



Europäisches
Patentamt
European
Patent Office
Office européen
des brevets

A view on some international initiatives on data acquisition and standardisation

Raw Data Day 2019



Nikolaos Chardalias

IT project co-ordinator – European & International Co-operation

March 2019

QaS achievements - benefits for Espacenet

Agenda

- QaS scope
- QaS benefits
- QaS status
- Authority File
- Bibliographic data coverage
- Full-text data coverage
- IP5 cooperation

Quality at Source

Objectives:

- Establish harmonised **front-file flows** of patent data from all EPO member states
- Digital, **medialess** deliveries following the **preferred standards** (XML ST.36, PDF) and best practises
- Eliminate gaps in the data collection
- Focus on **quality** criteria: **completeness, correctness** and **timeliness**

QaS – benefits on data quality

Standardization of data (WIPO standards): automatic validation/correction and decrease of the manual corrections (publication/application/priority numbers, classification...)

- accuracy of application/priority numbers allows accurate and complete families (75% of the cited documents belong to family larger than 2)
- accuracy of the publication numbers allows accurate citations
- enriched citations support examiners and AI implementation
- full-text search in Espacenet a major improvement in patent searching

QaS – benefits

On average **3 days after delivery** the documents are available in Espacenet / OPS / EPOQUE Net

Access to data:

- FTP : All Offices
- Web service, in pilot. No extraction requested, delivery on-demand

Validation

- At receipt stage: XML validation
- At processing stage: Automated standardisation and correction

QaS – status

- Complete Digital document : focus on Images and Full text

	Biblio	Image	Full text
Number of standardised flows	35	33	23

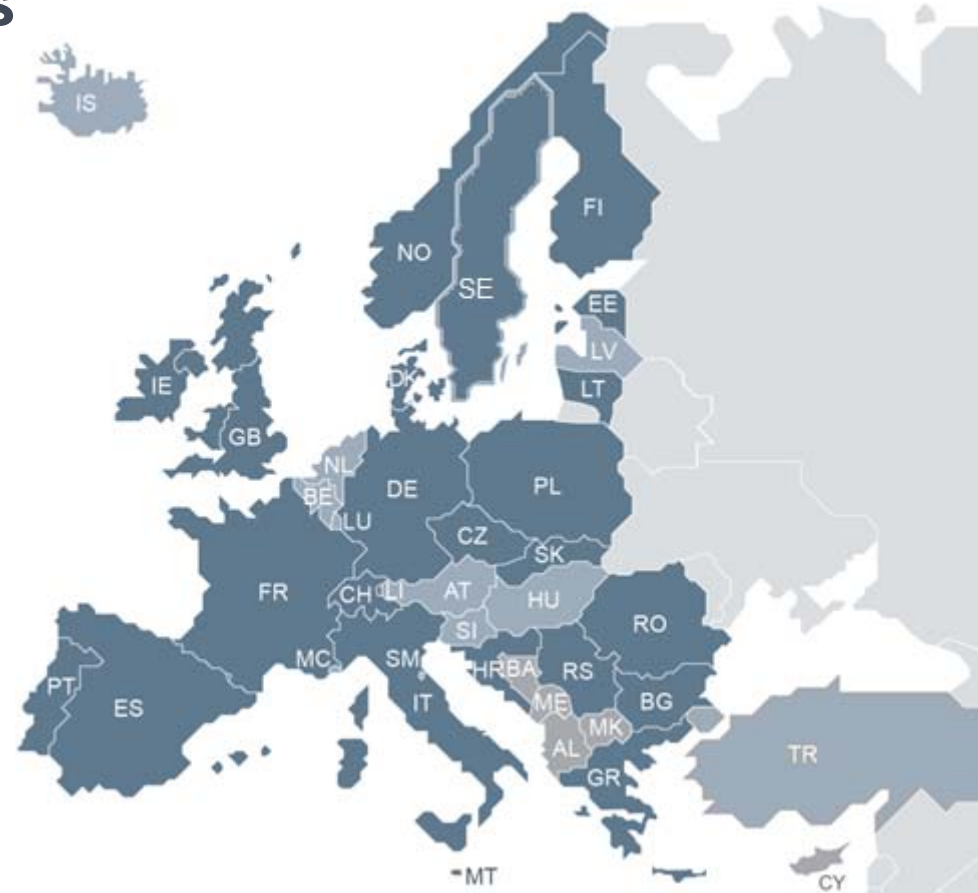
- Collection to be accurate & completed: Authority Files

	National offices
Authority File analysis	20

Quality at Source – status

- Data delivery: St. 36
- Data delivery: St. 36 + full-text

Batch 1	: AT, HR, IT, LT, PT, TR
Batch 2	: BE, EE, ES, HU, NO,
Batch 3	: AL, CH, CZ, FI, IE, NL
Batch 4	: BG, DK, GR, PL, RO, RS
Batch 5	: CY, IS, LV, SE, SK, SM
Batch 6	: DE, FR, GB, LU, MC, MK, MT, SI



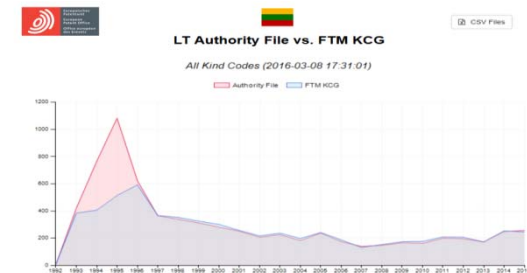
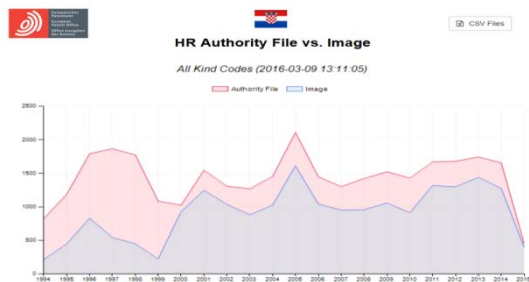
Authority File

An Authority File provides a definitive list of all patent documents issued by a patent office.

The purposes of authority files are:

Assess the completeness and correctness of National Offices collection against EPO collection.

- Report on gaps to be filled and corrections to be done
- Fill the gaps and correct data when requested



Authority File

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE authority-file SYSTEM "authority-file-v1-0.dtd">
<authority-file country="ZZ" date-produced="20150101">
  <authority-file-doc>
    <document-id>
      <country>ZZ</country>
      <doc-number>12345678</doc-number>
      <kind>B1</kind>
      <date>20011231</date>
    </document-id>
    <pub-ind/>
    <application-reference>
      <country>ZZ</country>
      <doc-number>999998</doc-number>
      <kind>A1</kind>
      <date>20100131</date>
    </application-reference>
  </authority-file-doc>
</authority-file>
```

Publication

Application

Authority File

```
<?xml version="1.0" encoding="UTF-8"?>  
<!DOCTYPE authority-file SYSTEM "authority-file-v1-0.dtd">  
<authority-file country="ZZ" date-produced="20150101">  
  <authority-file-doc>
```

Publication

```
<document-id>  
  <country>ZZ</country>  
  <doc-number>2233445566</doc-number>  
  <kind>B1</kind>  
  <date>20090321</date>  
</document-id>  
<pub-ind/>
```

Application

```
<application-reference>  
  <country>ZZ</country>  
  <doc-number>999999</doc-number>  
  <kind>A1</kind>  
  <date>20080321</date>  
</application-reference>
```

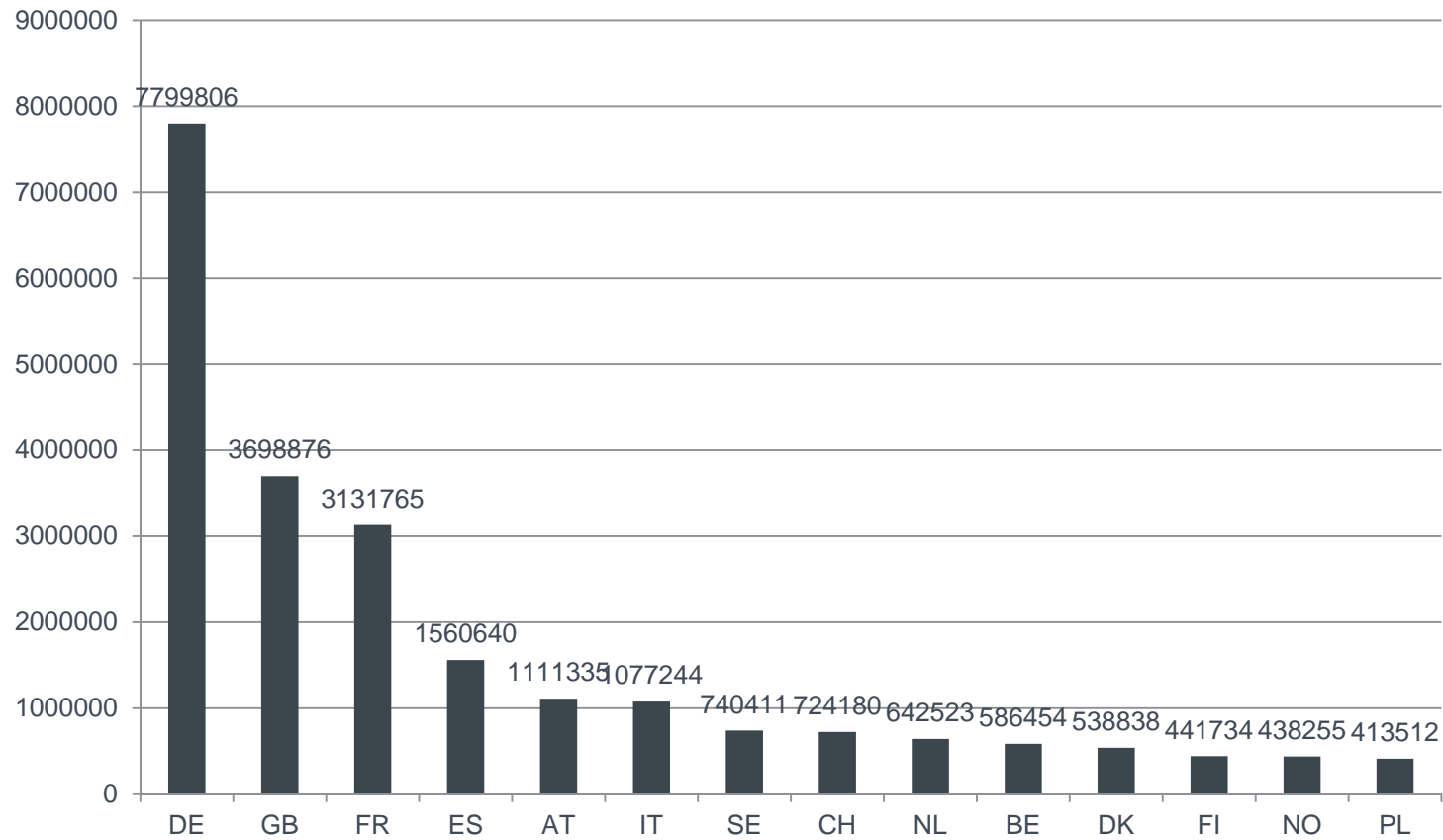
Priority

```
<priority-claims>  
  <priority-claim sequence="1" kind="national">  
    <country>XX</country>  
    <doc-number>111.111</doc-number>  
    <date>19980130</date>  
  </priority-claim>  
  <priority-claim sequence="2" kind="national">  
    <country>XX</country>  
    <doc-number>111.222</doc-number>  
    <date>19981229</date>  
  </priority-claim>  
  <priority-claim sequence="3" kind="national">  
    <country>YY</country>  
    <doc-number>1234567890.1</doc-number>  
    <date>19990320</date>  
  </priority-claim>  
</priority-claims>
```

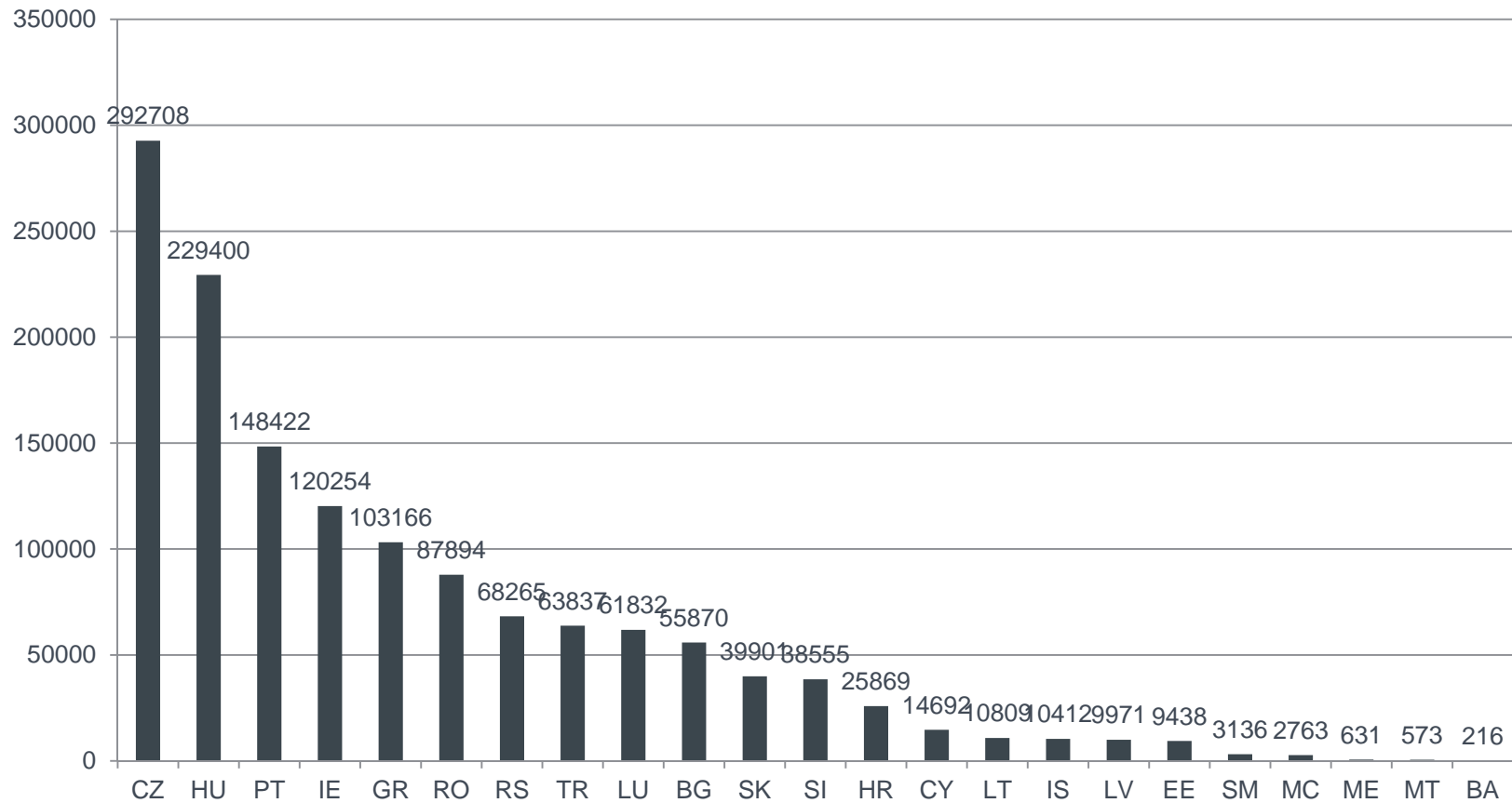
```
</authority-file-doc>
```

```
</authority-file>
```

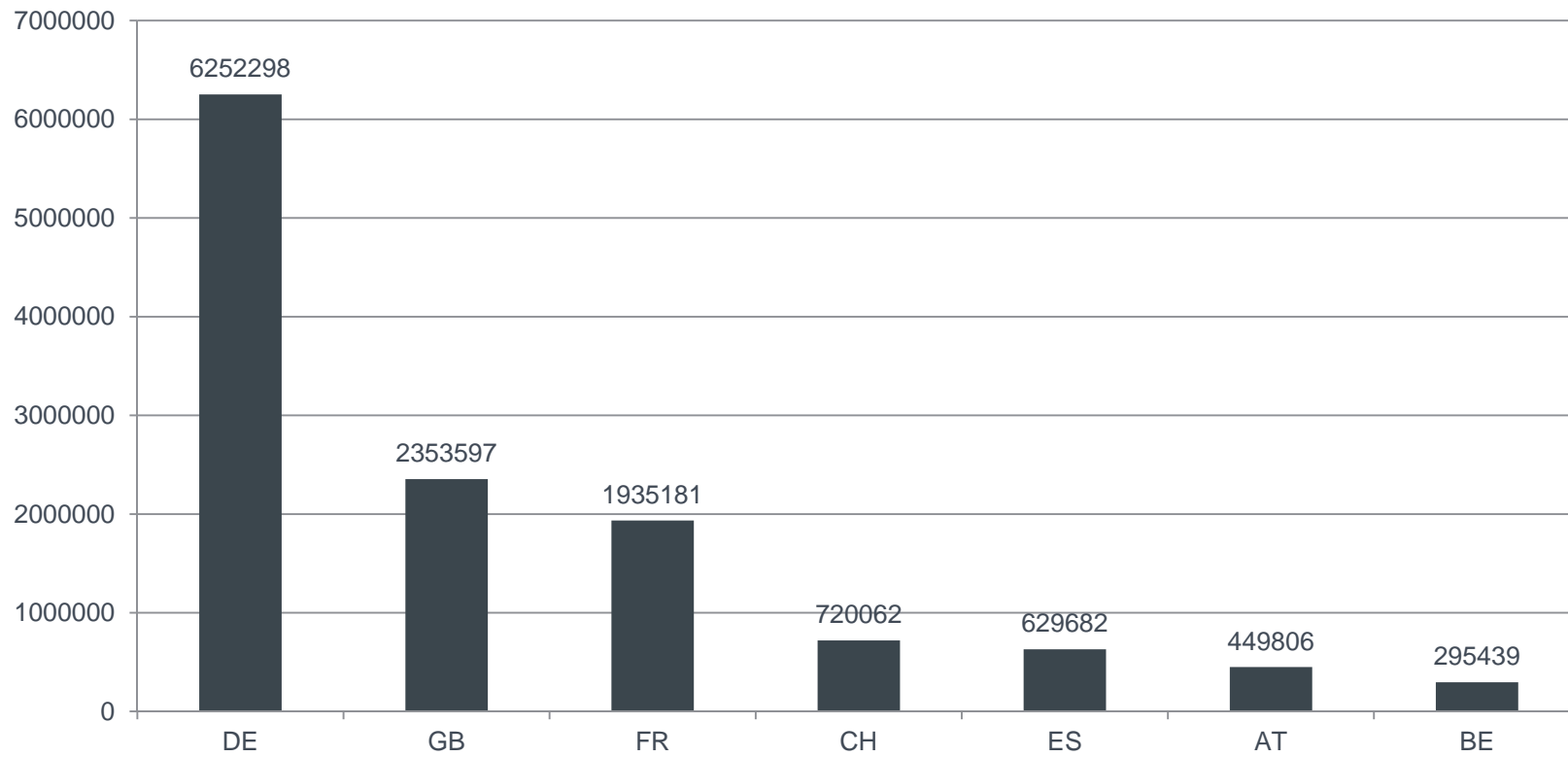
Bibliographic data coverage in Espacenet



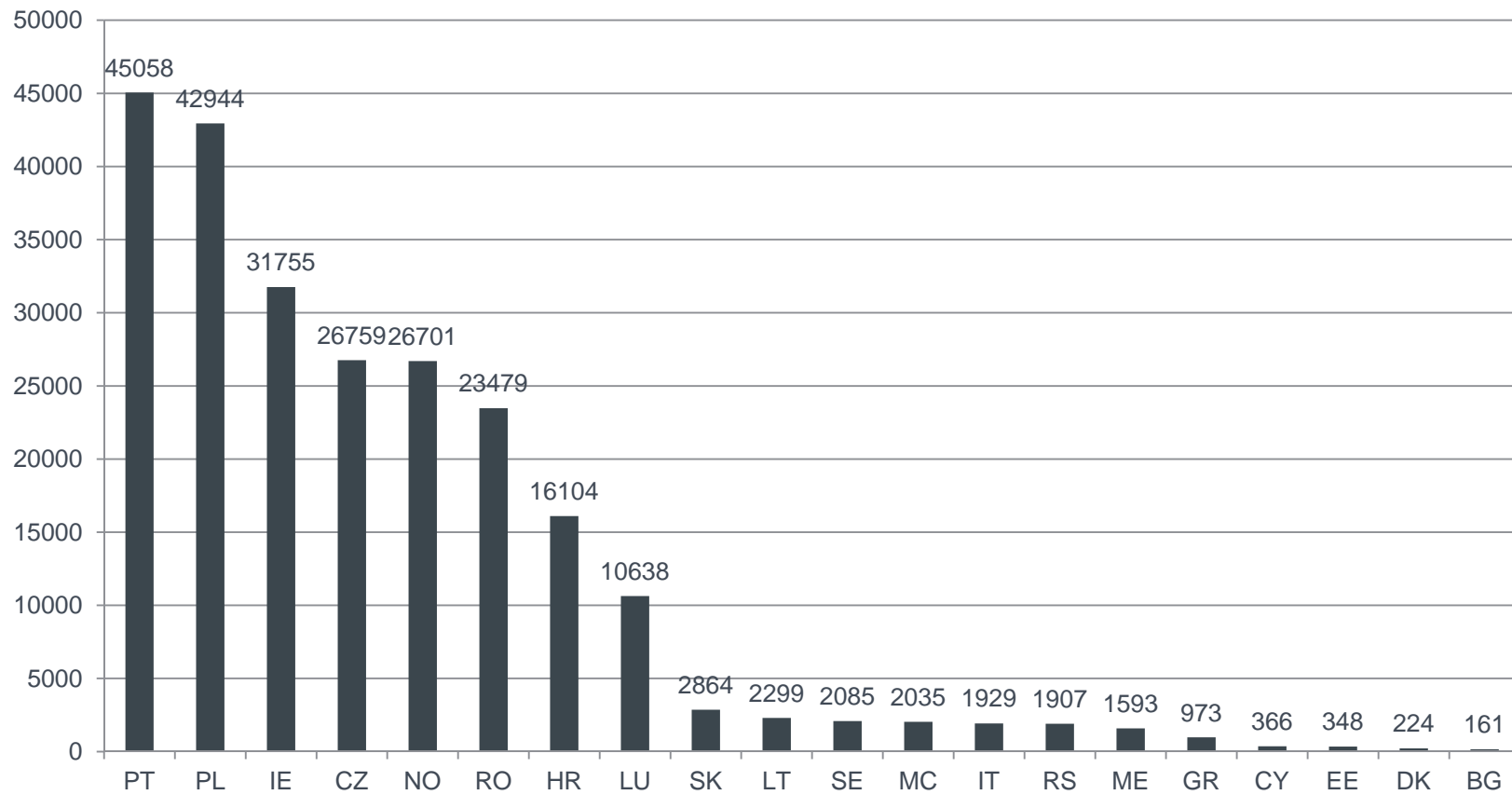
Bibliographic data coverage in Espacenet



Full-text coverage of EPO member states



Full-text coverage of EPO member states



QaS – Look to the future

Media

- colour PDF
- file-less web-service transfer

Scope

- MS, IP5, BRIC, Validation States, ASEAN, Latipat, Arabpat, India

Data quality

- correctness (supported by EPO web-services for number formats)
- timeliness
- completeness (supported by authority files)

IP5 data related activities

- USPTO provides enriched citation data
 - Extracted from Office Action Rejection Statements, Information Disclosure Statements, and Notices of References Cited by the Examiner
 - Enriched Citation information includes Relevancy (X/Y/A), associated Claims, and Relevant Passages
 - Combination of automated extraction of Enriched Citation information and manual Quality Assurance (QA) process for improving extraction algorithm
 - May enable the automated extraction of Enriched Citation information for back file applications

IP5 data related activities

- Access to NPL via DOIs
 - One of the main priorities for the IP5 industry.
 - Today, accessing IP5 file wrapper data is instantaneous and seamless but not cited NPL records.
 - This long standing issue has been tackled at the EPO by adding Digital Object Identifiers (DOIs) to the NPL records cited and contained in EPO databases.
 - A DOI is a permanent URL to a digital source:
<http://dx.doi.org/10.1097/CJI.0b013e3182594387>

IP5 Global Dossier improvements

- Alerting functionality
 - EPO and CNIPA providing RSS and email alerting
 - JPO in production as of May 2019
- Legal status
 - IP5 offices completed mapping of statuses with ST27
 - Next step: API for web-service based legal status information
- Applicant Name Standardisation
 - KIPO pilot project
 - Global Mapping Table with transliterations in all IP5 languages

- Thank you for your attention.

- Any questions

