

# Adding Value to INPADOC on STN

- Quality Assurance by FIZ Karlsruhe 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 90 patent authorities into a single patent data resource. Quality assurance processes at the EPO contribute significantly to the quality of the database<sup>(1)</sup>. 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 the national family.

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 those publications are rejected which do not match the required input standard. 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

(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) 279-28

of serial errors corrections are done automatically. Error corrections are typically online within one week the error has been detected. 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 90 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. The FIZ Karlsruhe editorial team takes great care to keep this table up to date and 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 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, 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. original sources from patent offices and related patent publications. Especially priority number corrections require a 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 or a family is split into different families (Fig. 1).

The two examples in Fig. 1 illustrate the benefit of priority number corrections of FIZ editorial. 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.

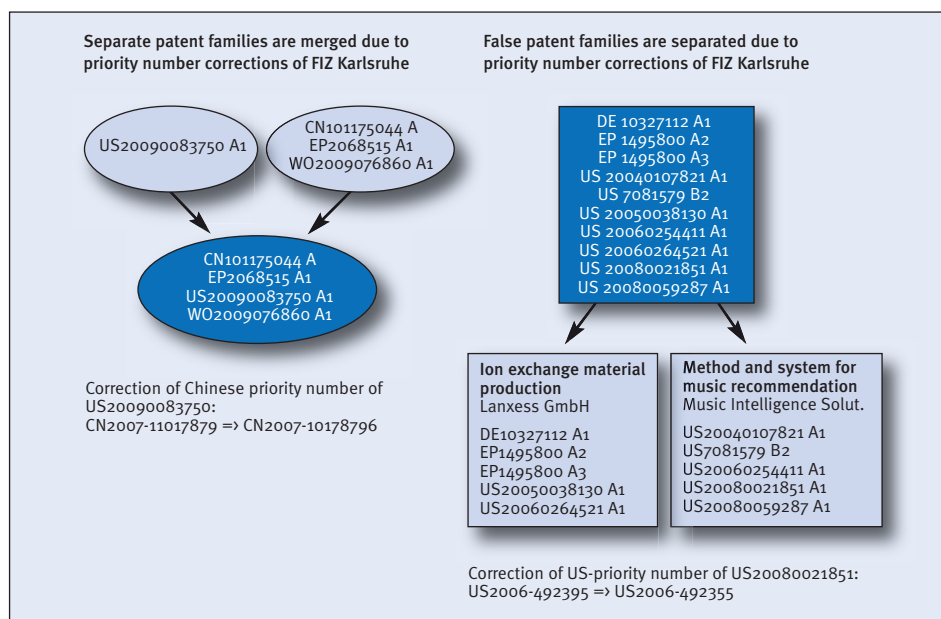


Fig. 1: Two examples of priority number corrections leading to more accurate patent families

## Statistics of FIZ Karlsruhe editorial corrections

The number of corrections performed by the FIZ Karlsruhe editorial team is constantly increasing. In 2010 the number of corrections was about 40.000, making the overall number of corrections more than 100.000 since the beginning of the statistics in 2008 (Fig. 2).

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 editorial, we make sure that our corrections are retained. In addition, all corrections from the EPO are checked against earlier corrections done by FIZ Karlsruhe editorial.

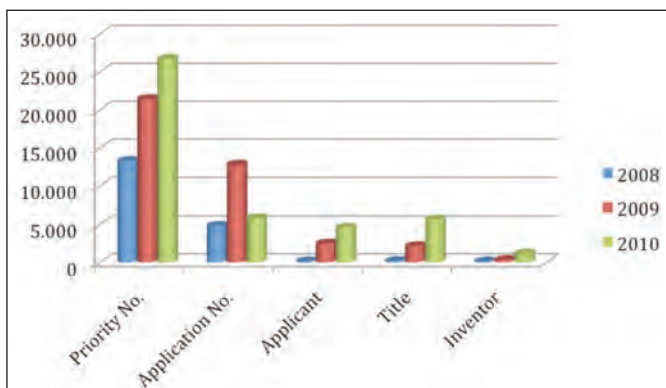


Fig. 2: Corrections of bibliographic data in INPADOC on STN from 2008-2010

## 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 2600 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.

The FIZ Karlsruhe editorial team reviewed the complete list of legal status codes and assigned legal status categories for half of the codes.

Following the discussion with customers, seven categories of major interest could be identified, (Fig. 3) 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.

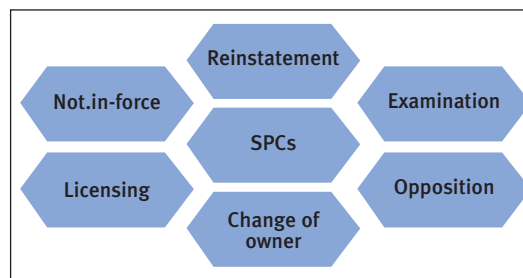


Fig. 3 Legal status categories assigned and maintained by FIZ Editorial

## 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 faced 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 fill coverage gaps where 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.