

Applicant name harmonization and classifying IP data by technology and institutional sectors

Date: September 14-16, 2016



OVERVIEW

Why do we need IP statistics by institutional sectors, industry, technology, etc.?

How to develop such data

Applicant's name

Field of technologies

Institutional sectors

Industries

WHY DO WE NEED SUCH DATA

To provide accurate data

Such data are important for economic analysis

Linking IP data with non-IP data

APPLICANT'S NAME – ISSUE AT HAND

KONINKIJKE PHILIPS ELECTRONICS N.V.
KONINKLIJKE PHILIPS ELECTRONICS N.V.
KONINKLIJKE PHILIPS ELECTRONICS NV
KONINLIJKE PHILIPS ELECTRONICS N.V.
KONINKLIJKE PHILIPS ELECTRONICS, N.V.
KONINKLIJKE PHILIPS ELECTRONICS N. V.
KONINKLIJKE PHILIPS ELECTRONICS N.V.,
KONINKLIKE PHILIPS ELECTRONICS N.V.
KONINKLIJKE PHILIPS ELECTRONICS N.V.
KONINKLIJKE PHILIPS ELECTRONIQUE N.V.
KONINKLIJKE PHILIPS ELECTRONICS
KONINKLIGKE PHILIPS ELECTRONICS N.V.
KONINKLIJKC PHILIPS ELECTRONICS N.V.
KONINKLIJKE PHILIPS ELECTRONICS NV
NINKLIJKE PHILIPS ELECTRONICS NV
KNONINKLIJKE PHILIPS ELECTRONICS, N. V
KONINKLIJKE PHILIPS ELECTRONICS N V
KONINKLIJKE PHILIPS ELECTRONICS, NV
PHILIPS ELECTRONICS N.V.
PHILIPS ELECTRONICS N.V

PHILIPS ELECTRONICS, N.V.
KONINKLIJKE PHILIPS ELECTRONICS, N. V.
KONINKLIJEKE PHILIPS ELECTRONICS N.V.
KONINKLIJKE PHILIPS ELECTZRONICS N.V.
KONINKLIJKE PHILIPPS ELECTRONICS N.V.
KONINKLIJKJE PHILIPS ELECTRONICS N.V.
KONINKLIJKE PHILLIPS ELECTRONICS N.V.
KONINKLIJKE PHLIPS ELECTRONICS N.V.
KONINGLIJKE PHILIPS ELECTRONICS N.V.
KONINKLIJKE PHILIPS ELECTRONIC NV
PHILIPS ELECTRONIC N.V.
KONINKLIJKE PHILIPS ELECTRONIC N. V.
KRONINKLIJKE PHILIPS ELECTRONICS
KONINKLIJKE PHILIPS ELECTRPMOCS. N.V.
KONINKLIJKE PHILIPS ELECTRONICS B.V.
KONINKLIJKE, PHILIPS
KONLNKLIJKE PHILIPS ELECTRONICS N.V.
KONINKLIJKE PHILIPS ELECTTRONICS N.V.
KONINKLIJKE PHILIPS ELECTRONICS N.V.
KONINKLIJKE PHILIPS ELECTRONICS N.V.

TOP 15 APPLICANT WORLDWIDE (2003-12)

Top 15	Original Applicant Name	Top 15	Applicant Standardized Name
1	SAMSUNG ELECTRONICS CO., LTD.	1	PANASONIC CORPORATION
2	CANON INC	2	SAMSUNG ELECTRONICS CO., LTD.
3	TOYOTA MOTOR CORP	3	CANON INC
4	TOSHIBA CORP	4	TOYOTA JIDOSHA KABUSHIKI
5	SEIKO EPSON CORP	5	KAISHA
6	LG ELECTRONICS INC.	6	TOSHIBA KK
7	MATSUSHITA ELECTRIC IND CO LTD	7	SEIKO EPSON CORP
8	RICOH CO LTD	8	LG ELECTRONICS INC
9	International Business Machines Corporation	9	SHARP CORP
10	SHARP CORP	10	RICOH CO LTD
11	MITSUBISHI ELECTRIC CORP	11	MITSUBISHI ELECTRIC CORP
12	SONY CORP	12	INTERNATIONAL BUSINESS
13	PANASONIC CORP	13	MACHINES CORPORATION
14	HITACHI LTD	14	SONY CORP
15	DENSO CORP	15	FUJITSU LTD
			HITACHI LTD
			DENSO CORP

WIPO's APPROACH

Keyword match

Utilize keywords or parts of words

- Example: International Business Machines
- Use keyword: Int + Busi + Mach
 - Capture typo errors: **intrenational** business **machnies**
 - Miss out acronyms – IBM
 - Capture subsidiaries – International Business Machines Japan
 - Need to eliminate false negative and false positive

Algorithmic matching

Establish most likely match (similar characters & tokens)

- Capture typo errors: **intrenational** business **machnies**
- Miss out acronyms – IBM
- Low matching scores for commonly re-ordered name structures (California, University of' vs '**University of California**)
- Capture subsidiaries – International Business Machines Japan
- Need to eliminate false negative and false positive

STEP BY STEP GUIDE

1. Formatting:

- a. Capitalizing.
- b. Removing special characters, such as “.”, “-“ etc.
- c. Removing common words, such as “the”, “of”, “and” etc.
- d. Standardizing company type, such as “Ltd” to “limited” etc.
- e. Reversing some names, such as “university of xxx” to “xxx university”.
- f. Calculating similarity between all name pairs (Levenshtein distance)
- g. Calculating similarity between all address pairs (Levenshtein distance)

2. Matching:

- a. Automatic matching by standard names.
- b. Automatic matching by matched name variations.
- c. Semi-automatic (user confirmation) matching by maximum similarity of names.
- d. Semi-automatic matching by maximum similarity of addresses.
- e. Semi-automatic matching by key words assigned to top applicants.

3. Issues:

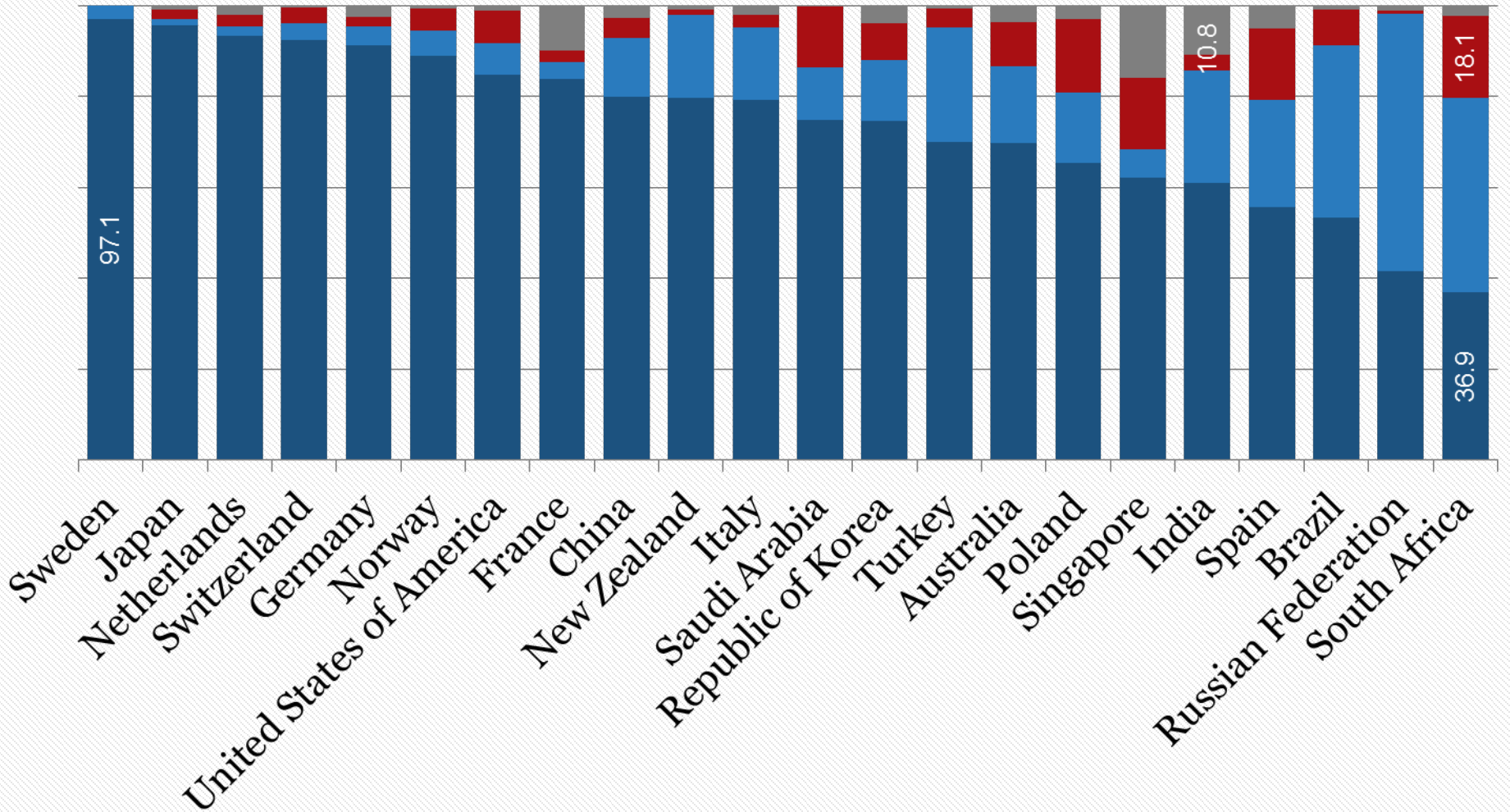
- a. Do not match names with different company type.
- b. Check address before matching names with different entity names.
- c. Do not set up similarity threshold and apply automatic matching according to similarity.

EXAMPLE: TOP PCT APPLICANTS

Ranking	Applicant's name	Origin	PCT 2015
1	HUAWEI TECHNOLOGIES CO., LTD.	China	3,898
2	QUALCOMM INCORPORATED	United States of America	2,442
3	ZTE CORPORATION	China	2,155
4	SAMSUNG ELECTRONICS CO., LTD.	Republic of Korea	1,683
5	MITSUBISHI ELECTRIC CORPORATION	Japan	1,593
6	TELEFONAKTIEBOLAGET LM ERICSSON (PUB	Sweden	1,481
7	LG ELECTRONICS INC.	Republic of Korea	1,457
8	SONY CORPORATION	Japan	1,381
9	KONINKLIJKE PHILIPS ELECTRONICS N.V.	Netherlands	1,378
10	HEWLETT-PACKARD DEVELOPMENT COMPAI	United States of America	1,310
11	SIEMENS AKTIENGESELLSCHAFT	Germany	1,292
12	INTEL CORPORATION	United States of America	1,250
13	ROBERT BOSCH CORPORATION	Germany	1,247
14	BOE TECHNOLOGY GROUP	China	1,227
15	TOYOTA JIDOSHA KABUSHIKI KAISHA	Japan	1,214
16	PANASONIC CORPORATION	Japan	1,185
17	HITACHI, LTD.	Japan	1,165
18	HALIBURTON ENERGY SERVICES, INC.	United States of America	1,121
19	SHARP KABUSHIKI KAISHA	Japan	1,073
20	TENCENT TECHNOLOGY (SHENZHEN) COMPA	China	981
21	FUJIFILM CORPORATION	Japan	947
22	NEC CORPORATION	Japan	895

EXAMPLE: INSTITUTIONAL SECTORS

■ Business ■ Individual ■ University ■ Government



PATENT DATA BY INDUSTRY

Methods



Concordance
Probabilistic match
Examination of patents
Applicant's industry code
Etc.

Results



Yale
MERIT
OTAF (USPTO)
Dan Johnson (OECD)
Brian Silverman (U. of Toronto)
Ulrich Schmoch, *et. al.* (EC)

No satisfactory solution

PATENT DATA BY TECHNOLOGY

Methods



**Keywords search
Mapping IPC to technology
Tagging by examiners**

Results

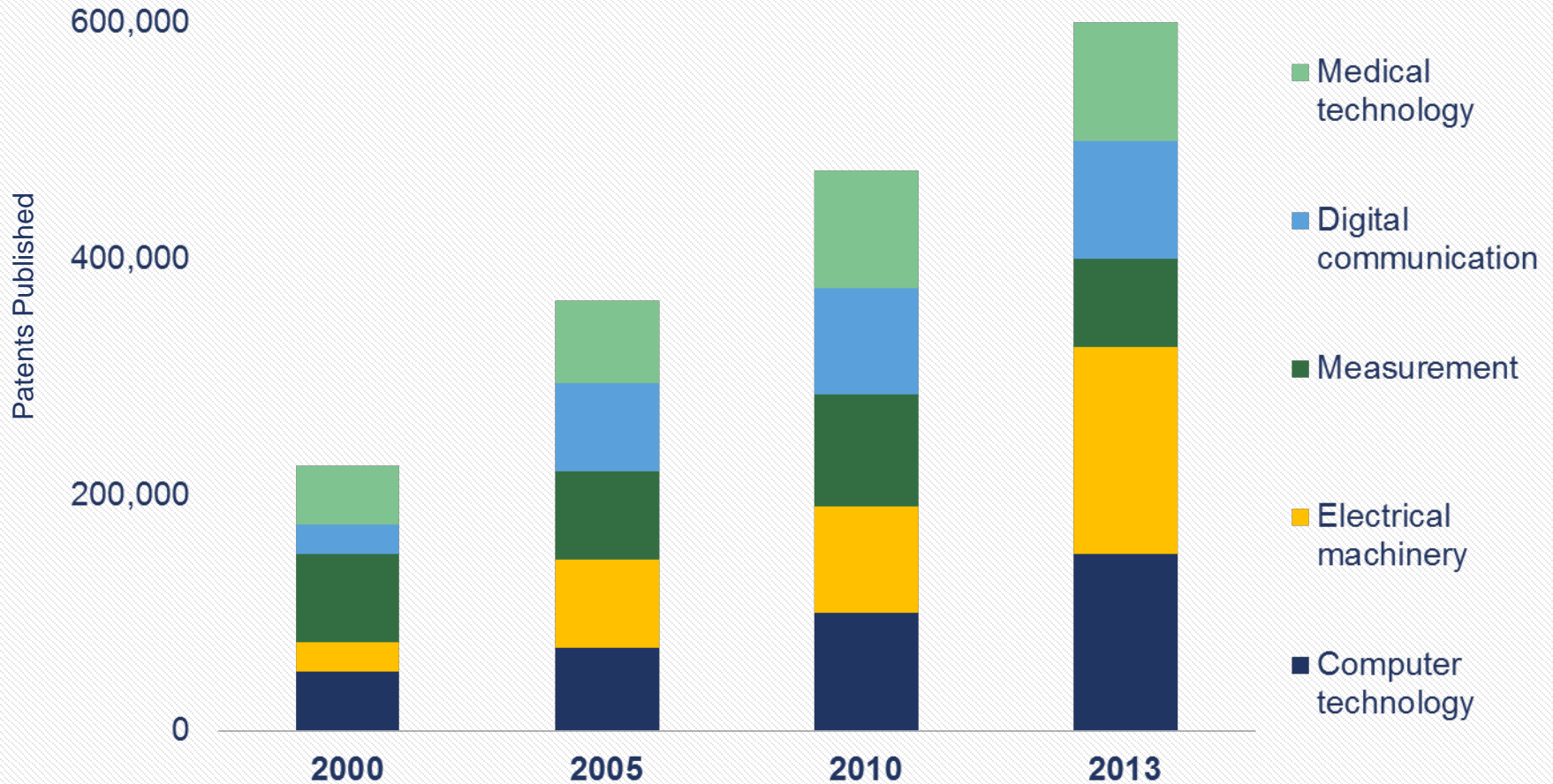


**IPC-Technology (WIPO)
Energy related technology (WIPO)
Nanotech (EPO)
Environmentally sound
technologies (EPO)
Etc.,**

Satisfactory results

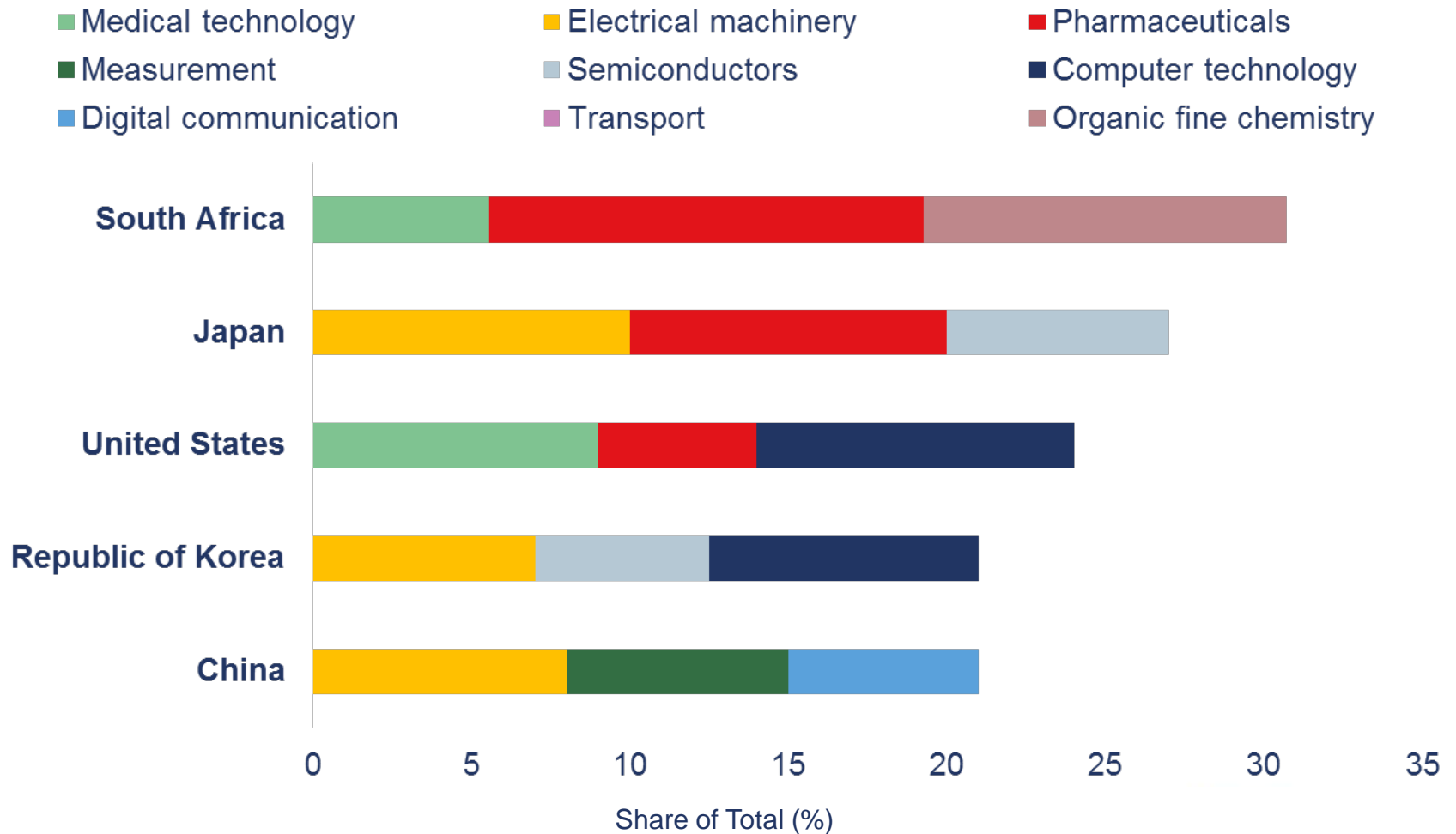
5 TECHNOLOGIES

account for 30% of all published applications



TOP TECHNOLOGIES

vary by country



KEY MESSAGES

Need to provide data that are policy relevant.

Challenging and resource intensive tasks, but possible to achieve good results.

Need to develop **international standard** and work on other IP types.

Mosahid Khan
Head, IP Statistics
WIPO

mosahid.khan@wipo.int

www.wipo.int/ipstats