CLAIMS

What is claimed is:

1. A method of determining a record format for a dataset, the dataset indicated by a plurality of bytes, the method comprising, with at least one computing device:

parsing the dataset using a first record format to determine a sequence of characters represented by the plurality of bytes and determining values of one or more data fields in accordance with the first record format;

displaying at least some of the values of the one or more data fields in accordance with the first record format via a user interface;

displaying a plurality of characters from the sequence of characters via the user interface as a sequence of user interface elements, wherein each of a plurality of characters is presented as a separate user interface element;

receiving user input selecting a user interface element in the sequence of user interface elements, the selected user interface element being associated with a character of the sequence of characters; and

generating a second record format based on the received input, wherein the second record format is generated to include a data field delimited by the character associated with the selected user interface element, wherein displaying the plurality of characters from the sequence of characters comprises:

displaying a contiguous subset of the sequence of characters via the user interface as the sequence of user interface elements, wherein each character of the subset is presented in sequence as a separate user interface element.

2. The method of claim 1, further comprising:

parsing the dataset using the second record format; and

displaying results of said parsing of the dataset using the second record format via the user interface.

3. The method of claim 2, further comprising determining that the second record format does not fully parse the dataset, and wherein displaying the results of the parsing of the dataset

using the second record format via the user interface comprises displaying an alert that the second record format does not fully parse the dataset.

4. The method of claim 1, further comprising determining the first record format based at least in part on one or more heuristics to identify one or more characters as a potential delimiter.

5. The method of claim 4, wherein determining the first record format comprises identifying a character of the dataset that is not alphanumeric, a space, a quote, a period, a forward-slash or a hyphen, and generating a data field of the first record format that is delimited by the identified character.

6. The method of claim 1, wherein the first character in the sequence of the characters is a non-printable character.

7. The method of claim 1, wherein the first record format includes at least one delimited data field.

8. The method of claim 1, wherein the user input causes the at least one computing device to alter the selected user interface element’s appearance in the user interface.

9. The method of claim 1, further comprising displaying results of said parsing of the dataset using the first record format via the user interface, comprising displaying a list of records of the dataset and data field values of the records.

10. The method of claim 1, wherein the first record format includes a plurality of delimited data fields having a plurality of different delimiters.

11. A computer system comprising:
at least one processor;
at least one user interface; and

at least one computer readable storage medium comprising processor-executable instructions that, when executed, cause the at least one processor to:

parse a dataset indicated by a plurality of bytes using a first record format to determine a sequence of characters represented by the plurality of bytes and determining values of one or more data fields in accordance with the first record format;

display, via the at least one user interface, at least some of the values of the one or more data fields of the first record format;

display, via the at least one user interface, a plurality of characters from the sequence of characters as a sequence of user interface elements, wherein each of a plurality of characters is presented as a separate user interface element;

receive, via the at least one user interface, user input selecting a user interface element in the sequence of user interface elements, the selected user interface element being associated with a character of the sequence of characters; and

generate a second record format based on the received input, wherein the second record format is generated to include a data field delimited by the character associated with the selected user interface element, wherein displaying the plurality of characters from the sequence of characters comprises:

displaying a contiguous subset of the sequence of characters via the user interface as the sequence of user interface elements, wherein each character of the subset is presented in sequence as a separate user interface element.

12. The computer system of claim 11, wherein the processor-executable instructions further cause the at least one processor to:

parse the dataset using the second record format; and

display, via the at least one user interface, results of said parsing of the dataset using the second record format via the user interface.

13. The computer system of claim 12, wherein the processor-executable instructions further cause the at least one processor to determine that the second record format does not fully parse the dataset, and wherein displaying the results of the parsing of the dataset using the second

record format via the user interface comprises displaying an alert that the second record format does not fully parse the dataset.

14. The computer system of claim 11, wherein the processor-executable instructions further cause the at least one processor to determine the first record format based at least in part on one or more heuristics to identify one or more characters as a potential delimiter.

15. The computer system of claim 14, wherein determining the first record format comprises identifying a character of the dataset that is not alphanumeric, a space, a quote, a period, a forward-slash or a hyphen, and generating a data field of the first record format that is delimited by the identified character.

16. The computer system of claim 11, further comprising determining the first record format, wherein determining the first record format comprises identifying a delimiter.

17. The computer system of claim 11, wherein the user input causes the at least one processor to alter the selected user interface element’s appearance in the user interface.

18. The computer system of claim 11, wherein displaying the results of said parsing of the dataset using the first record format via the at least one user interface comprises displaying a list of records of the dataset and data field values of the records.

19. The computer system of claim 11, wherein the first record format includes a plurality of delimited data fields having a plurality of different delimiters.

20. A computer system comprising:

at least one processor;

means for parsing a dataset comprising a plurality of bytes using a first record format to determine a sequence of characters represented by the plurality of bytes and determining values of one or more data fields in accordance with the first record format;

means for displaying at least some of the values of the one or more data fields of the first record format via at least one user interface;

means for displaying a portion of the sequence of characters via the at least one user interface as a sequence of user interface elements, wherein each character of the portion of the sequence of characters is presented in sequence as a separate user interface element;

means for receiving user input associated with a first user interface element of the sequence of user interface elements, the first user interface element associated with a first character of the sequence of characters; and

means for generating a second record format based on the received input, wherein the second record format is generated to include a data field delimited by the first character, wherein the means for displaying the portion of the sequence of characters further display a contiguous subset of the sequence of characters via the user interface as the sequence of user interface elements, wherein each character of the subset is presented in sequence as a separate user interface element.

21. The method of claim 1, wherein the plurality of characters is presented in an order matching that order in which the characters appear in the dataset.

22. The method of claim 1, wherein the selected user interface element represents a selection of the character of the sequence of characters as a delimiter, the second record format is generated to include the data field delimited by the delimiter, and the method further comprising:

displaying at least some of the values of the one or more data fields in accordance with the second record format via the user interface; and

receiving feedback via the user interface representing that the second record format corresponds to a canonical record format of the dataset

23. The method of claim 1, further comprising:

changing a state of the user interface element in response to the selecting of the user interface element; and

re-parsing the dataset using the second record format after the changing of the state of the user interface element.