

<?xml version="1.0" encoding="UTF-8"?>

<!--This entity may be identified by the PUBLIC identifier:

\*\*\*\*\*  
PUBLIC "-//WIPO//ST26 SEQUENCE LISTING 1.0//EN" "ST26SequenceListing-v1-0.dtd"  
\*\*\*\*\*

SEQUENCE LISTING XML STANDARD FOR THE PRESENTATION OF NUCLEOTIDE  
AND AMINO ACID SEQUENCE LISTINGS IN PATENT APPLICATIONS

\*\*\*\*\*  
\* CONTACTS  
\*\*\*\*\*

Stephane Nauche (European Patent Office)  
Raul Suarez Y Gonzales (European Patent Office)  
Bruce B Cox (US Patent & Trademark Office)

Date draft created: 2012-02-16

\*\*\*\*\*  
\* NOTES  
\*\*\*\*\*

The DTD for the general information part intentionally departs from WIPO  
Standards ST.36 and ST.96 to ensure management independently of those standards.

The purpose of the bibliographic information contained in the general  
information part is solely for association of the sequence listing to the  
patent application for which the sequence listing is submitted.

The sequence data part is composed of one or more sequence data elements  
containing information about one sequence. The sequence data elements include  
various feature keys and subsequent qualifiers agreed upon by the International  
Nucleotide Sequence Database Collaboration (INSDC) and the UniProt Consortium.  
The sequence data part is a subset of the complete INSDC DTD that only covers  
the requirements of WIPO ST.26.

\*\*\*\*\*  
VERSION 1.0

Initial Version of the DTD.

\*\*\*\*\*  
-->

<!-- ST26SequenceListing

\*\*\*\*\*  
\* ROOT ELEMENT  
\*\*\*\*\*

-->  
<!ELEMENT ST26SequenceListing ((ApplicantFileReference |  
    (ApplicationNumber, ApplicationFilingDate) ),  
    PriorityApplicationNumber?, EarliestPriorityDate?,  
    ApplicantName, ApplicantNameCharacters?,  
    InventorName, InventorNameCharacters?,  
    InventionTitle, InventionTitleCharacters?,  
    SequencesTotalNumber, SequenceData+) >

```

<!ATTLIST ST26SequenceListing
    dtdVersion      CDATA #IMPLIED
    fileName        CDATA #IMPLIED
    softwareName    CDATA #IMPLIED
    softwareVersion CDATA #IMPLIED
    productionDate  CDATA #IMPLIED >

<!-- ApplicantFileReference
Applicant's or agent's file reference, mandatory if application number not provided.
-->
<!ELEMENT ApplicantFileReference (#PCDATA) >

<!-- ApplicationNumber
Application information, when available.
-->
<!ELEMENT ApplicationNumber (FilingOfficeCode,ApplicationNumberText) >

<!-- FilingOfficeCode
WIPO ST.3 code
-->
<!ELEMENT FilingOfficeCode (#PCDATA) >

<!-- ApplicationNumberText
The application number without filing office code.
-->
<!ELEMENT ApplicationNumberText (#PCDATA) >

<!-- ApplicationFilingDate
The date of filing of the application.

ISO8601 long format:
yyyy-mm-dd
-->
<!ELEMENT ApplicationFilingDate (#PCDATA) >

<!-- PriorityApplicationNumber
Application Number of the earliest claimed priority.
Also contains FilingOfficeCode and ApplicationNumberText elements,
see ApplicationNumber for details.
-->
<!ELEMENT PriorityApplicationNumber (FilingOfficeCode?,ApplicationNumberText) >

<!-- EarliestPriorityDate
Where the international application claims the priority of several earlier
applications, the filing date of the earliest application whose priority is
claimed must be indicated as the priority date.

ISO8601 long format:
yyyy-mm-dd
-->
<!ELEMENT EarliestPriorityDate (#PCDATA) >

<!-- ApplicantName
The name of the first mentioned applicant in the Latin Alphabet.
-->

```

<!ELEMENT ApplicantName (#PCDATA) >

<!-- ApplicantNameCharacters

Name of the first mentioned applicant in characters other than the Latin Alphabet.

-->

<!ELEMENT ApplicantNameCharacters (#PCDATA) >

<!-- InventorName

The name of the first mentioned inventor in the Latin Alphabet.

-->

<!ELEMENT InventorName (#PCDATA) >

<!-- InventorNameCharacters

The name of the first mentioned inventor in characters other than the Latin Alphabet.

-->

<!ELEMENT InventorNameCharacters (#PCDATA) >

<!-- InventionTitle

Title of the invention in the Latin Alphabet.

Preferably two to seven words when in English or translated into English.

-->

<!ELEMENT InventionTitle (#PCDATA) >

<!-- InventionTitleCharacters

Title of the invention in characters other than the Latin Alphabet.

-->

<!ELEMENT InventionTitleCharacters (#PCDATA) >

<!-- SequencesTotalNumber

Indicates the total number of sequences in the document.

Its purpose is to be quickly accessible to some applications.

-->

<!ELEMENT SequencesTotalNumber (#PCDATA) >

<!-- SequenceData

Data for individual Sequence.

Intentionally skipped sequences see ST.26.

-->

<!ELEMENT SequenceData (INSDSeq) >

<!ATTLIST SequenceData sequenceIDNumber CDATA #REQUIRED >

<!--

\*\*\*\*\*

\* INSD Part

\*\*\*\*\*

The purpose of the INSD part of this DTD is to define a customized DTD for sequence listings to support the work of IP offices while facilitating the data exchange with the public repositories.

The INSD part is subset of the INSD DTD v1.4 and as such can only be used to generate an XML instance as it will not support the complete INSD structure.

This part based on:

The International Nucleotide Sequence Database (INSD) collaboration  
<http://www.insdc.org/page.php?page=home>

INSDSeq provides the elements of a sequence as presented in the  
GenBank/EMBL/DDBJ-style flatfile formats. Not all elements are used here.

-->

<!-- INSDSeq

All mandatory elements must be populated.

For intentionally skipped sequences please refer to ST.26.

The following elements are not necessary for ST.26 and were removed:

- INSDSeq\_locus ,
- INSDSeq\_strandedness ,
- INSDSeq\_topology ,
- INSDSeq\_update-date ,
- INSDSeq\_create-date ,
- INSDSeq\_update-release ,
- INSDSeq\_create-release ,
- INSDSeq\_definition ,
- INSDSeq\_primary-accession ,
- INSDSeq\_entry-version ,
- INSDSeq\_accession-version ,
- INSDSeq\_secondary-accessions ,
- INSDSeq\_project ,
- INSDSeq\_keywords ,
- INSDSeq\_segment ,
- INSDSeq\_source ,
- INSDSeq\_organism ,
- INSDSeq\_taxonomy ,
- INSDSeq\_comment ,
- INSDSeq\_primary ,
- INSDSeq\_source-db ,
- INSDSeq\_database-reference ,
- INSDSeq\_contig,
- INSDSeq\_references

INSDSeq\_sequence has been made mandatory

-->

<!ELEMENT INSDSeq (INSDSeq\_length,INSDSeq\_moltype, INSDSeq\_division,  
INSDSeq\_other-seqids?, INSDSeq\_feature-table?,INSDSeq\_sequence) >

<!-- INSDSeq\_length

-->

<!ELEMENT INSDSeq\_length (#PCDATA) >

<!--INSDSeq\_moltype

Admissible values: DNA, RNA, AA

-->

<!ELEMENT INSDSeq\_moltype (#PCDATA) >

<!-- INSDSeq\_division

Must be populated with the value PAT.

-->

<!ELEMENT INSDSeq\_division (#PCDATA) >

<!-- INSDSeq\_other-seqids

```

-->
<!ELEMENT INSDSeq_other-seqids (INSDSeqid*) >

<!-- INSDSeqid
Intended for the use of Patent Offices in Data Exchange only.

Format:

PAT|{country code}|{publication number}|{Sequence identification number}

-->
<!ELEMENT INSDSeqid (#PCDATA) >

<!-- INSDSeq_feature-table
Whenever the element INSDSeq_feature-table is used, it must contain at least
one feature.
-->
<!ELEMENT INSDSeq_feature-table (INSDFeature+) >

<!-- INSDFeature
The following elements were removed from the feature for the purpose of ST26
    INSDFeature_intervals ,
    INSDFeature_operator ,
    INSDFeature_partial5 ,
    INSDFeature_partial3
-->
<!ELEMENT INSDFeature (INSDFeature_key,INSDFeature_location,INSDFeature_qual?) >

<!-- INSDFeature_key
-->
<!ELEMENT INSDFeature_key (#PCDATA) >

<!-- INSDFeature_location
-->
<!ELEMENT INSDFeature_location (#PCDATA) >

<!-- INSDFeature_qual
-->
<!ELEMENT INSDFeature_qual (INSDQualifier*) >

<!-- INSDQualifier

Coding sequences and Variants see ST.26.
-->
<!ELEMENT INSDQualifier (INSDQualifier_name,INSDQualifier_value?) >

<!ELEMENT INSDQualifier_name (#PCDATA) >

<!ELEMENT INSDQualifier_value (#PCDATA) >

<!-- INSDSeq_sequence
-->
<!ELEMENT INSDSeq_sequence (#PCDATA) >

```