

Adding Value to INPADOC on STN – Quality Assurance by FIZ Karlsruhe’s Patent Editorial –

INPADOC on STN is based on the world’s largest collection of patent bibliographic and legal status data from the European Patent Office. The EPO performs a tremendous task to compile and harmonize patent data from more than 100 patent authorities into a single patent data resource. Quality assurance processes at the EPO significantly contribute to the quality of the database.¹ Nevertheless, FIZ Karlsruhe identified areas for which the quality of the database could be further improved and established its own quality management for INPADOC. A key aspect of these activities relates to corrections of patent bibliographic data, providing patent information professionals with more accurate and comprehensive patent family information.

The importance of high-quality patent numbers

Patent numbers (publication, application and priority numbers) are substantial bibliographic elements which are of high importance for building accurate patent families.

- Correct **priority numbers** are indispensable to build reliable **INPADOC families**: all patent publications referring to the same invention and having one priority number in common (directly or indirectly) constitute one INPADOC family. INPAFAMDB, the INPADOC family file on STN, is designed according to this family concept, i.e., one INPAFAMDB record includes all bibliographic and legal status data of all patent publications referring to the INPADOC family.
- **Application numbers** are the key elements for linking all patent publications of a **national family**. INPADOCDB is the INPADOC database on STN which features an application-based file design, i.e., one INPADOCDB record compiles all bibliographic details and legal status data available for a single patent application.

The high quality of patent numbers is also essential for the interaction of INPADOC with other patent files on STN: efficient crossfile searching between different patent files requires correct and highly standardized patent numbers. The STN patent standard creates consistency across the various patent databases on STN and harmonizes patent data from different producers, in particular from Chemical Abstracts Service, Thomson Reuters and the European Patent Office.

Standardization efforts and corrections of patent numbers

Every week about 60,000 new and 140,000 updated patent publications enter the INPADOC databases, comprising a huge amount of diverse patent numbers. All of these numbers are validated against a standardization table and converted to the STN patent standard format. At the validation stage publications which do not match the required input standard are rejected. Typical errors include missing or incorrect priority or application numbers, incorrect publication numbers or kind codes. Errors identified for the weekly INPADOC update are corrected intellectually by the FIZ Karlsruhe editorial team or in case of serial errors corrections are done automatically. Error corrections are typically online within one week after the error has been detected.

¹ Albrecht M A, Bosma R, van Dinter T, Ernst J-L, van Ginkel K, Versloot-Spoelstra F. Quality assurance in the EPO Patent Information Resource. World Patent Information, 32 (2010) 279-286.

The standardization table is the core module of the whole validation and standardization procedure. For historical reasons the patent numbering formats applied by the more than 100 patent authorities are highly inconsistent. Each patent office uses different numbering formats and kind codes for different patent publication types and time ranges. As a result, the standardization table covers more than 2,200 different numbering formats for patent publication and application numbers. FIZ Karlsruhe's editorial team takes great care to keep this table up to date and to create a consistent STN standard for new numbering formats.

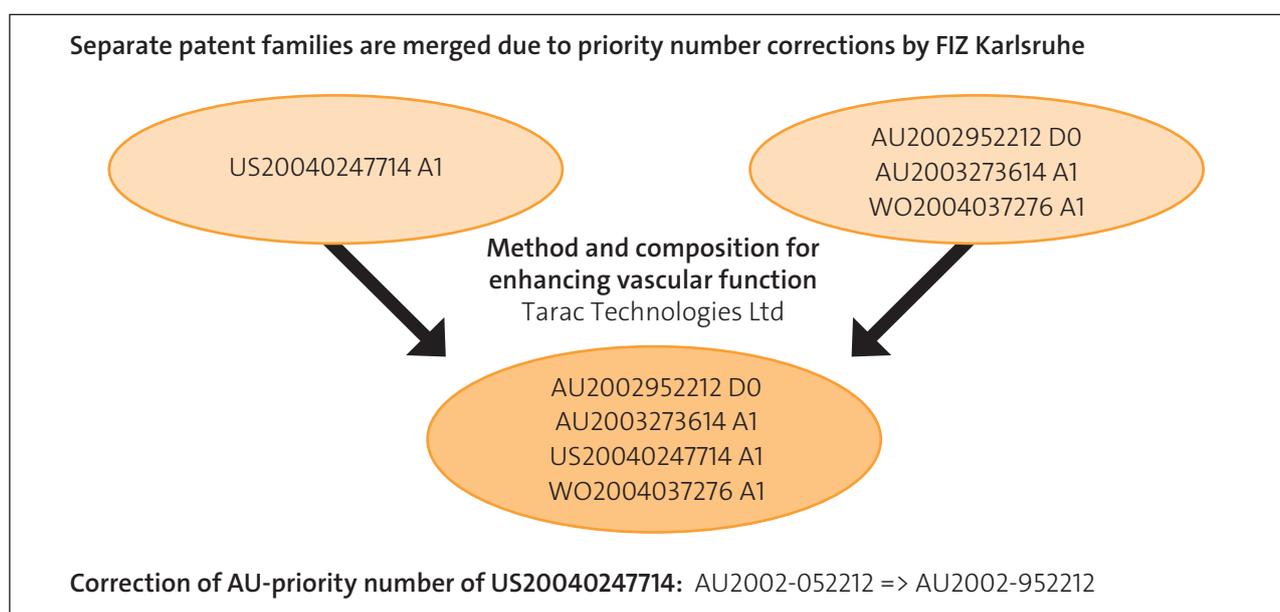
Apart from the weekly update routines, quality management for INPADOC also means that extra data deliveries from the EPO are monitored and that plausibility checks are performed for the entire database.

The editorial activities also involve corrections of single errors reported by customers or STN staff, e.g. wrong patent families, and missing or incorrect patent assignee names or titles. These errors are corrected on a case by case basis, and often various different sources need to be consulted, e.g., the original sources from patent offices and related patent publications. Especially priority number corrections require special expertise and careful error analysis.

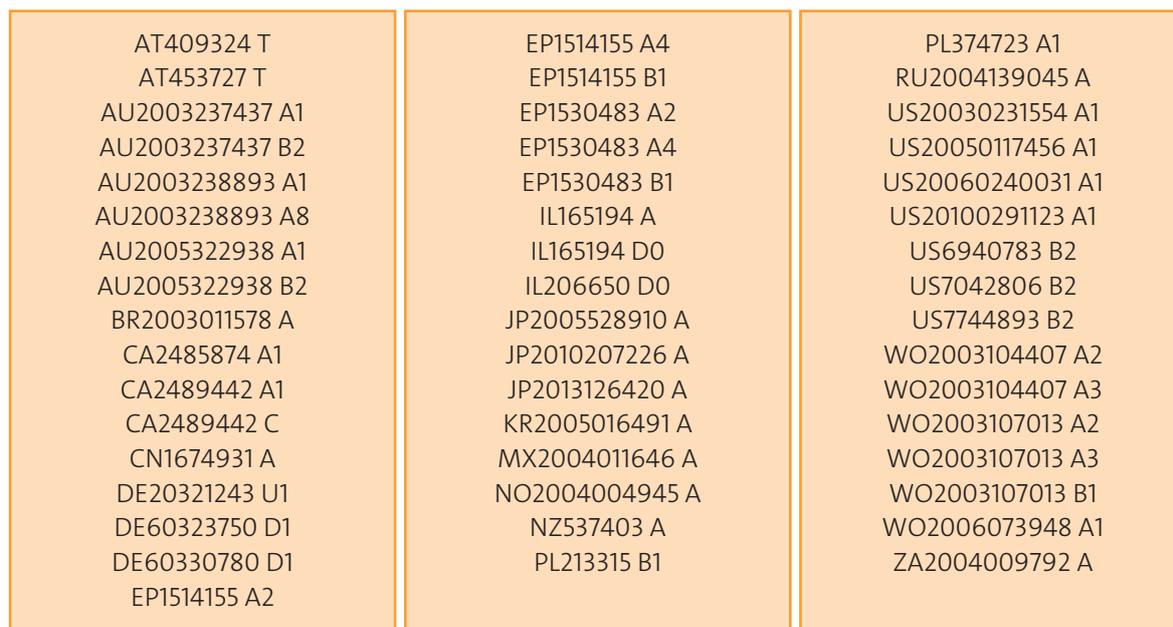
More accurate patent families through priority number corrections

Family building in INPADOC is a dynamic process which takes into account new priority relationships. The INPADOC family could be reassembled when a priority number is corrected or added to an existing document. This means that separate families merge and build a new family (example on the left) or a family is split into different families (example on the right).

The following two examples illustrate the benefit of priority number corrections by the FIZ editorial team. In both cases the priority information on the original US published application was wrong. The example on the left, a US publication joining a patent family, clearly demonstrates that correct priority information is essential for the comprehensiveness of a patent family.



False patent families are separated due to priority number corrections by FIZ Karlsruhe



Mat for timing competitions
Speed Stacks Inc.

AT409324 T	EP1514155 A4
AU2003238893 A1	US20030231554 A1
AU2003238893 A8	US20050117456 A1
AU2005322938 A1	US6940783 B2
AU2005322938 B2	US7042806 B2
CA2489442 A1	WO2003107013 A2
CA2489442 C	WO2003107013 A3
DE20321243 U1	WO2003107013 B1
DE60323750 D1	WO2006073948 A1
EP1514155 A2	
EP1514155 B1	

T cell receptor CDR3 sequences and methods for detection
Opexa Pharmaceuticals Inc.

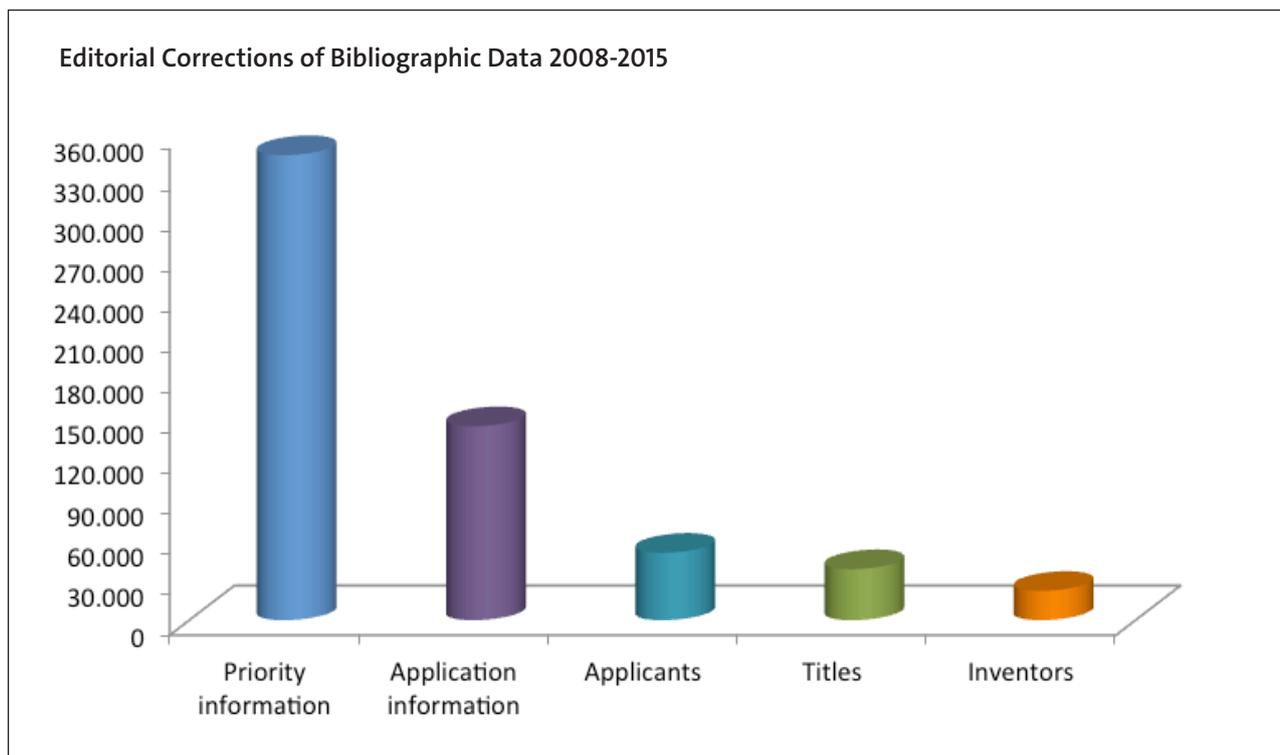
AT453727 T	JP2013126420 A
AU2003237437 A1	KR2005016491 A
AU2003237437 B2	MX2004011646 A
BR2003011578 A	NO2004004945 A
CA2485874 A1	NZ537403 A
CN1674931 A	PL213315 B1
DE60330780 D1	PL374723 A1
EP1530483 A2	RU2004139045 A
EP1530483 A4	US20060240031 A1
EP1530483 B1	US20100291123 A1
IL165194 A	US7744893 B2
IL165194 D0	WO2003104407 A2
IL206650 D0	WO2003104407 A3
JP2005528910 A	ZA2004009792 A
JP2010207226 A	

Correction of WO-priority number of US20060240031 A1: WO2003-US17673 => WO2003-US17873

Statistics of corrections made by FIZ Karlsruhe

The number of corrections performed by FIZ Karlsruhe's editorial team is constantly increasing. From the beginning of the statistics in 2008 until the end of 2015, more than 600,000 corrections were made, the major part being corrections of the priority information.

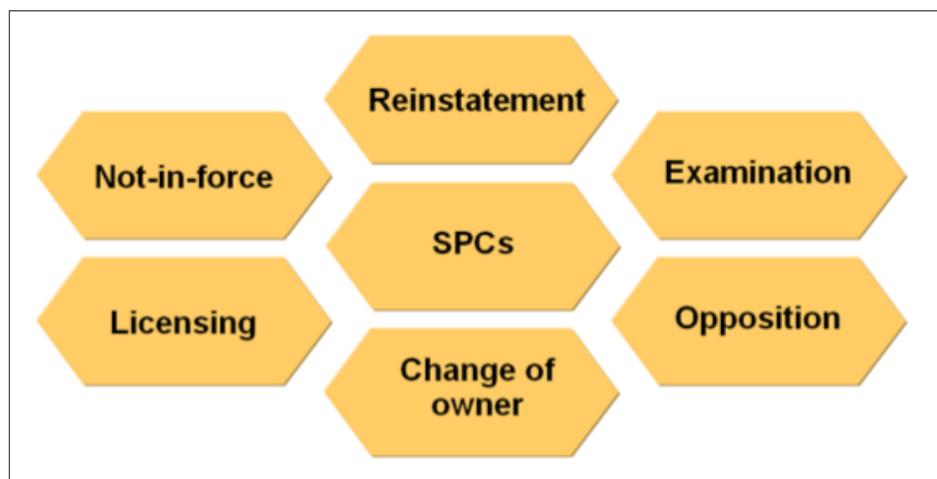
The editorial corrections are very well integrated in the weekly update routines for INPADOC. Whenever updated bibliographic information from the EPO includes the same errors as previously corrected by FIZ Karlsruhe's editorial team, we make sure that our corrections are retained. In addition, all corrections from the EPO are checked against earlier corrections done by FIZ Karlsruhe's editorial team.



Adding value to legal status information

The INPADOC collection is the largest repository of legal status data worldwide, comprising selected legal status events from currently 59 patent authorities. The EPO applies more than 2,600 legal status codes, each code representing a specific legal status event from a particular patent office. Worldwide legal status searching is a rather challenging task, as various different codes from different countries need to be searched for a particular legal request.

FIZ Karlsruhe introduced a set of legal status categories to simplify legal status searches. FIZ Karlsruhe's editorial team reviewed the complete list of legal status codes and assigned legal status categories to half of the codes. Following the discussion with customers, seven categories of major interest could be identified, e.g., applicant reassignments, oppositions or "not in force" (including codes like lapse or expiry of patents). The editorial team is in charge of these categories and continuously monitors any changes to the legal status codes and makes sure that new codes are captured in the relevant category.



Legal status categories assigned and maintained by FIZ Editorial

Outlook

In the fast growing patent information landscape the role of INPADOC as a central resource for worldwide patent information will become even more important in the future. More and more patent offices from Asia, the Middle East and Latin America contribute to INPADOC and the EPO is confronted with an increasing number of heterogeneous sources. Especially the patenting activities in Asia account for a rapid growth of the INPADOC data collection. Furthermore, the EPO makes a considerable effort to extend the backfile coverage for several patent authorities and to fill coverage gaps wherever possible. All of these coverage enhancements go along with additional standardization work and increase the potential for bibliographic errors.

In the wake of these developments, the thorough quality assurance by FIZ Karlsruhe remains a considerable value-add to INPADOC on STN.