# Using the SRA Toolkit to convert .sra files into other formats

Sequence Read Archive Submissions Staff

**What is the purpose of the SRA toolkit?**

The SRA Toolkit, and the source-code SRA System Development Kit (SDK), will allow you to programmatically access data housed within SRA and convert it from the SRA format to the following formats:

- ABI SOLiD native (colorspace fasta / qual)
- fasta
- fastq
- sff
- sam (human-readable bam, aligned or unaligned)
- Illumina native

You can also use the toolkit to convert from the formats listed below into the SRA format (not required for submission, but will allow you to use the SRA Toolkit to archive or analyze your data):

- fastq or fasta/qual pairs
- AB SOLiD-SRF
- AB SOLiD-native
- Illumina SRF
- Illumina native
- sff
- Aligned bam

The SRA toolkit is available in versions compatible with Linux, Windows and Mac operating systems.

**How do I download and install the SRA Toolkit?**

The SRA Toolkit can be obtained from SRA Software page. Please note that as of version 2.3.2, only 64-bit versions of the Toolkit are being produced. The reasons for this decision are manifold, but are primarily due to the limited memory and processing capacities of 32-bit operating systems, which are insufficient for handling large SRA data files. Legacy versions of the Toolkit, including previous 32-bit versions, are available here, but please note that we are serving these files "as is" – we are happy to assist with usage (email sra@ncbi.nlm.nih.gov), but bugs (known and unknown) will not be addressed. It is strongly recommended that you configure the Toolkit prior to using it to extract data.

**How do I use the SRA Toolkit to convert data into a particular format?**

The SRA Toolkit contains a series of independent data-"dump" utilities that will allow you to convert SRA data into different file formats. As of version 2.3.2, the list of "dumpers" that are included with the toolkit include:

- fastq-dump: Converts data to fastq and fasta format.
- sam-dump: Converts data to sam (human-readable bam). Data submitted as aligned bam are output as aligned sam, while other formats are output as unaligned sam.
- sff-dump: Converts data to sff format. Note that only data submitted as sff can be converted back to this format.
- abi-dump: Converts data to csfasta/csqual format. Note that data submitted in base-space can be represented in color-space, but please be aware of the advantages / disadvantages of converting between different encodings.
- illumina-dump: Converts data to Illumina native and qseq formats.
- vdb-dump: Exports the vdb-formatted data of the .sra file.

Each of the above links will open the current documentation / help page for the respective utility, which include frequently used options and their definitions, usage examples, and common errors messages / solutions. Please send all Toolkit questions to: sra@ncbi.nlm.nih.gov

**I'm having problems using the toolkit, and the documentation doesn't cover the problem I'm having. Who do I contact for help?**

Send any toolkit questions you have to: sra@ncbi.nlm.nih.gov

Be sure to provide as much detail as possible so that we may more quickly diagnose your problem: Your operating system, Toolkit version, the command that you are attempting to execute, error messages and/or the "ncbi_error_report.xml" (if one was generated).