

Patent Quality Chat: Sequence listings under ST.26 and new USPTO Rules

Kathleen Kalafus, Technical Information, Scientific and Technical Information Center (STIC)

Mary Till, Senior Legal Advisor, Office of Patent Legal Administration (OPLA)

June 15, 2022

UNITED STATES
PATENT AND TRADEMARK OFFICE



To send in questions or comments during the webinar, please email:

PatentQuality@uspto.gov

Agenda

- Background of World Intellectual Property Office (WIPO) Standard ST.26
 - What is WIPO Standard ST.26
 - Why change to WIPO Standard ST.26
 - Benefits of WIPO Standard ST.26
 - Differences between WIPO Standards ST.25 and ST.26
- Federal Register Notice
 - Synopsis of rule changes
 - Transition to ST.26 (effective date and applicability date provisions for new U.S. regulations)
- Creating an ST.26-complaint sequence listing
- Filing an ST.26 sequence listing
- ST.26 resources



What is WIPO Standard ST.26?

- WIPO Standard ST.26 is a new set of rules for the presentation of nucleotide and amino acid sequences in sequence listings using eXtensible Markup Language (XML) format.
- The Committee for WIPO Standards (CWS) established an international task force in October 2010 to draft the revised standard.
- WIPO Standard ST.26 version 1.5 was approved by the CWS on November 5, 2021.
- WIPO Standard ST.26 will go into effect worldwide on July 1, 2022.
- <https://www.wipo.int/export/sites/www/standards/en/pdf/03-26-01.pdf>



Why change to Standard ST.26?

- ST.25 format is not compatible with International Nucleotide Sequence Database Collaboration (INSDC) requirements so some data is lost when entered into public databases, such as GenBank.
- ST.25 requirements are not precise and intellectual property (IP) offices worldwide have interpreted and enforced those requirements differently.
- Certain sequence types that are common today are not covered by ST.25 rules (i.e., branched sequences and sequences that contain nucleotide analogs or D-amino acids) and therefore are not present in searchable databases.
- ST.25 sequence data is unstructured making it difficult to automate validation and data exchange.



Benefits of WIPO Standard ST.26

- Acceptance of a single sequence listing worldwide*.
- Serves as guidance to ensure agreement amongst IP offices on application of sequence listing requirements.
- Clarifies what sequence disclosures are required or permitted to be included in a sequence listing and how these sequences must be represented.
- Enhances submission quality due to the structure of a sequence listing in XML format.
- Increases automation of data validation and streamlines processing by IP offices.

* except for required translations of language-dependent free text qualifiers into the language of filing for certain IP offices, which may require replacement sequence listings.



Benefits of WIPO Standard ST.26 (cont.)

- Data compatibility with the INSDC* requirements, namely sequence annotations and organism names will be included in publically searchable databases.
- Standardization of feature keys, feature locations, qualifiers, qualifier values, and sequence variant presentation.
- Requirement for inclusion of additional types of sequences (i.e., branched sequences and sequences that contain nucleotide analogs or D-amino acids) means more sequence data will be searchable.

*INSDC includes the DNA Databank of Japan (DDBJ), the European Bioinformatics Institute (EMBL-EBI), and the National Center for Biotechnology Information (NCBI or GenBank); these databases are publicly searchable, free and utilized by EPO, JPO, KIPO, and USPTO.



WIPO ST.25 versus ST.26

General information

ST.25	ST.26
ALL priority application information may be included	ONLY the earliest priority application can be included
ALL applicant and inventor names may be included	ONLY one applicant AND optionally ONE inventor may be included
ONE invention title permitted	MULTIPLE invention titles permitted, each one in a different language
Applicant/inventor names and invention titles must be in basic Latin characters	Applicant/inventor names may be included using any valid Unicode character along with a basic Latin translation or transliteration; invention titles may be included using any valid Unicode character



WIPO ST.25 versus ST.26

Sequence information

ST.25	ST.26
ASCII text with numeric identifiers	XML with elements and attributes
Not required to include: <ul style="list-style-type: none">– Linear portions of branched sequences– Nucleotide analogs– D-amino acids	Must include: <ul style="list-style-type: none">– Linear portions of branched sequences– Nucleotide analogs– D-amino acids
Annotation of sequences for feature keys only	Annotation of sequences for feature keys and qualifiers
Permitted to include sequences: <ul style="list-style-type: none">– <10 specifically defined nucleotides– <4 specifically defined amino acids	Prohibited sequences: <ul style="list-style-type: none">– <10 specifically defined nucleotides– <4 specifically defined amino acids



WIPO ST.25 versus ST.26

Sequence data

ST.25	ST.26
Sequences only identified as DNA, RNA or PRT (protein)	Sequences identified as DNA, RNA, or AA (amino acid) along with a mandatory mol_type qualifier to further describe the molecule
Organism names: <ul style="list-style-type: none">– Latin genus/species– virus name– “artificial sequence”– “unknown”	Organism names: <ul style="list-style-type: none">– Latin genus/species– virus name– “synthetic construct”– “unidentified”
“u” represents uracil in nucleotide sequences	“t” represents uracil in RNA sequences and thymine in DNA sequences
Amino acids in protein sequences are represented by three letter abbreviations	Amino acids in protein sequences are represented by one letter abbreviations



WIPO ST.25 versus ST.26

Sequence data (cont.)

ST.25	ST.26
"n" and "Xaa" variables must have a definition provided in a feature key	Default value assumed for "n" and "X" variables with no definition
Feature location format not clearly defined	Strictly defined feature location formats permits the use of: <ul style="list-style-type: none">- "<" and ">" in all sequence types and- "^", "join", "order" "complement" in nucleotide sequences
"Mixed mode" sequences permitted (i.e., nucleotide sequence with amino acid translation shown below)	No "mixed mode"; nucleotide translations are included in "translation" qualifiers only



Federal Register Notice - final rule

- On May 20, 2022, the USPTO published a final rule to implement WIPO Standard ST.26 for applications filed on or after July 1, 2022 with the USPTO, along with conforming amendments to other rules. See [87 Fed. Reg. 30806](#)
- This final rule creates new rules, 37 CFR 1.831-1.835, based on and expressly incorporating relevant portions of WIPO Standard ST.26.



Synopsis of new 37 CFR provisions

- 37 CFR 1.831-1.835 were created for implementing the provisions of WIPO Standard ST.26. These regulations generally correspond to the content of the “Sequence Listing” regulations found currently at 37 CFR 1.821-1.825 but have been updated for implementing WIPO Standard ST.26.
- Conforming amendments to other rules have been made as well, for example, 37 CFR 1.77(b)(5) was revised to provide for an incorporation by reference statement included in the specification regarding the material in an XML file for a “Sequence Listing XML.”
- 37 CFR 1.839 is a new regulatory provision that provides express authorization for the USPTO rules to incorporate by reference certain WIPO Standard ST.26 provisions as found in 37 CFR 1.831(d) through (i), 1.832, 1.833, and 1.834(a)(2).



Rules to implement ST.26

37 CFR 1.831

- Provides the requirement for a sequence listing in XML format using the term "Sequence Listing XML."
- Defines which nucleotides and amino acids must be included and which are prohibited from being included in the "Sequence Listing XML."
- Incorporates certain definitional paragraphs from WIPO Standard ST.26.
- Provides the requirement to use a sequence identifier, i.e., "SEQ ID NO:" or the like.



Rules to implement ST.26

37 CFR 1.832

- Provides that sequences in the listing must be assigned a separate sequence identifier beginning with 1 and increasing sequentially by integers.
- Provides for the required representation and symbols to be used for identifying nucleotide and amino acid residues in a "Sequence Listing XML."
- Governs representation and symbols used for nucleotide and/or amino acid sequences that may be:
 - 1) a single continuous sequence derived from one or more non-contiguous segments of a larger sequence,
 - 2) sequences that contain regions of specifically defined residues separated by one or more regions of contiguous specified "n" or "X" residues or
 - 3) a sequence that contains regions of specifically defined residues separated by one or more gaps of an unknown or undisclosed number of residues.



Rules to implement ST.26

37 CFR 1.833

- Requires that the “Sequence Listing XML” is in a single file encoded using Unicode UTF-8.
- Requires that the “Sequence Listing XML” be valid according to the DTD (document type definition) of WIPO Standard ST.26 in Annex II.
- Specifies the components of the “Sequence Listing XML,” as set forth in WIPO Standard ST.26, by incorporating certain paragraphs of the standard as:
 - 1) an XML declaration;
 - 2) a document type declaration;
 - 3) a root element;
 - 4) a general information part and
 - 5) a sequence data part.
- Requires that language-dependent, free text elements are in the English language.



Rules to implement ST.26

37 CFR 1.834

- Provides the form and content of the “Sequence Listing XML.”
- Requires that the “Sequence Listing XML” is encoded using Unicode UTF-8 and is created by any means, like a text or nucleotide/amino acid sequence editor.
- Identifies the computer compatibility as well as file extension, naming convention parameters to be used and how and when file compression may be used.
- Provides for the number of permitted printable characters.
- Sets forth the filing options, on a single read-only optical disc or via the USPTO patent electronic filing system (for “Sequence Listing XML” submissions only Patent Center has compatibility for XML files).
- Sets forth the requirement for an incorporation by reference statement in the specification for content of the “Sequence Listing XML.”



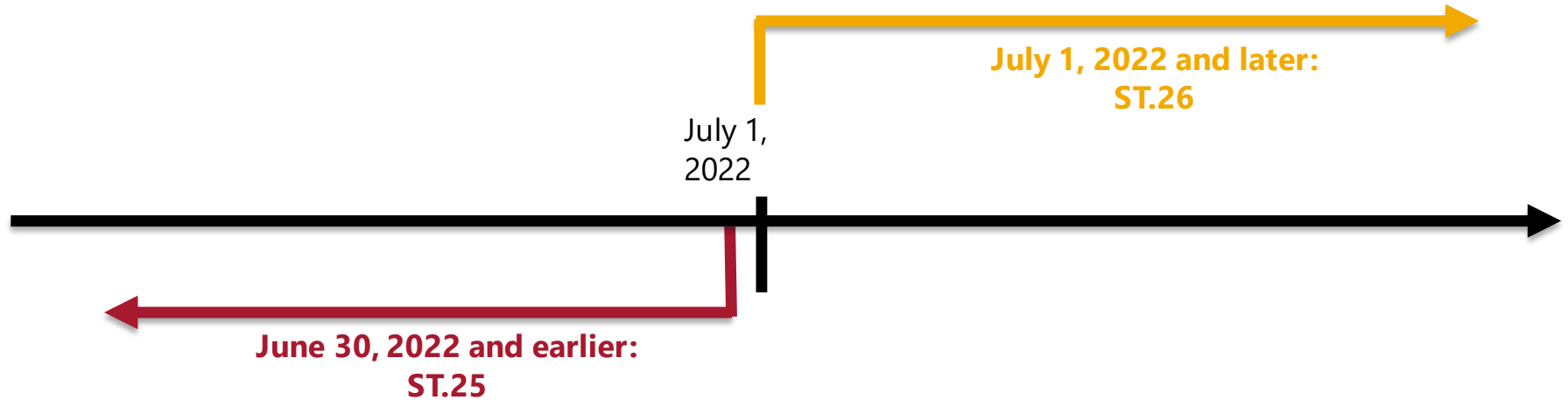
Rules to implement ST.26

37 CFR 1.835

- Provides the procedure for amending an application to add a “Sequence Listing XML” or replace a “Sequence Listing XML” along with the necessary identification of support for additions, deletions or replacements of sequence information.
- Makes a requirement for a statement of no new matter, as well as a statement showing support for the changes made.
- Provides the basis for the issuance of notices when non-compliance is raised, and the rules for responding to such notices.



Transition to WIPO Standard ST.26



Transition to WIPO Standard ST.26 (cont.)

- All applications with a filing date or international filing date **BEFORE** July 1, 2022 MUST file sequence listings in **ST.25 format**
 - For 111(a) applications, the relevant date is the “official filing date” i.e., the date all the requirements for granting a filing date are met.
 - For U.S. national phase (371) applications, the relevant date is the PCT filing date, NOT the 371(c) date.
 - Cannot choose to file in ST.26.
- All applications with a filing date or international filing date **ON OR AFTER** July 1, 2022 MUST file sequence listings in **ST.26 format**
 - An application with benefit or priority to an earlier filed application (under 35 USC 119, 120, 121 or 365) that may have contained a sequence listing in accordance with ST.25 will nonetheless be **REQUIRED** to submit a compliant sequence listing in XML file format in accordance with 37 CFR 1.831-1.835 (i.e., be in ST.26 format, there will no “grandfathering”).
 - Provisional applications are not required to file a sequence listing, however, after July 1, 2022, if an applicant chooses to submit a sequence listing in provisional application, such sequence listing must be comply with 37 CFR 1.831-1.835 (i.e., be in ST.26 format).



Creating an ST.26 sequence listing using WIPO Sequence

- **WIPO Sequence** is a desktop tool developed by WIPO to support authoring, validation, and generation of ST.26-compliant sequence listings.
- Member states requested WIPO develop this common tool for all offices and applicants.
- Use of **WIPO Sequence** simplifies ST.26 XML creation with a user-friendly interface wherein there is no need to ever directly edit an XML file.
- Available to download for free from

<https://www.wipo.int/standards/en/sequence/index.html>



WIPO Sequence functions

- Accept and store application and sequence information for multiple projects.
- Validate project data and generate a compliant XML sequence listing.
- Validate an existing XML sequence listing.
- Generate a “human readable” version of project data for easy review.
- Store custom applicant and inventor information.
- Store custom organism names.
- Import data from multiple file types – ST.25 sequence listings, ST.26 sequence listings, ST.26 projects, .raw files, multisequence format, and FASTA files.



WIPO Sequence

Projects page

WIPO Sequence 2.0.0

WIPO | Sequenc **PROJECTS** PERSONS & ORGANIZATIONS ORGANISMS HELP

PREFERENCES ENGLISH

NEW PROJECT IMPORT PROJECT IMPORT SEQUENCE LISTING VALIDATE SEQUENCE LISTING

PROJECTS

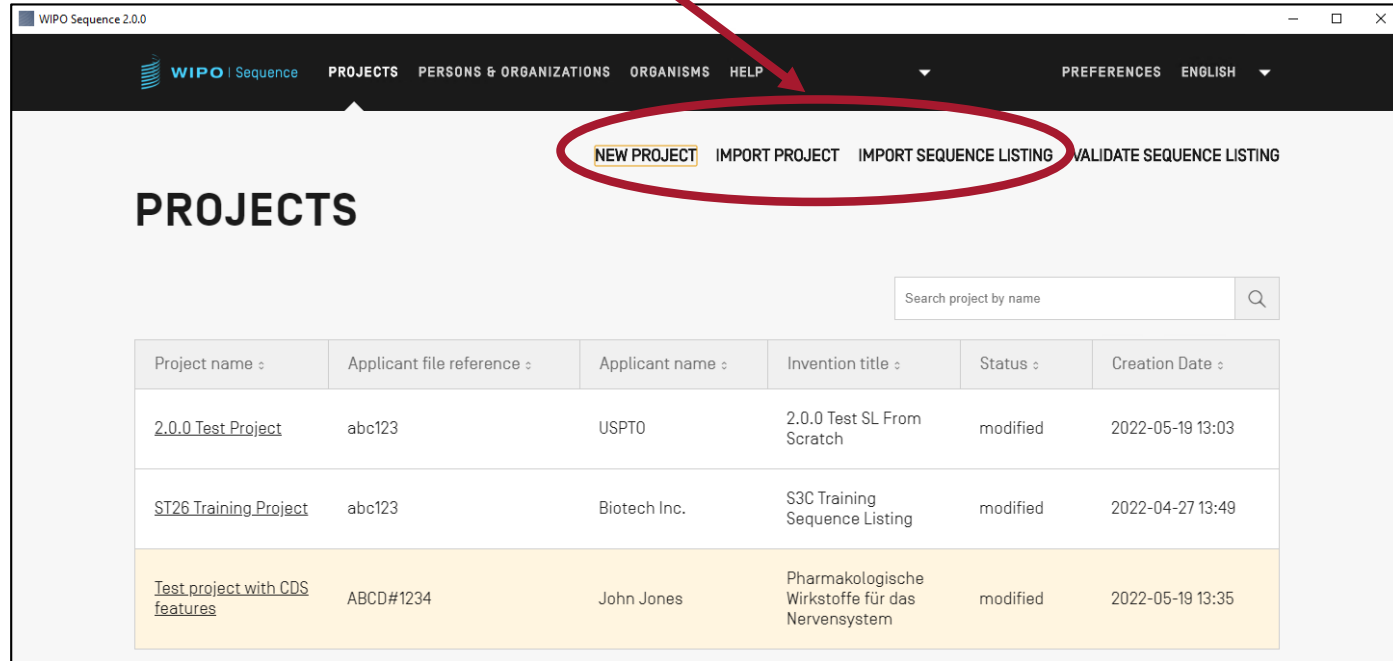
Search project by name

Project name :	Applicant file reference :	Applicant name :	Invention title :	Status :	Creation Date :
2.0.0 Test Project	abc123	USPTO	2.0.0 Test SL From Scratch	modified	2022-05-19 13:03
ST26 Training Project	abc123	Biotech Inc.	S3C Training Sequence Listing	modified	2022-04-27 13:49
Test project with CDS features	ABCD#1234	John Jones	Pharmakologische Wirkstoffe für das Nervensystem	modified	2022-05-19 13:35



WIPO Sequence

Projects page – to start a new project



WIPO Sequence 2.0.0

WIPO | Sequence PROJECTS PERSONS & ORGANIZATIONS ORGANISMS HELP PREFERENCES ENGLISH

NEW PROJECT IMPORT PROJECT IMPORT SEQUENCE LISTING VALIDATE SEQUENCE LISTING

PROJECTS

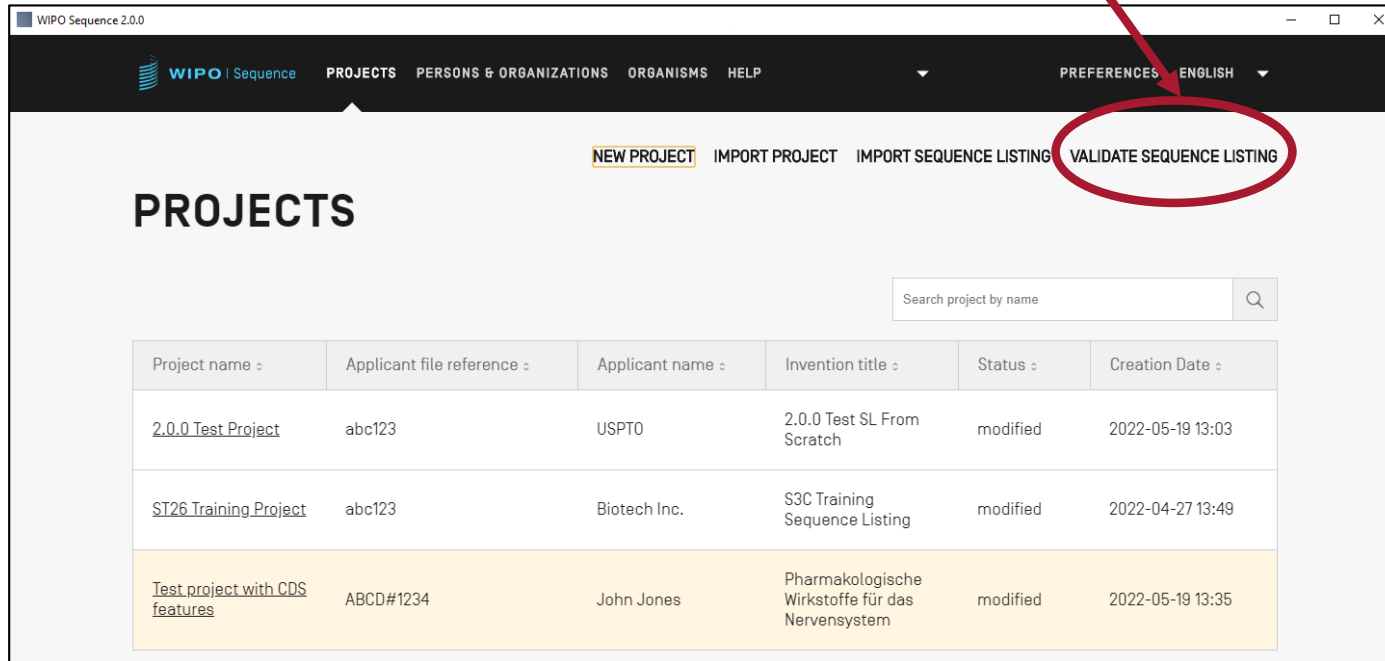
Search project by name

Project name :	Applicant file reference :	Applicant name :	Invention title :	Status :	Creation Date :
2.0.0 Test Project	abc123	USPTO	2.0.0 Test SL From Scratch	modified	2022-05-19 13:03
ST26 Training Project	abc123	Biotech Inc.	S3C Training Sequence Listing	modified	2022-04-27 13:49
Test project with CDS features	ABCD#1234	John Jones	Pharmakologische Wirkstoffe für das Nervensystem	modified	2022-05-19 13:35



WIPO Sequence

Projects page – to validate an existing XML sequence listing



The screenshot shows the WIPO Sequence 2.0.0 web application. The top navigation bar includes the WIPO logo, 'Sequence', and menu items: PROJECTS, PERSONS & ORGANIZATIONS, ORGANISMS, HELP, PREFERENCES, and ENGLISH. Below the navigation bar, there are buttons for 'NEW PROJECT', 'IMPORT PROJECT', 'IMPORT SEQUENCE LISTING', and 'VALIDATE SEQUENCE LISTING'. The 'VALIDATE SEQUENCE LISTING' button is circled in red, and a red arrow points to it from the top right. Below the buttons is a search bar labeled 'Search project by name'. The main content area displays a table of projects.

Project name :	Applicant file reference :	Applicant name :	Invention title :	Status :	Creation Date :
2.0.0 Test Project	abc123	USPTO	2.0.0 Test SL From Scratch	modified	2022-05-19 13:03
ST26 Training Project	abc123	Biotech Inc.	S3C Training Sequence Listing	modified	2022-04-27 13:49
Test project with CDS features	ABCD#1234	John Jones	Pharmakologische Wirkstoffe für das Nervensystem	modified	2022-05-19 13:35

WIPO Sequence

General information view

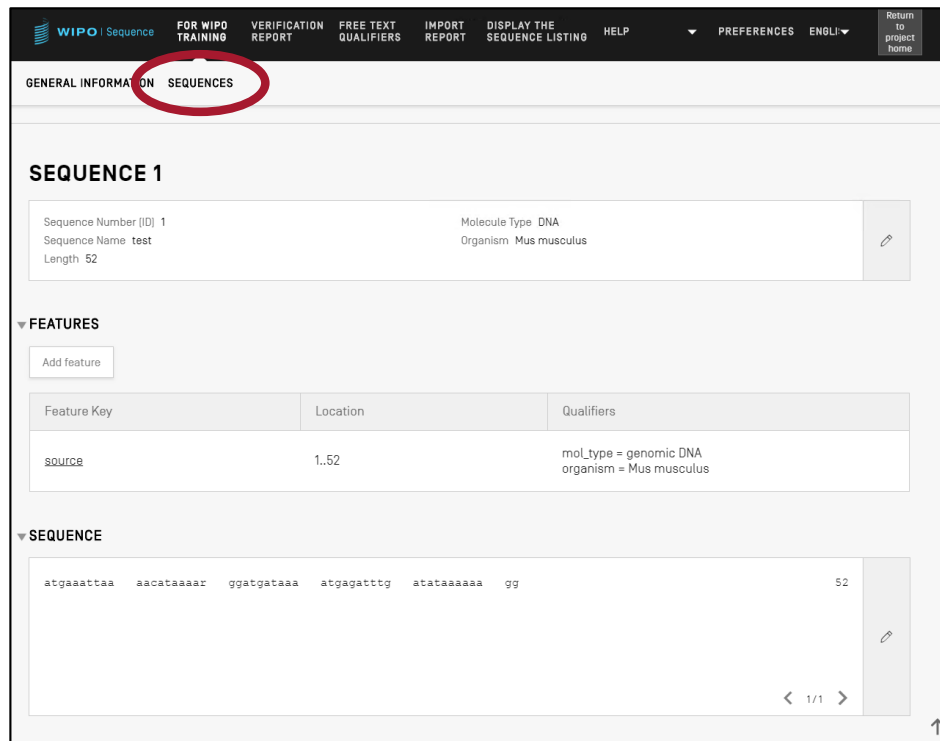
The screenshot displays the 'WIPO Sequence' web interface. The top navigation bar includes links for '2.0.0 TEST PROJECT', 'VERIFICATION REPORT', 'LANGUAGE DEPENDENT QUALIFIERS', 'IMPORT REPORT', 'DISPLAY THE SEQUENCE LISTING', 'HELP', 'PREFERENCES', 'ENGLI', and a 'Return to project home' button. The main content area is titled 'GENERAL INFORMATION SEQUENCES' (with 'GENERAL INFORMATION' circled in red) and '2.0.0 TEST PROJECT'. It features buttons for 'Print', 'Export', 'Import Another Project', 'Validate', and 'Generate Sequence Listing'. Below these are two main sections: 'GENERAL INFORMATION' and 'APPLICATION IDENTIFICATION'. The 'GENERAL INFORMATION' section contains fields for Project Name, Status, Description, Original free text language code, Creation date, File Name, Sequences, and Non English free text language code. The 'APPLICATION IDENTIFICATION' section contains fields for Application Identified, Applicant file reference, IP Office, Application number, and Filing date. Below this is a 'PRIORITY IDENTIFICATION' section with an 'Add Priority Application' button and a table of priority applications.

IP Office	Application Number	Filing date	Selected Earliest Priority Application
US - United States of America	XXXXXXX	2021-05-19	Yes



WIPO Sequence

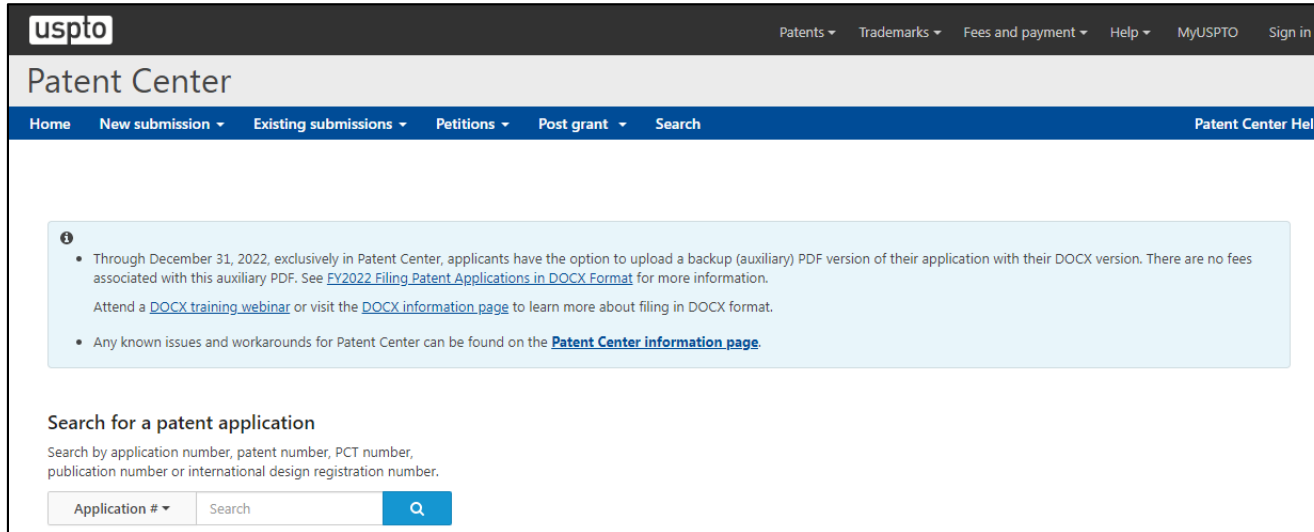
Sequence information view



The screenshot displays the WIPO Sequence application interface. The top navigation bar includes links for 'WIPO Sequence', 'FOR WIPO TRAINING', 'VERIFICATION REPORT', 'FREE TEXT QUALIFIERS', 'IMPORT REPORT', 'DISPLAY THE SEQUENCE LISTING', 'HELP', 'PREFERENCES', 'ENGLU', and 'Return to project home'. The 'SEQUENCES' tab is highlighted with a red circle. Below the navigation bar, the 'GENERAL INFORMATION' section is visible, followed by 'SEQUENCE 1'. The sequence details include: Sequence Number (ID) 1, Sequence Name test, Length 52, Molecule Type DNA, and Organism Mus musculus. The 'FEATURES' section contains an 'Add feature' button and a table with one entry: 'source' at location '1..52' with qualifiers 'mol_type = genomic DNA' and 'organism = Mus musculus'. The 'SEQUENCE' section shows the sequence text: 'atgaaattaa aacataaaaar ggatgataaa atgagatttg atataaaaaa gg' with a length of 52. Navigation arrows and a '1/1' indicator are present at the bottom of the sequence view.

Filing an ST.26 “Sequence Listing XML” at the USPTO

- ST.26 sequence listings in XML format MUST be filed electronically with the USPTO through Patent Center or on physical media.
- EFS-Web will not permit submission of ST.26 XML files.
- Patent Center is available at <https://patentcenter.uspto.gov/#/>



The screenshot shows the USPTO Patent Center website. At the top, there is a navigation bar with the USPTO logo and links for Patents, Trademarks, Fees and payment, Help, MyUSPTO, and Sign in. Below this is a secondary navigation bar with links for Home, New submission, Existing submissions, Petitions, Post grant, Search, and Patent Center Help. The main content area features a light blue notification box with an information icon and two bullet points: one about uploading a backup PDF version of applications by December 31, 2022, and another about finding known issues and workarounds on the Patent Center information page. Below the notification is a search section titled "Search for a patent application" with a subtext "Search by application number, patent number, PCT number, publication number or international design registration number." and a search input field with a dropdown menu for "Application #" and a search button.



Uploading an ST.26 “Sequence Listing XML” in Patent Center

uspto

Patent Center

Home New submission Existing submissions Petitions Post grant Search

Utility Nonprovisional

Nonprovisional Application under 35 USC 111(a)

Application data Upload documents Calculate fees Review & submit

Upload documents

Add files that you wish to accompany your patent application submission. To add more files, use the file upload interaction or drag & drop them into the area below.

Upload documents

Help with document description

Total documents uploaded: 2 [Download all documents](#)

Document Name	Size	Pages	Preview
generatedADS60680980.pdf	118 KB	5 pages	Preview
nucleotide variants no note.xml	3 KB		Preview

Back [Cancel submission](#) [Save](#) [Continue](#)

Use the “Select file(s)...” button to navigate to your sequence listing XML file

Patent Center will recognize the file format and add the document description automatically



XML filing requirements

- Sequence listings in ST.26 XML format may be filed:
 - via the USPTO Patent Center electronic filing system if the file is 100MB in size or less or
 - on physical media using read-only, optical discs (CDs or DVDs). Sequence listings submitted on discs may be compressed if the resulting compressed file is non-self extracting.
- Sequence listings in ST.26 XML format must be accompanied by an incorporation by reference statement (for 111(a) applications), which includes:
 - the name of the file,
 - the date of creation, and
 - the size of the file in bytes.
- Paper or pdf copies of the sequence listing are NOT required; any paper or pdf copies are considered part of the specification for calculating application size fees.



ST.26 resources

USPTO Resources:

- Sequence Listing Resource Center:
<https://www.uspto.gov/patents/apply/sequence-listing-resource-center>
- Sequence Listing Help Desk via SequenceHelpDesk@USPTO.GOV or 571-272-2510
- PCT Help Desk via 571-272-4300
- Patent Legal Administration Help Desk via 571-272-7701

WIPO Resources:

- Recorded training sessions:
https://www.wipo.int/meetings/en/topic.jsp?group_id=330&items=30
- WIPO Standard ST.26 Knowledge Base (coming soon):
<https://www.wipo.int/standards/en/sequence/>



LET'S CHAT about

Sequence listings under ST.26 and new USPTO Rules

with

Kathleen Kalafus, Technical Information Specialist,
Scientific and Technical Information Center (STIC)

Mary Till, Senior Legal Advisor, Office of Patent Legal
Administration (OPLA)



