



Using PCT EDI and the Minimal Specifications

Peter Waring
Head, PCT Technical Cooperation Section

PCT-EDI MAIN TOPICS

- Introduction
- PCT-EDI deployment
- Sending documents to the IB
- Sending data to the IB
- Receiving documents from the IB
 - Systematic communication from IB
 - Communication from IB on request
- Further technical Information

PCT-EDI Introduction



- PCT-EDI is the secure electronic means by which Offices can send and receive documents to/from the International Bureau of the PCT
- PCT-EDI connects systems holding PCT documents in electronic form
- Started in 2007 the service made available
 - Upload to the IB
 - Priority document transfer
- Extended to
 - Upload and download of all PCT documents
 - Good quality control



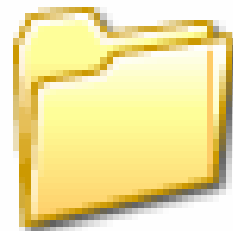
PCT-EDI office deployment

PCT-EDI how to deploy in your office

- Making a connection
 - See <http://www.wipo.int/patentscope/en/pct-edi/documents/>
 - Install a SFTP client (WinSCP or ...)
 - Install PuttyGen (generate key pair)
 - Send public key to WIPO (pct.edi@wipo.int)
 - Make 1st connection (using private key)
- Decide automation objective for example
 - sending record copies + forms
 - receiving PCT forms
- Exchange samples / testing
- Agree production and stop paper transmission

PCT- EDI what to send to IB

- Record copy
- Forms
- Search reports
- Other documents
- Fee data
- National phase entry data
- National patent information
- DAS – Priority documents (via dedicated acct.)



Upload

PCT-EDI – Sending documents to the IB

(presumes PCT-EDI connection already established)

References:

<http://www.wipo.int/pct/en/texts/>

<http://www.wipo.int/patentscope/en/pct-edi/documents/>

PCT- EDI how to send documents

- Build a system to create batches of documents
 - Ranging from a simple manual export function to a full workflow integration
 - For example included in PCT-ROAD
- Batch creation system must
 - Send all the intended documents
 - Prevent duplicate transmissions wherever possible
 - Name batches sequentially
 - Follow the Minimal Specifications rules for
 - Naming of the batches / documents
 - File formats of the documents

PCT- EDI how to send documents

- The IB supports the following transmission means:
 - PCT EDI Internet
 - PCT EDI Trinet
 - CD/DVD
- Frequency:
 - Batches of documents should be generated grouping by International Application number based on the overall volume
 - Practically more that 10 batches a day and batches sized in Gbs are undesirable.
- Neither the frequency of transmission nor the means of transmission influence the format of the contents of the transmissions

PCT- EDI how to send documents

■ Format:

- The PCT Minimal Specifications for Transmitting Documents to the IB should be adhered to, with text based electronic content conforming to the PCT Annex F, Standard for Filing and Processing in Electronic Form.
- All documents that are sent in XML format (or for which XML format is intended to be used in the future) must be sent conforming to Annex F, e.g. as a .wad file. For Office-Office communication a .wad file is sufficient (no need for a .wsp).

■ Current Annex F practice:

- E-filed record copies are Annex F wsp files grouping all parts of the record copy into a single wsp file.
- Other document types (e.g. pdocs, translations of application body) are often packaged as Annex F wad files representing a single document.

PCT- EDI how to send documents

EXAMPLE FILE

- An example file can be sent simply following minimum specifications:

- PCTIB2009006952-isre-000001-EN-JP-20100202.pdf

ALTERNATIVELY

- using Annex F packaging the same file can be sent as follows:

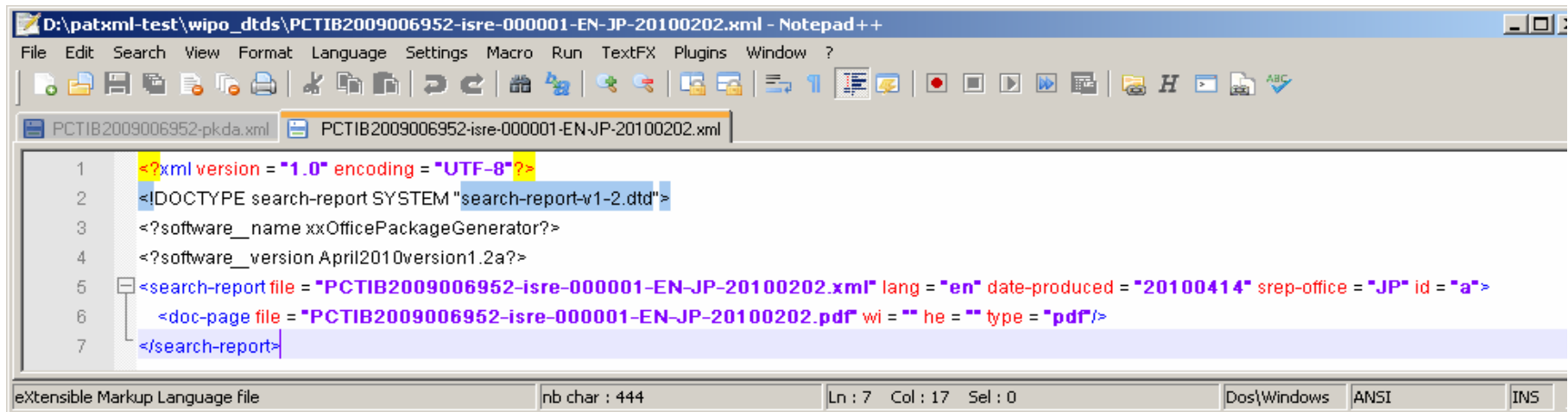
- PCTIB2009006952-isre-000001-EN-JP-20100202.wad containing:

- PCTIB2009006952-pkda.xml

- PCTIB2009006952-isre-000001-EN-JP-20100202.xml

- PCTIB2009006952-isre-000001-EN-JP-20100202.pdf

- Contents of the ISR XML:



The screenshot shows a Notepad++ window with the following XML content:

```
<?xml version = "1.0" encoding = "UTF-8"?>
<!DOCTYPE search-report SYSTEM "search-report-v1-2.dtd">
<?software_name xxOfficePackageGenerator?>
<?software_version April2010version1.2a?>
<search-report file = "PCTIB2009006952-isre-000001-EN-JP-20100202.xml" lang = "en" date-produced = "20100414" srep-office = "JP" id = "a">
  <doc-page file = "PCTIB2009006952-isre-000001-EN-JP-20100202.pdf" wi = "" he = "" type = "pdf"/>
</search-report>
```

The status bar at the bottom indicates: eXtensible Markup Language file, nb char : 444, Ln : 7 Col : 17 Sel : 0, Dos\Windows, ANSI, INS.

PCT- EDI how to send documents

EXAMPLE FILE

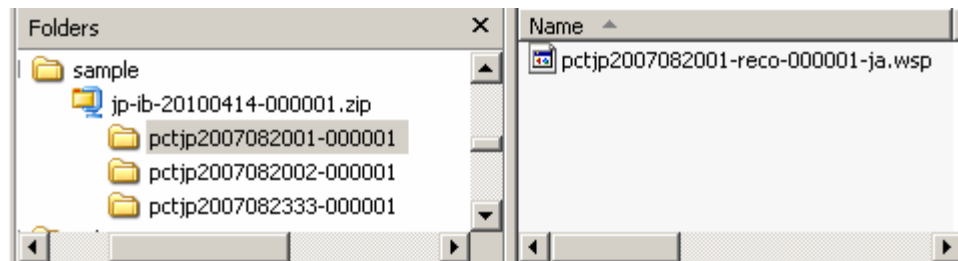
Contents of the
package data:

```
1 <?xml version = "1.0" encoding = "UTF-8"?>
2 <!DOCTYPE package-data SYSTEM "package-data-v1-6.dtd">
3 <?software__name xxOfficePackageGenerator?>
4 <?software__version April2010version1.2a?>
5 <package-data produced-by = "RO" dtd-version = "1.6" lang = "en" date-produced = "20100414">
6 <transmittal-info>
7 <office-to-office>
8 <application-reference>
9 <document-id>
10 <country>JP</country>
11 <doc-number>PCT/IB2009/006952</doc-number>
12 </document-id>
13 </application-reference>
14 </office-to-office>
15 </transmittal-info>
16 <signatories>
17 <signatory>
18 <name/>
19 <electronic-signature date = "20100414">
20 <basic-signature>
21 <text-string>/jpf</text-string>
22 </basic-signature>
23 </electronic-signature>
24 </signatory>
25 </signatories>
26 <search-report-doc file = "PCTIB2009006952-isre-000001-EN-JP-20100202.xml" lang = "en" />
27 </package-data>
```

PCT- EDI how to send documents

EXAMPLE BATCH

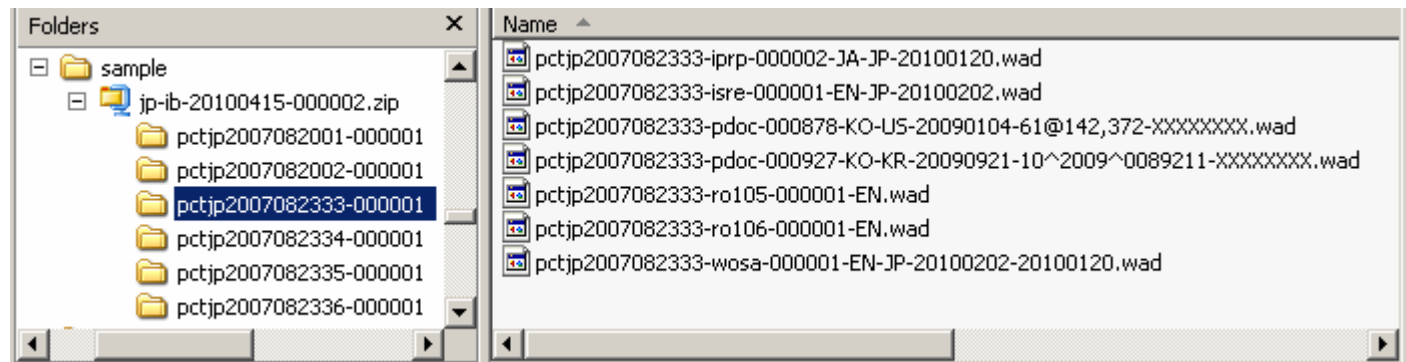
- Regardless of the transmission means, each batch sent by an Office should be sequentially numbered. This greatly aids traceability and maintenance.
- Office batches typically by office role / document type (but not required by the IB).
- A record copy batch would look something like



- The batch file jp-ib-20100414-000001.zip can then be transmitted to WIPO by the chosen transmission method
- This batch would represent the first ever batch (sequence number 000001) from JP to IB. The date 20100414 is merely an indication of when the batch was sent.

PCT- EDI how to send documents BATCHES

- There is no limit to the number of documents / IAs in a single batch. For practical reasons a batch should not be designed to be over a Gb.
- Below is an example that shows multiple documents being sent for the same IA in the same batch:



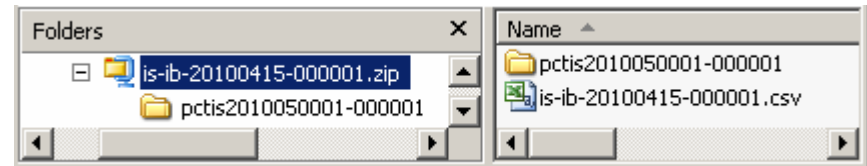
- Note: In the example above there is one Annex F wad per document. It would be perfectly acceptable to provide a single Annex F wad for a set of documents.
- In both examples files are grouped into folders by IA number.

PCT- EDI how to send documents

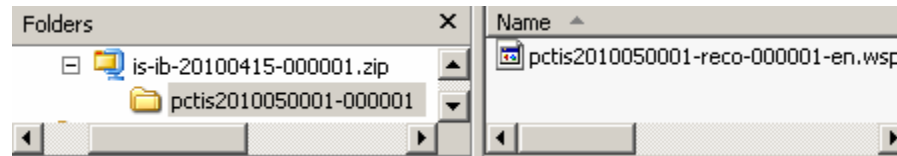
Sample minimal batch

- PCT EDI –The smallest batch that can be transmitted at a time is a single document. The example below is for one record copy.

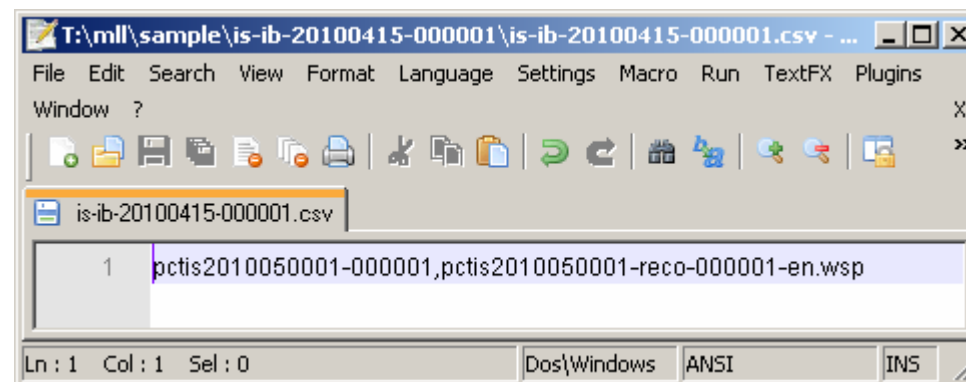
- Structure (index csv location):



- Content:



- CSV file index content:



PCT- EDI how to send documents

Document batch names (to IB)

- A zip file is named as follows:

RO-`{OO}`-yyyymmdd-nnnnnn`{-ROLE-CODE}`.zip

- Where

- RO is the WIPO Standard ST.3 code of the transmitting office. Optionally “`{OO}`” may be included as destination e.g. “US-IB-...”

- yyyymmdd is the date of transmission in local office time, not the date that the package was put together

- nnnnnn is a sequential counter incremented with each wrapper file sent by the office

- And optionally ROLE-CODE (RO|DO|EO|COR|ISA|IPEA|IB) may be included to direct the package to the correct operational unit within the office.

PCT- EDI how to send documents

file naming (to IB)

- The PCT Number
- The document type code in lower case (see the MinSpec document types)
- A numeric string NNNNNN to make the filename unique within its directory NNNNNN is a number right justified and padded with leading zeros
- The upper case ISO639 code representing the language of the document or XX if unknown
- Any optional additional bibliographic data information (format depending on the document type, see Annex of the MinSpec)
- An optional document status code "-corrected" to be used where a corrected version document previously supplied is furnished (corrected in business terms).
- The appropriate file extension (zip, wad, wsp).

PCT- EDI how to send documents

Record Copy WASP notes

- Please refer to Annex F for greater detail

- RO–WASP files
 - Include applicant WASP as submitted
 - Included RO converted and processed application files
 - PDF to TIFF conversion
 - Office updates reflected in XML files
 - Additionally including ex-officio corrections in Office section if required

- RO-WASP with Forms (PDF or WADs)
 - In Office part of WASP
 - OR alongside in PCT-EDI batch

Figure 14bis - Example of RO Package containing a record copy where the text of the description, claims and abstract is not in character coded format (but in PDF format)

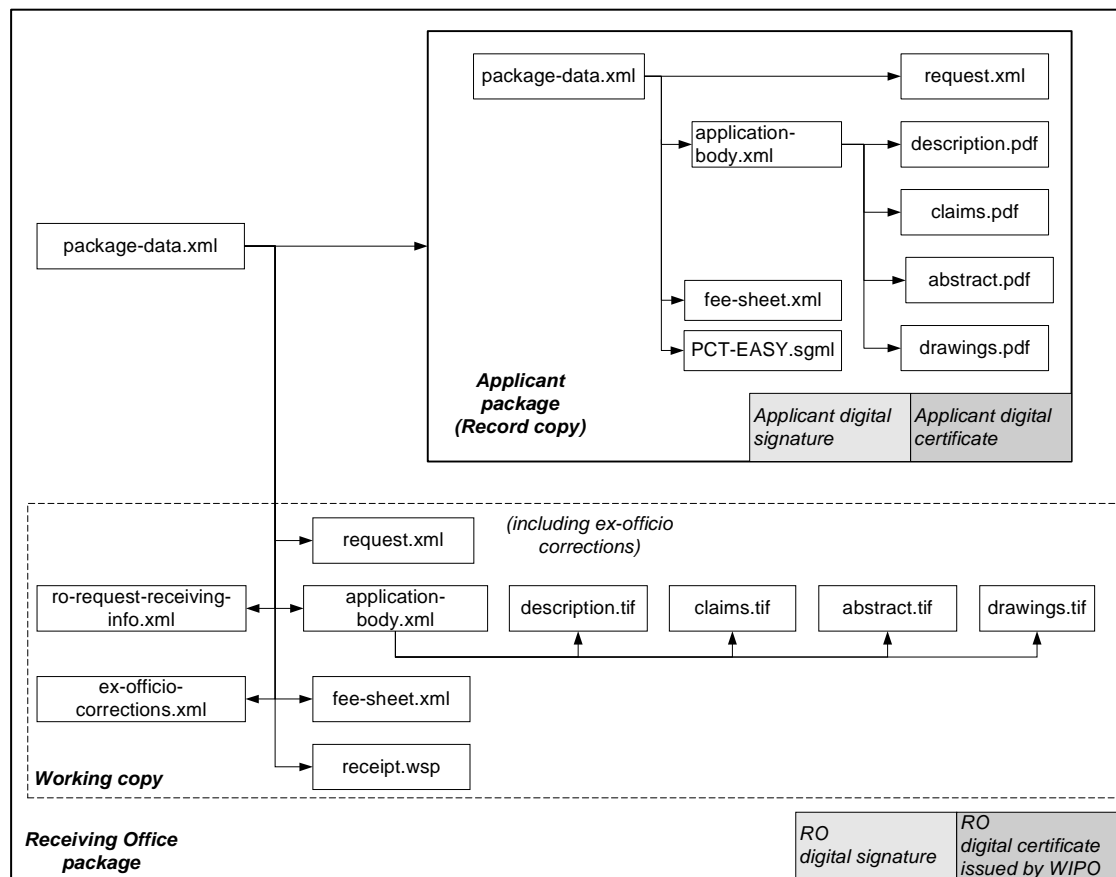
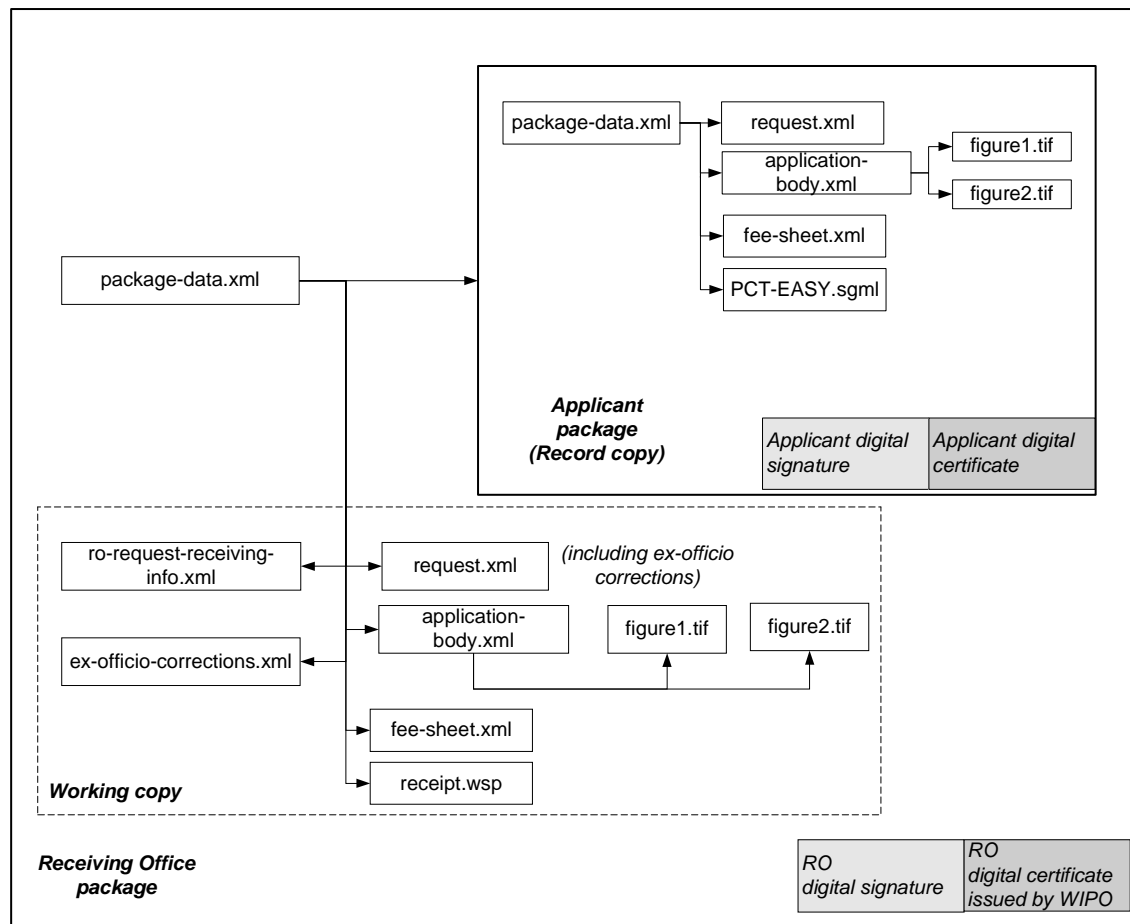


Figure 14ter - Example of RO Package containing a record copy where the text of the description, claims and abstract is in character coded format (in XML format)



PCT- EDI how to send documents

Sequence Listing content notes

Sequence listings and Sequence listings 13ter

- The content takes the form of zip file containing an ST.25 / Annex C TEXT file with an .app or .txt extension
- These files should not be converted by the RO in any way from the files submitted by the applicant

PCT-EDI – how to send documents 'after batches uploaded'

- Uploaded batches are:
 - moved from the office upload directory after session disconnect
 - initial acknowledgement is placed in the file

receive-at-ib.file-history.csv
 - copied into the IB 'as received store'
 - processed into the IB systems
- An acknowledgement CSV file placed in the office download directory

receive-at-ib.file-history.csv

- This is a plain csv file (it can be opened in Excel).
- The columns are as follows:
 - date
 - time
 - size bytes
 - name
 - Comments (ad hoc IB operational comments)
- e.g. A row such as:
 - 20091001,092843,71473009,fr-ib-20090930-000001.zip
 - is the acknowledgement that PCT-EDI picked up a file called fr-ib-20090930-000001.zip on the 1 Oct 2009 at 09:28:43 a file that is approximately 70Mb
- Note: it does not mean the file was understood or processed.

The acknowledgement csv file

- According the minimum specifications every batch should be sent with a csv file listing the contents of a batch as a means of cross checking with the actual contents.
- This CSV file is used to form the basis of the acknowledgement csv content.
- The intention of the acknowledgment file details is to provide feedback per file contained within the batch.
- The columns in the file are as follows:
 - ia number
 - file name
 - status
- e.g. pctfr2009051026,pctfr2009051026-reco.wsp,RECEIVED_NOT_LISTED

All possible values of status

- RECEIVED
 - BAD_NAME
 - BAD_CONTENT
 - BAD_FOLDER_NAME
 - DUPLICATE
 - RECEIVED_NOT_LISTED
 - BAD_NAME_NOT_LISTED
 - BAD_CONTENT_NOT_LISTED
 - BAD_FOLDER_NAME_NOT_LISTED
 - DUPLICATE_NOT_LISTED
 - MISSING
- RECEIVED and RECEIVED_NOT_LISTED - file has been received and initial validation is ok.
 - DUPLICATE and DUPLICATE_NOT_LISTED - file has already been sent in the past and is a real duplicate (md5 checksum)
 - BAD_NAME, BAD_CONTENT, BAD_FOLDER_NAME (the with xx_NOT_LISTED) - file sent is not automatically recognized and processible by PCT-EDI. There could be many reasons for this, typically reasons are:
 - file name used is not recognized (e.g. wrong document type code used)
 - pdf content is corrupt in some way
 - files are not grouped into folders by IA
 - MISSING - file listed in the csv file that was sent but is not actually present in the batch.
 - the xx_NOT_LISTED status messages mean that the file indicated has been sent, however it was not listed in the csv file that was sent with the batch.
 - Header info:
 - Feedback generated, source batch does not contain a csv, please refer to the minimal specifications for EDI

PCT- EDI Logs / Acknowledgements

- Office MONITORS:
 - receive-at-ib.file-history (CSV)
 - acknowledgement CSV files
- Office CLEANS UP:
 - contents of feedback folder





Sending data to the IB

PCT-EDI – sending data to IB

- National Phase Information

- XML file - DTD – v1.3

- CSV file

- Specification at

- <http://www.wipo.int/patentscope/en/pct-edi/documents/>

- Fee Data

- XML file - DTD – v1.3

- Specification at

- <http://www.wipo.int/patentscope/en/pct-edi/documents/>





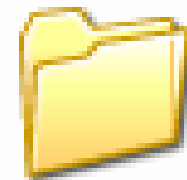
PCT-EDI Receiving documents

(presumes PCT-EDI connection already established)

PCT EDI Systematic Communication from IB

- IB Forms
- ISA Search Copies (from RO-IB)
- Supplementary International Search

- Excludes
 - PCT publication
 - Rule 87 / Article 20
 - Sequence listing portal



Download

PCT EDI download procedure

- Batches received in the download folder
- Office should check batches are:
 - named / sequenced correctly
 - valid and can be opened
- Office processes downloaded batch into local system
- Office deletes processed batch from the download directory on PCT-EDI



Download

PCT-EDI batch names (from the IB)

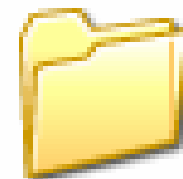
All batches are named according to the following pattern:

IB-`{office}`-`{date}`-`{6 digits office iterator}`-`{extra}`.zip.

`{office}` ST.3 code of the recipient Office.
`{date}` “yyyymmdd” is the date of sending

`{6 digits office iterator}` is a sequence specified as six numeric characters incremented from “000001” per Office, and is not intended to be reset to zero.

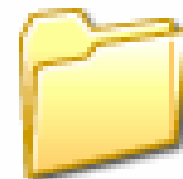
`{extra}` (Role code) In the case of batches destined to a specific Office role such as International Searching Authority the value will be ISA etc.
Possible values are: **RO, DO, ISA, IPEA, COR**



Download

Batch name examples

- IB-CN-20080717-000042-RO.zip
- IB-CN-20080720-000043-DO.zip
- IB-CN-20080720-000044-EO.zip
- IB-CN-20080721-000045-ISA.zip
- IB-CN-20080723-000046-IPEA.zip
- ...



Download

Content files (I)

- ***ZIP format***

- The content takes the form of A4 TIFF image files containing CCITT group 4 reverse bit order 300dpi black and white images.

- The files are named in their page order sequence:

000001.TIF

000002.TIF

000003.TIF

... etc.

Content files (II)

- WAD format
 - The specification of the WAD files is taken from ANNEX F, and specifically contains a structure of files accessed via an index file called:

packagedata-pkda.xml

- This file contains references to all other content within the WAD.

XML and TIFF image files
Black and white images

Forms WAD naming (sample batch)

- IB-CN-20080717-000003-ISA.csv
- PCTCN2008000951/PCTCN2008000951-ib307-000002-EN-20080717-ISA^CN-copy.wad
- PCTCN2008000969/PCTCN2008000969-ib301-000037-EN-20080717-ISA^CN-copy.wad
- PCTCN2008000970/PCTCN2008000970-ib301-000042-EN-20080717-ISA^CN-copy.wad
- PCTCN2008000989/PCTCN2008000989-ib301-000002-EN-20080717-ISA^CN-copy.wad
- PCTCN2008001108/PCTCN2008001108-ib371-000057-EN-20080717-ISA^CN-copy.wad

Form WADs

- WAD file typically contains the following entry in the packagedata-pkda.xml file, entry:

```
<notification file="pct-form-subset.xml" notf-type="ib317"/>
```

Search Copies

- Sample Batch Contents

IB-CN-20080715-000004-ISA.csv

PCTIB2007003218/PCTIB2007003218-isco-000024-EN-20080710.wad

PCTIB2007004447/PCTIB2007004447-isco-000018-EN-20080709.wad

...

Search Copy WADs

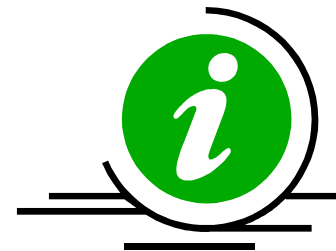
- Each WAD file typically contains the following entries in the packagedata-pkda.xml file:
- `<application-request status="search-copy" file="pctIB2007003218-requ.xml"/>`
- `<application-body-doc status="search-copy" file="pctIB2007003218-appb.xml"/>`
- `<other-documents>`
- `<other-doc file="wo-other-doc.xml" file-type="xml">`
- `<document-name>index file for printing purpose</document-name>`
- `</other-doc>`



Communication on request from IB

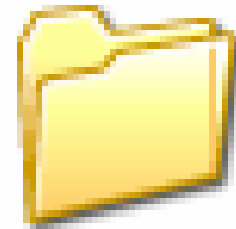
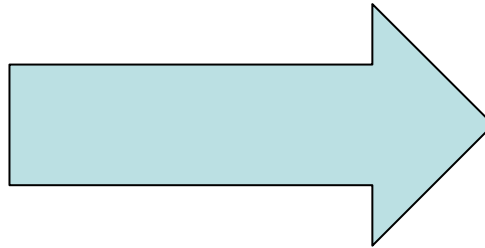
COR Communication

- Upload order
 - Choices for media types
 - Download documents
-
- COR communication is scheduled to change during 2010 and 2011 and it is recommended that any implementation be made using a new version, shortly to be announced



COR order XML

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE cor-order SYSTEM "http://www.wipo.int/it/en/projects/impact/cor-order.dtd">
<cor-order>
  <requested-media>
    <cd-dvd content-type="tiff"/>
  </requested-media>
  <from>
    <country>US</country>
  </from>
  <cor-request ia-number="AT2006/000351">
    <cor-doctype kind-of-doc="pamphlet"/>
    <cor-doctype kind-of-doc="et-iper"/>
    <cor-doctype kind-of-doc="p-doc"/>
    <cor-doctype kind-of-doc="declaration"/>
    <cor-doctype kind-of-doc="iper"/>
  </cor-request>
  <cor-request ia-number="ZA2007/000012">
    <cor-doctype kind-of-doc="pamphlet"/>
    <cor-doctype kind-of-doc="et-iper"/>
    <cor-doctype kind-of-doc="p-doc"/>
    <cor-doctype kind-of-doc="declaration"/>
    <cor-doctype kind-of-doc="iper"/>
  </cor-request>
</cor-order>
```

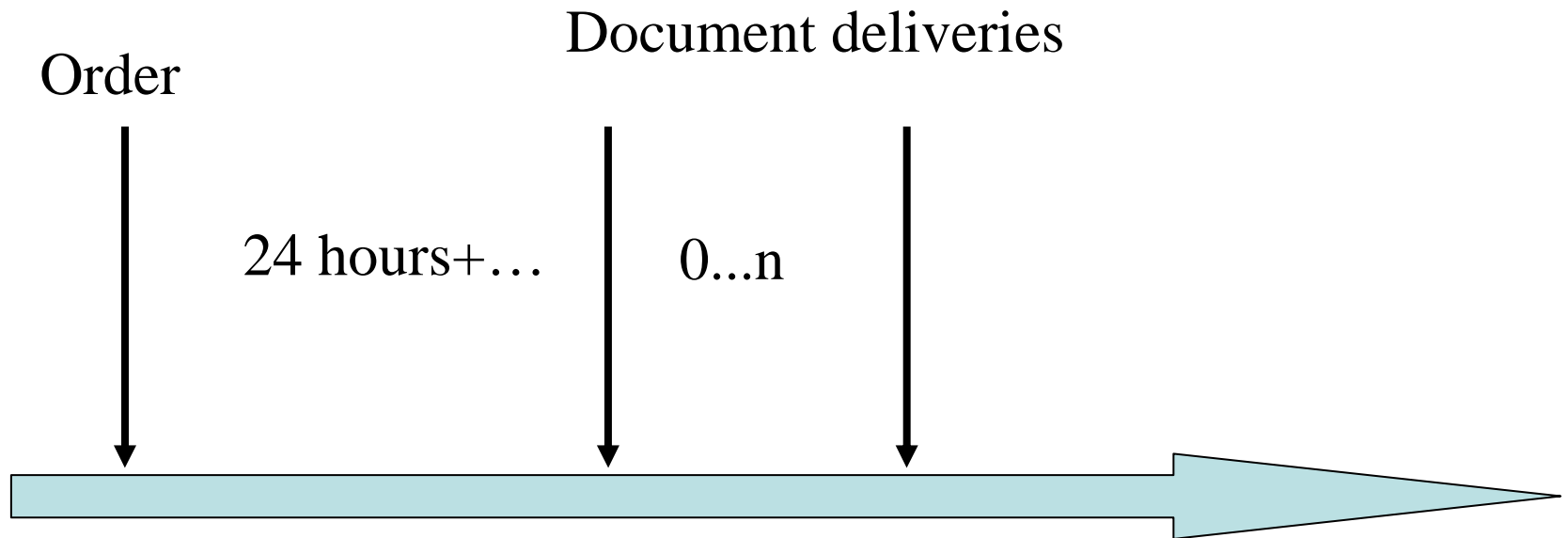


Request



feedback

COR ordering timelines



COR Batches (output)

An example batch file name is as follows:

IB-CN-20080710-000002-COR.zip



Download

The Specific COR document delivery is consistent with the existing DVD format, that is supported by the COR browser application.

Specific COR documents are dispatched using one batch per Office per role per day where there are forms to be sent on a given day; the contents of each batch file are as follows:

- a CSV index file listing the contents;

- sub-directories, named with the complete IA number format, one per IA;

- transmitted documents, each as ANNEX F WAD files.

COR Batches

- For example a batch, the second in the sequence, transmitted on July 10th 2008 to the Chinese Patent Office (IB-CN-20080710-000002-COR.zip) the content listing might appear as follows:
- IB-CN-20080710-000002-COR.csv
- index.xml
- CN02/000951/CN08_000951-pdoc-12345678.zip
- CN02/000969/CN02_000969-decla-12345678.zip
- CN05/000970/CN05_000970-wosa-12345678.zip



Download



Questions?



PCT-EDI further technical information

PCT-EDI further technical information

- IB implementation overview
- Office implementation outline
- Office software choices

PCT-EDI IB Implementation overview

- Infrastructure
 - Pair of redundant servers
 - running SFTP service
 - Connected to the Internet, TriNet and PatNet
 - Public / Private Key pair authentication only

- PCT-EDI Applications
 - Automated document transmission to Office download folder
 - Automated document collection from Office Upload folder
 - Automated COR order collection from Office Request folder

- At the IB Connected to
 - PCT COR Service
 - PCT weekly/daily media batch processing
 - Document capture

PCT-EDI Office Implementation outline

- Infrastructure
 - SFTP Client installation
 - Connect via Internet or TriNet

- At an Office, connected to:
 - COR (Upload / download)
 - Document batch reception process
 - Document batch transmission process

PCT-EDI office software choices

- Office - clients
 - WinSCP
 - Tectia SSH
 - Putty
 - Open SSH
 - Lftp
 - SFTP Perl module
 - JAVA SFTP etc

