

# The EPO Publication System

## overview

Guido Szamocki

Workshop with ROSPATENT  
20-22 May 2010



# Overview of the presentation:

- The departments involved
- Data volumes, Statistics
- The publication process
- Standards

*EPO organisational structure (for publication):*



## **DG 2 -The Hague**

### **A Documents**

- **Post room**
  - **Data entry (EPASYS)**
  - **Scanning (Phoenix)**
  - **Search Reports**
- Publication Group:**
- **Abstracts**
  - **Quality control**
  - **Contractor liaison**

*EPO organisational structure (for publication):*



## DG 2 -The Hague Information Systems

### PD 2 IM

- System support
- Operational systems
- EPASYS, Trimaran etc.
- *eBusiness*:
  - Phoenix
  - ePublication

*EPO organisational structure (for publication):*

# DG 2 - Munich

## B Documents

- Post room
- Data entry (EPASYS)
- "Druckexemplar"



*EPO organisational structure (for publication):*



**Dir 4.5.2 -  
Vienna**

- **Dir. of Publication**
- **Publication server**
- **Data distribution**
- **EP Bulletin**
- **MIMOSA - ESPACE**
- **Finance**
- **Quality control  
(B documents)**
- **Contractor liaison**

# Data volumes (approx)

- A documents processed: 63 000
  - B documents processed: 63 000 =
  - 89 000 pages captured per week
- We no longer print paper documents

## Current full-text database statistics:

- A documents = 514 000 (1996 - 2006)
- B documents =

(Note: backfile date to 1987 available in SGML)

# Extraction of images from PHOENIX

Madras: Szamocki, Guido | Phoenix: Phoenix General | Version: epo-PROD-3.51.00

EP 06 007 608.0 (11-04-2006) : Phoenix General

Dossier Document Message Windows Help ADM JDEWP MUSE POST NUF

Cover TOC Checklist Status Structure History

Search/Exam All

Show: Active

| Date       | Pages         | Code | PackageID        | SeqNr. | LogPkg.    | Handled by             | Annotation |      |
|------------|---------------|------|------------------|--------|------------|------------------------|------------|------|
| 11-04-2006 | 6 1001        |      | E1 0600760804P1  | 1      | 4 DD50333  | /                      |            | ⇒ SC |
| 11-04-2006 | 2 1002        |      | E1 0600760804P1  | 2      | 4 DD50333  | /+ add. sh.            |            | ⇒ SC |
| 11-04-2006 | 11 DESC       |      | E1 0600760804P1  | 3      | 4 DD50333  | /                      |            | ⇒ SC |
| 11-04-2006 | 5 CLMS        |      | E1 0600760804P1  | 4      | 4 DD50333  | 37-paid                |            | ⇒ SC |
| 11-04-2006 | 1 ABST        |      | E1 0600760804P1  | 5      | 4 DD50333  | fig. 1                 |            | ⇒ SC |
| 11-04-2006 | 12 DRAW       |      | E1 0600760804P1  | 6      | 4 DD50333  | see 1150               |            | ⇒ SC |
| 14-04-2006 | 2 1014        |      | E1 0600760804P1  | 7      | 4 DD50333  | /                      |            | ⇒ DE |
| 14-04-2006 |               |      |                  |        | DD50333    |                        |            | ⇒ PM |
| 20-04-2006 |               |      |                  |        | DD50333    | ok                     |            | ⇒ EP |
| 10-05-2006 |               |      |                  |        | DD50333    |                        |            | ⇒ SC |
| 10-05-2006 |               |      |                  |        | DD50333    | prio05                 |            | ⇒ SC |
| 12-05-2006 | 2 1150        |      | E1 0600760809P1  | 1      | 9 DD50333  | fora                   |            | ⇒ SE |
| 01-06-2006 | 1 APPL        |      | E1 060076080DDP1 | 1      | 13 BN21414 | re draw                |            | ⇒ SC |
| 01-06-2006 | 12 DRAW       |      | E1 060076080DDP1 | 2      | 13 BN21414 | /checked and ok        |            | ⇒ SC |
| 02-06-2006 | 1 1048        |      | E1 060076080BP1  | 1      | 11 BN21414 |                        |            | ⇒ EP |
| 02-06-2006 |               |      |                  |        | BN21414    |                        |            | ⇒ EP |
| 11-07-2006 |               |      |                  |        | DD50333    | AM                     |            | ⇒ SC |
| 24-07-2006 |               |      |                  |        | DD50333    |                        |            | ⇒ DE |
| 24-07-2006 |               |      |                  |        | DD50333    |                        |            | ⇒ CC |
| 24-07-2006 |               |      |                  |        | DD50333    |                        |            | ⇒ SD |
| 24-07-2006 | 1 1503        |      | E1 060076080JP1  | 1      | 19 DD50333 |                        |            | ⇒ SD |
| 14-09-2006 | 1 RPUB01-EP   |      | E1 060076080KP1  | 1      | 20 DD50333 | coded                  |            | ⇒ EP |
| 15-09-2006 | 1 APPL        |      | E1 060076080LP1  | 1      | 21 DD50333 |                        |            | ⇒ SC |
| 15-09-2006 |               |      |                  |        | DD50333    | /                      |            | ⇒ SC |
| 27-09-2006 |               |      |                  |        | ML23609    |                        |            | ⇒ EP |
| 30-10-2006 |               |      |                  |        | ML23609    |                        |            | ⇒ EP |
| 25-04-2007 |               |      |                  |        | ML23609    |                        |            | ⇒ SC |
| 08-06-2007 |               |      |                  |        | LS23933    | Exre1 sent & coded /is |            | ⇒ SE |
| 08-10-2007 | 5 EXRE3       |      | E11115943AE912P1 | 1      | 27 LS23933 | sent to ex /is         |            | ⇒ SC |
| 08-10-2007 | 3 CLMS        |      | E11115943AE912P1 | 2      | 27 LS23933 |                        |            | ⇒ SC |
| 08-10-2007 | 3 DESC        |      | E11115943AE912P1 | 3      | 27 LS23933 | HWA                    |            | ⇒ SC |
| 23-07-2008 | 1 2596L       |      | E1211B4F26C317P1 | 1      | 28 YS23511 |                        |            | ⇒ EP |
| 17-02-2009 | 1 2395L       |      | E11115943AE912P1 | 1      | 28 YS23511 |                        |            | ⇒ JC |
| 17-02-2009 |               |      |                  |        | YS23511    | coded&sentYys          |            | ⇒ JC |
| 17-02-2009 |               |      |                  |        | YS23511    |                        |            | ⇒ JC |
| 17-02-2009 |               |      |                  |        | YS23511    |                        |            | ⇒ JC |
| 20-05-2009 | 2 IGRA7       |      | E1212180F2E6E5P1 | 1      | 34 AN50951 | +spre coded /an        |            | ⇒ SC |
| 20-05-2009 | 3 CLMSTRAN-DE |      | E1212180F2E6E5P1 | 2      | 34 AN50951 | clms 1-15 ok           |            | ⇒ SC |
| 20-05-2009 | 3 CLMSTRAN-FR |      | E1212180F2E6E5P1 | 3      | 34 AN50951 | clms 1-15 ok           |            | ⇒ SC |
| 28-05-2009 | 2 2070        |      | E11121878034DCP1 | 1      | 36 AN50951 | rpud coded /an         |            | ⇒ SE |
| 05-06-2009 | 2 2006A       |      | E12121898C3D70P1 | 1      | 37         |                        |            | ⇒ EP |
| 18-06-2009 | 1 2031        |      | E12121F483DE42P1 | 1      | 39         |                        |            | ⇒ EP |
| 23-06-2009 | 1 2047        |      | E12121F4833BEAP1 | 1      | 38         |                        |            | ⇒ EP |
| 06-05-2010 | 1 2057        |      | E111285DD68902P1 | 1      | 40         |                        |            | ⇒ EP |

/checked and ok

Search/Exam

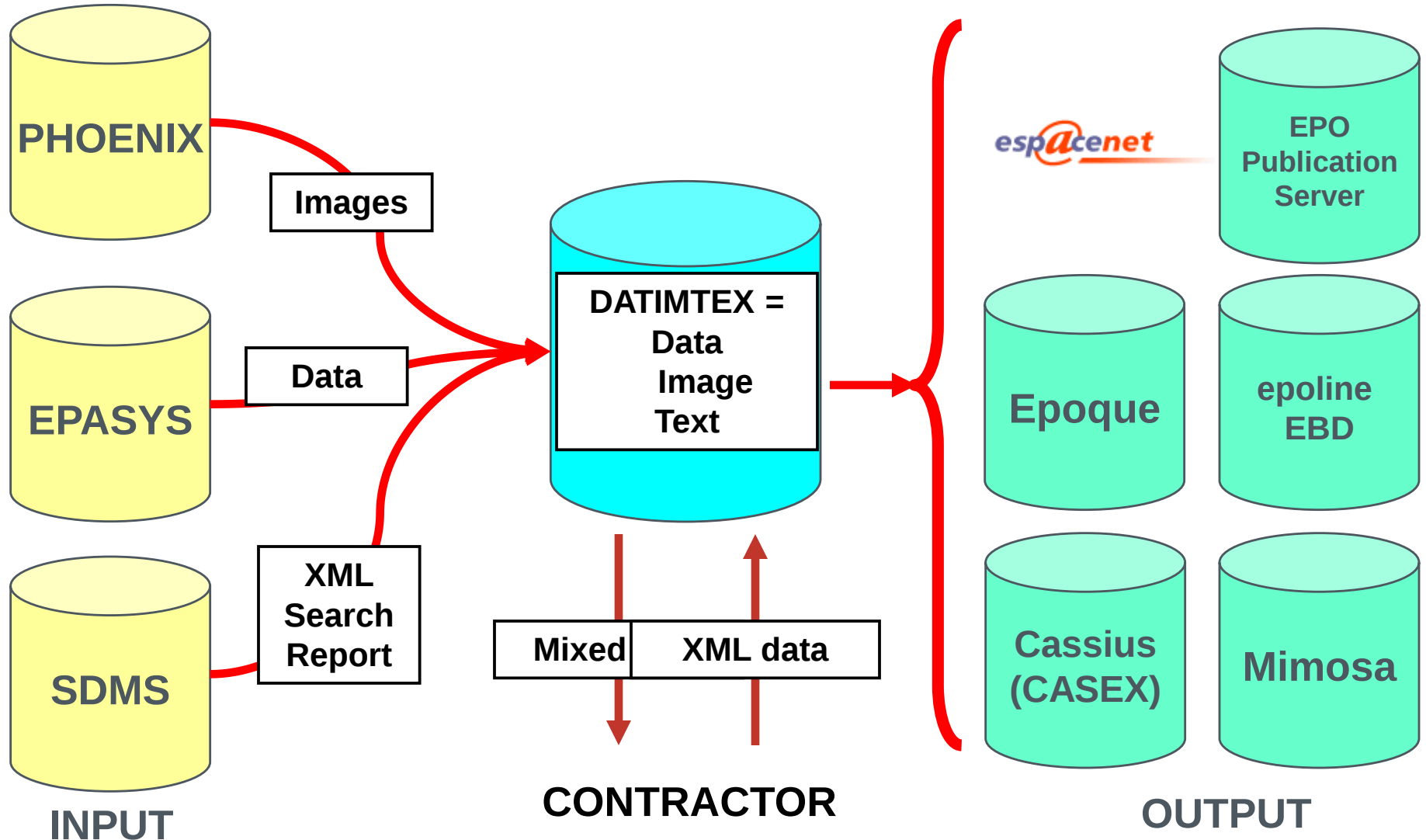
Menu EP 06 007 608.0 ...

106A 2031 2047 2057 DESC DRAW DRAW

Tag 1 of 12 3 Sec

FIG. 1

# Publishing System



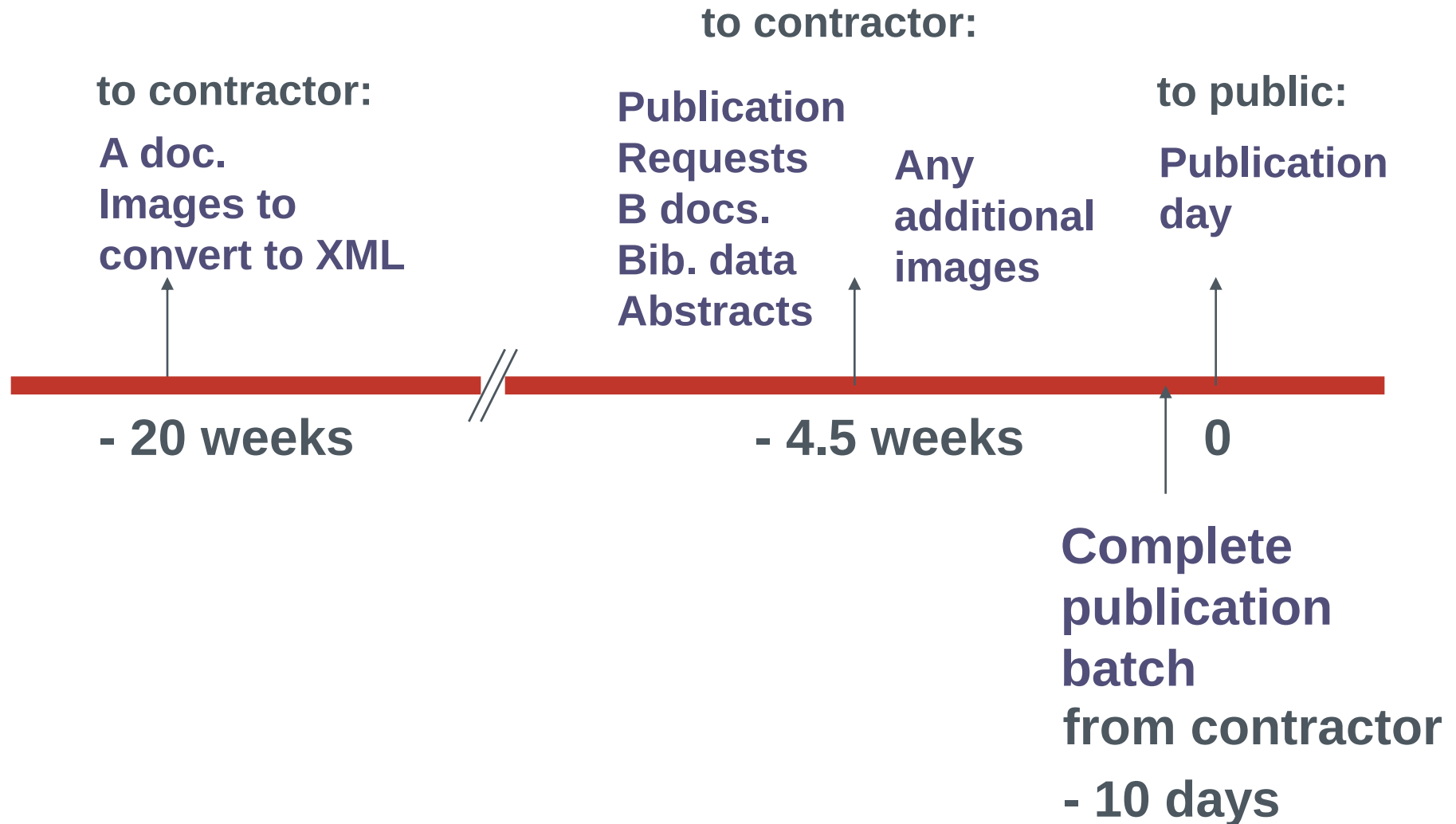
# Functions of DATIMTEX

- Manage the publication process
- Extraction, loading and storage of all data required for publication of A and B documents
- Source for published EP ST.36 documents
- Source for bibliographic data (EBD, Bulletin)
- Editorial system to modify abstracts
- Support Quality Control

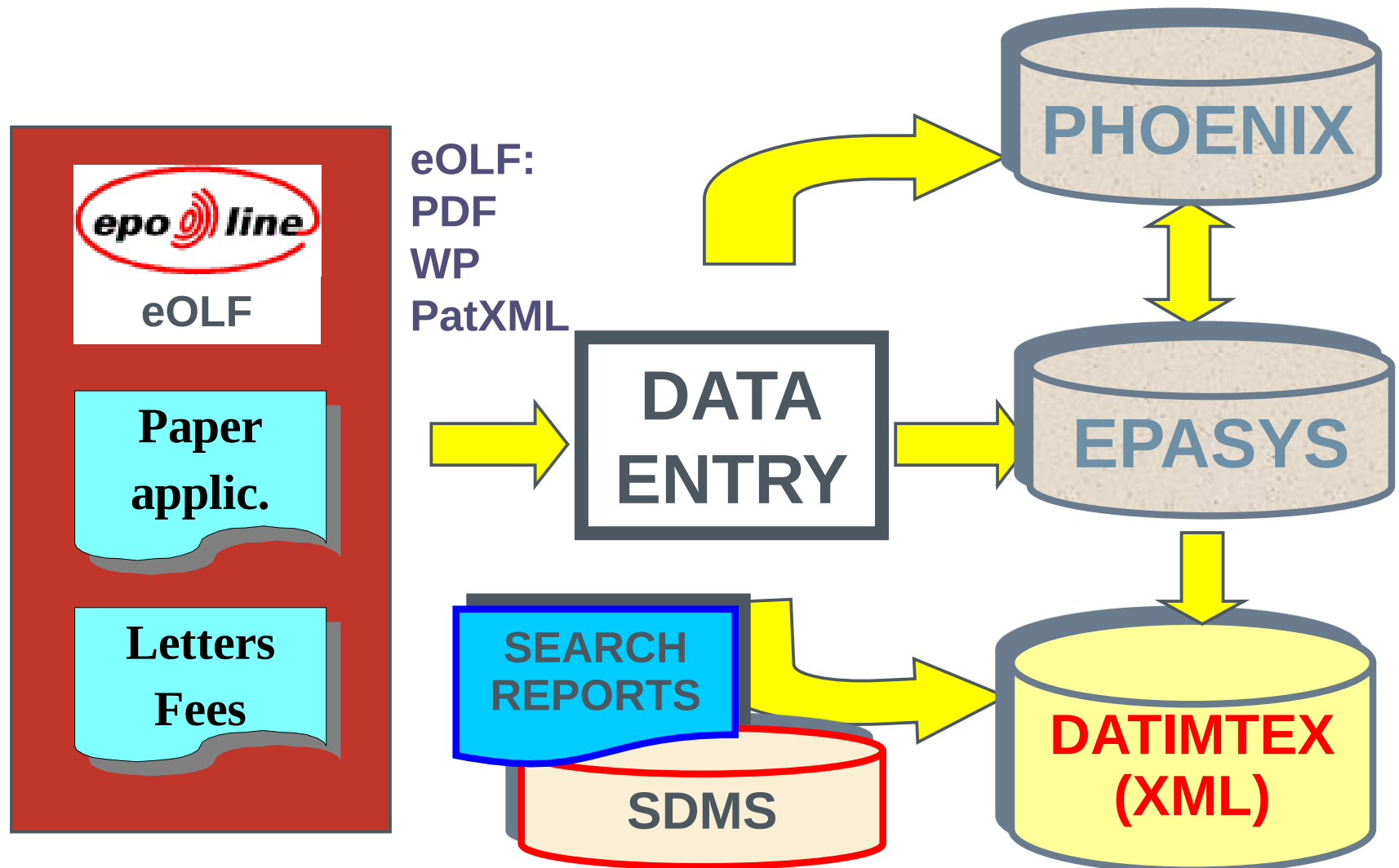
# Technical aspects

- Storage: PXI/Phoenix on Mainframe
- Application:
  - Java on AIX
  - Websphere / JBoss
- Interfaces ranging from:
  - DVD disks to
  - FTP and
  - SOAP web services

# Publishing Timeline



# Data input



## *The work of the EPO contractor (1):*

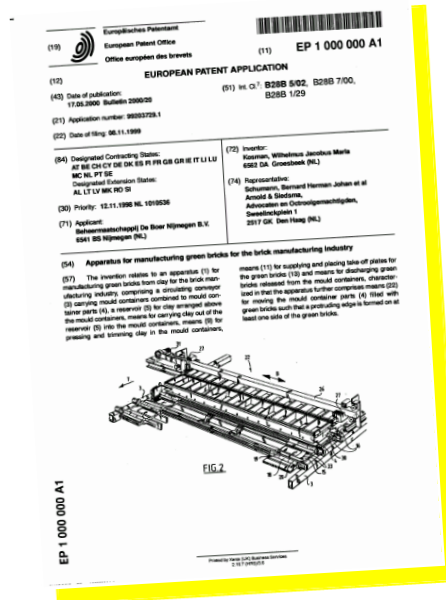
- Reception and checking of all files
- Image segmentation (to TIFF/Group 4)
- Intelligent Character Recognition (ICR) of scanned pages
- ICR editing and correction
- XML mark up (automatic where possible) of general text
- Special sections for XML of: tables and mathematics
- Processing bibliographic data (Title page)



## The work of the EPO contractor (2):

- ☐ Document composition (using Framemaker)
- ☐ Proof reading, layout check, QA, etc.
- ☐ Bit-map capture of composed document, automatic indexing (for facsimile ESPACE CD-ROM, BNS and espacenet)
- ☐ Tape output: ST.33 (facsimile) for BNS
- ☐ DVD/FTP: ST.36 (XML + image) and PDF

PDF

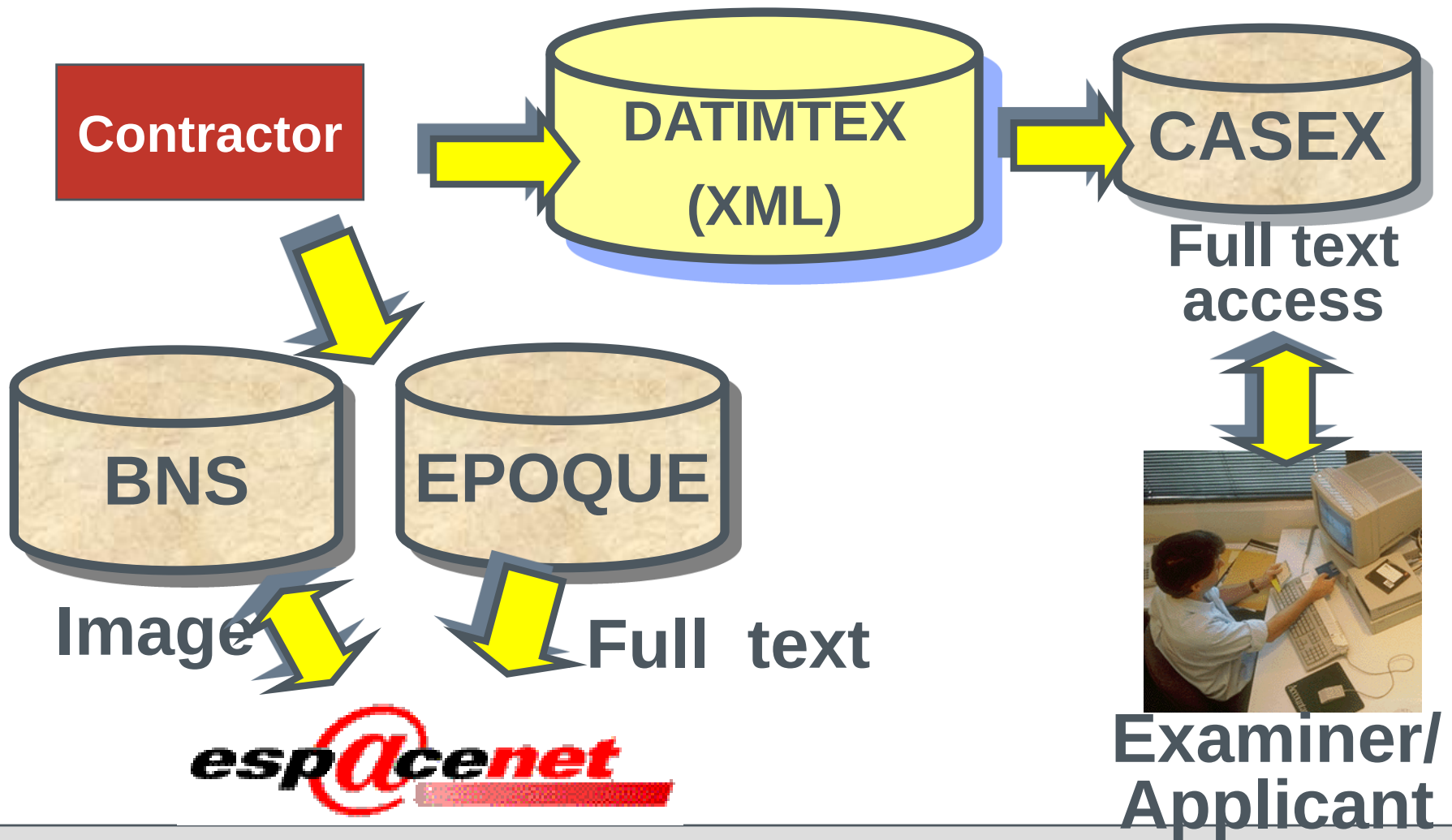


ST.33

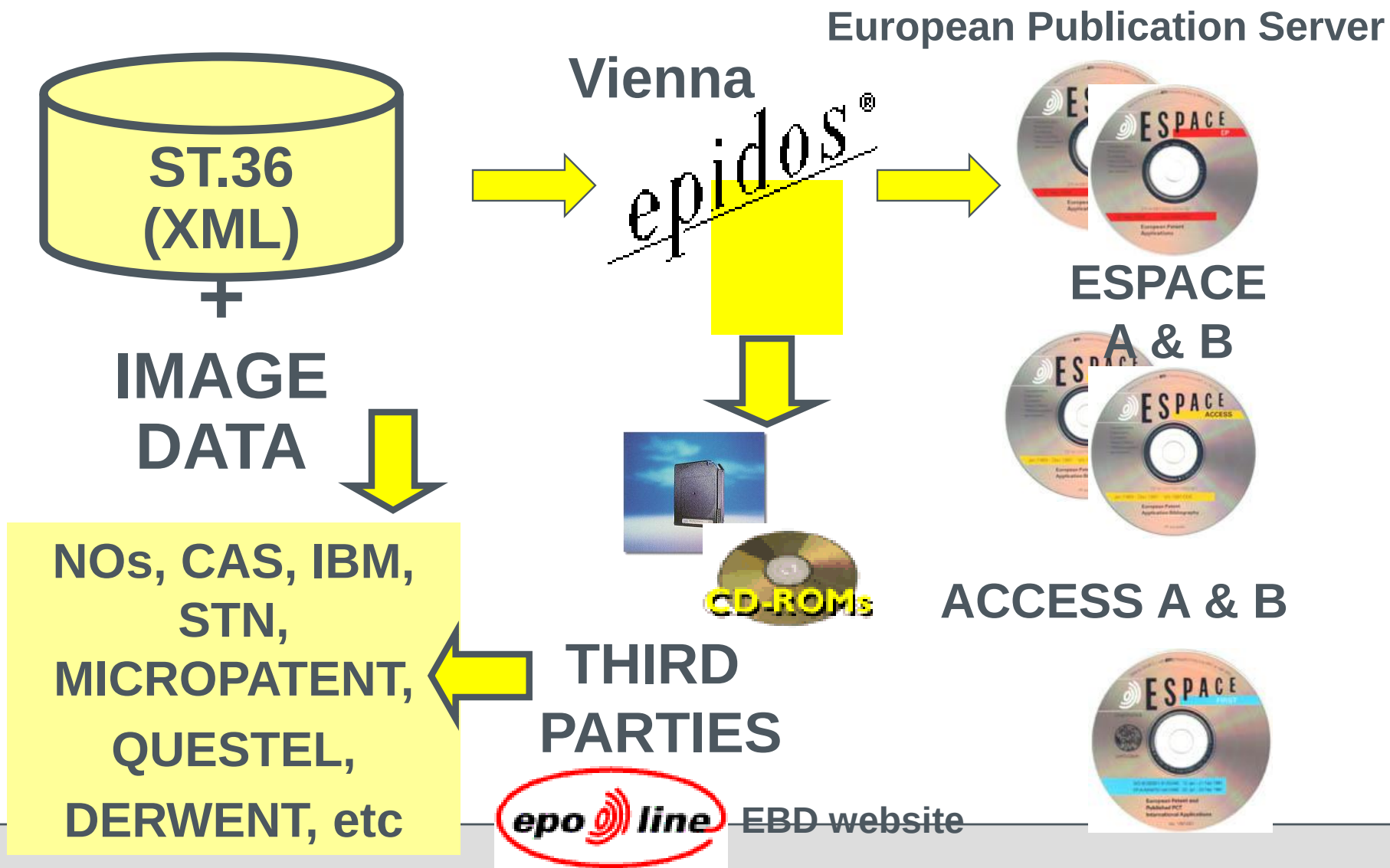
# Quality Control

- There are strict contract conditions
- Small teams in the EPO - The Hague and Vienna - check a selection from each batch
- There is daily contact between these teams and the contractors
- Detailed "*Rules for Contractors*" have been written

## Data output and re-use (internal)



## Data output (external)



# European Publication Server

<https://publications.european-patent-office.org/PublicationServer/main.jsp>

**European  
Patent Office**

**Patent information**

Deutsch **English** Français

Patents –  
knowledge for ev

**European Publication Server: Search**

[Home search](#) [Legal information](#) [Data coverage](#) [Help](#)

Publication number

EP 0893565

Application number

EP 98113349

IPC symbol

E05D

Year

2004

Week

1

Publication date

Between

and

Year

2006

Week

7

Kind

☒ Select all kind codes

☒ A1/A2 ☒ B1

☒ A3 ☒ B2

☒ A8/A9 ☒ B8/B9

Search

Reset

Message from Curt  
Edsall

# European Publication Server

Name 

-  DOC00001.xml
-  Document.pdf
-  imgaf001.tif
-  imgb0001.tif
-  imgb0002.tif
-  imgf0001.tif
-  imgf0002.tif
-  imgf0003.tif
-  srep0001.tif
-  toc.xml

[Home search](#)
[Result list](#)
[Legal information](#)
[Data coverage](#)
[Help](#)

Publication Date = 2006/02/15

Kind Code = A1 or A2

74 document(s) found

[View the hit-list](#)
[Select all](#)
[Unselect all](#)
[Download selection as ZIP](#)
[Download selection as PDF](#)

4

 First
  Previous
  Next
  Last

| Publication number | Kind code | Publication date |                                                                                     |                                                                                     |           |          |
|--------------------|-----------|------------------|-------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------|-----------|----------|
| 626605             | A1        | 2006/02/15       |  |  | esp@cenet | Register |
| 626606             | A1        | 2006/02/15       |  |  | esp@cenet | Register |



esp@cenet
 Register



esp@cenet
 Register

# EBD Website



European Patent Office  
*epoline*<sup>®</sup> - Bibliographic Data and Abstracts

[Home](#)
[Contact](#)

[EBD Main](#)
[Info](#)
[Notification of Change](#)

Selection screen for EBD ST.36 format data

Weeks currently available for download:

| Week                                                                                                             | File size | Notes |
|------------------------------------------------------------------------------------------------------------------|-----------|-------|
|  <a href="#">06-04</a> 6.2 MB   |           |       |
|  <a href="#">06-05</a> 6.4 MB  |           |       |
|  <a href="#">06-06</a> 6.1 MB |           |       |
|  <a href="#">06-07</a> 3.8 MB |           |       |

**Download complete bib. data for one week; also used for the production of the EP Bulletin (PDF only). NO's can download data earlier than publication day on a secure server**

# EBD Website - zip data file

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE ep-bulletin PUBLIC "-//EPO//EP PATENT BULLETIN 1.0//EN" "ep-bulletin-
v1-0.dtd">
<ep-bulletin id="EP00100986B1" file="00100986.9" lang="en" country="EP" doc-
number="1023965" kind="B1" date-publ="20060215" status="r" dtd-version="ep-
bulletin-v1-0"><SDOBI lang="en"><B000><eptags>
<B001EP>ATBECHDEDKESFRGBGRITLILUNLSEMCPTIESILTLVFIROMKCYAL.....
.....</B001EP><B002EP></eptags></B000>
<B100><B110>1023965</B110><B120><B121>EUROPEAN PATENT
SPECIFICATION</B121></B120><B130>B1</B130><B140><date>20060215</date></B
140><B190>EP</B190></B100><B200><B210>00100986.9</B210><B220><date>2000
0119</date></B220>. . .</SDOBI></ep-bulletin>
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE ep-bulletin PUBLIC "-//EPO//EP PATENT BULLETIN 1.0//EN" "ep-bulletin-
v1-0.dtd"><ep-bulletin id="EP00101835B1" file="00101835.7" lang="de"
country="EP" doc-number="1024053" kind="B1" date-publ="20060215" status="r"
dtd-version="ep-bulletin-v1-0"> . . .
<B500><B510EP><classification-ipcr sequence="1"><text>B60R 11/02
19680901AFI20000324BHEP      </text></classification-ipcr> . . .
```

**Need to extract each XML instance to have a valid XML file**

# Standards

- **PCT/AI/Annex F**  
Standard for the Electronic Filing and Processing of International Applications
- **WIPO ST.36**  
Recommendation for the Processing of Patent Information Using XML (Extensible Markup Language)  
  
Written, agreed and used by: EPO, JPO, USPTO and WIPO in publication and filing.
- **Unicode (UTF-8) - character set**

# Thank you