

**MEETING OF INTERNATIONAL AUTHORITIES
UNDER THE PATENT COOPERATION TREATY (PCT)**

Thirty-Second Session
Geneva, October 29 to 31, 2025

STANDARDS FOR DATA FORMAT DEVELOPMENT

Document prepared by the International Bureau

SUMMARY

1. Extensible Markup Language (XML) data exchanged for the purpose of the PCT is required to be generated in formats defined in [Appendix I to Annex F](#) of the PCT Administrative Instructions, based on [WIPO Standard ST.36](#). However, national Offices are increasingly indicating an intention to build national systems using output based on [WIPO Standard ST.96](#) ("Processing of Intellectual Property Information using XML" or "XML4IP") and/or [ST.97](#) ("Processing of Intellectual Property Data using JSON¹").
2. The International Bureau has no existing capability to process ST.96 XML or ST.97 JSON and is concerned about the cost and risks of maintaining processing capabilities in multiple formats for a long transitional period if these formats become used for the PCT international phase.
3. The Meeting is invited to share national plans for development of XML-based documents and to consider whether standards and usages of such documents could be improved in ways that would encourage national Offices to migrate rapidly to new standards, or other actions that could be taken to reduce the risks of data being shared in inconsistent formats.

BACKGROUND

PCT XML DEFINITIONS

4. Appendix I to Annex F of the PCT Administrative Instructions defines the standards for XML documents to be created as part of PCT processing. Of particular significance to the work

¹ JavaScript Object Notation (JSON)

of International Authorities are the following Document type Definitions (DTDs) and the associated stylesheets:

- (a) request.dtd
- (b) demand.dtd
- (c) application-body.dtd
- (d) search-report.dtd
- (e) written-opinion.dtd
- (f) other RO and IB (to the extent containing information for the Authority), ISA and IPEA forms DTDs.

5. These DTDs are based on WIPO Standard ST.36. The International Bureau and many national Offices have an enormous investment in systems for the production and processing of data in that format, as well as collections of data in that format. The International Bureau is keen to increase the amount and quality of documents and data received in XML format in order to help automate processing and to reduce errors and costs, especially with regard to documents and data that require translation (abstracts, search reports and international preliminary reports on patentability).

RELEVANT WIPO STANDARDS

WIPO Standard ST.36

6. WIPO Standard ST.36 provides recommendations for the processing of Patent Information using XML. It takes a significantly different design approach from WIPO Standards ST.66 and ST.86, which provided the first recommendations for the processing of trademark and design information, respectively. The current version (1.2) of the main body of the Standard was approved by the Committee on WIPO Standards (CWS) in 2007, and the International Common Elements were most recently updated in 2010. WIPO Standard ST.36 has not been maintained since then and is considered by some Offices as a legacy format. The International Bureau (including the Annex F DTDs for PCT) and various national and regional Offices have nevertheless continued to use and develop their own XML implementations in accordance with this Standard, and it has recently been agreed that the XML4IP Task Force may consider changes to WIPO Standard ST.36 if required.

WIPO Standard ST.96

7. The XML4IP Task Force was established by the CWS in 2007 to prepare a proposal for a dictionary of International Common Elements (ICE), and model XML schemas and DTDs to be used across all patents, trademarks and industrial designs data. This resulted in WIPO Standard ST.96. The first version was approved by the CWS in 2012 at its second session (see document CWS/2/4 Rev.) and the Standard has been updated continually since. Many of the data structures are based on PCT Forms and requirements, in some cases by adapting the structures in ST.36 to meet the ST.96 design rules. However, only a limited number of patent document types have been developed within this Standard and no stylesheets or processing arrangements have been developed for PCT use.

WIPO Standards ST.90 and ST.97

8. [WIPO Standard ST.90](#) provides recommendations for processing and communicating intellectual property data using Web APIs (application programming interfaces). The first version of the standard, adopted in 2020 by the eighth session of the CWS, indicated in paragraph 33 that:

“APIs must support XML and JSON requests and responses. For XML, responses must be compliant with WIPO Standard using XML such as ST.96. A consistent mapping between these two formats should be used.

“ ...

“[RSG-21] JSON format MUST be assumed when no specific content type is requested.”

9. At the time, no WIPO Standard existed for use of JSON and it was recommended that the BadgerFish convention (a widely used set of rules for converting XML content to JSON) be used to define the appropriate JSON content, based on ST.96 XML.

10. In 2022, the tenth session of the CWS adopted [WIPO Standard ST.97](#) on processing of intellectual property data using JSON. The CWS noted that ST.97 would require continuous revision to track updates to ST.96 and authorized the API Task Force (subject to safeguards concerning any controversial revision proposals) to update the Standard as required, using a fast-track procedure. Since that time, the API Task Force has continued its efforts to enhance the Standard. In the course of reviewing the existing ST.97 schemas, automatically generated using the XSD2JSON schema transformation tool (see the Appendix to Annex I of WIPO Standard ST.97), a number of issues were identified. These issues resulted from fundamental differences between XML Schema (XSD) and JSON Schema, as well as from structural aspects inherited from ST.96. Some of the identified issues include:

(a) some rules concerning the cardinality of elements (such as where an application number, a registration number or both can be provided) are difficult to transform automatically; and

(b) JSON objects are generally unordered and do not easily handle “mixed content” (types of tree structures which may contain character data as well as child elements, most importantly in the context of IP documents covering runs of formatted text). There are two recommendations proposed for handling such content: either to treat it as a special case where a fragment of XML is embedded as a text item within the relevant property of the JSON structure, or to prepare a pure JSON schema handling “mixed content”.

11. The API Task Force has documented the identified issues and is currently working on improving both the transformation rules and the associated tool. The next version of WIPO Standard ST.97, expected to be published in 2026, aims to address all of the above-mentioned technical issues and adopt a structure that is more native to JSON.

WIPO Standard ST.26

12. The XML format defined in [WIPO Standard ST.26](#) for the presentation of sequence listings is based on a combination of WIPO Standard ST.96 and industry standards for sequence listing markup, but sequence listings are treated as a distinct unit within the application body. The systems for preparing and using sequence listing data are typically separate from any procedural systems and do not need to be considered as part of this issue since ST.26 is already used mostly consistently by Offices otherwise using ST.36 XML, ST.96 XML or with little or no implementation of other XML systems at all.

NATIONAL SYSTEM DEVELOPMENT

13. National Offices are increasingly developing national processing systems that generate XML based on WIPO Standard ST.96. This uses more modern element naming conventions and data structures and seeks to align data standards across all types of intellectual property, whereas WIPO Standard ST.36 was exclusively aimed at patents. There is more strict validation and consistency between instances of ST.96 XML schemas. These Offices wish to use common systems for the production of national and international Forms and data, and to avoid creating systems for PCT that output data in what they regard as a legacy format.

14. However, although WIPO Standard ST.96 includes a number of elements closely based on PCT Forms, documents and data structures, these elements have not been scrutinized in detail. No stylesheets have been developed to render PCT Forms and related documents in ST.96 to human-readable format. In addition, few Offices are known to have implemented systems based on this Standard for documents that are shared with other Offices, with the exception of Application Body and related components for the preparation and publication of the patent specification.

EXCHANGE AND USE OF DATA

15. Most documents exchanged between Offices (or between applicants and Offices) in XML format for PCT purposes do not currently make significant use of the XML data. The items referred to in items (a) to (e) of paragraph 4, above, are exceptions. These are used very heavily by the International Bureau (including national Offices using ePCT for PCT data processing as well as presumably national Offices with their own systems) for automation of application processing, as well as for reducing the cost and increasing the quality of translation of abstracts, search reports and international preliminary reports on patentability.

16. Other documents do not currently have their XML data significantly used once the information has been rendered into a form for human reading. However, it is noted that:

(a) The International Bureau is gradually bringing systems online to allow XML to be used to render the relevant information into other languages of publication. This produces “perfect” results for the parts generated directly by stylesheets and enables effective understanding of free text parts by machine translation. This service is under development specifically for the existing Annex F DTDs and stylesheets. It would need significant further development and ongoing maintenance to handle ST.96 alternatives.

(b) It is desired to encourage producers of patent management systems to support the ePCT APIs to retrieve documents and data. Such systems would be most effective if they were able easily to use both the document metadata and structured content from the document itself received in addition to noting any associated metadata provided directly by the API.

17. As such, while it is believed that the items listed in paragraph 4 are the most important for Office processing purposes, consistency is desirable in all document types to the extent possible in order to make reuse of the data easy. Most national Offices’ systems for national and PCT processing are fairly distinct so it is not generally essential that the XML formats used are identical. But in general, it is important to ensure accurate data flows between national and international systems where relevant. It is also desirable to have a sufficiently high degree of similarity between systems to be able to use as many common services as possible in creating and using the different streams of data.

18. Some national Offices appear to be considering data services based primarily on JSON APIs, at least for trademark and designs purposes. At present, there is no known use of JSON as the format for patent document exchange and only limited use of JSON for retrieving bibliographic data through APIs.

DEALING WITH MULTIPLE STANDARDS IN PARALLEL

19. The International Bureau currently maintains DTDs and stylesheets for approaching 200 Forms, as well as for other key documents, including the application body, international publication and the international application status report. The stylesheets are maintained in up to 10 languages. The XML data is used:

- (a) to generate human-readable views of the corresponding forms and documents in their original languages,
- (b) to validate inputs and help automate processing at the International Bureau and national Offices using ePCT,
- (c) to provide data for use by other Offices in the international and national phase,
- (d) to support and increase the efficiency of producing official human translations (this is particularly important in the case of international search reports, written opinions and international preliminary reports on patentability), and
- (e) to generate high quality machine translations of forms on demand, using the alternative language stylesheets to ensure a high quality and consistent result.

20. Attempting to support ST.96 XML and ST.97 JSON in parallel with the current Annex F XML would require enormous resources for initial development and ongoing maintenance even for basic operations of generating human-readable views, without significant other use of the content. Moreover, enabling reliable conversion between the formats to allow Offices to communicate freely using either Standard is almost certainly not practical for all situations. Any conversion would therefore need to be limited to use cases where this was sufficiently important to justify the effort involved, having considered alternative approaches.

21. More complete validations of the ST.96 data structures, preparation of stylesheets and the development and maintenance of systems to import and export ST.96 format data would be an enormous investment. Moreover, unless moving to ST.96 offers a clear business benefit, national Offices with existing ST.36 implementations would have no incentive to transition. The International Bureau would consequently need to maintain parallel data import and processing systems for many years at great cost, with the risks of errors and inconsistencies in working with different systems.

STRATEGY FOR USE OF STANDARDS WITHIN THE PCT

22. The International Bureau's current strategies for data exchange and use are based primarily on the XML DTDs in Appendix I to Annex F of the Administrative Instructions, which in turn is based on WIPO Standard ST.36. The ePCT service includes APIs which offer a choice of JSON or XML payloads, but these relate to small snippets of processing information or bibliographic data and does not follow WIPO Standard ST.97. Exchange of data representing complete documents is always as a file, whose contents may simply be XML, though more typically, documents created from XML are exchanged as a package showing a human-readable rendering, accompanied by the XML version.

23. If the International Bureau were to move to the receipt and processing of data according to a new standard, it is highly desirable that this is not done simply to adhere to a standard perceived to be more modern, but to deliver an actual improvement in PCT services. The new standard should also be supported by national Offices such that any transitional period where different standards are used would be manageably short.

24. To encourage national Offices to implement or improve provision of PCT data in structured data formats, the International Bureau wishes to clarify the long-term vision for data exchange formats, including:

- (a) consideration of whether the current formats effectively meet the end-to-end business requirements;
- (b) opportunities for greater alignment and possible use of common services;
- (c) the costs and risks of attempting to maintain parallel processing systems; and
- (d) incentives for Offices to update their implementations of XML data creation.

25. To this end, the International Bureau seeks information on national plans or goals that could contribute to adjusting the strategies for PCT data exchange in a useful and practical way.

26. Some issues that could be considered to make improvement of systems worthwhile for Offices, or to assist the production of consistent XML include:

- (a) Optimization of data standards and their implementation to improve data flow from a first application, into the international phase and onwards to the national phase at different Offices.
- (b) Development of clear examples of data instances to highlight best practices in generating and using data and to allow the exchange of data between different Offices to be simpler and more reliable.
- (c) Development of common systems to assist the production of data to identical standards, as has been done in different ways with WIPO Sequence software and the PCT DOCX application body converter API, in addition to services available through the ePCT browser-based interface.

27. International Authorities are invited to:

- (i) share information concerning their plans to use WIPO Standards ST.96 and ST.97 for the production and sharing of patent data of types relevant to or equivalent to aspects of PCT processing;*
- (ii) comment on the desirability of supporting WIPO Standard ST.96 or ST.97 data exchange within the PCT system for some or all purposes; and*

(iii) if ST.96 or ST.97 data exchange is supported for a particular purpose, comment on ways to make implementation in national Offices a sufficient priority in order to avoid long transitional dual processing arrangements.

[End of document]