C

CWS/14/7
原文：英文
日期：2026 年 9 月 23 日

# 产权组织标准委员会（标准委）

**第十四届会议**
2026 年 11 月 23 日至 27 日，日内瓦

知识产权文档数字化调查结果分析

秘书处编拟的文件


## 概　要

1.　国际局分析了关于主管局在工业产权文档数字化过程中所遇挑战的调查答复。调查分析表明，主管局在数字化过程中面临的主要挑战在于从专利文档中提取复杂内容。


## 背　景

2.　在产权组织标准委员会（标准委）第十三届会议上，国际局提交了一份关于其注意到的数字化挑战的文件，特别是国际检索单位在努力满足《专利合作条约》（PCT）最低限度文献要求期间的挑战（见文件 CWS/13/28）。

3.　标准委批准了通过标准委和 PCT 联合通函，向标准委成员和 PCT 缔约方分发一份调查问卷，调查主管局在对其工业产权文档集进行数字化时遇到的挑战（见文件 CWS/13/33 第 212 段）。

## 调查结果

4.　2026 年 5 月，秘书处编拟了调查问卷，并通过标准委和 PCT 联合通函向标准委成员以及 PCT 受理局、国际检索单位、国际初步审查单位和/或指定局或选定局分发（见 C.CWS 206/C.PCT 1701）。通函收件人可于 2026 年 5 月 15 日至 6 月 15 日期间答复调查。

5.　共有 56 个主管局或组织答复调查，包括以下成员国：阿尔及利亚（DZ）、澳大利亚（AU）、阿塞拜疆（AZ）、巴林（BH）、比利时（BE）、波斯尼亚和黑塞哥维那（BA）、巴西

（BR）、保加利亚（BG）、加拿大（CA）、智利（CL）、中国（CN）、哥斯达黎加（CR）、克罗地亚（HR）、捷克（CZ）、丹麦（DK）、赤道几内亚（GQ）、爱沙尼亚（EE）、法国（FR）、格鲁吉亚（GE）、德国（DE）、匈牙利（HU）、印度（IN）、爱尔兰（IE）、以色列（IL）、牙买加（JM）、日本（JP）、拉脱维亚（LV）、立陶宛（LT）、墨西哥（MX）、莫桑比克（MZ）、新西兰（NZ）、挪威（NO）、阿曼（OM）、巴拉圭（PY）、秘鲁（PE）、菲律宾（PH）、波兰（PL）、大韩民国（KR）、摩尔多瓦共和国（MD）、罗马尼亚（RO）、俄罗斯联邦（RU）、沙特阿拉伯（SA）、塞尔维亚（RS）、新加坡（SG）、斯洛伐克（SK）、西班牙（ES）、斯里兰卡（LK）、瑞典（SE）、土耳其（TR）、联合王国（GB）、坦桑尼亚联合共和国（TZ）和美利坚合众国（US）；以及以下组织：非洲地区知识产权局（AP）、欧亚专利组织（EA）、欧洲专利局（EP）和欧洲联盟知识产权局（EM）。

6. 对调查问卷的各份答复和汇总结果已在《产权组织手册》第 7 部分公布。2026 年 7 月举办了一次讲习班，以分享调查结果和秘书处所作的初步分析。有关此次讲习班成果的更多详情见文件 CWS/14/8。

## 对调查的分析

7. 国际局开展本次调查，是希望更清楚地了解主管局在对其知识产权文档集进行数字化时遇到的挑战，以便就如何应对这些挑战提出建议，从而支持主管局开展数字化活动。

8. 虽然本次调查涉及知识产权文档数字化方面的挑战，但结果显示，这些挑战主要与专利文档有关，超过四分之三的调查答复者指出存在问题。相比之下，商标方面存在问题的主管局占 7%，工业品外观设计方面占 4%。

9. 调查向各主管局提出的第一个问题，旨在确定向由国际局托管的全球数据库提供数据是否可能成为开展数字化活动的推动因素。然而，开展数字化活动的主管局与向由国际局托管的数据库提供数据的主管局之间没有任何相关性。结果如下图 1 所示。

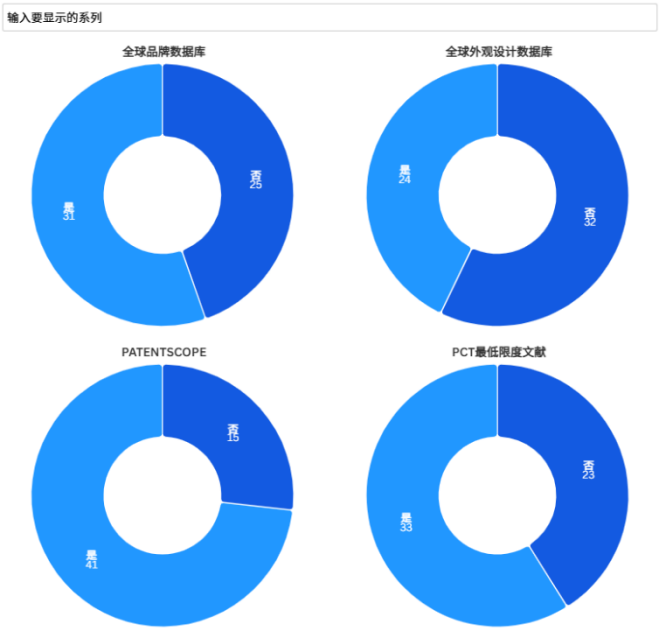



**图 1:调查收件人向产权组织数据库提供数据的比例。**

10. 接下来的问题涉及主管局在受理申请时接受的格式和公布时所采用的格式。如果两者存在不同，则说明主管局正在进行转换，并在某些情况下进行数字化。专利申请受理时最常接受的三种格式依次为 PDF（84%）、纸件（82%）和 DOCX（47%）。商标和外观设计最常见的三种格式依次为 PDF（70%至 80%）、图像格式（60%至 70%）以及工业品外观设计的纸件（约 70%）。关于为何有大量主管局接受纸件申请，一种假设是，可能因为这些主管局是《专利法条约》（PLT）的缔约方。然而，该条约共有 45 个缔约方，与在调查中表示接受纸件专利申请的调查收件人并不对应。

11. 就主管局采用的公布格式而言，专利绝大多数（超过 90%）以 PDF 格式公布，同时有很大比例（超过 70%）在公布时遵照产权组织标准 ST.36 或 ST.96 之一。就商标和工业品外观设计而言，PDF 格式（占 80%以上）和图像格式（占 40%以上）是最常用的公布格式，超过 50%的主管局按照至少一项产权组织标准进行公布。结果表明，商标和工业品外观设计在数字化方面挑战较少，部分原因在于受理申请时的格式与公布时采用的格式相同。

12. 关于著录项目数据，所考察的每类知识产权最常用的三种公布格式见下表 1。这些结果表明，无论知识产权类型如何，主管局都倾向于以 PDF 或 XML 格式公布著录项目数据。

| 知识产权 | 格式 | 采用该格式进行公布的主管局数量 |
|---|---|---|
| **专利** | PDF | 43 |
| | XML | 34 |
| | 纯文本 | 14 |
| **商标** | PDF | 40 |
| | XML | 26 |
| | 纯文本 | 13 |
| **外观设计** | PDF | 37 |
| | XML | 23 |
| | 其他 | 14 |

**表 1：按知识产权类型划分的三大著录项目格式。**

13. 接下来的问题涉及公布语言。这是为了确定申请语言是否会影响它们对知识产权文档进行数字化的就绪程度。最常用的申请语言见图 2。

14. 除三个主管局表示在西里尔文字方面存在一些有限问题外，各主管局均未报告其他挑战。这表明，数字化挑战与申请语言无关。

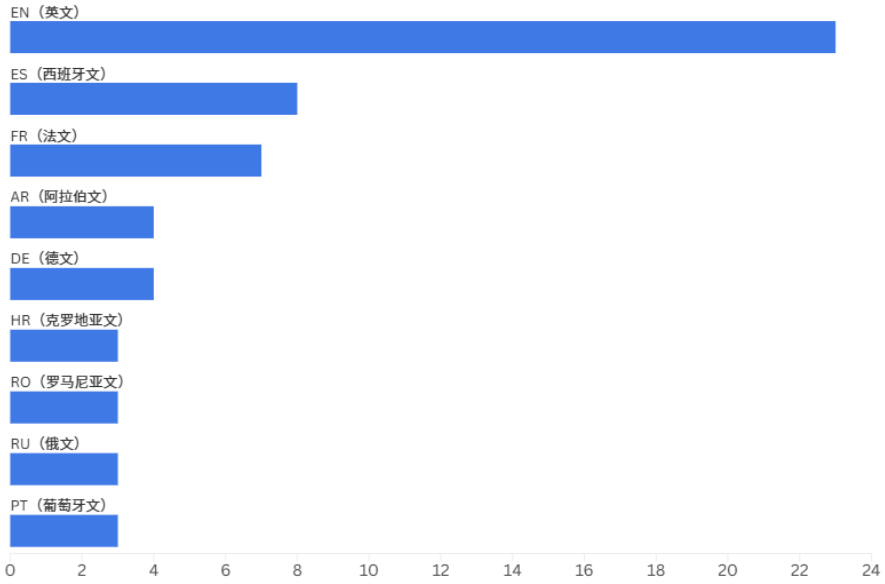



**图 2：调查答复者中最常用的申请语言。**

15. 鉴于 PCT 文本处理工作队的工作，PCT 体系特别关注的问题是，主管局是否公布其收到的彩色附图，还是仅以黑白形式公布。如果申请以 XML 格式提交，国际局已经能够公布彩色附图，但在提议修改 PCT 细则 11 和《PCT 行政规程》附件 F 之前，希望更好地了解用户的需求。结果显示，在允许提交彩色附图的主管局中，只有一半左右会公布彩色附图。结果如下图 3 所示。

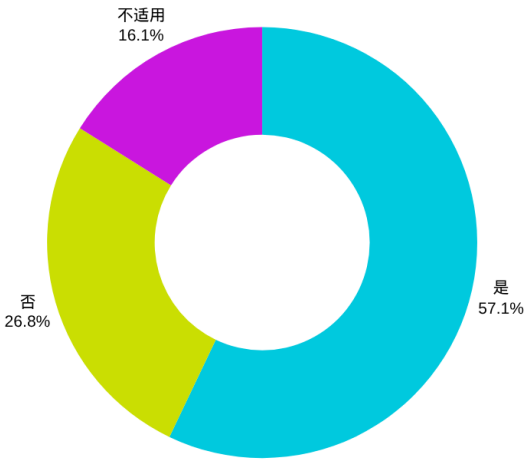



**图 3：主管局公布彩色附图吗？**

16. 调查的最后一部分所含的问题涉及主管局专门开展的数字化活动及其使用的流程和平台。在此语境下，数字化被定义为“将数据转录为可由计算机直接处理的数字形式的过程”[1]。超过80%的调查答复者表示开展数字化活动。其他未开展数字化活动的主管局则只是按申请时收到的格式予以公布。

17. 根据数字化流程，主管局可分为以下五级：

– 第 1 级：基本扫描（不进行光学字符识别（OCR））；

[1] 《柯林斯英语词典》

- 第 2 级：扫描后进行 OCR，生成 PDF；
- 第 3 级：扫描、OCR 以生成结构化输出（例如 XML）；
- 第 4 级：同上，但采用人工智能增强的处理流程；和
- 第 5 级：从申请到公布，全程采用纯数字化工作流程。

18. 如果主管局将数字化活动外包，主要原因通常有：需要数字化的文件数量庞大，尤其是专利历史档案；开展这项工作需要专门知识；外包具有成本效益；以及外部方执行时能更快完成。

19. 针对调查的最初问题和主要关注点，答复显示，主管局面临的最大挑战在于从专利文件中提取复杂内容。就连第 4 级主管局也表示即便使用人工智能提取内容仍会遇到问题，尤其是提取化学式和数学表达式时。其他主要挑战包括：纸件原件质量较差、与文件结构相关的问题（例如段落和权利要求编号），以及从周围文本中提取附图。

## 下一步工作

20. 根据对调查的答复和上文第 7 段至第 18 段所述的分析，国际局将考虑编拟一套数字化准则，尽可能就如何应对本次调查指出的挑战提供建议，或建议使用产权组织 OCR 解决方案等共同工具。关于这些下一步工作的更多详情，见文件 CWS/14/8。

21. 关于未来的计划，11%的主管局在调查答复中表示计划今后实施产权组织各项标准，或扩大产权组织标准的使用。鼓励有意实施产权组织标准的主管局通过电子邮件 cws.mail@wipo.int 联系国际局，寻求技术援助。

22. 请标准委员会：

(a) 注意本文件的内容；

(b) 审议并批准第 7 段至第 18 段所载的调查分析，并批准在《产权组织手册》第 7 部分中予以公布；并

(c) 鼓励有意将结构化输出作为其数字化活动组成部分加以实施的成员联系国际局寻求技术援助。

[文件完]