

中国发明专利申请公布标准化全文图像数据

(一) 数据编号

CN-PA-IMGS-10-A

(二) 数据名称

中国发明专利申请公布标准化全文图像数据

(三) 文献语种

中文

(四) 更新周期

周更新

(五) 数据内容及格式

数据内容：扉页、权利要求书、说明书、说明书附图、索引信息。

数据格式：PDF、XML

(六) 文件结构

1. 概述

本数据的文件结构包括两层文件夹，实体数据存储在第一层文件夹下。

第一层文件夹：数据编号与数据名称，例如 CN-PA-IMGS-10-A_中国发明专利申请公布标准化全文图像数据。

第二层文件夹：公布日期，8 位数字，YYYYMMDD（4 位年+2 位月+2 位日），例如 20160106。

2. 数据编号说明

编号方式：

国别代码-业务名称代码-数据资源名称代码-数据内容属性-专利类型代码-专利文献种类代码；

国别代码-业务名称代码-数据资源名称代码-数据内容属性-数据种类代码；

国别代码-业务名称代码-数据资源名称代码-数据内容属性-集成电路设计数据种类代码；

代码说明:

代码内容	编号方式	编号说明	参考依据
国别	两位大写字母	CN: 中国 EP: 欧洲专利局 JP: 日本 KR: 韩国 US: 美国	WIPO ST.3 标准
业务名称	两位大写字母	PA: 专利 TM: 商标 GI: 地理标志 IC: 集成电路布图设计	《知识产权基础信息数据规范（试行2020版）》
数据资源名称	2-4位大写字母	BIB: 著录项目 ABS: 摘要 TXT: 全文文本 IMG: 全文图像 PUB: 公告 PRS: 法律状态 FIFT: 日本 FI、Fterm 分类 USPC: 美国专利分类 ECLA: 欧洲 ECLA 分类 ICR: 国际专利分类 DEC: 决定 RE: 复审案件 IN: 无效宣告案件 CIT: 引文	《知识产权基础信息数据规范（试行2020版）》
数据内容属性	一位大写字母	O: 表示原始数据 T: 表示翻译数据 S: 表示标准化数据	
专利类型	两位数字	10: 发明专利 20: 实用新型 30: 外观设计	
专利文献种类	一位大写字母	A: 发明专利申请; B: 发明专利授权; U: 实用新型专利授权（韩国为实用新型专利申请）; Y: 实用新型专利授权（日本2000年以前, 中国1993-2010年4月及韩国为实用新型专	

		利授权)； S: 外观设计专利授权；	
数据种类	2-4 位大写字母	FILE: 表示为数据； DIC: 表示为字典、定义、说明文件等；	
集成电路布图设计数据种类	2-4 位大写字母	ER: 专有权公告； ERA: 专有权事务公告； TER: 专有权终止公告； RE-RE: 复审及撤销案件；	

(七) 数据说明

文件层级结构说明

包括两部分内容：

第一部分实体数据文件夹，文件层级结构为

YYYYMMDD-N1-N2. ZIP/N1/C1K1YYYYNNNNNNNNNC2NNNNNNNNNNNNNK2SRCLLYYYY
MMDDSSBKX/C1K1YYYYNNNNNNNNNC2NNNNNNNNNNNNNK2SRCLLYYYYMMDDSSBKX. PDF+
*. XML

YYYYMMDD-N1-N2. ZIP: 其中 YYYYMMDD 为公开（公告）日，如 20160106；N1 为文件夹流水号，每个流水号中不超过 2000 件专利文献，如本期数据超过 2000 件则按顺序流水编号，如 20160106-1-001. ZIP 为第一个 2000 件，20160106-2-001. ZIP 为第二个 2000 件…以此类推；N2 表示每 2000 件数据按容量是否超过 2G 分隔的压缩包个数，每个压缩包的数据容量以 2G 进行划分，超过 2G 后顺序排列，如 20160106-1-002. ZIP 为 20160106 期的第一个 2000 件中的第二个 2G 压缩包…以此类推。

N1 为文件夹流水号，用法同上。

C1K1YYYYNNNNNNNNNC2NNNNNNNNNNNNNK2SRCLLYYYYMMDDSSBKX 为标准化的文件夹名称。

C1K1YYYYNNNNNNNNNC2NNNNNNNNNNNNNK2SRCLLYYYYMMDDSSBKX. PDF 为标准化的文件名称，即全文图像数据。

第二部分为索引文件：其中“公开（公告）日-INDEX-流水号”命名的文件，

提供数据配套的路径索引文件；“公开（公告）日-INDEX-文献类型-流水号”命名的文件，提供号单索引文件。