
PCT International Cooperation Division

Data Format Specifications for documents delivered by the International Bureau to Offices through PCT- EDI (FTP)

*Version Number 1.0
October 24, 2008*



**WORLD INTELLECTUAL PROPERTY
ORGANIZATION**

GENEVA

Document Information

| | |
|---------------------|---|
| Document title: | Data Format Specifications for documents delivered by the International Bureau to Offices through PCT-EDI (FTP) |
| Document file name: | ib_ftp_package_formats.doc |
| Issued by: | Mr. Waring, |
| Reviewed by | Messrs. Richardson, Waring, Song, Blanc, Leach |
| Issue Date: | October 24, 2008 |
| Status: | Approved |

References

1. PCT EDI Cooperation Project.
2. The Minimal Specification for Transmitting Documents to the IB (MINSPEC).
3. WIPO Standard ST.36 (ST.36).
4. Annex F of the PCT Administrative Instructions (ANNEX F)

Table of Contents

- References 2
- Table of Contents 3
- 1. Introduction 4
- 2. Batch files and their naming..... 5
- 3. Content file formats..... 6
 - 3.1. ZIP format 6
 - 3.2. WAD format..... 6
- 4. Search copies 7
- 5. Daily forms..... 8
- 6. Specific COR documents 9
- Appendix I..... 10
 - A. Formats used prior to January 1, 2009 10

1. Introduction

The purpose of this document is to describe the technical specification of the formatting of batches of documents transmitted as part of the PCT process by FTP from the International Bureau (IB) to Offices.

This transfer is within the scope of the PCT Electronic Data Exchange Project (PCT-EDI) in order to facilitate the internet-based exchange of a wide range of electronic documents and data between WIPO and IP Offices.

This document describes the technical format of batches of documents transmitted by the IB to offices in their roles of receiving Office, International Authority and/or designated Office under the Patent Cooperation Treaty (PCT). These batches of documents can be grouped into three broad categories: “Search Copies” of international Applications received by the receiving Office at the IB (RO-IB) being transmitted to International Searching Authorities (ISAs), “Daily Forms” being transmitted to Offices in their roles of receiving Office, International Authority and/or designated Office and “Specific COR documents” being transmitted to Offices in their role of designated Office.

It should be noted that the DVD formats used for the transmission of documents under Article 20 or Rule 87 are not within the scope of this document.

/...

2. Batch files and their naming

All batches are transmitted as a single ZIP file with the extension “.zip”. It should be noted that prior to January 1, 2009 batches were split into 100MB chunks and that this is no-longer the case.

All batches are named according to the following pattern:

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

Where the following rules apply to the codes, separated by the “-“ character, specified between the brackets “{}”:

| | |
|----------------------------|---|
| {office} | is the WIPO Standard ST.3 code of the transmitting office |
| {date} | in format “yyyymmdd” is the date of transmission in local office time, not the date that the package was put together (in this case 20070213) |
| {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} | is an additional field provided to address specific needs. 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 and the value COR to denote that the batch is related to COR orders. |

/...

3. Content file formats

Content files are transmitted in either ZIP format or WAD format.

3.1. ZIP format

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

The files are named in their page order sequence:

000001.TIF
000002.TIF
000003.TIF
... etc.

3.2. 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. The other content will normally take the form of XML and TIFF image files where the TIFF image files contain CCITT group 4 300dpi reverse bit order black and white images.

/...

4. Search copies

Search copies are transmitted from the IB to the International Searching Authorities and Supplementary International Searching Authorities thus the value of the {extra} field in the batch name is either "ISA" or "SISA". For example:

batch file name: IB-CN-20080715-000004-ISA.zip

The contents of the 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.

For example a batch transmitted on July 15th 2008 to the International Searching Authority in China the content listing might appear as follows:

```
IB-CN-20080715-000004-ISA.csv
PCTIB2007003218/PCTIB2007003218-isco-000024-EN-20080710.wad
PCTIB2007004447/PCTIB2007004447-isco-000018-EN-20080709.wad
PCTIB2007004461/PCTIB2007004461-isco-000025-EN-20080710.wad
PCTIB2007004563/PCTIB2007004563-isco-000020-EN-XXXXXXXXX.wad
PCTIB2008000270/PCTIB2008000270-isco-000028-EN-20080712.wad
```

Please note that in the case where the date of receipt of the search copy at the receiving office is not available it will be replaced by the string 'XXXXXXXXX'.

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>
```

These entries constitute references to the ANNEX F XML files that contain the references that identify the pages of the documents transmitted.

/...

5. Daily forms

Daily forms are transmitted from the IB to offices in their capacities as receiving Office, International Authority and/or designated Office under the Patent Cooperation Treaty (PCT) and thus the value of the {extra} field in the batch name can be one of “RO”, “DO”, “EO”, “ISA”, “IPEA” or “SISA”. For example:

```
IB-CN-20080717-000004-RO.zip
IB-CN-20080717-000001-DO.zip
IB-CN-20080717-000002-EO.zip
IB-CN-20080717-000003-ISA.zip
IB-CN-20080717-000003-IPEA.zip
```

Daily forms 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.

For example a batch, the third in the sequence, transmitted on July 17th 2008 to the International Searching Authority in China (IB-CN-20080717-000003-ISA.zip) the content listing might appear as follows:

```
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
```

It should be noted that the “ISA^CN” translates to ISA-CN using the MINSPEC substitution rule and refers to the destination.

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

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

These entries constitute references to the ANNEX F XML files that contain the references that identify the pages of the document transmitted.

/...

6. Specific COR documents

Specific COR documents are transmitted from the IB to offices in their capacities as receiving Office, International Authority and/or designated Office under the Patent Cooperation Treaty (PCT). However the office orders the documents for delivery without specifying the purpose for which it is ordering the documents (typically as Designated Office), and thus the {extra} field contains the code “COR” rather than the role of the Office receiving the batch. An example batch file name is as follows:

IB-CN-20080710-000002-COR.zip

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.

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

[end of document, annex follows]

Appendix I

A. Formats used prior to January 1, 2009

A1. Search copies old format:

Archives split:
IB2557_CN_aa
IB2557_CN_ab
IB2557_index.xml
IB2557_merge.bat
IB2557_merge.sh

Content of archive:
IB07/003218/IB07_003218-isco-14466343.zip
IB07/004447/IB07_004447-isco-14466157.zip
IB07/004461/IB07_004461-isco-14451454.zip
IB07/004563/IB07_004563-isco-14145741.zip
IB08/000270/IB08_000270-isco-15478124.zip

A2. Daily forms old format:

Archives split:
IB2241_CN_aa
IB2241_CN_ab
IB2241_index.xml
IB2241_merge.bat
IB2241_merge.sh

Sample of content structure:
CN08/000951/CN08_000951-ib307-CN-ISA-copy-16461387.zip
CN08/000969/CN08_000969-ib301-CN-ISA-copy-16154124.zip
CN08/000970/CN08_000970-ib301-CN-ISA-copy-16148752.zip
CN08/000989/CN08_000989-ib301-CN-ISA-copy-11298745.zip
CN08/001108/CN08_001108-ib371-CN-ISA-copy-25487452.zip

A3. Specific COR output old format:

Archives split:
COR1234_CN_aa
COR1234_CN_ab
COR1234_index.xml
COR1234_merge.bat
COR1234_merge.sh

Content:
index.xml
CN02/000951/CN08_000951-pdoc-12345678.zip
CN02/000969/CN02_000969-decla-12345678.zip
CN05/000970/CN05_000970-wosa-12345678.zip