Use this form to tell us important information about this document, then start the text on the following page. All information you give in this form will appear in the document.

Document Information

Document Title	Created (YYYY-MM-DD)	Updated (YYYY-MM-
SRA XML Schema 1.4 Release Notes Draft D	2012-02-09	2011-04-20

Author Information

Given Name(s)	Last Name	Suffix	Degrees	Affiliation	Email

Use one row for each author. List authors in order of appearance in the document. Add rows to add more authors.

SRA XML Schema 1.4 Release Notes

Draft D – 20 Apr 2012

Status	Active
Active Date	2012-05-01
Inactive Date	
Scope	INSDC SRA

Table Of Contents

Overview

This document summarizes the proposed changes for Release 1.4 of the Sequence Read Archive (SRA) schemas governing XML metadata. This schema will be used by the SRA archive instances and has been developed under the auspices of the International Nucleotide Sequence Database Collaboration (INSDC, insdc.org).

Release 1.4 is an expansionary change over Release 1.3, which was in introduced in August 2011. These changes are being introduced with the objective of not invalidating any currently valid XML documents.

Major new features in this release are:

- Addition of new instrument platform, CAPILLARY, and new instrument models
- Enhancement of choices for library and experiment
- Introduction of an IDENTIFIERs block to track multiple active and inactive accessions and IDs.
- Support for BAM file submission through SRA Run

Notice

The features described in the SRA XML schema DO NOT constitute a statement of features and mechanisms available in the SRA. The schema changes frequently must precede actual implementation. New feature rollouts and functionality changes are made asynchronously with XML schema changes. Each SRA implementation by INSDC partners may impose additional business rules not reflected in the schema.

Related Documents

The SRA schema for this release can be obtained from this site:
http://www.ncbi.nlm.nih.gov/viewvc/v1/trunk/sra/doc/SRA_1-4d

Using_the_SRA_Identifier_Block.pdf

Revision History

Drafts D- 2012-04-20 submitted for approval by INSDC partners

Explanation of Changes

Introduction of IDENTIFIERS block to all documents

An IDENTIFIERS block of IdentifierType has been added to all documents. This block is intended to give more flexibility in how IDs are tracked. IDs include primary and secondary accessions, equivalent records in other databases, submitter primary and secondary names for records. The number of IDs is unbounded. Whether they are active or not (replaced or deprecated) can be indicated. A uuid (universally unique ID) ID type is supported, although this will not be used by INSDC SRA archives.

Changes to PlatformType

Add new platform CAPILLARY

This platform choice is intended to support handling of Traces in the SRA. The instrument model choices are:

- AB 3730xL Genetic Analyzer
- AB 3730 Genetic Analyzer
- AB 3500xL Genetic Analyzer
- AB 3500 Genetic Analyzer
- AB 3130xL Genetic Analyzer

- AB 3130 Genetic Analyzer
- AB 310 Genetic Analyzer

Add new instrument models

New instrument values have been added to Platform block :

- Remove “none” as an instrument model for Complete Genomics, PacBio
- Correct name for AB 5500, 5500xl instruments
- Add 454 FLX+
- Add AB SOLiD 3.0 plus
- Add Illumina HiSeq 2500
- Add Illumina HiScanSQ
- Add Ion Proton

Changes to LibraryDescriptorType

Added Library Strategies

- WGA (whole genome amplification) to replace some instances of RANDOM
- Added miRNA-Seq for micro RNA and other small non-coding RNA sequencing

Added Library Selections

- Added MDA (multiple displacement amplification)
- Added Padlock Probes capture strategy to be used in conjuction with Bisulfite-Seq

Make LibraryName optional

The LibraryName field is not needed except from bulk submitters who may submit multiple experiments per library.

Added options to TARGETED_LOCI

The PROBE_SET block was made optional (a technical change). In addition, the following items were added to the locus attribute:

- 18S ribosomal RNA
- RBCL

- matK
- COX1
- ITS1-5.8S-ITS2

Changes to RunType

New Filetypes Added

In order to support submission reference alignments in BAM format, the filetypes table has been augmented with these new filetypes:

- BAM header
- Reference fasta
- Complete Genomics native

Title block added

The TITLE block will allow for expansion of run to include all sequencing for the experiment or to include a certain logical fraction. The TITLE block can be used to distinguish which fraction.

Changes to ExperimentType

SPOT_DESCRIPTOR made optional in Experiment

The SPOT_DESCRIPTOR block is used by the loader to cognate the input data during load into the SRA. If the data are never transformed, then it can serve as the permanent map of the layout of the reads in the run. In order to refactor information needed for loading or interpreting the read layout, this block should be used in the RUN instead. For BAM loads it is not needed at all.

Modify GapDescriptorType

Some changes to the schema for the GapDescriptor have been implemented in order to better support Complete Genomics libraries. There are as yet no deposited experiments with GapDescriptor blocks in them, so this change will be benign.

Changes to Submission

The SUBMISSION/FILES block has been deprecated. Use the DATA_BLOCK/FILES instead.

Deprecated Fields

SRA 1.4 contains the following fields, branches, and options that should no longer be used in current submissions.

Field	Notes
/STUDY/DESCRIPTOR/CENTER_NAME	1
/STUDY/DESCRIPTOR/PROJECT_ID	2
/EXPERIMENT/@expected_number_reads	
/EXPERIMENT/@expected_number_spots	
/EXPERIMENT/DESIGN/LIBRARY_DESCRIPTOR/LIBRARY_SOURCE[NON GENOMIC]	4
/EXPERIMENT/DESIGN/SPOT_DESCRIPTOR/SPOT_DECODE_METHOD	
/EXPERIMENT/DESIGN/SPOT_DESCRIPTOR/SPOT_DECODE_SPEC/NUMBER_OF_READS_PER_SPOT	
/EXPERIMENT/LIBRARY/LIBRARY_DESCRIPTOR/LIBRARY_SOURCE[NON GENOMIC]	4
/EXPERIMENT/LIBRARY/SPOT_DESCRIPTOR/SPOT_DECODE_METHOD	
/EXPERIMENT/LIBRARY/SPOT_DESCRIPTOR/SPOT_DECODE_SPEC/NUMBER_OF_READS_PER_SPOT	
/EXPERIMENT/PLATFORM/ABI_SOLID/COLOR_MATRIX	
/EXPERIMENT/PLATFORM/ABI_SOLID/COLOR_MATRIX_CODE	
/EXPERIMENT/PLATFORM/ABI_SOLID/CYCLE_COUNT	
/EXPERIMENT/PLATFORM/ABI_SOLID/INSTRUMENT_MODEL[AB SOLiD 5500]	5
/EXPERIMENT/PLATFORM/ABI_SOLID/INSTRUMENT_MODEL[AB SOLiD 5500xl]	5
/EXPERIMENT/PLATFORM/ABI_SOLID/SEQUENCE_LENGTH	
/EXPERIMENT/PLATFORM/HELICOS/FLOW_COUNT	
/EXPERIMENT/PLATFORM/HELICOS/FLOW_SEQUENCE	
/EXPERIMENT/PLATFORM/ILLUMINA/CYCLE_COUNT	
/EXPERIMENT/PLATFORM/ILLUMINA/CYCLE_SEQUENCE	
/EXPERIMENT/PLATFORM/ILLUMINA/SEQUENCE_LENGTH	
/EXPERIMENT/PLATFORM/LS454/FLOW_COUNT	
/EXPERIMENT/PLATFORM/LS454/FLOW_SEQUENCE	
/EXPERIMENT/PLATFORM/LS454/KEY_SEQUENCE	
/EXPERIMENT/PROCESSING/BASE_CALLS	6
/EXPERIMENT/PROCESSING/BASE_CALLS/BASE_CALLER	
/EXPERIMENT/PROCESSING/BASE_CALLS/SEQUENCE_SPACE	
/EXPERIMENT/PROCESSING/QUALITY_SCORES	6
/EXPERIMENT/PROCESSING/QUALITY_SCORES/@qtype[other]	
/EXPERIMENT/PROCESSING/QUALITY_SCORES/@qtype[phred]	
/EXPERIMENT/PROCESSING/QUALITY_SCORES/MULTIPLIER	
/EXPERIMENT/PROCESSING/QUALITY_SCORES/NUMBER_OF_LEVELS	
/EXPERIMENT/PROCESSING/QUALITY_SCORES/QUALITY_SCORER	
/RUN/@instrument_model	7
/RUN/@run_file	
/RUN/@total_data_blocks	
/RUN/DATA_BLOCK/@format_code	
/RUN/DATA_BLOCK/@number_channels	
/RUN/DATA_BLOCK/@total_reads	
/RUN/DATA_BLOCK/@total_spots	
/RUN/PLATFORM/ABI_SOLID/COLOR_MATRIX	
/RUN/PLATFORM/ABI_SOLID/COLOR_MATRIX_CODE	
/RUN/PLATFORM/ABI_SOLID/CYCLE_COUNT	
/RUN/PLATFORM/ABI_SOLID/INSTRUMENT_MODEL[AB SOLiD 5500]	
/RUN/PLATFORM/ABI_SOLID/INSTRUMENT_MODEL[AB SOLiD 5500xl]	
/RUN/PLATFORM/ABI_SOLID/SEQUENCE_LENGTH	
/RUN/PLATFORM/HELICOS/FLOW_COUNT	
/RUN/PLATFORM/HELICOS/FLOW_SEQUENCE	
/RUN/PLATFORM/ILLUMINA/CYCLE_COUNT	

/RUN/PLATFORM/ILLUMINA/CYCLE_SEQUENCE	
/RUN/PLATFORM/ILLUMINA/SEQUENCE_LENGTH	
/RUN/PLATFORM/LS454/FLOW_COUNT	
/RUN/PLATFORM/LS454/FLOW_SEQUENCE	
/RUN/PLATFORM/LS454/KEY_SEQUENCE	
/RUN/SPOT_DESCRIPTOR/SPOT_DECODE_METHOD	
/RUN/SPOT_DESCRIPTOR/SPOT_DECODE_SPEC/NUMBER_OF_READS_PER_SPOT	
/SUBMISSION/ACTIONS/ACTION/HOLD/@HoldForPeriod	
/SUBMISSION/FILES	8

Notes

1. Use document header attribute @center_name
2. Use STUDY/RELATED_STUDIES/RELATED_STUDY
3. n/a
4. Use TRANSCRIPTOMIC or METAGENOMIC or METATRANSCRIPTOMIC
5. Use AB 5500 Genetic Analyzer or AB 5500xl Genetic Analyzer
6. Use PIPELINE
7. Use PLATFORM/*/INSTRUMENT_MODEL
8. Use DATA_BLOCK/FILES/FILE/filetype, DATA_BLOCK/FILES/FILE/checksum

Future Planned Revisions

The next revision, SRA 1.5, will be contracting revision (one that potentially invalidates current documents). The main changes will be to remove deprecated fields. This will involve migration of data in anticipation of future schema changes. The SRA 1.5 schema release will follow soon after SRA 1.4 is deployed.

Figures, Tables and Boxes Appendix (do not delete)

Place numbered figures, tables and boxes (referred to from the main text) below.

“In-line” figures (e.g. equations) and tables should be placed within the main text in their desired final location.

Boxes can have a single level of sections; the titles for these sections should be marked up in “Box subhead” style.