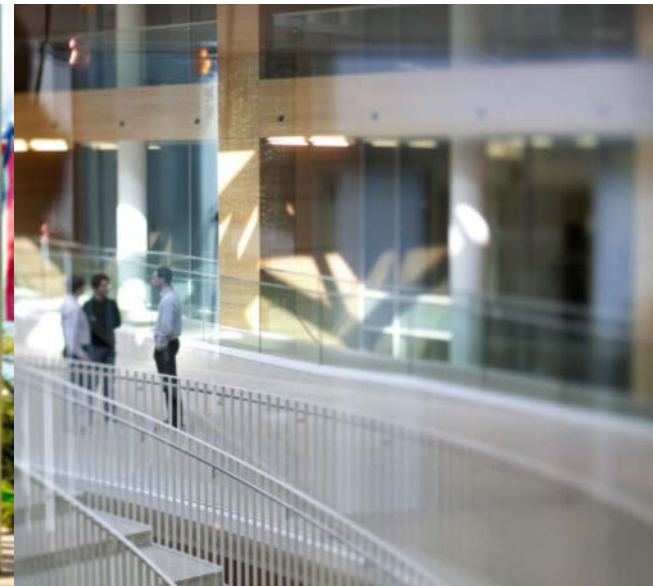




Europäisches
Patentamt
European
Patent Office
Office européen
des brevets

WIPO Workshop on Applicant Name Standardization

Name harmonization activities at the European Patent Office



Outline

- Customer data harmonization for the EP patent granting procedure
- Standardization of applicant names for DOCDB worldwide bibliographic database
- Harmonized names for the PATSTAT product line

Main difficulties → different forms of the name (address) to indicate the same applicant¹

- **Spelling variations**, e.g. “IBM” and “I.B.M.”, or “BAIN & CO” and “BAIN AND COMPANY”;
- **Typographical errors**, e.g. “INTERNATIONAL BUSINESS MACHINES” and “INTERAATIONAL BUSINES MACHINES”;
- **Addition of the legal form** (again with possible acronyms, spelling variations, mistakes and typographical errors in the legal form) e.g. “IBM”, “IBM CORP”, “IBM CORPORATION” and “IBM COPRORATION”, or “BAYER”, “BAYER A.G.” and “BAYER AG”;
- **Errors** e.g. “INTERNATIONAL BUSINESS MACHINES” and “INTELLIGENT BUSINESS MACHINES”;
- **Other additions** like establishment, business unit, department, subsidiary name or geographic identifier, e.g. “IBM” and “IBM JAPAN”;
- **Use of acronyms**, e.g. “IBM” and “INTERNATIONAL BUSINESS MACHINES”.

¹ “Patent Statistics at Eurostat: Methods for Regionalisation, Sector Allocation and Name Harmonisation”. Eurostat : Methodologies & working papers. 2011 edition. ISSN 1977-0375.

EPO efforts in Customer data standardization



- Customer data relating to the Patent Granting procedure is maintained in a central Database containing many duplicates (ca. 20%)
- Duplicate customer data may arise from minimal differences in names or addresses
- The EPO will “clean” this database to create a central Master Data repository for EP applicants, with **harmonized names** and **addresses** (NACM Project)
- Aiming at single, **unique identification** per legal entity or individual

Quality at source!

- Cleaning strategy and plan is currently being developed to establish:
 1. Name and address standardization
 2. Deduplication of names
- Data Quality standards:
 - Addresses standardized to local Postal Authority databases
 - Names standardized based on syntactic analysis
- Data quality policies:
 - Data Steward role
 - Customer data process flows
- **Challenges ahead:**
 - Legal frameworks EPC, PCT
 - Validation of standardizations with applicants



DOCDB Standardization of Applicant Name - STAN

- DOCDB is the **EPO Worldwide Bibliographic Database**
- **STAN** since more 30 years to improve patent search results

- **Coverage:**
 - For the **100 million patent documents** present in **DOCDB**:
 - +/- 250.000** standardized applicant names covering
 - +/- 945.000** variations
 - Variations include typo/spelling errors, acronyms, etc.
 - Standardization is only for **companies, universities, etc.** not for individuals

Actual Challenges:

- Achieve automated harmonization process of applicant names
- Maintain data quality using applicant name harmonization standards for the cleaning process

STAN in **Espacenet** for the Public

STAN names visible

Inventor(s): GILLESPIE JOHN JR [US]; LABEDZ RALPH H [US]; PLATT MICHAEL KENNETH [US]; SPANG RONALD H JR [US]; BERRILL JAMES FREI [US]; VOGEL MATTHEW STEPHEN [US]; GREANEY MICHELLE KOWALSKI [US] ±

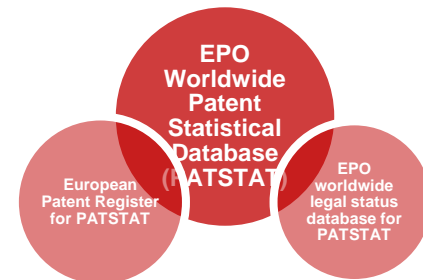
Applicant(s): **BAXTER** INT INC [US] +

STAN + nonstandardized names visible (clicking “+”)

Inventor(s): GILLESPIE JOHN JR [US]; LABEDZ RALPH H [US]; PLATT MICHAEL KENNETH [US]; SPANG RONALD H JR [US]; BERRILL JAMES FREI [US]; VOGEL MATTHEW STEPHEN [US]; GREANEY MICHELLE KOWALSKI [US] +

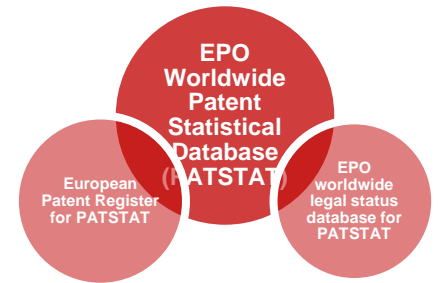
Applicant(s): **BAXTER** INT INC [US] ± (**BAXTER** INTERNATIONAL INC)

The EPO's PATSTAT product line



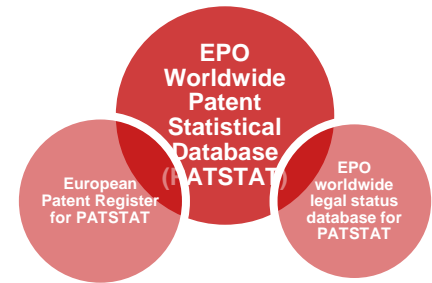
- Unique basis for conducting **sophisticated statistical analyses of patent data**
- Bibliographic data on **nearly 100 million patent documents** of leading industrialised and developing countries (DOCDB)
- **Legal event data** of patent documents in many countries worldwide (INPADOC)
- **European patent register data:** Bibliographic, legal and procedural information on published European and EURO-PCT applications (REGISTER)

The EPO's PATSTAT product line: Added value on Applicant names



- **Harmonization of names** to increase usability and connectivity with other data sources
- Applicant and inventor names options in PATSTAT:
 - **Original names**
 - **DOCDB** standardized name (twice/year)
 - **HAN** applicant name (OECD) (twice/year)
 - **PATSTAT harmonized name:** Harmonized applicant and inventor names (twice/year)

The EPO's PATSTAT product line: PATSTAT harmonized name

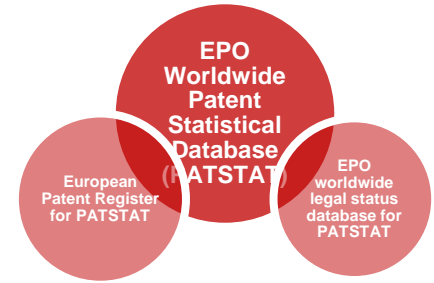


- Procedure (developed by KU Leuven): **Two-step process¹**:
 - methodology layer 1: **automated**; focus on identification of name variations, using data cleaning and processing
 - methodology layer 2: **semi-automated**; increasing completeness, with focus on top 500 applicants

¹ "Patent Statistics at Eurostat: Methods for Regionalisation, Sector Allocation and Name Harmonisation". Eurostat : Methodologies & working papers. 2011 edition. ISSN 1977-0375.

Harmonised Names in PATSTAT

- original name, DOCDB names, externally standardized names



	original name	DOCDB standardised	EEE_PPAT (Uni. Leuven)	HAN (OECD)
1	TOYOTA MOTOR CORP	TOYOTA MOTOR CORP	TOYOTA MOTOR CORPORATION	TOYOTA MOTOR CORP
2	Toyota Jidosha Kabushiki Kaisha	NISSAN MOTOR	HONDA MOTOR COMPANY	TOYOTA JIDOSHA KK
3	NISSAN MOTOR CO LTD	HONDA MOTOR CO LTD	NISSAN MOTOR COMPANY	NISSAN MOTOR CO LTD
4	HONDA MOTOR CO LTD	TOYOTA MOTOR CO LTD	HYUNDAI MOTOR COMPANY	HONDA MOTOR CO LTD
5	Hyundai Motor Company	HYUNDAI MOTOR CO LTD	ROBERT BOSCH	NISSAN MOTOR CO
6	Robert Bosch GmbH	BOSCH GMBH ROBERT	AISIN AW COMPANY	HYUNDAI MOTOR CO
7	HONDA MOTOR CO., Ltd.	AISIN AW CO	MAZDA MOTOR CORPORATION	BOSCH ROBERT
8	Nissan Motor Co., Ltd.	MAZDA MOTOR	DENSO CORPORATION	HONDA GIKEN KK
9	MAZDA MOTOR CORP	FUJI HEAVY IND LTD	FUJI HEAVY IND	DENSO CORP
10	HONDA GIKEN KOGYO KABUSHIKI KAISHA	DENSO CORP	KUBOTA CORPORATION	HONDA MOTOR CO

Reflections for the future

In a digital world, name standardization is superseded by IDs.

Should we consider applicant IDs in the IP community?

Is **ECLI** (not centralized) **ORCID** (central registry)
a model?

or

better follow (sparse) national info?

THERE IS NO LOW HANGING FRUIT

Thank you for your attention!

John Potts, Anne-Marie McConnell, Ignacio Muñoz Ozores, Christian Soltmann
European Patent Office

jpotts@epo.org

amcconnell@epo.org

imunozozores@epo.org

csoltmann@epo.org