

中国专利全文图像数据

（一）数据名称

中国专利全文图像数据

（二）文献语种

中文

（三）更新周期

周更新

（四）数据内容及格式

1、基本内容

扉页，权利要求书，说明书，说明书附图，索引信息

2、数据格式

XML、PDF

（五）文件存储层级结构

1、概述

本数据的文件存储层级结构包括三层文件夹，实体数据存储在第三层文件夹下。

第一层文件夹：数据类别。数据类别由数据内容描述（小写英文字母）表示。
例如：full image。

第二层文件夹：文献种类描述，由小写英文字母组成，例如：invention publication。

第三层文件夹：公开日期，8位数字，YYYYMMDD（4位年+2位月+2位日），例如：20160330。

2、相关说明

1) 数据内容描述说明

bibli and abstract: 著录项目与文摘数据

full text: 全文文本

full image: 全文图像
legal status: 法律状态
enriched citation: 引文
family: 同族
reexamination: 复审
invalidation: 无效
verdict: 判决书

2) 文献种类描述说明:

invention publication: 发明专利申请
invention grant: 发明专利授权
utility model: 实用新型专利授权
design: 外观设计专利授权

(六) 数据说明

1、文件包层级结构说明

示例: YYYYMMDD-N1.ZIP/C1NNNNNNNNNK1/ C1NNNNNNNNNK1. XML

YYYYMMDD-N1.ZIP

YYYYMMDD 为公开(公告)日, 如 20160330; N1 表示任一公开(公告)日的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 20160330-002.ZIP 为 20160330 期的第二个 2G 压缩包...以此类推。

C1NNNNNNNNNK1

C1 为公开国别代码(二位大写字母), NNNNNNNNN 为公开流水号, K1 为文献类型, C1NNNNNNNNNK1 为文件夹名称。该文件夹中包括: 文件名称为 C1NNNNNNNNNK1.PDF 的全文图像数据文件及对应的 C1NNNNNNNNNK1PTD.XML 索引文件。

2、索引文件主要数据元素说明

元素名称	元素内容
TitlePages	扉页起始及终止页
ClaimPages	权利要求起始及终止页
DescriptionPages	说明书起始及终止页
DrawingPages	说明书附图起始及终止页