C

CWS/14/6
原文：英文
日期：2026 年 10 月 19 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

将知识产权文档转为 XML 格式

数字转型工作队牵头人和秘书处编拟的文件

概　要

1.　根据第 62 号任务，数字转型工作队一直开展工作，为知识产权文档从微软 Word 开放 XML 文档（DOCX）到可扩展标记语言（XML）的转换解决方案（DOCX2XML 转换器）制定共同功能要求和统一的输出配置。尽管各知识产权局目前使用不同的转换工具，对复杂特征（如表格和公式）采用不同的处理规则，并对哪个文件具有法律权威作出不同认定，但制定一个共同的基准将确保XML 的互操作性、呈现结果的一致性和数据交换的可靠性，而无需统一主管局各自的法律框架或业务框架。为支持这项工作，并探讨开发或复用共同的转换工具，工作队牵头人和秘书处提议与感兴趣的知识产权局开展结构化磋商，收集用例、技术要求和测试数据，目标是提交一份经修订的《DOCX2XML 转换器功能规范文件》，供产权组织标准委员会（标准委）第十五届会议审议。

背　景

2.　在第十一届会议上，标准委批准对第 62 号任务的说明作如下修订：

“着眼于电子申请和知识产权文献的公布和交换，审查基于纸质或图像的产权组织标准，在必要时提出对这些标准的修订或新建议；编写一份关于 DOCX 到 XML（DOCX2XML）转换器共同要求规范建议的提案。”

（见文件 CWS/11/28 第 59 段。）

3. 在第十三届会议上，标准委注意到数字转型工作队提交的《DOCX2XML 转换器功能规范文件》工作草案，该草案转录于文件 CWS/13/12 附件。除制定共同功能规范外，若干代表团还表示有兴趣开发一个共同的 DOCX2XML 转换器，用于实施所商定的功能规范。标准委指出，若干主管局，尤其是技术资源有限的主管局，将从这一共同解决方案中受益（见文件 CWS/13/33 第 74 和 75 段）。

DOCX2XML 做法和共同输出要求

DOCX2XML 转换器功能规范文件

4. 继标准委第十届会议达成一致后，数字转型工作队即着手编拟 DOCX2XML 转换器功能规范文件。鉴于此后若干主管局要求提供一种共同的 DOCX2XML 转换器，工作队扩大了讨论范围，既考虑可能开发或复用此类工具，也考虑统一若干主管局所用转换器产生的输出。

5. 这一更广泛的做法旨在确保，即使各主管局继续使用不同的转换工具，所产生的 XML 也是统一的且具有互操作性。因此，该功能规范文件不提交标准委本届会议作最后审议，以便工作队有更多时间，在考虑现有解决方案和主管局做法的基础上，纳入对共同工具的要求和统一的输出要求。数字转型工作队请标准委鼓励其成员审查转录于文件 CWS/13/12 附件的《DOCX2XML 转换器功能规范文件》工作草案并提出反馈意见。

6. 对共同 DOCX2XML 转换器的考虑并不取代对共同功能要求和输出要求的需要。相反，工作队意在确立一个共同的基准，该基准可通过一种共同的解决方案加以实施，也可通过主管局已经运行的转换器加以实施。这将使主管局既能保留适合其法律、技术和业务环境的 DOCX2XML 转换器，又能产生符合约定结构、可在各主管局系统之间一致地交换、验证和复用的 XML。

7. 主管局之间的协作对于这项工作至关重要。已经开发或实施 DOCX2XML 转换器的主管局可贡献技术知识、转换规则、有代表性的测试用例和经验教训，而正在考虑新解决方案的主管局则可提供其业务要求和预期用例。这些协作努力将有助于工作队查明可复用的组件，依据一套共同的示例对输出进行比较，并逐步统一转换结果，同时避免重复劳动并顾及技术能力的不同水平。

现行 DOCX2XML 转换做法

8. 数字转型工作队查明了中国国家知识产权局（国知局）、大韩民国知识产权部（MOIP）、美国专利商标局（美国专商局）、欧洲专利局（欧专局）以及国际局用于《专利合作条约》（PCT）申请的 DOCX2XML 转换器。工作队对所收集信息的审查表明，转换器可能以不同方式处理相同的 DOCX 特征，并产生不同的输出。例如，公式可能转换为 MathML 或图像；复杂表格可能以 XML 表示，也可能被拒绝，或被转换为图像；不支持的特征可能被删除、作为警告报告或作为错误处理。

9. 除数字转型工作队所开展的工作外，并根据标准委第十三届会议上关于共同 DOCX2XML 转换的请求，秘书处于 2026 年 9 月为感兴趣的知识产权局举行了一次在线会议。会议介绍了将 DOCX 专利申请转换为 XML 文档的解决方案和主管局做法，并探讨了合作机会。国际局报告称，PCT DOCX 转换器已运行十多年。该转换器集成于 PCT 申请流程中，可作为独立工具和网络服务使用，能够生成符合产权组织标准 ST.36 的 XML 文档、所引用的图像和便携式文档格式（PDF）的预览文件，支持多种语言和文字，并提供编号和验证信息。法国国家工业产权局报告称，其将 PCT 转换器作为网络服务用于本国专利申请。实施工作需要试点、指导、培训和专门支持，同时 DOCX 底层结构

的差异仍不断产生新的转换情形。美国专商局介绍称，其内部 DOCX2XML 转换器用于将专利申请转换为 XML 文档和其他下游格式，可执行提交前验证并识别错误和警告。该转换器能够处理常见的文档对象，并对无法可靠映射的内容加以标示。

10. 各主管局报告了对转换所得 XML 输出的地位和使用所采取的不同做法。在 PCT 下，XML 输出构成法律副本。法国国家工业产权局保留 DOCX 文件用于审查，并将 XML 和 PDF 输出用于传播和公布。美国专商局将经更正的 DOCX 文件视为法律副本，并生成 XML 和 TIFF 文件用于下游处理。尽管存在这些差异，共同的技术配置仍可界定可互操作的 XML 结构、验证结果和呈现方式。

11. 会议查明了可构成统一转换输出基础的若干要素：各系统间转换和呈现的一致性；明确的验证和预览机制；所提交内容的保全；法律副本或参考副本的确定；以及符合产权组织标准ST.36或 ST.96。就这些要素达成一致意见将促进合作，同时又能兼顾不同的国家实施情况和做法。

共同 DOCX2XML 输出的理由、范围和要求

12. 上述结果表明了各局加强合作以界定共同、统一的 DOCX2XML 输出和做法所具有的价值。尽管各局可能继续使用不同的转换工具，并对源文件和转换后的文件赋予不同的法律作用或业务作用，但就共同的输出配置达成一致意见，将提高所产生的 XML 文件在各局系统和工作流程中的一致性、互操作性和复用性。

13. 作为初始范围，合作应侧重于专利申请，包括说明书、权利要求书、摘要和附图。为这些内容提供一致的结构化 XML 文件，将有助于申请和审查程序、PCT 最低限度文献的准备、数字化、公布、传播、数据交换以及在后续工作流程中的复用。该范围日后可扩大到主管局认为具有明确业务效益的其他基于文本的知识产权文档。

14. 在此范围内，共同的做法应界定约定的用例和专利申请正文的最低结构，包括转换器必须识别的章节和内容以及任何可选要素或主管局特有要素。它还应规定对复杂内容的预期处理方式，例如表格、公式、化学结构、图像和嵌入对象。对于其他 DOCX 要素，包括隐藏文本、批注、页眉、页脚、内容控件、对象链接与嵌入（OLE 对象）、修订痕迹、形状、超链接、书签、水印和图表，功能规范应指明每一要素是予以保留、删除、转换、作为警告报告还是作为错误处理。对此类内容作可预期的处理，对申请人、审查程序以及数据传播和交换至关重要。

15. 这些要求还应处理以下方面：DOCX 结构和撰写做法的差异；文件章节的识别；生成符合业务需要的有效 XML 文件；更正和修改的处理；大文件和申请高峰量的处理；以及维护、指导和支持方面的安排。由于人工智能辅助撰写可能引入不受支持或未预期的结构，这一点也值得认真考虑。

16. 这些要求还应兼顾主管局对源文件和转换后输出所赋予的不同作用。因此，共同要求应确定权威副本或参考副本、需保留的文件、申请人预览和差异解决程序，以及用于审查、公布、传播和交换的格式，但不规定各局所赋予的法律地位。

拟议后续行动

关于共同要求的磋商

17. 在上述理由、范围和要求的基础上，秘书处和数字转型工作队牵头人提议与感兴趣的主管局开展结构化磋商。该磋商将使数字转型工作队能够验证拟议范围，查明共同用例，并界定统一DOCX2XML输出的共同要求，同时顾及现有的转换解决方案、各国做法和技术能力的不同水平。

18. 提议数字转型工作队处理的问题包括但不限于：

(a) 主管局为什么需要DOCX2XML转换器，输出的XML将用于哪些流程，包括申请提交、审查、PCT最低限度文献、数字化、交换、公布或传播？

(b) 哪些知识产权权利和文档类型落入范围之内：仅限专利申请正文，还是也包括已确定的商标和工业品外观设计文档？

(c) 应识别何种最低限度的专利申请正文结构和章节标题，需要针对特定主管局作出哪些调整？

(d) 必须支持哪些生成DOCX的应用程序、语言、文字、字体、图像、表格、公式、化学结构和其他对象？

(e) 应如何处理不受支持的内容、验证失败、警告、更正和修改？

(f) 哪个版本的文档是正式文件或权威文件：原始DOCX文档、经更正的DOCX文档、转换后的XML文档、呈现的PDF文档还是其他表示形式？

(g) 对文件大小、图像大小、页数、处理时间和峰值业务量适用哪些限制？

(h) 在测试、培训、申请人支持、审查员审查、维护和人工智能编拟的申请方面取得了哪些经验教训？

(i) 现有DOCX2XML转换器能否复用？如能，需要哪些部署、保密、治理安排以及针对特定主管局的调整？

(j) 哪些共同的XML输出和PDF或HTML呈现要求能够在各局继续使用不同转换器的情况下仍实现互操作性？

19. 为支持此次磋商，提议请主管局分享以下方面的信息：其对转换器的要求、转换工作流程、功能要求、实施经验、测试用例和预期输出，以及任何可复用或调整的现有工具或组件。工作队将汇总这些意见，以确定使用转换工具的领域、主管局有正当理由的特定差异，以及在不同DOCX2XML转换器之间实现可互操作的XML输出和一致呈现所需的最低限度要求。

修订DOCX2XML转换器功能规范文件

20. 在开展磋商并汇总各局所提意见的基础上，数字转型工作队将修订《DOCX2XML转换器功能规范文件》工作草案，以纳入约定的用例和共同要求，包括最低限度的可互操作输出配置。

21. 经修订的文件将提交标准委第十五届会议审议。

22. 请标准委：

(a) 注意本文件所载的内容，包括现有的DOCX2XML转换器、主管局的转换做法以及统一转换输出的理由；

(b) 如上文第5段所述，审查文件CWS/13/12附件中所载的《DOCX2XML转换器功能规范文件》工作草案并提出评论意见；

(c) 如上文第17段和第18段所述，请数字转型工作队就统一和可互操作的DOCX2XML输出的共同用例、范围和要求与感兴趣的主管局进行磋商；

(d) 如上文第19段所述，邀请其成员通过分享转换工作流程、功能要求、实施经验、测试用例和预期输出，并在可行的情况下分享关于可复用或调整的现有工具或组件的信息，对此次磋商予以支持；

(e) 如上文第20段所述，请数字转型工作队根据所收到的评论意见和磋商结果，更新《DOCX2XML转换器功能规范文件》工作草案；并

(f) 如上文第21段所述，请数字转型工作队呈交经修订的《DOCX2XML转换器功能规范文件》以及主管局之间磋商与合作的结果，供标准委第十五届会议审议。

[文件完]