

WHAT IS CLAIMED IS:

1. A computer-implemented method of generating a custom file, comprising:
 - receiving a query for information associated with a desired sample;
 - determining a schema for structuring the custom file;
 - obtaining, according to the schema, a plurality of nucleic acid sequencing analysis files, wherein each one of the plurality of nucleic acid sequencing analysis files comprises nucleic acid sequence information, genetic variant information, gene expression information, or any combination thereof, of a plurality of biological samples, wherein the plurality of biological samples comprise the desired sample;
 - for each one of the plurality of nucleic acid sequencing analysis files:
 - determining, according to the schema, a plurality of data objects in the nucleic acid sequencing analysis file to be stored in the custom file;
 - determining, according to the schema, a plurality of custom data fields in the custom file to store the data objects; and
 - storing the data objects in the custom data fields;
 - generating a checksum by evaluating a cryptographic hash function for a portion of the custom file according to the schema; and
 - storing the checksum in the custom file.
2. The method of Claim 1, wherein determining a schema for structuring the custom file comprises:
 - choosing a schema from a plurality of pre-defined schemas;
 - optionally, receiving user modifications for modifying the schema; and
 - storing the user modifications and a version value associated with the schema in the custom file.
3. The method of Claim 1, wherein obtaining, according to the schema, a plurality of nucleic acid sequencing analysis files comprises:
 - searching a database for a plurality of files comprising one or more keywords specified by the schema; and
 - copying the plurality of files.

4. The method of Claim 1, wherein determining, according to the schema, a plurality of data objects in the nucleic acid sequencing analysis file to be stored in the custom file comprises:

parsing the nucleic acid sequencing analysis file;

identifying, according to the schema, the plurality of data objects to be stored;

and

extracting the plurality of data objects.

5. The method of Claim 1, wherein each of the nucleic acid sequencing analysis files further comprises at least one of: sequencing device condition, sequencing related data, analysis software information, analysis pipeline information, base calls, run quality control metrics, DNA quality control metrics, RNA quality control metrics, DNA small variants outputs, copy number variant outputs, RNA fusion outputs, DNA fusion outputs, splice variant outputs, tumor mutational burden biomarker outputs, and microsatellite instability biomarker outputs.

6. The method of Claim 5, wherein the sequencing device condition comprises sequencing parameters and/or information about errors in the sequencing device.

7. The method of Claim 1, wherein each of the nucleic acid sequencing analysis files further comprises at least one of: sample preparation related data, sample identification number, sample manifest, patient identify, tissue type, genomic area of interest, disease information, and treatment information.

8. The method of Claim 1, further comprising:

receiving a user input associated with the desired sample;

determining, according to the schema, a plurality of data objects in the user input to be stored in the custom file;

determining, according to the schema, a plurality of custom data fields in the custom file to store the data objects; and

storing the data objects in the custom data fields.

9. The method of Claim 8, wherein the user input associated with the desired sample comprises at least one of: sample preparation related data, sample identification number, sample manifest, patient identify, tissue type, genomic area of interest, disease information, and treatment information.

10. The method of Claim 1, wherein the cryptographic hash function is a MD5 hash function, a MD6 hash function, a SHA-1 hash function, a SHA-256 hash function, or a SHA-512 hash function.

11. The method of Claim 1, further comprising:
generating a verification value by adding or multiplying the checksum by a number; and
storing the verification value in the custom file.

12. The method of Claim 11, wherein the number is π .

13. The method of Claim 1, wherein the portion of the custom file according to the schema comprises a plurality of custom data fields declared by the schema as not permitting user corrections.

14. The method of Claim 13, further comprising:
generating an additional checksum by evaluating a cryptographic hash function for an additional portion of the custom file according to the schema, wherein the additional portion of the custom file comprises a plurality of custom data fields declared by the schema as permitting user corrections; and
storing the additional checksum in the custom file.

15. The method of Claim 1, further comprising:
receiving and storing a plurality of user changes to a plurality of custom data fields;
updating the checksum by re-evaluating the cryptographic hash function for the portion of the custom file according to the schema; and

storing the updated checksum in the custom file.

16. The method of Claim 1, wherein some of the nucleic acid sequencing analysis files are compressed.

17. The method of Claim 1, further comprising compressing and/or encrypting the custom file.

18. The method of Claim 1, wherein the custom file is in text-based JavaScript Object Notation (JSON) format or binary JSON format.

19. The method of Claim 1, wherein each of the nucleic acid sequencing analysis files is in one of JSON, CSV, TSV, XML, NirvanaJSON, VCF, CSVVCF, or SpliceJSON format.

20. The method of Claim 1, wherein the method is implemented in a cloud computing environment.

21. A database comprising a plurality of files, wherein each of the plurality of files is generated according to the method of Claim 1.

22. A system for generating a custom file, comprising:
a memory storing instructions to implement the method of Claim 1; and
one or more processors configured to execute the instructions.

23. A computer program product for generating a custom file, comprising a computer readable storage medium having program instructions to implement the method of Claim 1.