
Frequently Asked Questions

Changes possibly affecting usage of the reloaded Derwent World Patents Index[®] on STN[®].

The reloaded database (files WPINDEX, WPIDS and WPIX) offers many new and enhanced features for patent searchers, including all the benefits of the recently introduced International Patent Classification (IPC) Reform. Thomson Scientific and FIZ Karlsruhe have taken great care to ensure that existing uses of the *Derwent World Patents Index*[®] (DWPISM) will not be disrupted, while at the same time offering an array of new ways in which to retrieve and display the enhanced database content. Yet a few small changes were unavoidable as part of the reload which may in rare circumstances affect the use of the database. These changes are outlined below together with more general questions on the content and technical implementation of the reloaded file.

More detailed information on the reload can be found in the fully revised *DWPI* on STN online user guide: http://www.stn-international.com/training_center/patents/stn_guide.pdf.

Content

Q: What is the coverage of the new First Level data within the file?

A: Coverage varies by data element, patent kind (e.g. applications, granted patents and utility models) and time. In general though:
Original titles: Germany since 1968; EPO since 1978; US since 1975; PCT since 1978; Japan since 1975; Australia since 2004 and UK since 2004.
Original abstracts: Germany since 2000; EPO since 1978 in English and 2000 in French or German; US since 1975 and PCT since 1978.
First claim: Germany since 1968; EPO since 1991; US since 1993 and UK (database update 198409 to 199751).
Inventor full name and associated address and original patent assignee and associated address: Germany since 1968; EPO since 1978; US since 1975; PCT since 1978 and Japan since 1977 (no addresses)
Agent and associated address: Germany since 1968; EPO since 1978; US since 1975 and PCT since 1999.

At present the first claim is NOT available for PCT applications. In addition machine assisted translations of the JP original abstract and first claim are also NOT available.

Q: Historically Thomson Scientific included abstracts for many equivalent members of the patent family. Is this data still available ?

A: Yes, although this data is only available within the ABEQ field at the Member Patent Level. However in the case of GB-B, EP-B and US-A documents the equivalent abstract for certain time periods was actually the first or main claim and not a Thomson Scientific value-add abstract. For consistency this data has been moved to the new claims field at the Member Patent Level and so no longer qualifies as an equivalent abstract (ABEQ).

Q: What additional Thomson Scientific value-add abstracts are available ?

A: Approximately 750,000 additional abstracts have been made available in a text searchable format for the first time. The time period covered is 1995-1999. These “Documentation Abstracts” were the longest form of value-add abstracts prepared at the time and have been indexed in the ABDT field (Abstract, Documentation). In 1999 Thomson Scientific introduced the Extension Abstract which has been available since that time in the ABEX field on STN.

Q: Is it possible to search using the US national patent classification system ?

A: Yes. Original (Initial) US national classes as issued on the US document at the time of publication are available for US documents since 1975. This data is available at the Member Patent Level. At present Thomson Scientific is NOT updating the US national classes following any USPTO reclassification cycle. The introduction of this reclassification data is a planned future enhancement. When this data is introduced the Current US classes will be available at both the Invention and Member Patent Levels.

Q: Is it possible to search using the ECLA and F-terms patent classification systems?

A: No not at present. The introduction of this data are planned future enhancements.

Q: What are the Inventor Total (IN.T), Original Patent Assignee Total (PA.T) and Agent Total (AG.T) fields ?

A: Where possible Thomson Scientific has split inventor full names, original patent assignees and agent information from the associated inventor address, patent assignee address and agent address details. However, where it has not been possible to identify the constituent elements of the data the complete name and address information has still been made available within the corresponding “Total” field. In such cases the data is more suited to display rather than search.

Q: What is the Patent Assignee Limitation (PA.LIM) field ?

A: When creating the standardized version of the patent assignee (PA) Thomson Scientific does not take into account any country specific limitations on the assignees. This information is however now present as part of the Member Patent Level within the PA.LIM field. Additional fields available at the Member Patent Level are the original patent assignee (PAO), patent assignee address (PAA), patent assignee residence (PA.RES) and patent assignee nationality (PA.NAT) fields.

Q: How is Thomson Scientific implementing the IPC Reform within *DWPI*?

A: Thomson Scientific will be embracing all aspects of the IPC Reform within *DWPI*. At the Invention Level the current IPCs for the patent family will be provided. At the Member Patent Level both the current and original IPCs for each member patent will be provided.

Q: Is IPC Reform reclassification data available within *DWPI* yet ?

A: No, not yet. This reclassification data will be added to the file in the coming months. IPC Reform data on documents issued since January 1, 2006 is available within the file.

Q: How can the IPC Reform codes be searched and displayed ?

A: At the Invention Level the current IPC Reform IPCs for the patent family are displayed in the fields IPCI and ICPR based on the source level attribute. IPCI contains IPCs with a source level attribute of “B” indicating “Original” and ICPR contains IPCs with a source level attribute of “R” indicating “Reclassification”.

Note: on STN the “B” attribute has been indexed as “O” within the IPC Reform keywords field (IPC.KW) to indicate “Original” to avoid confusion with the “Basic” concept within patent families.

At the Member Patent Level the current and original IPC Reform IPCs for each constituent member patent will again be displayed in fields IPCI and ICPR according to the level attribute.

IPCI and ICPR are display fields only.

The /IPC index contains all IPCs irrespective of their status (e.g. Invention and Member Patent Levels, original and current, IPC Reform and IPC 1-7).

It is also possible to selectively search IPC Reform IPCs using the IPC.REF index. However, this must be combined with the document level identifier if the search is to be restricted to the Member Patent Level only (Publication/DLVL) in more focused searches.

Q: How do I search the IPC Reform attributes ?

A: IPC Reform attributes are searchable using keywords in the IPC Reform keywords field (IPC.KW). They are linked to the respective IPC codes using (S) proximity.

Q: What about pre-IPC Reform codes ?

A: IPC 1-7 IPCs continue to be available. Once all the IPC Reform reclassification data has been loaded they are most likely to be found as original IPCs at the Member Patent Level. New search and display fields have been introduced for IPC 1-7 IPCs which constitute original IPCs. These are IIC (Original, Main and Secondary), IICM (Original, Main), IICS (Original, Secondary), IICA (Original, Additional) and IICI (Original, Index).

However, as the IPC Reform reclassification data has not yet been loaded into *DWPI* then IPC 1-7 IPCs will also be found as current IPCs at both the Invention and Member Patent Levels. These continue to be searchable and displayable using the IC (Main and Secondary), ICM (Main), ICS (Secondary), ICA (Additional) and ICI (Index) fields.

The IC, ICM, ICS, ICA and ICI fields must be combined with the document level identifier (Publication/DLVL) if the search is to be restricted to the Member Patent Level only in more focused searches.

Q: What is the origin of the additional application information within the application number (AP) field ? Why does the application number (AP) field contain foreign priority information associated with a particular patent application ?

A: Historical application information whose origin has been impossible to determine was previously not output within *DWPI*. However, for completeness it was decided to include this data as part of the reload. This data may comprise for example an application, continuation, division or even a foreign priority number.

Q: What data is available within the APTS (Application Number, Thomson Scientific) and PRTS (Priority Number, Thomson Scientific) fields ?

A: Complete application information has only been included within *DWPI* since 1992. Before then coverage varied according to patent issuing authority. Where possible historical gaps in the coverage of application and priority data have been filled using original data from DE, EP, US, PCT and JP sources. This additional data is available within the APTS and PRTS fields in combination with the application and priority information which has been recorded in *DWPI* over time (AP and PRN fields). AP and PRN do not contain this additionally sourced data.

Q: What additional DCR numbers are available and how have they been generated ?

A: Additional DCR numbers are available in the Enhanced Polymer Indexing field (auto-generated from the Specific Compound Numbers) and in the Chemical Coding field (auto-generated from the Specific Compound and Registry Numbers).

Q: There is a new search field /ANX - what does it comprise ?

A: Prior to 1970 Thomson Scientific had separate printed services operating independently. This newly available data allows for cross-referencing older documents which may have appeared in more than one printed service.

Q: The REP field is no longer available – where has it gone ?

A: References formerly listed in REP had not been updated since the end of 1997. This data has now been removed from *DWPI* as comprehensive and up-to-date citation searching is available within the *Patents Citation Index*TM, DPCI.

Technical Implementation

Q: How is the data organised within the new *DWPI*? What are the Invention and Member Patent (publication) Levels ?

A: Two “levels” of data are now available within *DWPI*. The Invention Level is the traditional view of *DWPI* comprising bibliographic data, Thomson Scientific value-add title and abstracts, and general and, where appropriate, in-depth indexing. The Member Patent Level allows users to search and display bibliographic data and general indexing information associated with individual documents that make up the patent family Invention Level. This can allow very specific searching of individual documents. Additional data elements such as original titles and abstracts, claims, addresses and agent information are also present at the Member Patent Level.

Q: How do I search these Invention and Member Patent Levels ?

A: You can continue to search *DWPI* as you have always done if you are only interested in the traditional view of *DWPI* (the Invention Level). No additional search or display qualifiers are required.

Data elements which are common between the two levels, such as Thomson Scientific standardized patent assignees (PA), have the same search qualifiers. To restrict searches to the Member Patent Level field only for more focused searches requires the use of (L) proximity in combination with the document level qualifier /DLVL (Publication/DLVL).

Data elements which are unique to the Member Patent Level such as original patent assignees have been assigned new fields and so can be searched using the field qualifiers only (e.g. PAO).

Q: May I still use the Link (L) proximity operator ?

A: Yes, you may, but it may act in a slightly different manner than before. The Link proximity level has been used to organize the data into Invention and Member Patent Levels: one Link is associated with the invention or with any of the individual publications. It is used in combination with the document level qualifier /DLVL to restrict searches to the Member Patent Level only (Publication/DLVL).

The previous use to indicate that the search terms in the basic index belong to a common display field is no longer supported. For example :

=> s chain (L) sprocket

required ‘chain’ and ‘sprocket’ to reside in a common display field like AB or TI. In the reloaded file this is no longer the case, instead the operator gives the same results as a Boolean ‘AND’.

Q: Why are there two “Basic Indexes”; the standard Basic Index and the new Extended Basic Index ?

A: In order to retain the ability to easily search *DWPI* with great precision using the Thomson Scientific value-add titles and abstracts the new First Level data elements do NOT form part of the standard Basic Index, instead this data has been indexed separately in an Extended Basic Index.

Q: How does STN index patent and priority information that may cause false hits due to an overlapping number series?

-
- A: Distinguishing marks have been added to certain numbers in the index in order to remove ambiguities with independently running number series. Cases where this was deemed necessary are listed below. Some of these were already in operation on STN and in *DWPI* previously.

Patent Numbers

1. Chinese C documents
CN1023374 C
- 2a. Japanese B documents
JP2500002 B
- 2b. Japanese X documents (PCT-Transfers)
JP2005502887 X
- 2c. Japanese utility model publications JP Y and JP Z
JP08500005 U (JPY)
JP04600008 U (JPZ)
3. Taiwanese B documents
TW171625 B

Application Numbers

Utility Model Applications

ARU	AR1998-591U
ATU	AT2004-918U
AUU	AU2001-5900U
BRU	BR2004-3214U
BYU	BY2004-474U
CHU	CH2004-80928U
CLU	CL1995-387U
CNU	CN2004-20102658U
CRU	CR2002-150U
CZU	CZ2004-15975U
DKU	DK2003-213U
ECU	EC2001-4158U
EEU	EE2004-34U
ESU	ES2004-2859U
FIU	FI2005-93U
FRU	FR1995-5728U
GEU	GE2002-1076U
GRU	GR2001-200167U
HKU	HK1996-240223U
HRU	HR1999-71U
HUU	HU2003-297U
ITU	IT2005-VR1U
JPU	JP1993-27378U
KRU	KR2005-60U
MXU	MX2004-233U
PHU	PH2004-165U
RUU	RU1999-12236U

SGU	SG2003-17U
SKU	SK2000-289U
TRU	TR2005-2557U
UAU	UA2004-80702U
UZU	UY2003-4022U
TWU	TW2005-200609U
ZAU	ZA1982-3613U

US provisional applications
US2000-184362P

Q: What fields make up the standard Basic Index and the new Extended Basic Index?

A: In general the standard Basic Index comprises Thomson Scientific value-add fields and the Extended Basic Index comprises First Level data.

The standard Basic Index comprises:

Thomson Scientific value-add title, title terms, additional words and all abstract words (including Equivalent Abstracts, Extension Abstracts and Documentation Abstracts).

The Extended Basic Index comprises:

Original title, abstract and first claim (in English, German and French).

Q: The standard Basic Index of WPINDEX (non-subscriber file) and WPIDS (subscriber file) comprises the Thomson Scientific value-add Extension Abstracts and Documentation Abstracts. Does this mean that I can now display these abstracts as well even though I am not a Thomson Scientific subscriber?

A: No, even though the text of these abstracts is now available for searching within WPINDEX and WPIDS the abstracts can still only be viewed by eligible subscribers within WPIX (subscriber file with Extension Abstracts). Please contact your nearest Thomson Scientific office if you wish to discuss becoming an eligible subscriber to this data.

Q: Can I search in German and French language?

A: Yes, German and French language First Level data is available for some patent issuing authorities (original titles, abstracts and claims).

Q: What will I retrieve and display if I use the new abstract AB index/display field?

A: The following data is indexed under /AB:

Thomson Scientific value-add abstracts available for the Basic and equivalents of the patent family plus any available original author abstracts.

The /AB index is the only index which contains a mixture of Thomson Scientific value-add and original First Level data.

However, the AB display qualifier will only display the Thomson Scientific value-add abstract which is available for the Basic of the patent family. This means that although equivalent and original abstracts are indexed under /AB they can NOT be displayed using AB.

The Thomson Scientific value-add abstract for the Basic and equivalents of the patent family can be displayed using the ABS display qualifier.

Thomson Scientific value-add Documentation and Extension Abstracts and the Technology Focus field are NOT indexed under /AB.

Q: I seem to receive more documents in the new file with identical queries. Why is that?

A: There may be various reasons for why more documents are retrieved, for example

- i. Documentation Abstracts (ABDT) and Extension Abstracts (ABEX) have been added to the standard Basic Index in both WPINDEX and WPIDS.
- ii. There is more application information in the file.
- iii. Additional DCR references will yield more bibliographic documents after a structure search.

Q: I run scripts depending on update/entry dates. Will I encounter problems resulting from the changeover ?

A: While it has not been possible to preserve the entire history of updates in the new file, the updates in the old and reloaded files have been in sync since update 200610. Therefore scripts reaching back in time till 200610 should not be affected, yet you may recognize that the volumes of the updates are not identical for particular updates, since numbers of documents may vary according to different quality assessment processes applied at Thomson Scientific. For less frequently executed scripts it is recommended that you execute a final run shortly before the changeover.

Q: I run scripts depending on update/entry dates carrying proximity information pertaining to the updated data elements. Will I encounter problems resulting from the changeover ?

A: The all-engulfing update and entry dates (UP and ED) themselves no longer carry the proximity information of the data elements required to link the dates with the updated data element. Yet there are various update dates available for individual items carrying the required information. Please enquire at our HelpDesk for further particulars.

Q: I need to see when a particular patent publication entered the *DWPI* database. Can I still use the UPP display for this purpose ?

A: Yes, in contrast to other update dates STN has assigned the historical update dates for patent publications to the individual publications in the reloaded database.

Q: What does it mean when you state there are no “stopwords” any longer ?

A: Improved searching technology now allows the indexing of all words whereas previously some very prevalent ones such as “from”, “is”, “which” etc. (stopwords) could not be indexed. Since all words are now indexed, precision and recall will be improved and some terms have been made searchable which were not accessible at all previously.

Q: I see that standard patent assignee codes (PACOs) now have a suffix ‘-C’ attached. Why is this ?

A: Standard PACOs now have a type ‘C’ assigned to provide a uniform patent assignee code format. The ‘C’ stands for ‘company’. The standard codes will still be searchable without the suffix though.

Q: Application data can no longer be linked to PD (publication date), PY (publication year) and ICM (main IPC1-7) with paragraph (P) proximity. How can I accomplish this in the new file ?

A: While PN (patent number) and PC (patent country) can be linked with paragraph (P) proximity to the application data, for PD, PY and PK a more elaborate search statement needs to be employed within the new Member Patent Level: use link (L) proximity and employ the publication document level qualifier for this purpose (Publication/DLVL).

Q: Using the (T) operator to link AP (application number) and AC (application country) with AD (application date) and AY (application year) is no longer available. How can that be accomplished in the new file ?

A: Use of paragraph (P) proximity is recommended for the same purpose.

Q: The *DWPI* format for application and priority numbers seems to have changed. Where can I obtain a specification document ?

A: *DWPI* standard application and priority numbers have assumed a new format where the serial is front filled with leading zeroes to nine digits unless it is an exception like the new German applications numbers or Indian PCT transfer documents. A description is given in the fully revised *DWPI* on STN online user guide. [URI for the manual.]

Q: The *DWPI* format for some patent numbers seems to have changed. Where can I obtain a specification document ?

A: *DWPI* standard patent numbers have changed in a few cases. For example US applications are now 11-digits in length and PCT applications are now 10-digits in length with a four digit year irrespective of the publication date. A description is given in the fully revised *DWPI* on STN online user guide.

http://www.stn-international.de/training_center/patents/stn_guide.pdf

Q: The STN standard format for US published applications seems to have changed. Why has this change occurred ?

A: The STN standard for US applications are now 11-digits in length as devised by the USPTO. US published applications were supposed to be of an eleven digit long format ever since, yet IT systems were largely not prepared to accept such a format when the new law US published application numbers came along. As an intermediate measure one zero was removed to accommodate the new numbers in a ten digit format. With the reload the opportunity was taken to reconstitute the eleven digit numbers both from the input and the database implementation side in order to comply with the USPTO devised format.

Q: In the index the pre-Y2K application numbers now seem to have a four digit year portion. Why has this been changed ?

A: Application and priority numbers are now indexed with a four digit year throughout for a more uniform appearance and to overcome sorting problems in the index. The same applies to AN (accession number), CR (cross-reference accession number), DNC (CPI secondary accession number) and DNN (non-CPI secondary accession number) . Old formats are searchable via a search field edit.

Q: Cross reference *DWPI* accession numbers do not carry the associated update (week) any longer. Why is that ?

A: It is no longer possible to populate this data due to changes in the Thomson Scientific production process.

Q: The 'structured' DCR numbers in the IT display field have shed their structure – why is that, and what are the implications for crossover between DCR and *DWPI* ?

A: Structured DCRs are no longer supplied for the bibliographic part of *DWPI* although they are still available in DCR. In the display in IT they have been replaced by a unique and unambiguous structure identifier, which is in the DCR segment in the AN.S display field, albeit the 'DCR-' prefix has been omitted for simplicity. The crossover employing the /DCR feature is not affected by this change.

Q: In the chemical coding paragraphs there are MCN: entries listed – what are they ?

A: These Markush compound numbers have been taken out of the realm of DCNs (Specific Compound Numbers) and are now labelled MCNs. They are also now searchable in a separate search field in order to avoid confusion.

Q: Where can I obtain details of the XML document structure for use with the XML download format ?

A: Thomson Scientific will provide DTDs corresponding to distributed XML documents. Please contact your nearest Thomson Scientific office for further details.

Q: I understand that the introduction of the IPC Reform requires current awareness searches to be adjusted. What do I need to particularly look at in terms of the new IPC Reform codes ?

A: By and large the new IPC Reform data have been implemented in a uniform way at all STN databases, though a few database specific issues had to be recognized. Hence measures to adjust to the new IPC Reform data apply as to other databases having adopted the new codes previously. For further particulars please refer to general information about the IPC Reform changes on STN.

[\[http://www.stn-international.com/stndatabases/details/ipc_reform.html\]](http://www.stn-international.com/stndatabases/details/ipc_reform.html)

For further *DWPI* specific details refer to the *DWPI* on STN online user guide

[\[http://www.stn-international.de/training_center/patents/stn_guide.pdf\]](http://www.stn-international.de/training_center/patents/stn_guide.pdf).

Q: I used to select the main IPC of the basic patent for statistical purposes. What can I do to accomplish the same task with the new setup ?

A: Since the new IPC Reform data doesn't have the concept of a 'main' IPC, the ICM.B became obsolete. In order to provide users with a single 'lead' IPC which can be utilized for statistical and sorting purposes, the IPC.F select and sort code has been made available. (Since it works its way through an elaborate algorithm it is currently not very fast, but we strive to improve on the performance).

Q: I understand that the introduction of the IPC Reform requires current awareness searches to be adjusted. But do I have to do anything about queries pointing to IPC 1-7 data ?

A: By and large nothing needs to be done to queries concerning IPC 1-7 data, but searches employing truncation symbols may give rise to problems in the new /IPC search field. Since IPC stems are being up-posted the use of truncation symbols should not be required, but if it is, it should be noted that searching with truncation in the /IPC search field can only be used for IPC Reform codes comprising a four digit main group. Please refer to general information about the IPC Reform changes on STN.

[\[http://www.stn-international.com/stndatabases/details/ipc_reform.html\]](http://www.stn-international.com/stndatabases/details/ipc_reform.html)

Q: How are German umlauts and French accents represented in the new *DWPI*?

A: For online representation and indexing purposes ASCII transliterations have been provided by Thomson Scientific. If not national characters have been transliterated according to the rules generally applying to FIZ Karlsruhe files. However due to variations in the source data it is recommended that you expand the relevant index before searching in order to include all potential options.

For many sections both marked-up and ASCII versions of the data have been provided to FIZ Karlsruhe by Thomson Scientific. These marked-up sections contain national characters and symbols in Unicode encoded in UTF-8. This data will be made available

in unadulterated form through an XML download option. Hence for post-processing purposes these characters are at your disposal in Unicode form if required.