

STN[®]

Making the Difference – Enhancing the Quality of First Level Patent Data

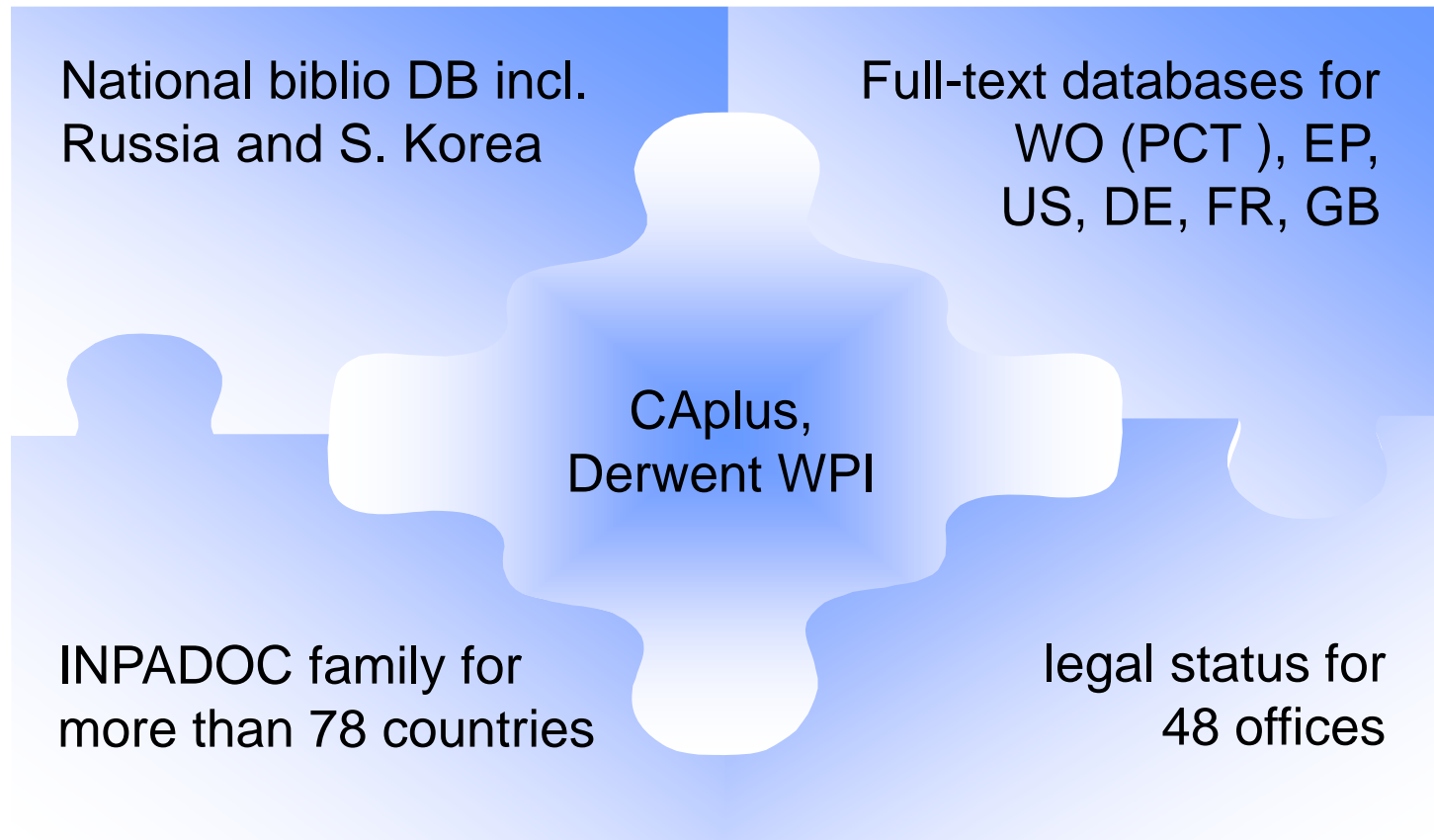
IPI-ConfEx 2007, March 4-7, Sorrento, Italy

Rainer Stuike-Prill, FIZ Karlsruhe

Agenda

- Introduction
- Enhancing the quality of INPADOC
 - Family corrections
 - Legal status categories
 - Company name standardization
- Adding chemical indexing to USPATFULL
- Summary

STN provides an unsurpassed portfolio of value-added and first level databases



Enhancing high quality first level data

- Formal quality checks of the delivered data
- Corrections of obviously wrong data by database producer or STN
- Standardization of data variations during file implementation or through an editorial process

Enhancing the quality of INPADOC

Corrections

INPADOC
family data

Categorizations

INPADOC
legal status
data

Standardization

company
names
(in first level
patent databases)

INPADOC Family corrections

- Challenges
 - identify missing family members
 - analyze reasons for incorrect or incomplete family information
 - assure reliable identification of patent families
- Solution
 - automatic input checking procedures
 - systematic handling of errors
 - checking multiple sources for error detection

INPADOC Family corrections

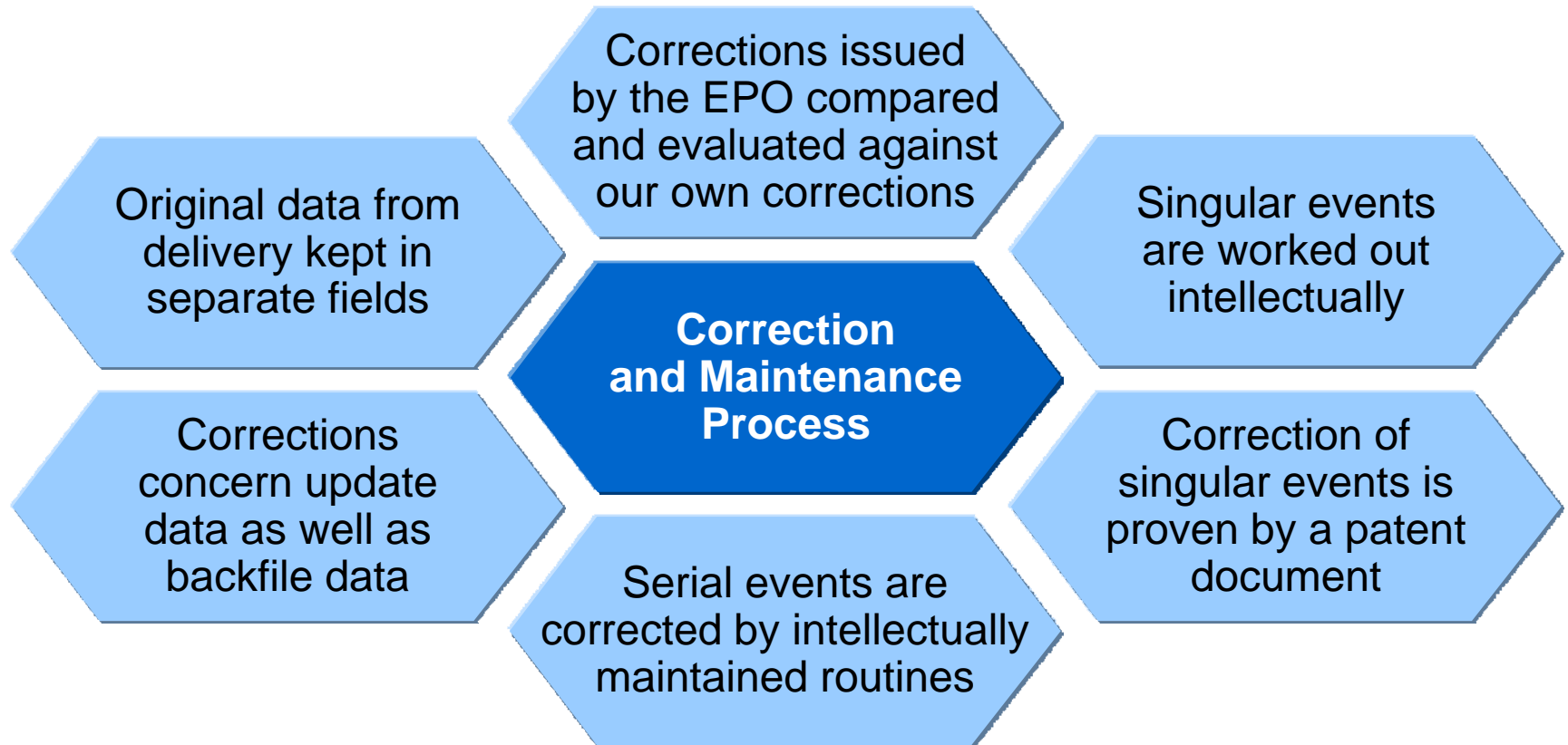
Types of errors

- errors in the original document
- errors caused during data processing
- singular errors (typos)
- serial errors

Detection of errors

- by users (internal, external)
- by routines that check input data against known standards

INPADOC Family corrections



INPADOC Family corrections

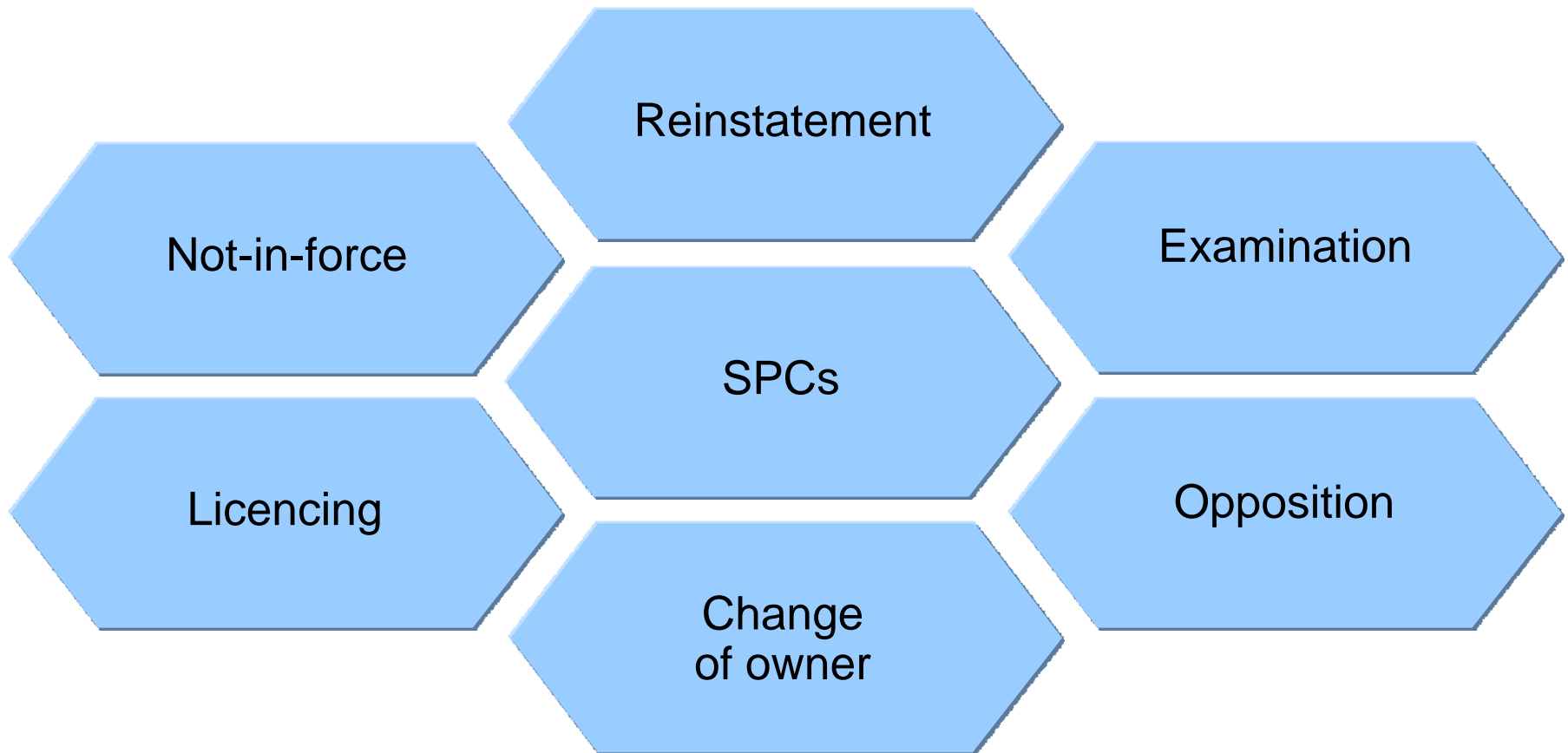
- Benefits
 - fast access to corrected data
 - more reliable identification of individual documents
 - more reliable and complete family information

INPADOC Legal Status categories

- Challenges
 - 2,600+ legal status codes for data from 48 patent offices exist
 - code definitions may change over time
 - comprehensive searches require sophisticated strategies and deep understanding

- Solution
 - establish and maintain a limited number of legal status categories
 - the 2,600+ codes have been reviewed and grouped into six categories

INPADOC Legal Status categories



INPADOC Legal Status categories

- Benefits
 - legal status categories will enable easy-to-use and efficient legal status searches
 - improved current awareness searching

Legal status categories will be available
with the new INPDOADB file in 2Q 2007

Company Name standardization

- Challenges
 - numerous name variations from first level sources complicate search and analysis
- Solution
 - standardization of company names across all first level sources as far as possible
 - starting to add standards for 90% of patent assignee entries in INPADOC/DOCDB
 - create standard name and address during update of database for known variations

Company Name standardization

- Benefits
 - efficient multiframe searching
 - more effective search and analysis
 - additional search and analysis capabilities through address data

Tools developed for CAplus enhance search and retrieval of first level patent data

- CAS-developed thesauri for patent classification schemes
 - IPC, NCL, F-Terms
- IPC thesaurus is available in 15 STN files
- NCL thesaurus and CAS controlled indexing terms available in USPATFULL

Chemistry and chemically relevant patents in USPATFULL enriched with CAS indexing for greater search precision

L1 ANSWER 404 OF 3407 USPATFULL on STN

[Full Text](#)

AN 2006:117421 USPATFULL

TI Methods and formulations for making controlled release oral dosage form

IN Chow, San-Laung, San Jose, CA, UNITED STATES

Wong, David, Milpitas, CA, UNITED STATES

Garcia, Damian, Sunnyvale, CA, UNITED STATES

PA BioKey, Inc. (U.S. corporation)

PI US 2006099262 A1 20060511

AI US 2005-295995 A1 20051206 (11)

...

DRWN 4 Drawing Page(s)

LN.CNT 876

CAS INDEXING IS AVAILABLE FOR THIS PATENT.

CT Antidepressants

CT Shellac

CT Behavior

CT Drug delivery systems

CAS controlled indexing terms

IPCI CODE	VERSION	POS	INV	LEVEL	CC	ASSIGNMENT	DATE	STAT
A61K0009-24	(200601)	F	I	Advanced	US	Human	20060511	O

IPCR CODE	VERSION	POS	INV	LEVEL	CC	ASSIGNMENT	DATE	STAT
A61K0009-24	(200601)	F	I	Advanced	US	Human	20060511	O
A61K0009-24	(200601)	L	I	Core	US	Human	20060511	O

IPC data including
reclassifications

Summary

STN adds value to first level patent data by

- correcting INPADOC family data to assure more complete family results
- categorizing INPADOC legal status codes to increase efficiency in legal status retrieval
- standardizing company information, thus providing better company name search and analysis capabilities
- utilizing IPC thesauri in all STN patent files
- enriching USPATFULL with CAS indexing

Thank you for your attention