Unfielded bibliographic data; ancient history or current challenge?





## Contents

International patenting activity and the Paris	
Convention of 1883	3
Standards for information exchange; the work	
of ICIREPAT and WIPO ST.9	5
Problem solved? – the issue of partial or	
incomplete adoption	9
"It's data – but not as we know it…"	16
About the author	18

### International patenting activity and the Paris Convention of 1883

In the centuries leading up to the Industrial Revolution, patents tended to be a strictly local affair. The idea that the same invention may already be known or protected elsewhere was generally immaterial. In fact, the definition of 'inventor' was applied equally to the very first originator of a product or process, or to the person responsible for first importing a known idea into their own country.

A non-national applicant could not always expect to be treated on an equal footing with a national one; indeed, under the law of the short-lived Republic of Texas, only citizens of the Republic or those who were seeking citizenship were eligible as patent applicants, and even if they were granted a patent it would only stay in force for as long as the proprietor remained within the country.

Throughout the nineteenth century, the concept that a single person (natural or legal) could hold a family of equivalent patents in multiple countries was much less recognised in national patent laws, and consequently the need for patent offices to exchange information about 'equivalent' applications was not of pressing concern.

On 11th April 1851, the United Kingdom Parliament passed the Protection of Inventions Act (14 & 15 Vict. Cap. VIII), a mere 19 days before the Great Exhibition of 1851 was due to open. This Act allowed exhibitors at the Exhibition to disclose their inventions until the end of that year without jeopardising their rights to a subsequent patent grant.

By the time of the corresponding 1873 Weltaustellung in Vienna, some exhibitors were still reluctant to attend due to fears of lack of protection for their products overseas, irrespective of whether they were patented in the exhibitors' home country; accordingly, the Austro-Hungarian Government passed a special law "for

the provisional protection of articles introduced at the Vienna Exposition". This stop-gap measure echoed the 1851 British practice and steps taken at later Exhibitions in 1855 (Paris), 1862 (London) and 1867 (Paris), but was clearly not satisfactory as a long-term solution.

Continuing concerns from industry and lawyers were voiced during the parallel International Patent Congress the same year, which was the starting point of negotiation towards the Paris Convention for the Protection of Industrial Property, ten years later. This laid down for the first time the agreed principles of national treatment (i.e. equally favourable to non-national and national applicants) and the mutual recognition of priority of applications made in any of the contracting states. Thus was established the means for identification of patent families, and the need for exchange of bibliographic data between national patent offices.

## Standards for information exchange; the work of ICIREPAT and WIPO ST.9

After the entry into force of the Paris Convention on 7th July 1884, and into the first half of the twentieth century, published patent documents began to carry additional information relevant to priority claims. However, the means of recording these data were not universally consistent; neither was the level of detail provided.

An example is shown at *Figure 1* and *Figure 2*; the British document refers to the priority date of a US filing, but fails to supply the priority application number. This makes it more difficult for present-day database producers to link together members of the same patent family with any degree of certainty, and certainly cannot be achieved using digitised data on a purely algorithmic basis.

N° 29,587	A.D. 1904
(Unde	er International Convention.)
Date claimed for Patent being date of first United States),	t under Patents Act, 1901, t Foreign Application (in
Date of Application (in th Accepted, 19th Jan., 19	he United Kingdom), 31st Dec., 1904 05
CO	MPLETE SPECIFICATION.
	s for Stretching Fence and other Wires.

Figure 1: United Kingdom equivalent patent, with incomplete priority data

No	. 765,523.			Patented July 19, 1904.	
	United	STATES	PATENT	OFFICE.	a.
	MARK	TURNBULL, OF	GODERICH, CAN	ADA.	
6) (M)		WIRE-ST	RETCHER.		
		AND DESCRIPTION OF A DE	s Patent No. 765,523, Serial No. 206,899. (No mode		

Figure 2: United States grant of priority application, with serial number.

Other bibliographic elements such as the inventor/applicant names were usually present on published documents of this era, but not necessarily separated out or presented in what we would recognise as distinct fields; *Figure 3* from the same two patents shows that each publishing authority recorded slightly different information about the inventor, and in both cases it is embedded in the first paragraph of the main text.

1, MARK TORNBULL, of St George's Rectory, Goderich, Ontario, Canada, Clergyman, do hereby declare the nature of this invention, and in what manner the same is to be performed, to be particularly described and ascertained in and by the following statement :---

To all whom it may concern: Be it known that I, MARK TURNBULL, a citizen of the Dominion of Canada, residing at Goderich, in the Province of Ontario, Domin-5 ion of Canada, have invented new and useful Improvements in Wire-Stretchers, of which the following is a specification.

Figure 3: Variable inventor information on GB specification (above) and US specification (below)

Equally important for information retrieval purposes is the lack of a distinct abstract – the nearest equivalent was sometimes found as a short paragraph at the beginning of the text, which could (as now) vary hugely in its informational value, as a means of informing third parties about the central features of the invention. The increasingly-global business of filing and obtaining patents brought new challenges to users in the form of multi-lingual texts, numbering systems and calendars. The Japanese granted patent at *Figure 4* may be clearly identified as number 131055 (red box), but only if the user is familiar with the appropriate sy mbol set; in the same way, the date of grant can be understood as 13th July, but in the year Showa 14, not 1939 (blue box).



Figure 4: Bibliographic data from Japanese grant document

By the 1960s, the issues of data exchange and information retrieval began to be addressed by a committee formed of the major examining (as opposed to registering) patent offices. The ICIREPAT group (International Cooperation for Information Retrieval among Examining PATent offices) developed a number of documentation standards over the period between 1962-1970. By 1967, an early version of bibliographic field identifiers, known as INID codes (ICIREPAT Numbers for the Identification of Data, later WIPO ST.9) was introduced as Recommendation STAC III No.62d, followed later by a Standard Code for Identification of Different Kinds of Patent Documents (ICIREPAT SI.8, later to be known as WIPO ST.16, the Kind of Document suffix codes).

From 1970 onwards, with the completion of the Patent Cooperation Treaty and in the run-up to its entry into force, the need for standardization – at least between the International Search Authorities – became more acute. The early version of INID codes (later re-designated as "International Numbers..." rather than "ICIREPAT Numbers...") had already been adopted by a number of major offices, appearing gradually on US documents after 4th August 1970 *(see Figure 5)*, and an expanded version of the codes was brought into force on 1st January 1973. Documents published at the time often emphasise that a primary motivation for the use of these new codes was to assist users to identify key bibliographic elements of patent documents published in unfamiliar languages or scripts; this need was recognised long before the 21st century tsunami of Chinese and other Asian languages began to make itself felt to the patent information industry.

Uni	ted States	s Patent	-		(11) 3,	522,778
[72]	P. 54	Villiam F. Porath O. Box 91, Gillett, Wisconsin 4124; filton A. Porath, Route 1, Suring, /isconsin 54174	3,001,843 3,180,642 3,263,996 3,312,438 3,366,079	9/1961 4/1965 8/1966 4/1967 1/1968	Davis Reineman Braun Goetz, et al Koransky. et al	108/ 108/151X 273/3. 248/188. 248/43
(21) (22) (45) Continu	Patented: A	14,629 pril 9, 1969 ug. 4, 1970 r. No. 549,962. May 3, 1966.	Primary Exa Assistant Ex Attorney—V	aminer-O		d Strampel
[54]	GAME TABLE 14 Claims, Drawin	ag Figs.			e table and leg structure	
[52]	U.S. Cl	108/153, 273/3, 248/188.1, 248/431	larly spaced	l brace mer	the same including a plur mbers in substantially abu- table and a plurality of ang	tting relatio
	Int. Cl					
[51]			loos alea in	arbetantially		
[51] [50]		108/1 to 108/1 to 188.1, 431; 273/3.3.1, 3.23.4, 3.3, 3.5, 11	bers at or n manently o removal an	ear the cen r detachabl d replacement	y abutting relation with th tral axis of the table. The ly secured to the brace a ent of three or more legs	e brace men legs are per members, th being accorr
	11:150 to 157; 24	108/1 to 48/188.1, 431; 273/3.3.1, 3.23.4,	bers at or n manently o removal an plished by a constructed	ear the cen r detachabl d replaceme a single nut as a pool ta	tral axis of the table. The ly secured to the brace r	e brace men legs are per members, th being accom in the table i or permanen

Figure 5: Early example of US use of INID codes

# Problem solved? – the issue of partial or incomplete adoption

The success of any standard lies in the hands of the actual or prospective adopters. Unless a standard is convenient and straightforward to implement, there is always the risk that it may become distorted in day-to-day practice, either through mis-application of the recommendations or adoption of only parts of the standard, thus creating areas of uncertainly for the subsequent searcher.

The latest version of the WIPO "<u>CWS Survey on the use of WIPO Standards</u>" lists returns from up to 44 patent offices worldwide, across all standards. The results for the 40 offices which returned the questionnaire with respect to their use of the bibliographic standard ST.9 show that 33 offices reported complete implementation, 2 partial implementation and 5 non-implementation. It is of interest to note that the two countries reporting only partial implementation (Saudi Arabia and El Salvador) both noted that production of a standardised front page was not mandatory under their patent law, which accounted for the non-implementation. It is perhaps a salutary reminder that users and standard-setting organisations alike should not assume that all national offices have adopted – or will adopt – the same basic documentation practices as the well-established or mature patent systems.

#### Progress towards completing the puzzle

The 1973 edition of the INID code listing contained 33 field codes. Notable by their absence is any method of recording information about designated states in regional patent systems or international patent applications (neither the EPO nor the PCT were operational at this time), and certain elements now considered mandatory, such as INID (13) for the Kind of Document code.

Only a few patent offices in this period published their specifications more than once (the Netherlands, West Germany, Japan) so the latter identifier was yet to

assume its current importance. The current edition of ST.9 has 59 available fields, including the new (80) and (90) series for regional and international data; the only codes to be deleted since 1973 are INID (53) for the Universal Decimal Classification and INID (55) for Keywords, both dropped in 1997.

Encouraging though it may be to observe an 80+ % report of 'complete implementation' in the CWS survey, this still leaves both user and database producer with no hard data on the situation of the remaining 100+ patent offices worldwide. These are generally smaller offices in terms of volume of patents granted per year, which are also less likely to originate documents or bibliographic data in electronic form. Consequently, converting the patent data from these smaller offices into standardised bibliographic databases is doubly complicated. Even today, it raises challenges to achieving the stated objective of the INID standard as per paragraph 2 – that of improving access to documentation "without knowledge of the language used and the industrial property laws, conventions or treaties applied".

Furthermore, a questionnaire report of 'complete implementation' does not always guarantee that all parts of the standard are used with equal efficiency. It was foreseen that smaller patent offices may not be accustomed to recording the level of detail which the current standard calls for, so it is permitted to group certain elements under a single code (e.g. under code (30) instead of defining elements (31), (32) and (33) separately). A similar concession is found under Paragraph 12 of the Standard, which recognises that not all codes are used on every document, and states that *"If bibliographic data to which INID codes are assigned...do not appear on the first page of a patent document...because they are not applicable...it is not necessary to call attention to the non-existence of such elements..."*. These variations are convenient but may lead to unintended consequences.

#### Variations in adoption

The same caveats which are designed to improve implementation can result in patent offices only gaining familiarity with a small selection of the available codes (those which are used regularly). When faced with an unusual data element, it may

not be recorded in compliance with the standard simply because the applicant, patent office or both are unaware that provision is made within the standard for it to be recorded systematically, rather than on an ad hoc basis.

One example of this problem is INID code (83), designed to capture information about microorganism deposits as part of the patent application process. Only a very small proportion of a typical year's filing activity will include documents where these data are present. Different offices seem to vary in their use of the available INID code and corresponding data tag, or recording the information in free text form elsewhere in the specification.

*Figure 6* shows an example of two documents from different offices, where the same data are treated differently.

(21)申请号 201880056878.3         (72)发明人 M. 马吉德 K. 纳加布芬都綱 L. 蒙德库尔 S. 马吉德           (22)申请日 2018.08.30         (74)专科代理机构 北京市總沈律师事务/ 1105           (30)优先权数据         (74)专科代理机构 北京市總沈律师事务/ 11105           (35)PCT国际申请的申请数据         (74)专科代理机构 北京市總沈律师事务/ 11105           (22)D130697 2017.08.31 IN FCT/IS2018/01865 2018.08.30         (74)专科代理机构 北京市總沈律师事务/ 11105           (65)PCT国际申请的申请数据 <i>A61K 8-9</i> 9(2017.01)           (67)PCT国际申请的合本数据 <i>A610 19:0</i> /02(2006.01)           W02019/046508 EN 2019.03.07 <i>A610 17:0</i> /02(2006.01)           (63)生物保藏信息         (74)专利保证机构 北京市總公本の批		(10)申请公布号 CN 111050748 / (43)申请公布日 2020.04.21
(87)PCT国际申请的公布数据         A610         17/04(2006.01)           W02019/016508         EN 2019.03.07         A610         17/00(2006.01)           (83)生物保藏信息	<ul> <li>(22)申请日 2018.08.30</li> <li>(30) 优先权数据 IN201741030867 2017.08.31 IN</li> <li>(85)PCT国际申请进入国家阶段日 2020.03.02</li> <li>(86)PCT国际申请的申请数据</li> </ul>	(72)发明人 M.马吉徳 K.納加布莎娜姆 L.蒙德库尔 S.马吉徳 (74)专利代理机构 北京市柳法律师事务所 11105 代理人 张文辉 余晓文 (51)Int.01. <i>A61K 8-99</i> (2017.01) <i>A510 19-00</i> (2006.01)
MTCC 5856 2013.09.19 (71)申请人 萨米莱布斯有限公司	W02019/046508 EN 2019.03.07 (83)生物保藏信息 MTCC 5856 2013.09.19	A610 17/04(2006.01)
地址 印度班加罗尔 权利要求113页 1717/80/134後4页 第66页	地址 印度班加罗尔	

Microbial Type Culture Collection and Gene Bank and was assigned the strain number MTCC 5856) exhibiting 99% genetic homology with the known bacterial strains

Bacillus coagulans ATCC 31284, Bacillus coagulans NBRC 3887 and Bacillus coagulans ATCC 7050;

Figure 6: Comparison of CNIPA (top) and USPTO (bottom) practice in capture of Budapest Treaty data

Although the documents are not equivalents, both include claims which involve the use of the same microorganism; one authority uses the ST.9 data field, the other records the information in a free text paragraph.

This inconsistency in data handling has the unfortunate consequence of rendering the designated field label unreliable, such that it cannot be used as a means of ensuring comprehensive retrieval of all the relevant documents. Both patent offices and patent information professionals need to ensure that they keep up to date with legal or regulation changes which could affect their documentation. Although the Budapest Treaty entered into force in 1980, it was not until 1992 that new PCT Regulation 48.2(viii) required the PCT publication to include "the indications in relation to deposited biological material furnished under Rule 13bis separately from the description, together with an indication of the date on which the International Bureau received such indications".

Other variations in national patent office practice may lead to variations in interpretation and application of ST.9. Although these may be superficially trivial, they can cause problems in creating algorithms or search protocols for handling large volumes of multi-national patent data, when data elements are tagged "in the wrong place" when compared with the majority opinion. For example, a user wishing to analyse the time interval between publication of an unexamined application and publication of a corresponding granted patent can search granted European Patents (EP-B1 documents) and extract the desired data from INID (45) (*date of making available to the public...a patent document on which grant has taken place...*) and INID (43) (*date of making available to the public...an unexamined patent document, on which no grant has taken place...*) respectively.

The same protocol could be used for documents from most of Western Europe and Asia. However, to obtain the same information for US patent documents, the user would initially have to limit their search to US-B2 documents (since US-B1 documents are grants with no preceding published application) and then search for data at INID (45) and INID (65), not (43). The reason for the change is that the USPTO records any corresponding early publication date in this different field which is defined as "*Number [sic] of a previously published patent document concerning the same application*" and (fortunately for the user) has chosen to include the date [sic] of early publication as well.

A further range of issues which affect bulk data preparation and analysis arise because ST.9 is itself dependent upon other WIPO Standards. If these related standards are inconsistently applied, the resulting ST.9 fields become less reliable for search purposes. For example, INID (13) is listed as one of the minimum front-page data elements, and contains the Kind of Document code from ST.16. However, ST.16 is worded in a way which allows for substantial variation in the choice of code. Consequently, a database producer or searcher wishing (for example) to identify utility models can only do so by first ascertaining how each office has chosen to code such documents at INID (13), and then combining this data element with the correct INID (19) code for the office or organization publishing the document. A search for all documents using the -U or -Y code alone will miss utility models from (amongst others) France (FR-A3, FR-B3), Hong Kong (HK-A2) and Slovenia (SI-A2).

#### Empty fields and data holes

The huge increase in the popularity of the PCT as a mechanism for multi-national patent filing has brought with it some increasing difficulties in establishing a standard way of linking members of a patent family. This is due in part to local differences in documenting the transition from the international to the national or regional phase, which must precede any grants arising from an international application.

Although the ST.9 standard provides for a range of different fields to be created and populated, some of which can be seen as redundant, it is necessary in practice to consider all possible fields before being able to comprehensively retrieve a complete family. In *Figure 7*, the Indonesian gazette notes an ultimate Japanese priority at INID (30) and an effective national filing date, derived from the intermediate PCT application, at INID (22). However, the fields (86) and (87) – which are available to record the application and publication data for the PCT publication – are not used. By contrast, the Mexican document, which is part of the same family, records only the Japanese priority details at INID (30) together with the PCT data at INID (86) and (87). However, the date at INID (22) for this record is different to that of the Indonesian example. On further investigation, it transpires that the date recorded here is the national phase entry date, for which an entirely different field (85) already exists but has not been used.



63			(11)	MX 2017015338 A
IMPRIME MERCENO INSTATUTO MERCENO INDUSTRIAL				
(12)	SOLICITU	ID de PAT	TENTE	
(43) Fecha de publicación:	28/03/2018	(51) Int. CI:	F02B 75/04	(2006.01) (2006.01)
(22) Fecha de presentación: (21) numero de solicitud:	29/11/2017 2017015338		F02B 75/32 F02D 15/02 F16H 1/28 F16H 1/32	(2006.01) (2006.01) (2006.01)
		<ul> <li>(86) Número de solicitud PCT: JP 2016/062730</li> <li>(87) Número de publicación PCT: WO 2016/194511 (08/12/2016)</li> </ul>		
(30) Prioridad(es): 02/06/2015 JP 201	5-111813			

Figure 7: Comparison of PCT filing data captured by Indonesia (top) and Mexico (bottom)

This is not an isolated example of non-use of parts of the Standard. Policy decisions at a given patent office, made at one point in time for very good reasons, may be inappropriate in later years (how many of today's searchers can recall the disruption of adapting publication numbers to Y2K-compatible formats?). As approaches to patent search and analysis change, so too does the desire on the part of the searching community for more consistent use of different data fields, and usage of elements which were at one time considered 'obvious' or 'redundant' can now benefit from being harmonised. For example, although the International Bureau of the PCT makes use of INID code (26) to capture the language of publication of its publications, the European Patent Office, which publishes its specifications in three different languages, does not do so on its specifications, even though the information is stored in the corresponding EPO Register entry.

## "It's data – but not as we know it..."

There is no doubt that the majority (by volume) of raw electronic patent data is much better organised today than in previous years. However, converting old back files into usable electronic databases remains a significant challenge. It can be difficult to track down exactly when a given patent office started to comply with sophisticated user demands for standardised front pages and well-fielded electronic data. Exceptions to "the general rule" will always occur, even amongst major offices, whenever a standard leaves certain choices open to discretion.

Users still face challenges to achieve comprehensive retrieval, notably when handling publications from any patent offices which – whilst developing in other areas – may still lack an effective IT infrastructure or other technical support. But even the best-equipped publishing authorities do not have a 100% record.

*Figure 8* shows two modern examples of unfielded information on patent front pages. The PCT published application is coded as a WO-A4, and consists of a new set of claims, which will be the text forwarded for examination under any corresponding national phase. The date at INID (43) is the date of publication of the corresponding WO-A1 document, with the original claims; the "true" date when the amended claims were made available to the public is listed in a free text note found at the bottom of the front page, after all the INID-coded fields.

The Egyptian granted front page likewise includes some free text, this time underneath the INID (54) title. Despite it being in Arabic, any user who is familiar with the numbering system will be able to identity two calendar dates – 5th August 2008 and 4th August 2015 i.e. exactly 7 years apart. Despite the fact that there is no INID (13) kind of document code used, this information would lead the skilled user to infer that this is a 7-year utility model and not a 20-year patent.

Wherever it occurs, in legacy pages or in modern electronic data, the challenge of organising user-critical information into precise searchable fields remains one

which is nearly – but not quite – overcome. As long as exceptions to the rule continue to be published, there will be a need for the data provider and the data user to be on their guard in the constant battle to achieve highly relevant, focussed results from amongst the extraordinary volumes of worldwide patent information.

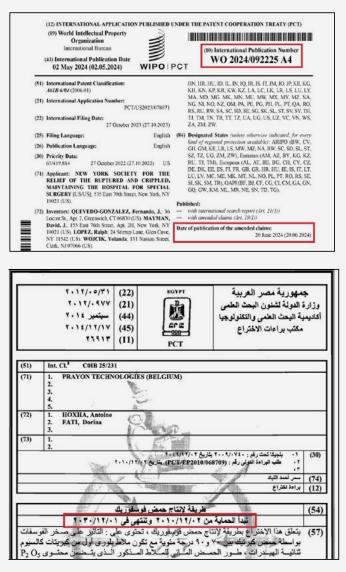


Figure 8: Unfielded data on a modern-day PCT application (top) and an Egyptian grant (bottom)

## **About the author**

Stephen Adams is the managing director of Magister Ltd., a UK-based consultancy specialising in patents information. Mr. Adams is a Qualified Patent Information Professional (number 20190044100092) and holds a B.Sc. in chemistry from the University of Bristol and an M.Sc. in Information Science from City University, London, as well as professional memberships of the Royal Society of Chemistry (RSC) and the UK's Chartered Institute of Library and Information Professionals (CILIP).

He is the author of three editions of "Information Sources in Patents", the latest published in 2020 by Walter de Gruyter KG, contributed several book chapters and written numerous articles in the field of patent information, including over 25 refereed papers for the Elsevier journal "World Patent Information".

His professional service includes the Editorial Advisory Board of "World Patent Information" between 2006-2020 and three terms on the Board of PIUG Inc., the International Society for Patent Information, as Director-at-Large (2002-2006) and Vice-Chair (2014-2016 and 2016-2018), as well as service on the management committee of the UK's Patent and Trade Mark Group over many years. He received the PIUG's Special Recognition Award in 2008 and the IPI Award in 2012 for outstanding contribution to patent information.



Tupolevlaan 81, 1119 PA Schiphol-Rijk The Netherlands Telephone +31 85 800 0024 info@lighthouseip.com

lighthouseip.com

The world's most complete IP data collection

