What is claimed is:

1. A method for determining a data quality rule for a field of a dataset, the method including:

analyzing one or more particular instances of the dataset, including determining a reference profile for the field of the one or more particular instances of the datasets; and

based on the determined reference profile, determining a data quality rule for the field of the dataset, in which the data quality rule for the field of the dataset is indicative of one or more of:

(i) an allowable deviation between the reference profile for the field and a profile for the field of one or more data records of an instance of the dataset,

(ii) one or more allowable values for a data element in the field of a data record of an instance of the dataset, and

(iii) one or more prohibited values for a data element in the field of a data record of an instance of the dataset.

2. The method of claim 1, in which analyzing the one or more particular instances of the dataset includes analyzing one or more historical instances of the dataset.

3. The method of claim 1, in which determining the reference profile for the field includes determining an historical average profile for the field.

4. The method of claim 3, in which analyzing the one or more particular instances of the dataset includes analyzing multiple instances of the dataset until a variation in the historical average profile for the data element in the field varies less than a threshold amount.

5. The method of claim 1, in which determining the reference profile for the field includes identifying an historical average value for a data element in the field.

6. The method of claim 1, in which determining the reference profile for the field includes identifying a standard deviation of values for a data element in the field.

7. The method of claim 1, in which determining the reference profile for the field includes identifying a number of distinct values for a data element in the field.

8. The method of claim 1, in which analyzing the one or more particular instances of the dataset includes analyzing a predefined number of instances of the dataset.

9. The method of claim 1, including analyzing the one or more particular instances of the dataset using machine learning techniques.

10. The method of claim 1, including applying the data quality rule to the one or more data records of a first instance of the dataset.

11. The method of claim 10, in which applying the data quality rule to the first instance of the dataset includes determining that the first instance of the dataset has an error or possible error.

12. The method of claim 11, in which determining that the first instance of the dataset has an error or possible error includes:

determining a deviation between the reference profile for the field and a profile for the field of one or more records of the first instance of the dataset and; and

determining that the deviation between the reference profile and the profile exceeds the allowable deviation.

13. The method of claim 11, in which determining that the first instance of the dataset has an error or possible error includes identifying a data element in a field of one or more data records of the first instance of the dataset that does not satisfy one or more of the allowable values or prohibited values.

14. The method of claim 11, including based on data lineage information relating to the first instance of the dataset, identifying a downstream dataset that depends on the first instance of the dataset.

15. A non-transitory computer readable medium storing instructions for causing a computing system to determine a data quality rule for a field of a dataset, the instructions including instructions to cause the computer system to:

analyze one or more particular instances of a dataset, including determining a reference profile for a field of the one or more particular instances of the datasets; and

based on the determined reference profile, determine a data quality rule for the field of the dataset, in which the data quality rule for the field of the dataset is indicative of one or more of:

(i) an allowable deviation between the reference profile for the field and a profile for the field of one or more data records of an instance of the dataset,

(ii) one or more allowable values for a data element in the field of a data record of an instance of the dataset, and

(iii) one or more prohibited values for a data element in the field of a data record of an instance of the dataset.

16. The computer readable medium of claim 15, in which analyzing the one or more particular instances of the dataset includes analyzing one or more historical instances of the dataset.

17. The computer readable medium of claim 16, in which determining the reference profile for the field includes determining an historical average profile for the field.

18. The computer readable medium of claim 17, in which analyzing the one or more particular instances of the dataset includes analyzing multiple particular instances of the dataset until a variation in the historical average profile for the data element in the field varies less than a threshold amount.

19. The computer readable medium of claim 15, in which determining the reference profile for the field includes identifying an historical average value for a data element in the field.

20. The computer readable medium of claim 15, in which determining the reference profile for the field includes identifying a standard deviation of values for a data element in the field.

21. The computer readable medium of claim 15, in which determining the reference profile for the field includes identifying a number of distinct values for a data element in the field.

22. The computer readable medium of claim 15, in which analyzing the one or more particular instances of the dataset includes analyzing a predefined number of instances of the dataset.

23. The computer readable medium of claim 15, the instructions causing the computing system to analyze the one or more particular instances of the dataset using machine learning techniques.

24. The computer readable medium of claim 15, the instructions causing the computing system to apply the data quality rule to the one or more data records of a first instance of the dataset.

25. The computer readable medium of claim 24, in which applying the data quality rule to the first instance of the dataset includes determining that the first instance of the dataset has an error or possible error.

26. The computer readable medium of claim 24, in which determining that the first instance of the dataset has an error or possible error includes:

determining a deviation between the reference profile for the field and a profile for the field of one or more records of the first instance of the dataset and; and

determining that the deviation between the reference profile and the profile exceeds the allowable deviation.

27. The computer readable medium of claim 24, in which determining that the first instance of the dataset has an error or possible error includes identifying a data element in a field of one or more data records of the first instance of the dataset that does not satisfy one or more of the allowable values or prohibited values.

28. The computer readable medium of claim 24, the instructions causing the computing system to, based on data lineage information relating to the first instance of the dataset, identify a downstream dataset that depends on the first instance of the dataset.

29. A computing system for determining a data quality rule for a field of a dataset, the computing system including:

one or more processors coupled to a memory, the one or more processors and memory configured to:

analyze one or more particular instances of a dataset, including determining a reference profile for a field of the one or more particular instances of the datasets; and

based on the determined reference profile, determine a data quality rule for the field of the dataset, in which the data quality rule for the field of the dataset is indicative of one or more of:

(i) an allowable deviation between the reference profile for the field and a profile for the field of one or more data records of an instance of the dataset,

(ii) one or more allowable values for a data element in the field of a data record of an instance of the dataset, and

(iii) one or more prohibited values for a data element in the field of a data record of an instance of the dataset.

30. The computing system of claim 29, in which analyzing the one or more particular instances of the dataset includes analyzing one or more historical instances of the dataset.

31. The computing system of claim 29, in which determining the reference profile for the field includes determining an historical average profile for the field.

32. The computing system of claim 31, in which analyzing the one or more particular instances of the dataset includes analyzing multiple instances of the dataset until a variation in the historical average profile for the data element in the field varies less than a threshold amount.

33. The computing system of claim 29, in which determining the reference profile for the field includes identifying an historical average value for a data element in the field.

34. The computing system of claim 29, in which determining the reference profile for the field includes identifying a standard deviation of values for a data element in the field.

35. The computing system of claim 29, in which determining the reference profile for the field includes identifying a number of distinct values for a data element in the field.

36. The computing system of claim 29, in which analyzing the one or more instances of the dataset includes analyzing a predefined number of instances of the dataset.

37. The computing system of claim 29, the one or more processors and memory being configured to analyze the one or more particular instances of the dataset using machine learning techniques.

38. The computing system of claim 29, the one or more processors and memory being configured to apply the data quality rule to the one or more data records of a first instance of the dataset.

39. The computing system of claim 38, in which applying the data quality rule to the first instance of the dataset includes determining that the first instance of the dataset has an error or possible error.

40. The computing system of claim 38, in which determining that the first instance of the dataset has an error or possible error includes:

determining a deviation between the reference profile for the field and a profile for the field of one or more records of the first instance of the dataset and; and

determining that the deviation between the reference profile and the profile exceeds the allowable deviation.

41. The computing system of claim 38, in which determining that the first instance of the dataset has an error or possible error includes identifying a data element in a field of one or more data records of the first instance of the dataset that does not satisfy one or more of the allowable values or prohibited values.

42. The computing system of claim 38, the one or more processors and memory being configured to, based on data lineage information relating to the first instance of the dataset, identify a downstream dataset that depends on the first instance of the dataset.

43. A computing system for determining a data quality rule for a field of a dataset, the computing system including:

means for analyzing one or more particular instances of a dataset, including determining a reference profile for a field of the one or more particular instances of the datasets; and

means for, based on the determined reference profile, determining a data quality rule for the field of the dataset, in which the data quality rule for the field of the dataset is indicative of one or more of:

(i) an allowable deviation between the reference profile for the field and a profile for the field of one or more data records of an instance of the dataset,

(ii) one or more allowable values for a data element in the field of a data record of an instance of the dataset, and

(iii) one or more prohibited values for a data element in the field of a data record of an instance of the dataset.

What is claimed is:

1. A method for determining a data quality rule for a particular field of a dataset, the dataset including data records having data elements for each of one or more fields, the method including:

analyzing data records in one or more particular instances of the dataset, including analyzing data elements for the particular field for the analyzed data records to determine a reference profile for the particular field for the analyzed data records in the one or more particular instances of the dataset; and

based on the determined reference profile, determining a data quality rule for the particular field of the dataset, in which the data quality rule for the particular field of the dataset is indicative of one or more of:

(i) an allowable deviation between the reference profile for the particular field and a profile for the particular field of an instance of the dataset,

(ii) an allowable value for a data element for the particular field of a data record of an instance of the dataset, or

(iii) a prohibited value for a data element for the particular field of a data record of an instance of the dataset.

2. The method of claim 1, in which analyzing data records in the one or more particular instances of the dataset includes analyzing data elements for the particular field for data records of one or more historical instances of the dataset.

3. The method of claim 1, in which determining the reference profile for the particular field includes determining an historical average profile for the particular field.

4. The method of claim 3, in which analyzing data records in the one or more particular instances of the dataset includes analyzing data elements for the particular field for data records of multiple particular instances of the dataset until a variation in the historical average profile for the data element in the particular field varies less than a threshold amount.

5. The method of claim 1, in which determining the reference profile for the particular field includes identifying an historical average value for a data element for the particular field.

6. The method of claim 1, in which determining the reference profile for the particular field includes identifying a standard deviation of values for a data element for the particular field.

7. The method of claim 1, in which determining the reference profile for the particular field includes identifying a number of distinct values for a data element for the particular field.

8. The method of claim 1, in which analyzing data records in the one or more particular instances of the dataset includes analyzing data records in a predefined number of particular instances of the dataset.

9. The method of claim 1, including analyzing data records in the one or more particular instances of the dataset using machine learning techniques.

10. The method of claim 1, including applying the data quality rule to data records of a second particular instance of the dataset.

11. The method of claim 10, in which applying the data quality rule to data records of the second particular instance of the dataset includes determining that the second particular instance of the dataset has an error or possible error.

12. The method of claim 11, in which determining that the second particular instance of the dataset has an error or possible error includes:

determining a deviation between the reference profile for the particular field and a profile for the particular field of the second particular instance of the dataset; and

determining that the deviation between the reference profile and the profile exceeds the allowable deviation.

13. The method of claim 11, in which determining that the second particular instance of the dataset has an error or possible error includes identifying a data element for the particular field for one or more data records of the second particular instance of the dataset that does not satisfy either the allowable value or the prohibited value.

14. A non-transitory computer readable medium storing instructions for causing a computing system to determine a data quality rule for a particular field of a dataset, the dataset including data records having data elements for each of one or more fields, the instructions causing the computing system to perform the method of any one of claims 1-13.

15. A computing system for determining a data quality rule for a particular field of a dataset, the dataset including data records having data elements for each of one or more fields, the computing system including:

one or more processors coupled to a memory, the one or more processors and memory configured to perform the method of any one of claims 1-13.