Priorities, Equivalents and Patent Value

Dietmar Harhoff

Institute for Innovation Research, Technology Management
and Entrepreneurship (INNO-tec)

Ludwig-Maximilians-Universität (LMU) München

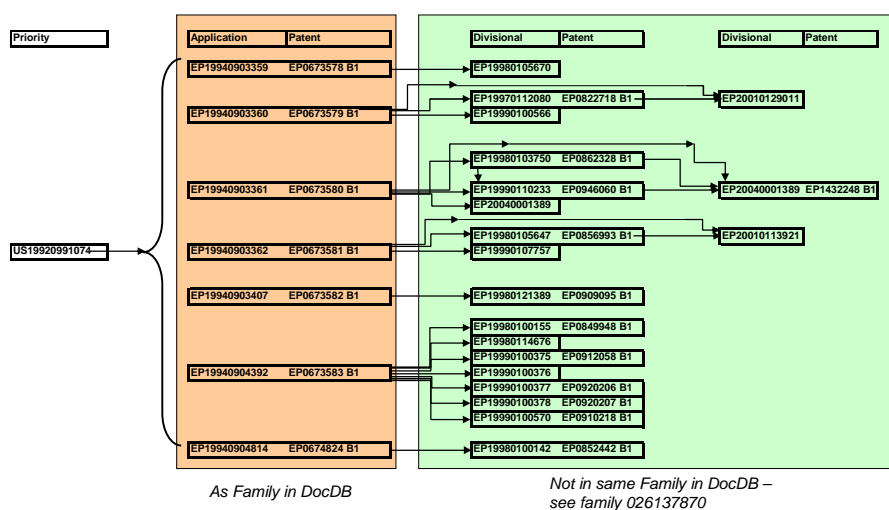**NBER Productivity Group Meeting**
December 5th, 2008

Research Issues

- Single patents as unit of analysis are becoming less relevant.
- Alternatives? - "patent families"
- Patent families have two dimensions
  - international (well-known, see Putnam's work)
  - national (not well known)
- What is the motivation behind the increasing size of national patent families (=related patents protecting the same invention)?
- How to get data and construct "patent families"?

Examples (example 1 in the paper)

- Priority application US19920991074 ("cable set-top box", 91 claims, 183 pages description) was used to file 7 (very similar) EPO applications.
- An additional 16 divisionals were filed based on the first 7 applications.
- Another 3 (second generation) divisionals were filed on the basis of the first round of divisionals.
- Summing up – 26 applications of which (to date) 18 were granted.

©Dietmar Harhoff 2008

Examples

As Family in DocDB

Not in same Family in DocDB – see family 026137870

©Dietmar Harhoff 2008

Examples

- publication number EP1286749, applicant: Genentech Inc.
- title: Secreted and Transmembrane Polypeptides and Nucleic Acids Encoding the Same
- characteristics: 22 claims, 11 inventors, 30 priorities (mostly US and WO)
- 16 divisional applications
    - EP1683864, EP1657253, EP1666597, EP1666490
    - EP1666596, EP1686174, EP1657254, EP1666491
    - EP1666492, EP1659177, EP1666493, EP1666497
    - EP1666594, EP1657251, EP1700867, EP1702928

Examples
Patents in Cosmetics (example 2 in the paper)

- EP1707181, EP1707182, EP1707183, EP1707184, EP1707190
- priorities:
  FR20050050835 20050331 - FR20050050838 20050331 - FR20050050841 20050331 - FR20050050837 20050331 - FR20050050842 20050331
- for all of these EPO filings:
    - application date: March 28, 2003
    - publication date: October 4, 2006
    - applicant: major player in the cosmetics industry
    - same inventor
    - between 24 and 28 claims (industry average 12.2 claims)
    - same examiner
- NOTE: not included in TLS205 as related to each other.

Examples
Patents in Cosmetics

- Title of EP1707181
  Dye composition with a reduced content of starting materials, and process for dyeing keratin fibres using the same
- Title of EP1707182
  Dye composition comprising a fatty acid and process for dyeing keratin fibres using the same
- Title of EP1707183
  Dye composition comprising a non-ionic associate polymer, process for dyeing keratin fibres using the same
- Title of EP1707184
  Dye composition with a reduced content of starting materials, process for dyeing keratin fibres using the same and device therefore
- Title of EP1707190
  Dye composition comprising a hydrophobically modified nonionic cellulose and method for dyeing keratin fibres using it

---

Data Issues

- Collecting data on families used to be tedious business (see Graham and Harhoff 2006).
- Much better data available now – PATSTAT.
- What follows: a short report on what the dataset can do, and what the drawbacks are.
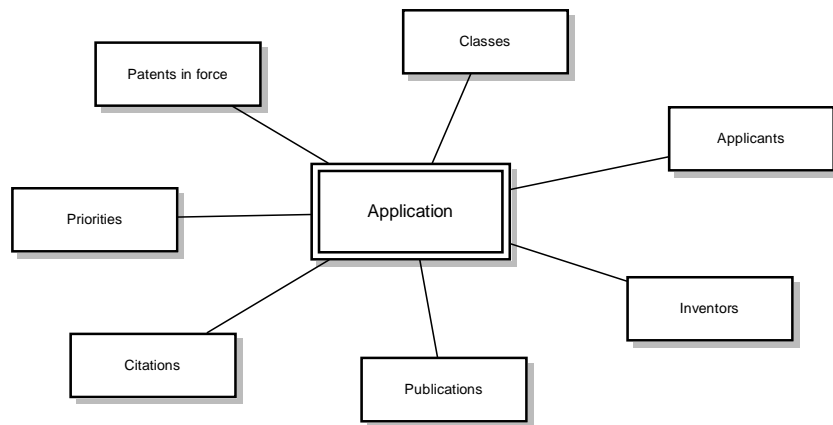
What is PATSTAT?

- OECD Taskforce on Patent Statistics - "EPO Worldwide Patent Statistical Database"
- distributed from the Vienna sub-office of the European Patent Office
- fee of 1000,00€ per year – 2 versions p.a. (April/October)
- most recent version: 2008-10
- http://www.epo.org/patents/patent-information/raw-data.html
- members of the Patent Statistics Task Force
  - Organisation for Economic Co-operation and Development (OECD)
  - World Intellectual Property Organisation (WIPO)
  - Japanese Patent Office (JPO)
  - US Patent and Trademark Office (USPTO)
  - US National Science Foundation (NSF)
  - European Commission (EC)
  - Eurostat and DG Research

©Dietmar Harhoff 2008

What is PATSTAT?

As of April 2008, PATSTAT consisted of 17 tables.

TLS201_APPLN	60,339,738
TLS202_APPLN_TITLE	43,489,959
TLS203_APPLN_ABSTR	13,300,418
TLS204_APPLN_PRIOR	25,334,588
TLS205_TECH_REL	2,073,752
TLS206_PERSON	33,664,886
TLS207_PERS_APPLN	118,126,651
TLS208_DOC_STD_NMS	15,955,146
TLS209_APPLN_IPC	260,101,620
TLS210_APPLN_N_CLS	21,520,217
TLS211_PAT_PUBLN	66,731,993
TLS212_CITATION	80,328,974
TLS214_NPL_PUBLN	10,312,865
TLS215_CITN_CATEG	15,551,918
TLS216_APPLN_CONTN	1,430,055
TLS217_APPLN_I_CLS	101,753
TLS218_DOCDB_FAM	53,357,975

©Dietmar Harhoff 2008

5

What is PATSTAT?

Source: Data Catalog v3.27 (in drive ../2008-04/DOC)

Comparative Advantages & Problems
Benefits

- comprehensive – data from more than 60 countries in one common format
- consolidated – priority numbers have been cast in a format that allows for easy analysis
- population data for long time horizons
- regular updates and error correction
- data on applicants and inventors
- setup of SQL database made easy – installation files come with data

Comparative Advantages & Problems
Drawbacks

- large – a monster of a database – just the unzipped text files consume 52.3 GB
- complex – institutional knowledge required
  - priority system (backbone of the international system)
  - PCT (most important application path)
  - citations (correcting double-counting, getting WO citations right, etc.)

LUDWIG-
MAXIMILIANS-
UNIVERSITÄT
MÜNCHEN

FAKULTÄT FÜR BETRIEBSWIRTSCHAFT
MUNICH SCHOOL OF
MANAGEMENT

13

©Dietmar Harhoff 2008

What can you do with the data?
Examples

- trace patent applications and grants in the international patent system
- study the patent portfolios of corporations
- compute correct citation counts for a number of countries (EP, WO, US, DE, FR, GB, …)
- study impact of IPRs on development and trade

LUDWIG-
MAXIMILIANS-
UNIVERSITÄT
MÜNCHEN

FAKULTÄT FÜR BETRIEBSWIRTSCHAFT
MUNICH SCHOOL OF
MANAGEMENT

14

©Dietmar Harhoff 2008

Priorities, Equivalents and Families
Types of Families

Application D1	**Priority P1**		
Application D2	**Priority P1**	**Priority P2**	
Application D3	**Priority P1**	**Priority P2**	
Application D4		**Priority P2**	**Priority P3**
Application D5			**Priority P3**

Source: Own computations based on PATSTAT April 2008.

©Dietmar Harhoff 2008

Priorities, Equivalents and Families
New Equivalents Data from 2008-04 (Table 1 in paper)

	Applications	Publications	Countries (Authorities)
n	27,069,434	33,434,129	24,856,791
families	7,307,933	7,307,933	7,307,933
p10	2	2	2
mean	3.70	4.58	3.40
median	2	3	2
p90	7	10	6
min (max)	2 (314)	2 (396)	1 (47)

Source: Own computations based on PATSTAT April 2008.

©Dietmar Harhoff 2008

Priorities, Equivalents and Families

example: equivgroup **C-A366628**, 12 publications, 7 applications

appln_id	equivgroup	publno
000963193	C-A366628	AU559533B2
000963194	C-A366628	AU1083583A
009409298	C-A366628	DE3204681C2
009409298	C-A366628	DE3204681A1
015038108	C-A366628	EP0086359B1
015038108	C-A366628	EP0086359A2
015038108	C-A366628	EP0086359A3
024240744	C-A366628	JP3048853B
024240744	C-A366628	JP58148730A
024240744	C-A366628	JP1684629C
048077031	C-A366628	US4555834A
050088543	C-A366628	US4649641A

©Dietmar Harhoff 2008

Some Results from the Paper

- background information – types of priorities
- size of equivalent families at EPO is increasing over time.
- technical focus: electrical engineering/IT, instruments
- strongly involved: applicants from JP, GB, US

©Dietmar Harhoff 2008

Some Results from the Paper

Table 5
Regression Analysis of Equivalents Family Size at the EPO

Variables	(1)	(2)	(3)	(4)
av. number of claims	0.000821***	0.000698***	0.000670***	0.000416***
av. number of citations (fwd)	0.000832***	0.000385***	0.000109	-0.00229***
av. number of X-type references	0.00144***	0.00107***	0.000976***	0.000933***
number of jurisdictions				0.00199***
DE		reference case	reference case	reference case
FR		-0.00445***	-0.00462***	-0.00700***
GB		0.0204***	0.0192***	0.00377***
JP		0.0308***	0.0293***	0.0217***
other		0.0113***	0.0108***	0.000817
US		0.0237***	0.0217***	0.00940***
electrical engineering			reference case	reference case
instruments			0.00420***	0.00118***
chemistry			0.00712***	-0.00918***
process engineering			0.000637	-0.00577***
mechanical engineering			-0.00696***	-0.00761***
consumption & construction			-0.00505***	-0.00728***
dummies for applications years	yes	yes	yes	yes
Observations	1,128,752	1,128,752	1,128,659	1,128,659
R-squared	0.023	0.032	0.034	0.132

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

FAKULTÄT FÜR BETRIEBSWIRTSCHAFT
MUNICH SCHOOL OF MANAGEMENT

©Dietmar Harhoff 2008

Wrap-up

- PATSTAT - comprehensive, but unwieldy database with untapped potential
- Biggest advantages:
  - coverage (time, countries)
  - consolidated priority linkages across countries and authorities
  - unified identification system for applications, publications, inventors, applicants
  - citations (PL and NPL)
- Some standardized output (equivalents, citations) available on my website.
- Why do firms construct complex families? Protect value vs. maintain questionable applications.

LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

FAKULTÄT FÜR BETRIEBSWIRTSCHAFT
MUNICH SCHOOL OF MANAGEMENT

©Dietmar Harhoff 2008

The PATSTAT Database – A Primer

Dietmar Harhoff

Institute for Innovation Research, Technology Management and Entrepreneurship (INNO-tec)

Ludwig-Maximilians-Universität (LMU) München

NBER Productivity Group Meeting
December 5th, 2008

Overview

- What is PATSTAT?
- Advantages & Problems
- What can you not do with the data? Examples
- How to best work with the data?
- Applications, Documents and Grants
- Priorities, Equivalents and Families
- Citations
- Applicants and Inventors
- Wrap-up

What is PATSTAT?

- OECD Taskforce on Patent Statistics - "EPO Worldwide Patent Statistical Database"
- distributed from the Vienna sub-office of the European Patent Office
- fee of 1000,00€ per year – 2 versions p.a. (April/October)
- most recent version: 2008-10
- http://www.epo.org/patents/patent-information/raw-data.html
- members of the Patent Statistics Task Force
  - Organisation for Economic Co-operation and Development (OECD)
  - World Intellectual Property Organisation (WIPO)
  - Japanese Patent Office (JPO)
  - US Patent and Trademark Office (USPTO)
  - US National Science Foundation (NSF)
  - European Commission (EC)
  - Eurostat and DG Research

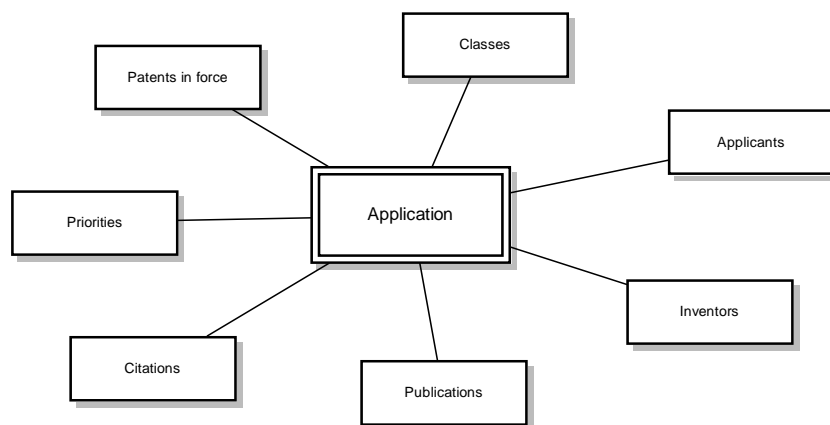©Dietmar Harhoff 2008

What is PATSTAT?

As of April 2008, PATSTAT consisted of 17 tables.

Table	Count
TLS201_APPLN	60,339,738
TLS202_APPLN_TITLE	43,489,959
TLS203_APPLN_ABSTR	13,300,418
TLS204_APPLN_PRIOR	25,334,588
TLS205_TECH_REL	2,073,752
TLS206_PERSON	33,664,886
TLS207_PERS_APPLN	118,126,651
TLS208_DOC_STD_NMS	15,955,146
TLS209_APPLN_IPC	260,101,620
TLS210_APPLN_N_CLS	21,520,217
TLS211_PAT_PUBLN	66,731,993
TLS212_CITATION	80,328,974
TLS214_NPL_PUBLN	10,312,865
TLS215_CITN_CATEG	15,551,918
TLS216_APPLN_CONTN	1,430,055
TLS217_APPLN_I_CLS	101,753
TLS218_DOCDB_FAM	53,357,975

©Dietmar Harhoff 2008

What is PATSTAT?

- The data come largely from the EPO's DOCDB database – this is probably the largest database for search and examination world-wide.
- Additional variables have been added by the EPO and other members of the Task Force.
- The data quality is very, very good – but not perfect.
- Using the data requires some setup costs – with the routines transferred to GIT, that should be a lot easier.
- Very attractive: update with new data every 6 months.
- Fee: 1000€ per year
- Available from: EPO, Vienna Branch Office.

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

FAKULTÄT FÜR BETRIEBSWIRTSCHAFT
MUNICH SCHOOL OF MANAGEMENT

25

©Dietmar Harhoff 2008

What is PATSTAT?

Source: Data Catalog v3.27 (in drive ../2008-04/DOC)

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

FAKULTÄT FÜR BETRIEBSWIRTSCHAFT
MUNICH SCHOOL OF MANAGEMENT

26

©Dietmar Harhoff 2008

Comparative Advantages & Problems
Benefits

- comprehensive – data from more than 60 countries in one common format
- consolidated – priority numbers have been cast in a format that allows for easy analysis
- population data for long time horizons
- regular updates and error correction
- data on applicants and inventors
- setup of SQL database made easy – installation files come with data

©Dietmar Harhoff 2008

Comparative Advantages & Problems
Drawbacks

- large – a monster of a database – just the unzipped text files consume 52.3 GB
- complex – institutional knowledge required
  - priority system (backbone of the international system)
  - PCT (most important application path)
  - citations (there is more than counting)

©Dietmar Harhoff 2008

What can you do with the data? Examples

- trace patent applications and grants in the international patent system
- study the patent portfolios of corporations
- compute correct citation counts for a number of countries (EP, WO, US, DE, FR, GB, …)
- look at international linkages in citations
- study impact of IPRs on development and trade

©Dietmar Harhoff 2008

How to best work with the data?

- DH's preferred version
  - UNIX machine with Stata MP
  - 16 GB of RAM
  - large hard disks
  - standard tables for a broad set of tasks
    - patent families
    - citation counts
    - structured tables with dates of application, search, grant etc.
  - files for reading data into Stata and setting up standard tables are available upon request
- SQL alternative – supported by the task force (installation files, etc.) – drawback: slow for non-standard tasks

©Dietmar Harhoff 2008

Applications, Documents, Grants

- do not use the term "patent" (unless you need to be broad and imprecise)
- The term „patent" does not tell you if you are talking about
  - a *patent grant*
  - a *patent application* in general
  - any of the *documents* published during the course of examination
  - **Note: little need for differentiation in the US system prior to AIPA in 2000.**
- One or multiple priorities
  leading to one or several application(s)
  leading to one or several publication(s)
  leading to one grant document each and/or
  leading to (possibly) further applications/documents
  (reexamination, re-issue, amended patent, divisional, continuation, continuation in parts)

©Dietmar Harhoff 2008

Applications, Documents, Grants

- Applications and publications in PATSTAT
  - applications – identified by **appln_id**
  - publications (multiples for each application) – identified by **pat_publn_id**
- These identifiers change with every release of the PATSTAT data – do not use them to select patents!
- If you need to identify patents across releases of the database – use the combination of
  - publn_auth (str2), publn_nr (str15), publn_kind (str2) for publications
  - appln_auth (str2), appln_nr (str15), appln_kind (str2) for applications

©Dietmar Harhoff 2008

Applications, Documents, Grants

- note: application numbers need not be identical with publication numbers
  - EP example - application id: 014847214, publications: EP1031782A2, EP1031782A3
  - US example – application id: 044078925, publications: US2003035113A1 - US6741357B2
- publications derived from the same application can have different publication numbers (as in the US)
- publication and application kind codes differ by country/office (read up in documentation)

Priorities, Equivalents and Families

- See paper for more on this topic.
- Paris Convention priority rights
  - backbone of the international patent system
  - option right for filing applications in foreign jurisdictions
- priority rights under multilateral agreements
  - Patent Cooperation Treaty (PCT)
- internal priority rights
  - divisionals
  - continuations

Priorities, Equivalents and Families

- Priority linkages can be used to identify "twins" within a patent domain (e.g., EPO) and across domains (e.g., EPO and USPTO).
- See Graham/Harhoff 2006 for an application.
- You can download a table with equivalents from my webpage.

©Dietmar Harhoff 2008

Priorities, Equivalents and Families
Some Statistics

- based on PATSTAT 2007-04 (but: new release available since October)
- 31,893,283 observations (documents)
- 7,108,969 **sets of equivalents** with
  - a maximum of 298 publications/246 applications
  - a mean of 4.48 publications/3.66 applications
- zipped dataset with ≈300 MB

URL for download of data
www.inno-tec.de/forschung/forschungsprojekte/
patent_cit_project/index.html

©Dietmar Harhoff 2008

Priorities, Equivalents and Families

example: equivgroup **C-A366628**, 12 publications, 7 applications

```
appln_id       equivgroup        publno
000963193      C-A366628        AU559533B2
000963194      C-A366628        AU1083583A
009409298      C-A366628        DE3204681C2
009409298      C-A366628        DE3204681A1
015038108      C-A366628        EP0086359B1
015038108      C-A366628        EP0086359A2
015038108      C-A366628        EP0086359A3
024240744      C-A366628        JP3048853B
024240744      C-A366628        JP58148730A
024240744      C-A366628        JP1684629C
048077031      C-A366628        US4555834A
050088543      C-A366628        US4649641A
```
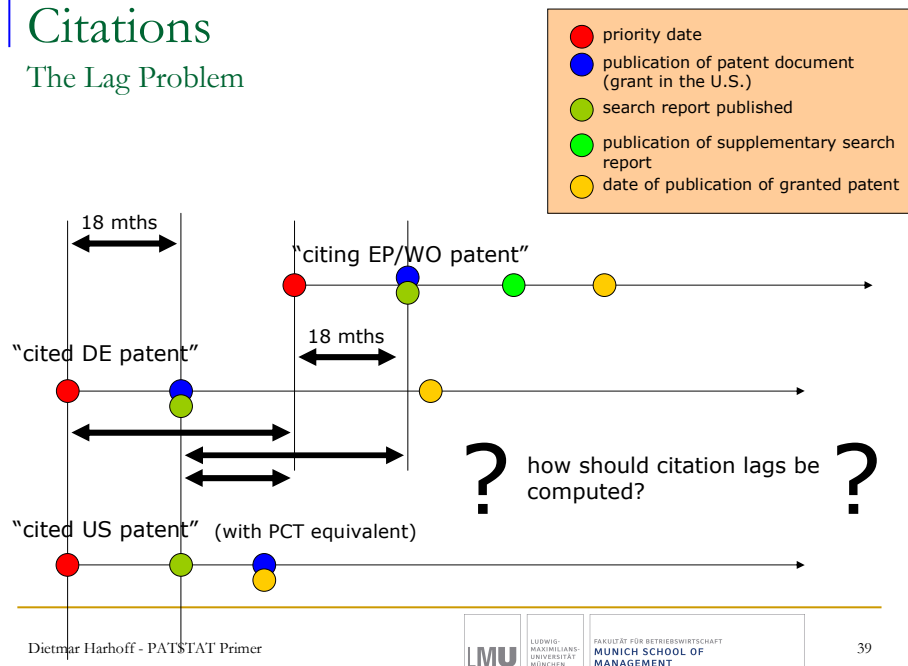
©Dietmar Harhoff 2008

Citations

- PATSTAT has the most comprehensive set of references available for research.
- All US and EP references (TLS212), also: DE, FR, GB, …
- All non-patent literature references (TLS214)
- Ready-to-use citation counts have been constructed and are available.
  - Citations are given for certain lag times and in total, e.g. *US_ncit_3yrs* or *EP_ncit_3yrs_x* or *EP_ncit_tot_self*
- Citation counts have been corrected for a number of problems, in particular an alias-problem.

©Dietmar Harhoff 2008

Citations

The Lag Problem

🔴	priority date
🔵	publication of patent document (grant in the U.S.)
🟢	search report published
🟢	publication of supplementary search report
🟡	date of publication of granted patent

18 mths

"citing EP/WO patent"

"cited DE patent"

18 mths

? how should citation lags be computed? ?

"cited US patent" (with PCT equivalent)

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

FAKULTÄT FÜR BETRIEBSWIRTSCHAFT
MUNICH SCHOOL OF MANAGEMENT

Citations

The Alias Problem

Excerpt from RepEC

Stuart J. H. Graham & Bronwyn H. Hall & Dietmar Harhoff & David C. Mowery, 2002. "Post-Issue Patent "Quality Control": A Comparative Study of US Patent Re-examinations and European Patent Oppositions," NBER Working Papers 8807, National Bureau of Economic Research, Inc. [Downloadable!]

Other versions:

* Stuart J. H. Graham & Bronwyn H. Hall & Dietmar Harhoff & David C. Mowery, 2003. "Post-Issue Patent "Quality Control": A Comparative Study of US Patent Re-examinations and European Patent Oppositions," Industrial Organization 0303009, EconWPA. [Downloadable!]

* Stuart Graham & Bronwyn Hall & Dietmar Harhoff & David Mowery, 2002. "Post-Issue Patent "Quality Control": A Comparative Study of US Patent Re-examinations and European Patent Oppositions," Department of Economics, Working Paper Series 1046, Department of Economics, Institute for Business and Economic Research, UC Berkeley. [Downloadable!]

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

FAKULTÄT FÜR BETRIEBSWIRTSCHAFT
MUNICH SCHOOL OF MANAGEMENT

Citations
The Alias Problem

Source of Reference

	DE	EP	FR	GB	JP	OT	US	WO
without	12.4	18.3	6.6	6.2	9.6	2.6	36.2	8.3
with	11.9	31.8	6.6	6.0	9.3	2.6	29.9	1.8

correction for equivalents

After correcting for equivalents, the share of within-EP referencing increases from 18.3% to 31.8% of all patent references.

In US citation data, about 20% of referenced documents are of non-US origin.

©Dietmar Harhoff 2008

Applicants and Inventors

- Each inventor and applicant has a *person_id* identification in the data.
- Most of these are not consolidated yet – the same inventor will have different person_id's.
- Similar with applicants – several projects for consolidation under way at EPFL, Bologna, OECD, Leuven, … see paper by Grid Thomas et al. (2008).
- Data quality can be messy in some countries. Best inventor address data at DPMA or EPO.

©Dietmar Harhoff 2008

Wrap-Up

- Unique, but unwieldy database with untapped potential
- Biggest advantages:
  - coverage (time, countries)
  - consolidated priority linkages
  - unified identification system for applications, publications, inventors, applicants
- Standard SQL installation and alternatives for Stata are available.
- Some standardized output (equivalents, citations) is available on my website.

LMU

LUDWIG-
MAXIMILIANS-
UNIVERSITÄT
MÜNCHEN

FAKULTÄT FÜR BETRIEBSWIRTSCHAFT
MUNICH SCHOOL OF
MANAGEMENT

©Dietmar Harhoff 2008