taken to achieve this.

The EPO explained that, following PatCom's suggestion, it had begun to spread the delivery of data across the week to avoid high loads on the system and to help ensure high data transfer speeds. It had also implemented "push" and "pull" information so that users were both actively informed and could also look up for themselves the status of the available data.

b)   Indian data

PatCom reported a problem with data from India, the delivery of which seemed to have stopped in 2004, following a change to Indian IP law. The EPO replied that there were indeed some issues to be resolved where data from India was concerned, most notably in terms of the format used for delivering the data. It stressed, however that the number of domestic Indian patent applications was still rather low, so the amount of information "lost" as a result of these problems was not huge. It recognised, though, that India would be a very important country in future and said that it had been given a high priority for future work.

c)   US reassignment data

PatCom asked about US reassignment data. The EPO said that there was a quality problem with the backfile of the reassignments, and that the information had some reliability problems, due to the need to do a lot of manual work on series numbers to avoid conflicts of duplicate numbers from different series. The inclusion of the series number in the data would avoid this problem, but this will only happen from 2010. The conflict concerns a small low number of cases from 2001 onwards.

d)   full-text data

The EPO said it was investigating how to make public data products from the new Full-text Master Database. It could already offer full-text data for EP and WO

documents and was currently working on converting it to ST 36 with UTF-8 coding. The data included embedded images. Some legal clarifications were necessary for other countries (e.g. DE, FR, GB).

e) **legal status data from China and Russia**

The EPO announced that it had been able to obtain the Chinese legal status back to 1985 and that this would be included in the INPADOC legal status files in April 2009. In terms of content, the data would report on approximately 40 legal status events, rather than the 16 available online from SIPO.

The EPO had received two test packages of Russian legal status data and said it was confident of soon being able to offer the front file for 2009 onwards.

G. *esp@cenet* **DEVELOPMENTS**

2008 had seen a number of new features introduced into esp@cenet, most visibly SmartSearch, currently available as a beta version, which includes command-line searching, albeit within the usual *esp@cenet* technical limitations. The EPO said that once it was satisfied that it had fine-tuned the tool as well as possible, it would make SmartSearch available as a download-able "widget".

Other new features include:

- an increase of the maximum number of documents in the "My patents" list to 100

- an export function for lists

- an increase the page limit for downloads to 500 pages for EP/WO docs and 250 pages for others.

Plans for 2009 include:

- fulltext searching of EP documents

- a sort function for the results list

- highlighting of search terms

- extended citation coverage to include applicant citations and citations in opposition procedures

- a "claim parser" which will visualise the relationship between claims in a document, identify dependent and independent claims

- an RSS feed to alert users whenever a document fulfilling certain criteria becomes available in the database (a very simple arrangement to allow users to get an information feed about particular technologies or a particular inventor, for example).

PatCom asked for more information on the claims parser. The EPO replied that the idea had started life as a private development by an EPO examiner, and at some point the feedback had come that it would be a very interesting thing to have in *esp@cenet*. This has been confirmed in feedback from the patent attorney community. PatCom remarked that this type of product already existed on the commercial market, so those asking for it could already get it. It seemed they now wanted it for free.

As in the past the EPO agreed to provide a release plan for *esp@cenet*. It said it would do this after the following week's *esp@cenet* meeting with national patent offices.

a)   Statistics on usage

The statistics from the first week of February showed the following usage levels for *esp@cenet*:

- 800k PDF pages per day

- 540k bibliographic views per day

- 26k IP addresses per day.

The EPO added that it was working on a way of producing more detailed statistical analyses on the use of *esp@cenet*. It said it would provide more information on this at next meeting.

H. OPS DEVELOPMENTS

The EPO reported that OPS (Open Patent Services) had been rebuilt in 2008 (albeit with slight delays due to problems with the EPOQUE search engine) and newly launched as version 2 in November 2008. Versions 1 and 2 were currently running in parallel and would continue to do so until September 2009. Some additional new features were available and the coverage of the fulltext data had been extended a little to include Austria, Switzerland and Canada (claims only).

There were some small problems with people who use programming packages, and the EPO was working on making OPS compatible with these standard tools.

To come in 2009:

- Register data via web services

- ECLA classes as XML file

- Extension of full-text coverage, depending on agreements with the national offices concerned.

a) Statistics on usage

OPS version 1:

- 1700 individual IP addresses per week

- 2 million requests per day;

OPS version 2:

- 300-400 IP addresses per week

- 200 000 requests per day.

I. DEVELOPMENT OF PUBLICATION SERVER, OPEN WEB SERVICE INTERFACE

The EPO reported that use of its Publication Server was very stable, with around 5000 visitors per day, and a slight increase in the data volume. The technical

solution for the Publication Server is available for other patent offices who are interested. Poland, for example, is using it. This seems to have led to some data standardisation which, in turn, had meant that the EPO receives more reliable data supply.

As stated in section III.D. above, the EPO plans to implement automatic translation for English, French and German for the Publication Server. It is enthusiastic to make progress on the provision of gene sequence data, but this is proving difficult.

J.   ANY OTHER BUSINESS

a)   Patstat

Responding to a question from PatCom, the EPO said that its PATSTAT database had been available for three years under restricted distribution to the OECD statistical task force members and academia. However, since the data was practically, an extraction from DOCDB, enriched with applicant and inventor address data from US/EP, it was entirely public, so there seemed no reason to restrict access any longer. PATSTAT, which was now generally available as a raw data product, grouped data into tables, making it easier to perform statistical analyses. Sample files were available on the internet.

b)   GPI (global patent Index)

PatCom asked about the concept of the EPO's planned "GPI" product and how it fitted with *esp@cenet* and "EPOQUE net". The EPO responded that GPI was a simple evolution of existing offerings. The online capability of MIMOSA had made it possible to take the existing Globalpat concept of some years ago and put a it on line, searchable with the MIMOSA search software. The EPO said it was important to think of solutions that did not impact the internal resources of the office (every user on esp@cenet put a load on the EPO's systems; this was not the case with GPI).

c)   date of next meeting

It was provisionally agreed to hold the next PatCom meeting on 15 September 2009. If not possible, the preferred alternative will be 22 September 2009.

K. LIST OF POINTS FOR FOLLOW-UP AT NEXT MEETING

- The EPO will provide a status report on IP5 projects.

- The EPO will provide a release plan for *esp@cenet*.

- The EPO will provide more information on its statistical analysis tools on the use of *esp@cenet*.

VI. ALTERNATIVES

None

VII. FINANCIAL IMPLICATIONS

None

VIII. LEGAL BASIS

Not applicable

IX. DOCUMENTS CITED

None

NOTE TO THE PRESIDENT
OF THE PATCOM GROUP

REPORT ON ACTIONS RESULTING FROM
THE PATCOM/EPO MEETING OF 26 SEPTEMBER 2008

1.   INTRODUCTION

During the meeting of 26.9.2009, the EPO promised to investigate various issues and either report back in writing, or provide a status report at the next meeting (i.e. on 17 March 2009). The document gives feedback on all the action points noted at the September meeting, namely:

- To issue a report on what developments we expect from China soon
- To collect input from PatCom for the patent information survey
- To send an abstract of the Automation Plan as soon as it is approved
- To include IPC-related issues in the next raw data day
- To investigate using ICO codes to identify "spam" patent applications
- To provide an update on the machine-translation dictionaries available as raw data

2.   NEWS ON DATA FROM CHINA

a)   English abstracts for Chinese utility models

The Chinese Patent Office (SIPO) has delivered English translations to the EPO of the official SIPO abstracts for Chinese utility models covering the period

1985-2007. This data will be loaded into Doc-DB and supplement our Chinese abstract coverage and will be available in *esp@cenet* and EPODOC under EPOQUE. A sample record is given below for information. Claims are not covered and there are no plans for SIPO to deliver this data.

Example record:

```xml
<?xml version="1.0" encoding="GB2312" ?>
- <Patent>
    <APNNO>200520122613</APNNO>
    <PUBNR>000000000</PUBNR>
    <APPNR>2784662</APPNR>
    <APD>2005/9/17</APD>
    <PUD />
    <GRD>2006/5/31</GRD>
    <GRPD>2006/5/31</GRPD>
    <APPD>2006/5/31</APPD>
    <NC>91</NC>
    <PNUM>4</PNUM>
    <FNUM>2</FNUM>
    <CNUM>1</CNUM>
    <AGENCY />
    <FIELDC />
    <ADDRESS />
    <AGENT />
    <TITLE>Electromagnetically controlled sliding buffer gasoline economizer for motorcycle</TITLE>
    <ZIP>116001</ZIP>
    <IPC>F16D 41/12</IPC>
    <IPC>F16D 41/30</IPC>
    <IPC>F16D 48/00</IPC>
  - <PRI>
      <CO>CN</CO>
      <NR>2005200384 0.X</NR>
      <DATE>2005/1/21</DATE>
  </PRI>
    <APPL>Liu Jie</APPL>
    <INVENTOR>Liu Jie</INVENTOR>
    <ABSTR>The utility model relates to an electromagnetically controlled sliding buffer gasoline economizer for a motorcycle, which comprises a chain wheel, a buffer and a bearing, wherein one side of the buffer is provided with a convex column and the other side is provided with an opening circular cavity. The perisporium of the circular cavity of the opening circular cavity is uniformly distributed with convex teeth, and a control wheel is arranged in the circular cavity of the buffer and is respectively connected with the outer side of a fixing sleeve and a lining by the bearing. The control wheel connected with one end of the chain wheel is provided with a current conducting bearing, and the control wheel is provided with an electromagnet, a capacitor and two sets of ratchets in opposite directions. The ratchet is respectively engaged with the convex tooth uniformly distributed on the perisporium of the circular cavity of the buffer, wherein one set of ratchet is connected with the electromagnet controlled by a control circuit, which can make the ratchet and the convex tooth in an engaged state or in a separate state. The utility model not only can transfer motor dynamic force, but also can make a transmission system do overrunning slide thereafter under the condition of an idle speed. When necessary, the rear part with the high-speed slide and the front part with the idle speed dynamic force can be connected at any time; the sliding function can be closed, and the utility model can be used as an ordinary buffer; thus the goals of energy saving and reduction of tail gas pollution are realized.</ABSTR>
</Patent>
```

b) Automatic translation

A joint EPO-SIPO working group is studying the quality of Chinese-English machine translation outputs. An evaluation framework has been put in place in order to provide quantitative quality measures corresponding to user's satisfaction and accuracy criteria. Most of the efforts are now focusing in finding technical solutions to improve the current SIPO/CPIC machine translation system. One possible improvement factor, as proven in the case of European languages, is the compilation of domain-based language terminology. Another factor remains the ability of the MT systems to constrain themselves during the segmentation process [the segmentation process consists in finding group of characters representing terms. If this process is properly constrained then the probability of finding the right group increases. The way of constraining the system is to look for term having a high probability of appearing, such as:
- those present in a dictionary
- those already validated within another process (manual?)
- those likely to appear in a known sentence context.
If the system is left without constraints, then it faces a combinational explosion.]

All these aspects are addressed on a regular basis during exchange sessions with SIPO counterparts.

c) SIPO's free web-based machine translation tools

Since 25 April 2008, SIPO has been offering two free online machine translation services for Chinese patent and utility model documents, which can be accessed via different English search interfaces. Both systems are currently still in a trial phase. The translation engine developed by SIPO and its subsidiary organisation "China Patent Information Center" (CPIC) is available at www.sipo.gov.cn/sipo_English/ and http://www.cnpat.com.cn., English machine translations are currently available via this source for documents published between 1985 and May 2007. The second translation engine was developed by SIPO's "Intellectual Property Publishing House" (IPPH) and can be accessed at http://english.cnipr.com/enpat. Currently, it allows the retrieval of machine translations for documents published between 1985 and December 2007 (patchy coverage between May and December 2007).

It is expected that SIPO representatives will shed some more light on the current status of their machine translation project and services they plan for the future during this year's "East meets West" forum in Vienna (23-24 April 2009).

d) Machine-translated abstracts for Chinese utility models and cross-lingual patent retrieval

SIPO's "Intellectual Property Publishing House" (IPPH) launched a new English search interface in April 2008: http://english.cnipr.com/enpat. Use of this database is currently free. Users can search in English abstracts for patents and machine-translated abstracts for utility models. The site also provides a cross-lingual patent retrieval tool whereby users can input English keywords which are automatically translated into Chinese and searched against Chinese full texts (patents and utility models). The result of the search is English abstracts or machine-translated abstracts with a possibility to display and download the Chinese original or a machine translation of the full text of the document.

3.   PATENT INFORMATION SURVEY

As follow-up research to its 2003 survey, the EPO had planned to carry out a comprehensive survey of how patent information is used in Europe. However, due to budget restrictions at the current time, this project has been put on hold.

4.   ABSTRACT OF THE AUTOMATION PLAN

The Office maintains its promise to provide PatCom with a copy of the Automation Plan as soon as it has been approved by the Office's governing bodies. The document is, however, not available yet.

5.   INCLUSION OF IPC-RELATED ISSUES IN THE NEXT RAW DATA DAY

The matter has been discussed in bilateral discussions between Rob Willows (for PatCom) and the EPO. Regrettably, due to time-tabling difficulties, the EPO's leading experts will not be available for the Raw Data Day on 18 March, but the Office will nevertheless prepare some basic information slides in time for the event.

6. USING ICO CODES TO IDENTIFY "SPAM" PATENT APPLICATIONS

A proposal was prepared and presented to DG 1 management. The discussion concluded in that, in view of the small number of known cases, the effort needed for such a procedure does not translate in appropriate benefits.

7. AVAILABILITY OF MACHINE-TRANSLATION DICTIONARIES AS RAW DATA

Upon closure of the first two projects (German/English and Spanish/English), the Office will plan the design and production of data products for the dictionary pairs. It is expected that these could be made available by mid 2009.