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

SRA XML Schema 1.3 Release Notes Draft A	Created (2011-01-14)	Updated (2011-04-08

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.

Please Enter Title here

Status	Active
Active Date	2011-05-01
Inactive Date	
Scope	INSDC SRA

1 Overview

This document summarizes the proposed changes for Release 1.3 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 Nuceotide Sequence Database Collaboration (INSDC).

Release 1.3 is a change over Release 1.2, which was in introduced in October 2010. While the schemas are incompatible, all data have been migrated so that documents submitted or modified before release remain valid. The goal of this release is to update choices, introduce new features, and specify a usable Analysis object usable for BAM file submissions. These changes are being introduced with the objective of not invalidating any current valid XML documents.

Major new features in this release are:

- Addition of new instrument models
- Require certain fields that have already been migrated
- Allow for modification of already-loaded analysis objects

1.1 Notice

The features and modalities described in the 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.

1.2 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_1-3

1.3 Revision History

Drafts A- … created 14 Jan 2011 and updated through …. Approved for release by INSDC partners …. Scheduled for release ….

2 Explanation of Changes

2.1 Changes to All Documents

2.1.1 Adjustment to import statements

All document importing SRA.common.xsd now point to a resolvable URL:
http://www.ncbi.nlm.nih.gov/viewvc/v1/trunk/sra/doc/SRA/SRA.common.xsd?view=co

2.2 Changes to SRA.Common.xsd

2.2.1 Add new instrument models

New instrument values have been added to Platform block :

- "Illumina HiSeq 1000" [Illumina],
- "Illumina MiSeq" [Illumina],

- AB SOLiD 5500xl SOLiD System
- AB SOLiD 5500 SOLiD System
- "PacBio RS" [Pacific Biosciences]
- "Complete Genomics" [Complete Genomics] (platform already exists)
- ION_TORRENT (new platform) and instrument models:
    - Ion Torrent PGM

2.2.1 Remove deprecated instrument models

- Solexa 1G Genome Analyzer (use Illumina choices)

2.2.2 Add GapDescriptor

A new structure called the GapDescriptor is introduced that will encode the placement of spot subsequences (tags) against a reference or assembly substrate. This structure encodes mate pair gaps and tandem read gaps. It is possible to express gaps distances in three ways: as mean/standard deviation, as min-max range, and as histogram. Orientation of the tag pairs can be described as “innie”, “outie”, “normal”, and strand-opposite “anti-normal”, following the nomenclature of the Celera Assembler.

Introduction of the GapDescriptor element was motivated by the need to describe CompleteGenomics platform sequencing. It is also intended that the GapDescriptor replace the LIBRARY_LAYOUT element in the LibraryType. The GapDescriptor can be specified at the level of Run in order to override any general settings at the level of experiment.

2.2 Changes to SRA Experiment

2.2.1 Require LIBRARY_STRATEGY

The design parameter LIBRARY_STRATEGY is now required.

2.3 Changes to Study

2.3.1 The STUDY_TYPE block is now optional

In preparation for migration to BioProjects, this block has been made optional, and will be deprecated.

2.3.1 The RELATED_STUDIES/STUDY block removed

In preparation for migration to BioProjects, this deprecated block has been removed.

2.4 Changes to Sample

2.4.1 TAXON_ID now required

The TAXON_ID field in the SAMPLE_NAME block is now required. All records already have this.

2.5 Changes to Submission

2.5.1 Submission handle removed

Submission handle attribute has been removed.

2.5.1 PROTECT action is now a complex type

This is a technical improvement requested by a major submitter.

2.6 Changes to Run

2.6.1 Replicated descriptors at Run level

- Replicated SAMPLE_DESCRIPTOR at the level of Run. If specified at Run, it will override the setting at the level of Experiment.
- Replicated GAP_DESCRIPTOR at the level of Run. If specified at Run, it will override the setting at the level of Experiment.
- Replicated SAMPLE_DESCRIPTOR at the level of Run. If specified at Run, it will override the setting at the level of Experiment.

2.7 Changes to Analysis

2.7.1 DATA_BLOCK not required for modification

The DATA_BLOCK is now required for add submissions, but no longer for modify submissions.

3 Deprecated Fields

SRA 1.3 contains the following fields, branches, and options that should no longer be used in current submissions.

SRA.common.xsd	SPOT_DECODE_METHOD	
SRA.common.xsd	NUMBER_OF_READS_PER_SPOT	

SRA.common.xsd	'454 Titanium'	use '454 GS FLX Titanium'
SRA.common.xsd	'GS 20'	use '454 GS 20'
SRA.common.xsd	'GS FLX'	use 'GS FLX'
SRA.common.xsd	'Solexa 1G Genome Analyzer'	use 'Illumina Genome Analyzer'
SRA.common.xsd	CYCLE_SEQUENCE	use SEQUENCE_LENGTH
SRA.common.xsd	CYCLE_COUNT	use SEQUENCE_LENGTH
SRA.study.xsd	CENTER_NAME	use STUDY@center_name
SRA.study.xsd	PROJECT_ID	use RELATED_STUDIES instead
SRA.study.xsd	RELATED_STUDIES/STUDY	use RELATED_STUDIES/RELATED_STUDY instead
SRA.experiment.xsd	LIBRARY_STRATGEY/BARCODE	use another library strategy
SRA.experiment.xsd	LIBRARY_SOURCE/NON GENOMIC	use METAGENOMIC or TRANSCRIPTOMIC instead
SRA.experiment.xsd	PROCESSING/BASE_CALLS	use PIPELINE instead
SRA.experiment.xsd	PROCESSING/QUALITY_SCORES	use PIPELINE instead
SRA.experimentxsd	@expected_number_spots	
SRA.experimentxsd	@expected_number_reads	
SRA.run.xsd	'_seq.txt, _prb.txt, _sig2.txt, _qhg.txt'	use 'Illumina_native' instead
SRA.run.xsd	@total_spots	
SRA.run.xsd	@total_reads	
SRA.run.xsd	@number_channels	
SRA.run.xsd	@format_code	
SRA.run.xsd	@instrument_model	use PLATFORM/INSTRUMENT_MODEL instead
SRA.run.xsd	@run_file	
SRA.run.xsd	@total_data_blocks	
SRA.submission.xsd	HoldForPeriod	
SRA.submission.xsd	@submission_id	use alias instead
SRA.submission.xsd	@handle	

4 Future Planned Revisions

The next revision is anticipated to 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.

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.