WHAT IS CLAIMED IS:

1. A method for generating, from a label of a data field, one or more rules for processing data values of the data field, the method being implemented by a data processing system and including:

generating a label index that associates a label with a set of one or more fields in a data record, wherein the label identifies a type of information expected in each field of the set of one or more fields, the generating including:

performing a plurality of different classification tests on a field of the data record; and

comparing results across the classification tests using corroboration logic to determine the label of the field;

accessing a data dictionary that associates the type of information indicated by the label with a set of attribute values representing requirements for values of the one or more fields associated with the label, the requirements including logical or syntactical characteristics of the values for the one or more fields; and

for a field of a particular data record:

identifying, by accessing the label index, a particular label associated with the field of the particular data record;

retrieving, from the data dictionary, an attribute value for the particular label, the attribute value specifying a particular requirement for the field; and

based on the attribute value, applying to the field one or more rules for processing data values of the field.

2. The method of claim 1, wherein at least one of the classification tests specifies logic for fuzzy matching of field names from a dictionary of terms, wherein the logic for fuzzy matching is configured to find a match between a name of the field and a term in the dictionary of terms when an exact match cannot be found.

3. The method of claim 1 or 2, wherein at least one of the classification tests specifies logic to determine whether any one term or word of a set of predetermined terms or words is included within the field.

4. The method of any one of the preceding claims, wherein the classification tests includes at least one pattern test that specifies pattern matching logic for recognizing one or more patterns in data of the field, preferably wherein the at least one pattern test includes multiple different pattern tests and the method includes selecting, based on the data in the field, a pattern test to be performed on the data in the field.

5. The method of any one of the preceding claims, wherein at least one of the classification tests specifies logic for comparing a known list of data against the data of the data field to determine if the data of the field correlates to the data at the list, wherein the known list is representing likely values that should appear in the data of the field.

6. The method of any one of the preceding claims, wherein the method is for selecting test data to cause execution of a processing rule during testing of a data processing application and the applying of the one or more rules includes:

generating a subsetting rule that, when executed, specifies whether a value of the field includes the attribute value;

selecting a subset of fields from the particular data record according to the subsetting rule, wherein each selected field includes values that have the attribute value; and

providing the selected subset to a data processing application for testing of the data processing application.

7. The method of any one of claims 1 to 5, wherein the method is for masking data of a data processing application and the applying of the one or more rules includes:

determining, based on the attribute value, that the field of the particular data record represents sensitive data; and

in response to the determining, performing a data masking function to transform values including the sensitive data of the field into masked values.

8. The method of any one of claims 1 to 5, wherein the method is for determining a schema of a data record and the applying of the one or more rules includes:

determining that the attribute value represents a schema feature for values included in the first field;

determining, based on the schema feature, that the first field having the label comprises key values being referenced by values of a second field; and

in response to the determining, updating schema data describing the particular data record to reference that the first field comprises the key values being referenced by the values of the second field.

9. The method of any one of claims 1 to 5, wherein the method is for determining a data quality rule for values in a field of a data record in a set of data records based on a label associated with the field, the label indicating characteristics of the values of the field, the applying of the one or more rules including:

generating a data quality rule that, when executed, is configured to:

validate whether a value of the field satisfies the particular requirement represented by the attribute value, and

generate output data indicating whether the particular requirement is satisfied.

10. The method of claim 9, wherein the data quality rule is indicative of one or more of:

an allowable deviation for the value in the field from the requirement,

one or more allowable values in the field, and

one or more prohibited values in the field.

11. The method of claim 9 or 10, wherein the field is a first field, wherein the data record comprises a second field, wherein the data quality rule is a first data quality rule, and wherein the method comprises:

determining, based on the attribute value for the particular label, that a relationship exists between the first field and the second field; and

generating a second data quality rule for the second field based on the relationship between the second field and the first field.

12. The method of any one of claims 9 to 11, comprising:

determining that the attribute value associated with the particular label indicates that the field comprises primary key values for the particular data record; and

configuring the data quality rule to require that the primary key values are each unique in the field.

13. The method of any one of claims 9 to 12, the attribute value for the particular label represents at least one of an average for the values in the field, a maximum length for the values, a minimum length for the values, a data type for the values, and a format for the values.

14. The method of any one of claims 9 to 13, wherein the particular data record is a first data record, and wherein the method comprises applying the data quality rule to another field associated with the particular label in a second data record that is different than the first data record.

15. The method of any one of claims 9 to 14, further comprising:

generating data for requesting approval of the data quality rule; and

approving the data quality rule in response to obtaining approval data indicative of approval of the data quality rule.

16. The method of any one of claims 9 to 15, wherein generating the data quality rule includes:

determining a historical trend for values in the field; and

generating a requirement based on the historical trend for the values in the field.

17. The method of any one of claims 9 to 16, wherein generating the data quality rule for the field includes:

identifying a historical average of a value in the field; and

generating a requirement based on the historical average of the value in the field.

18. The method of claim 9, comprising:

profiling, by the data processing system, values stored in a field included in data records of one or more datasets;

applying, by the data processing system, one or more classifiers to the profiled values;

identifying, based on applying the one or more classifiers, one or more attributes indicative of a logical or syntactical characteristic for the values of the field, with each of the one or more attributes having a respective confidence level that is based on an output of each of the one or more classifiers;

associating, with the field, one or more of the identified attributes for which the confidence level satisfies a threshold level;

determining, based on the one or more attributes associated with the field, one or more constraints for values included in the field; and

based on the determined one or more constraints, determining a data quality rule for the field of the dataset, the data quality rule indicating a relationship between values of the field and the one or more constraints.

19. A data processing system, the data processing system including:

at least one processing device; and

at least one memory in communication with the at least one processing device, the at least one memory storing instructions that, when executed by the at least one processing device, cause the at least one processing device to perform the method of any one of the preceding claims.

20. One or more non-transitory computer readable media storing instructions, the instructions being configured, when executed by at least one processing device, to cause the at least one processing device to perform the method of any one of the claims 1 to 18.