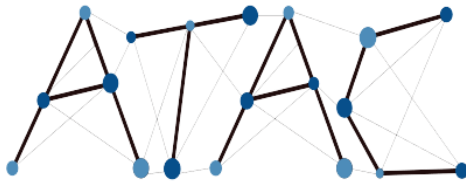# IP Offices, Global Databases and Machine Learning: What WIPO can offer you?

Side event WIPO General Assemblies

21-07-2022

Ken-Ichiro Natsume, Christophe Mazenc

Advanced Technology Application Center

WIPO
WORLD
INTELLECTUAL PROPERTY
ORGANIZATION

WIPO's Global Databases

Find if your new idea could become a patent:

PATENTSCOPE: Search 120 million patent applications from 75 countries in 34 different languages
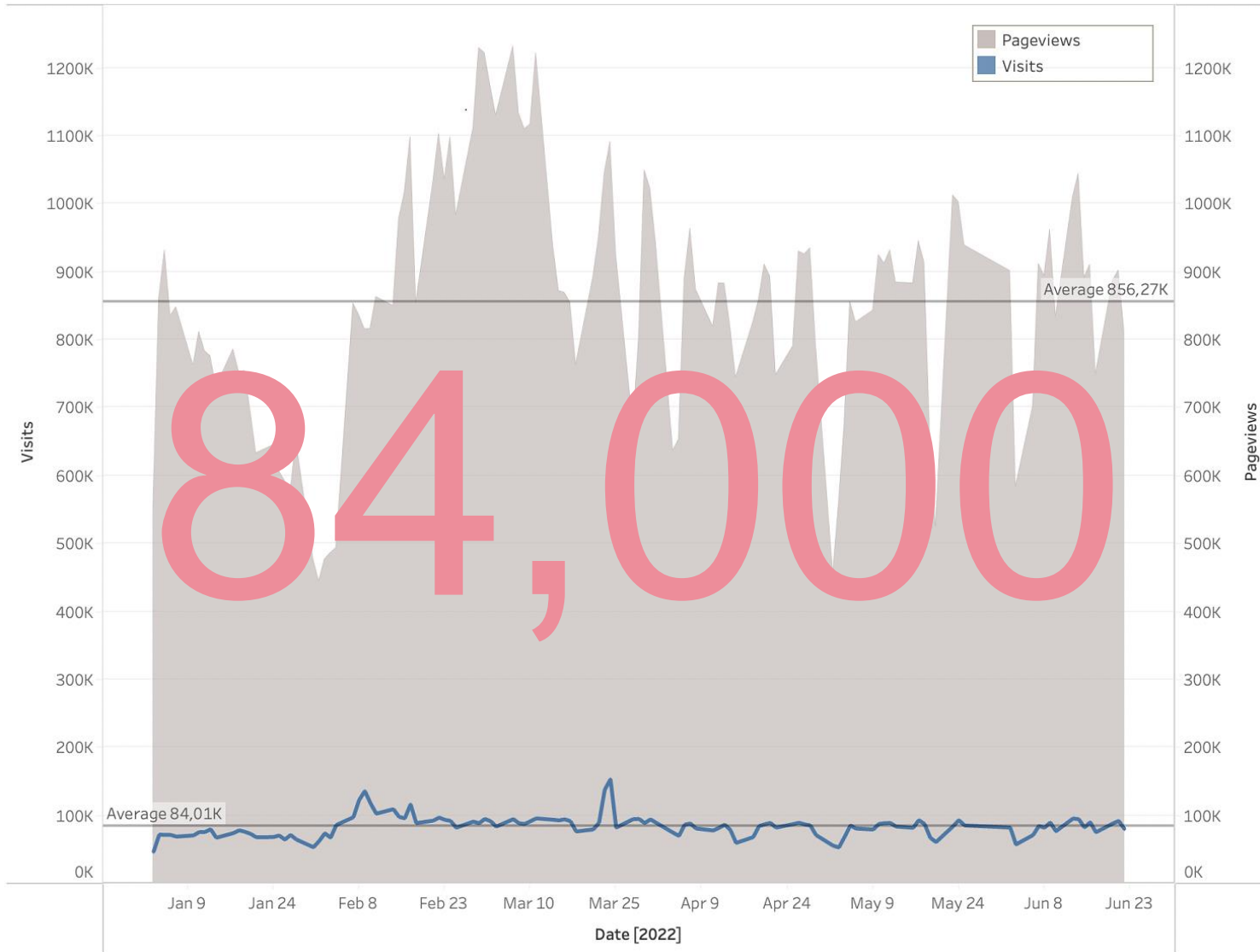
Find a name/logo as a trademark for your new business:

Global Brand Database: Search 52 million trademarks from 68 countries, as well as appellations of origins and emblems using Artificial intelligence

See what your competitors are working on:

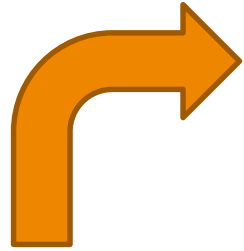Global Design Database: Search 14.5 million industrial designs from 36 countries
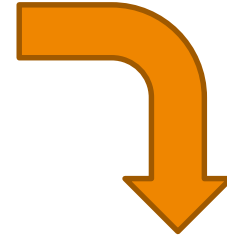
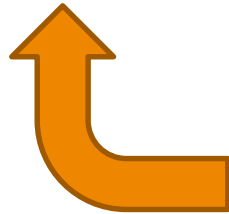# Main Metrics (Global Databases)

Improved visibility of national IP rights

Growth in IP filings

Stronger national IP system

more investments, more innovation

Economic growth
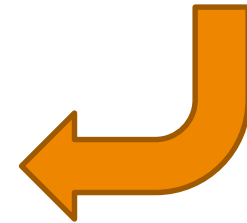
Other benefits for IP offices

- Improvement of quality of national IP data (data validation procedures, data improvement procedures, authority files)

- Connect your national IP rights with other countries (PATENTSCOPE patent families, Office of Second Filing examination)

- Redirect international users to your national IP registry

- Obtain free licenses of WIPO machine learning tools that are built by putting together the data sets of PCT, Madrid and Hague data with national patent, trademark and industrial designs data

Benefits for the IP users in your country

- Available for free: only an internet connection required

- State of the art search functions available for your national IP data (chemical search, language specific keyword search, image similarity search,…)

- One stop secure search

- State of the Art Machine translation that unlocks technological knowledge transfer in the area of patents

- For national IP owners: reduce the risk of conflicts with IP rights in other countries

Joining the Global Databases

1. Signature of bilateral data exchange & technical cooperation agreement
   No redistribution by WIPO

2. National IP data sent to WIPO in the format already available

Experiences by IP Offices

Mrs. María de los Angeles Sánchez Torres, Director General, Cuban Industrial Property Office

Mr. Anders Bruun, Patent Expert, Swedish Patent and Registration Office

Dr Klauss Plessner, Deputy Head of Online Services, German Patent and Trade Mark Office

IP offices, global databases and machine learning: what WIPO can offer you

Side event – WIPO General Assemblies

July 21, 2022

María de los Angeles Sánchez Torres

Director General, Cuban Industrial Property Office

maria@ocpi.cu

Challenges

Innovation

Computerization

Communication Dissemination of IP

OCPI

Policy on the industrial property system

Institutional mechanisms
Laws, international agreements and treaties, national development policy

Human resources Technological infrastructure
Personnel specializing in ICT and technology

Cooperation

WIPO, EUIPO, EP-OEPM
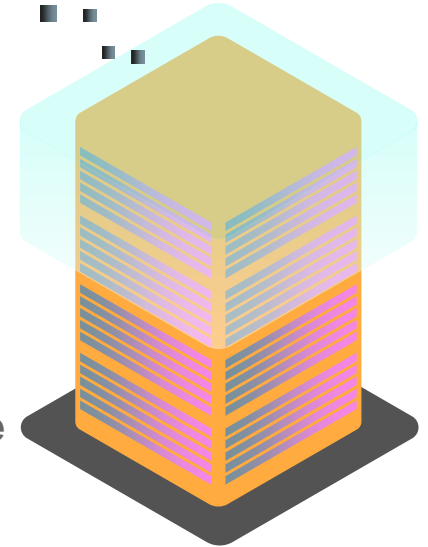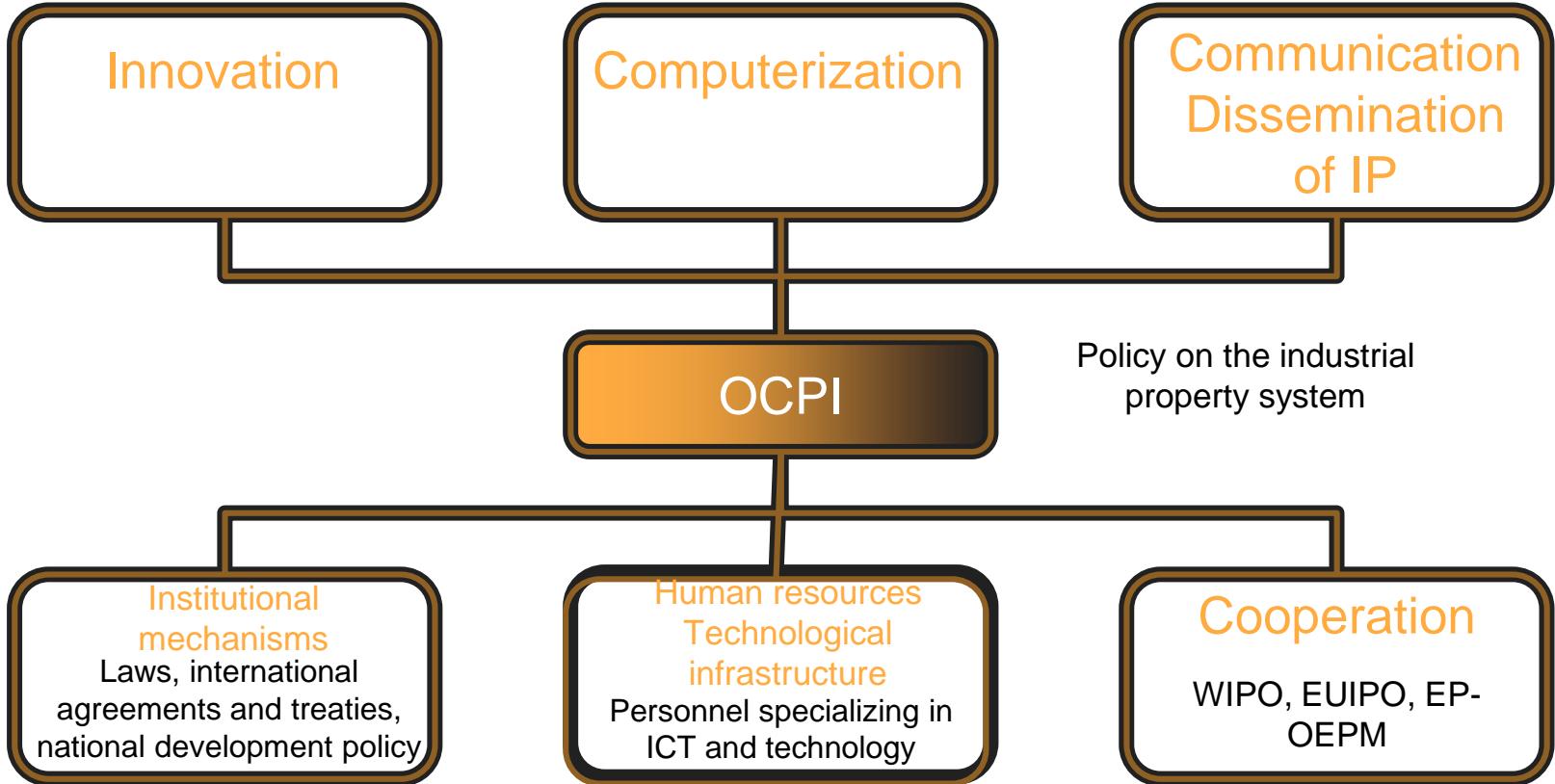
Limitations

Risks

Data migration process – automation system for IPOs

Planning   Data backup   Execution   Production

Quality control   Migration design   Validation
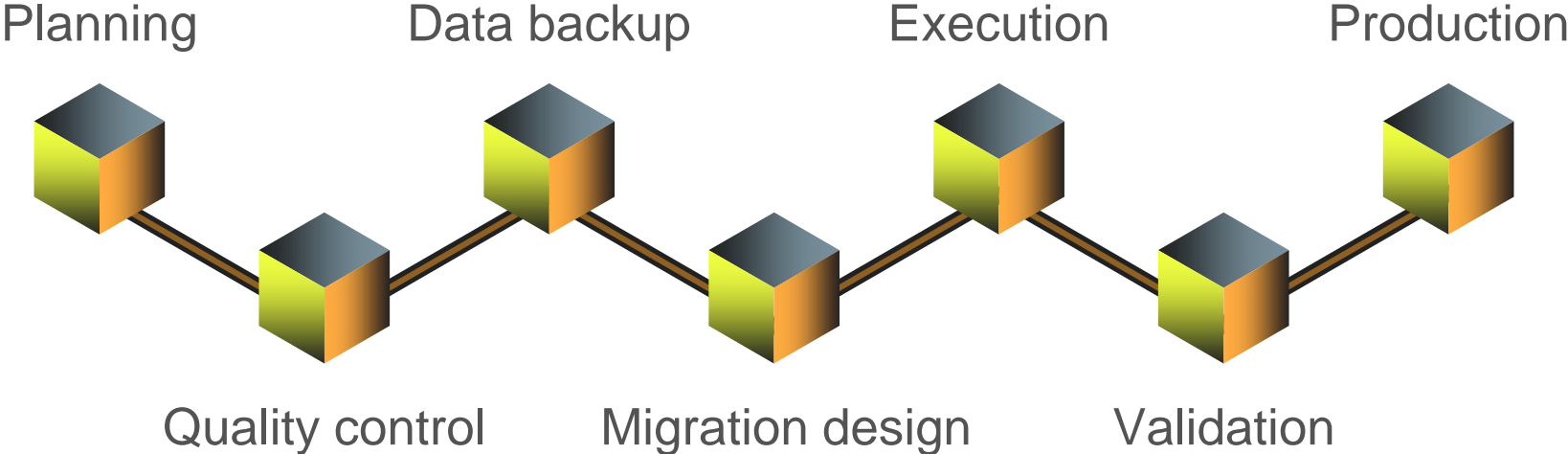
WIPO assurance – national office

Integration of ITC tools

WIPO OCR

Tools for producing full text in XML. ST. 36

Classification service

Automatic update of Nice, Vienna, IPC and Locarno classification systems

PATENTSCOPE

Access to PCT international applications in full text format and patent documents from participating national and regional patent offices

IPAS Madrid

Notifications relating to designations made under the Madrid System. Electronic exchange of documents

Global Brand Database

Global Design Database

WIPO Publish

Online sharing of information on official publications and other aspects of IP. Search, official publications, data extraction (XML)

Regional databases
LATIPAT, TMview, Design View, TMclass

IPAS

Registration process for trademark, patent and industrial design applications. Administration, workflow design, statistics

WIPO File

Online services for applicants and agents

MILESTONES

2003   2010   2012   2014   2018   2019   2021   2022

Global Brand Database

Coverage – Cuba
Years: 1917–2022
103,364
registrations

Benefits

- Multi-source trademark search
- Database links: WIPO Publish – CU, TMview
- AI-based search of image elements
- Report generation

Global Design Database

Coverage – Cuba

Years: 1980–2021
766 registrations

Benefits

- Multi-source design search
- Database links: WIPO Publish – CU, DESIGNview
- Report generation

PATENTSCOPE global database

Coverage – Cuba

Years: 1968–2022
4,795 registrations
2,079 images

Benefits

- Multi-source patent search
- Database links: WIPO Publish – CU, LATIPAT
- Full text search
- Chemical structure search
- Automatic translation
- Patent family
- Report generation
- Statistics

LATIPAT database

TMview database

DESIGNview database

Conclusions

- Industrial Property Administration System (IPAS)

- Flexibility (adaptation to domestic legislation)

- Integration with global and regional databases

- Integration of WIPO central services

- Use of WIPO standards for information exchange

- Technical assistance

- Technology transfer

- Incorporation of future developments

- Low resources

- Minimal risk

Challenges

Digital transformation

Leveraging developments in the cloud
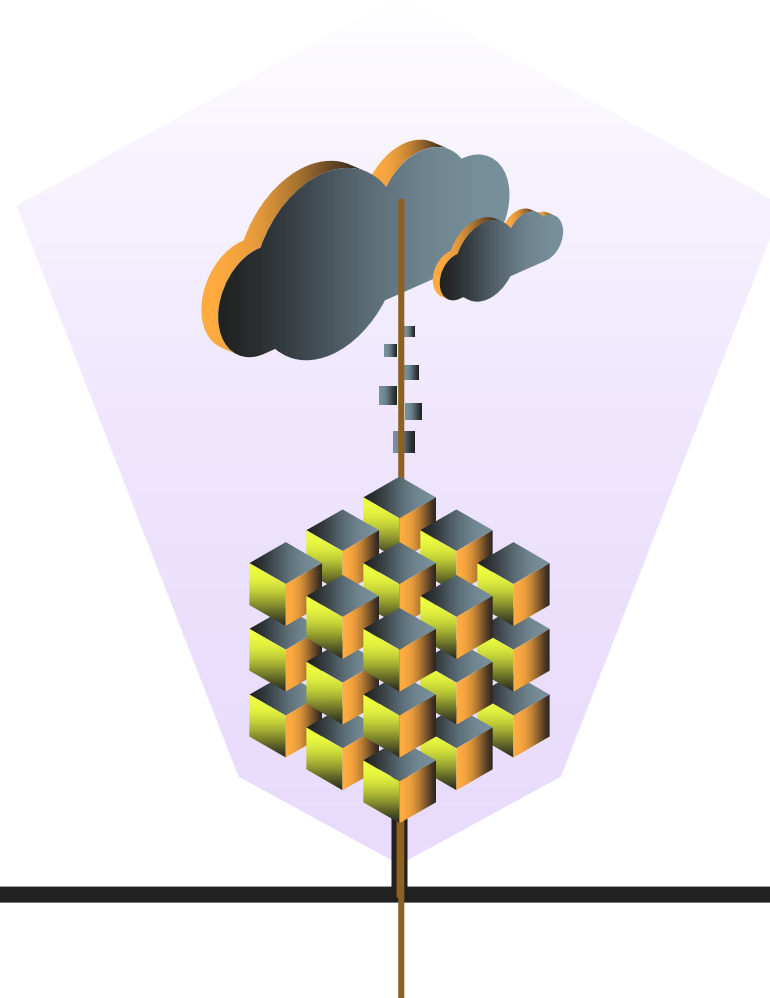
Better integration with global systems

Building national capacities and competencies

Thank you

Experiences by IP Offices

Mrs. María de los Angeles Sánchez Torres, Director General, Cuban Industrial Property Office

Mr. Anders Bruun, Patent Expert, Swedish Patent and Registration Office

Dr Klauss Plessner, Deputy Head of Online Services, German Patent and Trade Mark Office

Experiences of the WIPO OCR solution at PRV

Anders Bruun 2022-07-21

PRV in brief

- Around 350 employees in Stockholm and Söderhamn
- The patent division is located in Stockholm and has around 150 employees, of which around 110 are examining patent engineers
- We examine around 2200 national applications per year and around 900 PCT applications.

Photo: Anna Fanqvist

Photo: Yvonne

PRV

Background

The EPO "Quality at Source" project (2017-2019) included several requirements for member states:

- Complete data containing all relevant bibliographic elements, for example citations
- Timely and standard-compatible data
- Authority files of published documents
- The biggest challenge: Structured full text from patent documents

The WIPO OCR solution

- Proposed by the EPO and used by other EPO member states
- Produces results that are accepted by EPO
- Made available to PRV through an informal bilateral agreement with WIPO to deliver the data for inclusion in PatentScope

PRV

Training and installation

- Training session at WIPO for two persons in May 2018
- Installation package and adaptations for Swedish document kinds and languages were provided by WIPO
- Testing and development of an office-internal manual started testing in May 2018
- Production data to EPO from October 2018
- Production data to WIPO from June 2019
- Backfile to January 2018 processed

PRV

How does it work?

- PDF documents are converted to TIF and broken up into single pages
- OCR treatment and generation of binary files from TIFs
- Proofreading in a specially developed application – for example correction of errors and identification of embedded figures
- Documents reassembled from the single pages
- Generation of final XML and HTML result files

PRV
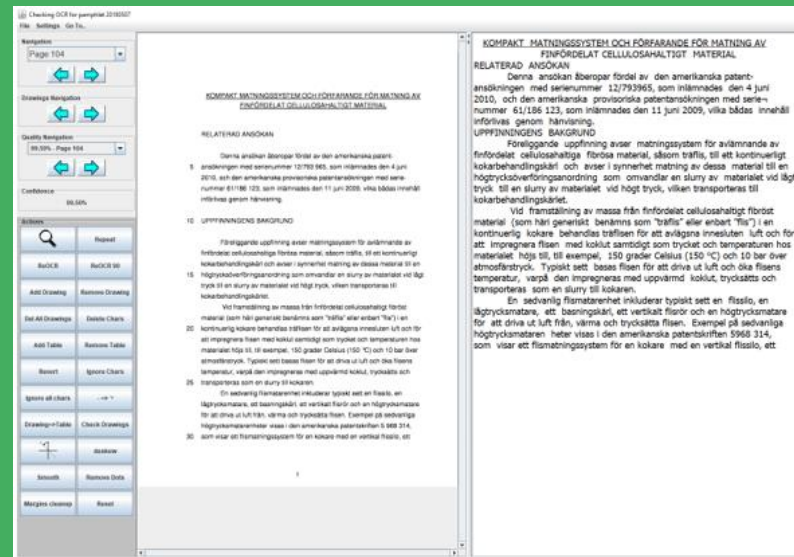
Experiences

- A very clever package with many features that are specially developed for patent documents
- Mainly trouble-free operation
- Some problems with updates and cache handling, caused by the different setups at WIPO and PRV
- WIPO staff has been extremely supportive and helpful!
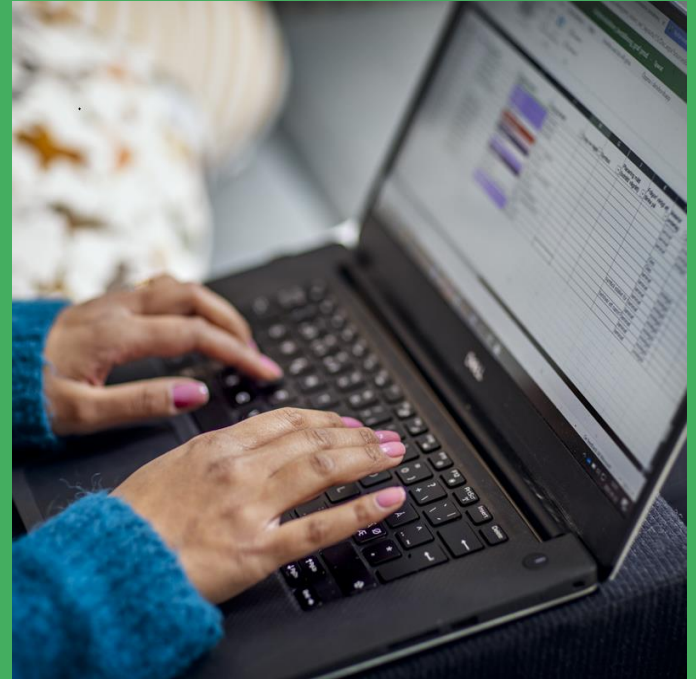
Some figures

- PRV publishes around 2,200 documents per year, totalling around 60,000 pages
- The regular proofreading work is performed by two persons who take turns
- An average weekly publication batch of 1200 pages takes less than half a week's work to process for one person
- Result files are delivered monthly via FTP

PRV
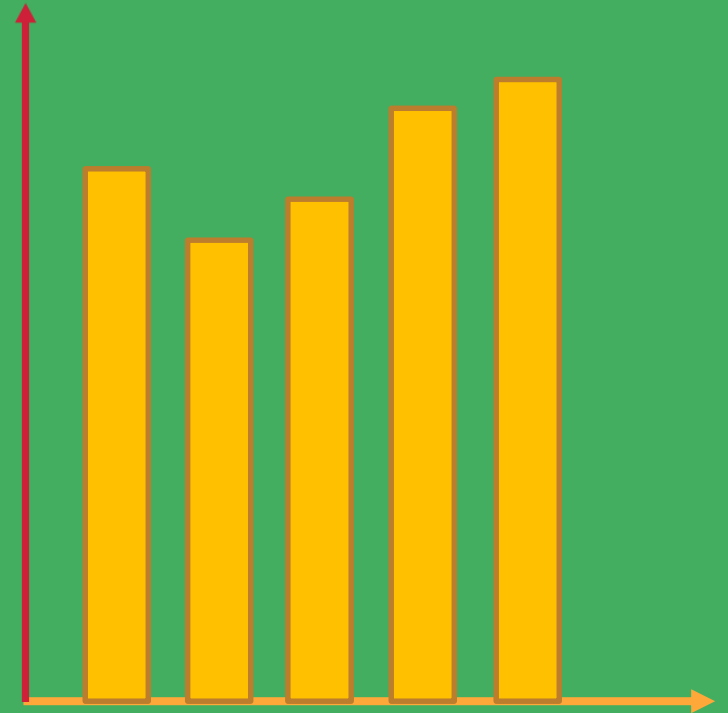
Advantages

- The full text of SE documents is searchable and translatable in PatentScope
- The full text of SE documents is searchable and translatable in Espacenet and in EPO's search systems
- The full text of SE documents has been made available for download from PRV's Open Data Portal

Thank you !

Experiences by IP Offices

Mrs. María de los Angeles Sánchez Torres, Director General, Cuban Industrial Property Office

Mr. Anders Bruun, Patent Expert, Swedish Patent and Registration Office

Dr Klauss Plessner, Deputy Head of Online Services, German Patent and Trade Mark Office

DPMA's Experiences with WIPO Global Databases

Side event WIPO General Assemblies

21.07.2022

Klaus Plessner
German Patent and Trade Mark Office

WIPO Translate – Motivation

- DEPATIS database (status in 2020)
  - More than 40 Mio full text documents from DE, US, EP, FR, GB, …
  - But: No full text documents from CN, JP and KR
- The Problem
  - In 2019 CN, JP and KR received 65 percent of the world's patent applications
  - Examiners had to search in external databases for Asian patents
  - Use of external databases generally causes loss of efficiency and data protection issues

WIPO Translate – Project Objectives

- Full access to Asian patent literature
  - Bulk translation enables full text search
  - Machine translation copes with the enormous amount of documents (over 60 Mio)
- Use of WIPO Translate
  - Analysis of translations should determines the best software
  - We tested products from WIPO, EPO and Google
  - Test criterium: How much of the translated features is comprehensible to a person of ordinary skill in the art?
  - Further criteria: OS, training data, languages, price

WIPO Translate – Realization

- **First Stage (Q1 2021 to Q1 2022)**
    - 1 Server with 4 GPUs: Intel Xeon Gold 6130 (192 GB RAM, 60 Cores) with Nvidia Quadro GV100 GPU (4 GPUs, 32 GB)
    - 2 Servers with 4 GPUs each: Intel Xeon Gold 6230 (192 GB RAM, 80 Cores) with Nvidia Quadro GV100 GPU (4 GPUs, 32 GB)
    - Medium Throughput of 1 Mio pages/day

- **Second Stage (from Q2 2022)**
    - 2 additional Servers with 4 GPUs each: AMD EPYC 7763 (1 TB RAM, 256 Cores) with Nvidia A100 (4 GPUs, 80 GB)
    - Overall medium Throughput of 1,7 Mio pages/day

WIPO Translate – Status Today

- Machine translation completed so far
  - Translation is done backwards in time (starts with new documents)
  - CN: 31.1 Mio of 36.9 Mio documents
  - JP: 14.3 Mio of 27.8 Mio documents
  - KR: none
- 1st translation round completed by end of 2022
  - CN: completion of translation expected in August
  - JP: 13.5 Mio documents (from before 1994)
  - KR: 6.7 Mio documents

WIPO Translate – Further Steps

- Regular translations
    - The translation software is continuously improved
    - To take advantage of the improvement, the data stock will be regularely retranslated

- Expansion of stock
    - Asian WO documents will be added next

WIPO Translate – Feedback

- Benefits for patent examiners
  - Full text search in Asian patent documents
  - Full functionality of DEPATIS (highlighting, search strategies)
  - No need to use external databases
  - No need for additional trainings

- Cooperation with WIPO team
  - Installation support, optimization, new translation models
  - Excellent communication and support

TM Image Similarity Search – Architecture

- ## Web Service input data
  - An image-file has to be specified
  - The image may be combined with further criteria (data origin DE/WO/EM, Nice class, image class, application date, registration date, search strategy)
- ## Web Service output data
  - The answer contains a list of trademarks sorted by relevance
- ## Integration in DPMAregister
  - DPMAregister GUI contains a search form and displays the hit list

TM Image Similarity Search – Feedback

- ## Web Service implementation phase
  - WIPO offered a 1st version and added features according to DPMA's suggestions
  - Support and communication were excellent
- ## Production and improvement
  - Trademark examiners have approved the quality of results
  - Maximum number of results will be increased

German Patent
and Trade Mark Office

DPMA's Experiences with WIPO Global Databases

Side event WIPO General Assemblies

21.07.2022

Klaus Plessner
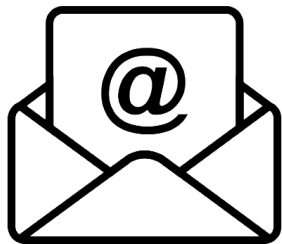German Patent and Trade Mark Office

Q&A - Brainstorming

- Possible international collaboration to produce patent full text backfiles for offices where it is not available?

- Suggestions to increase the collaboration level between your office, WIPO and other offices on data quality and machine learning?

- Other use cases for applying Machine Learning to facilitate IP administration procedures in the future?

Machine-learning WIPO tools

- PATENTSCOPE OCR suite

- WIPO Translate

- Global Brand Database Image Similarity Search API for trademarks (for the trademarks of your jurisdiction)

- ATAC IPC Classifier

- Vienna Classification Assistant (public web interface)

christophe.mazenc@wipo.int