[0038] In some embodiments, the electronic processor 300 may use one or more libraries or web services to infer additional dimensions for a data graph. For example, in some embodiments, the electronic processor 300 uses a different library or web service to infer whether each of a plurality of potential additional dimensions are applicable for a particular node or data graph. For example, the electronic processor 300 may be configured to execute or access a first library or web service to determine whether weather information could be added to one or more nodes of a data graph and may execute or access a second library or web service to determine whether traffic information could be added to one or more nodes of a data graph. Each such library or web service may be configured to infer, from data included in data graph, such as the format of data, a type of data, or a name of data (feature name or identifier) whether a particular dimension is relevant for the data graph.

[0039] It should be understood that the additional dimension identified by the electronic processor 300 may include the addition of an additional feature or feature value to an existing node of the data graph, the addition of a new node to the data graph, the addition of a new connection between new or existing nodes of the data graph, or the addition of a plurality of new nodes and connections (such as an existing data graph) to the data graph. For example, through the expansion process, the electronic processor 300 may be configured to identify an existing provenance chain (previously created by the electronic processor 300 or a separate system or application) and add the existing provenance chain as a new connection at block 405.

[0040] In some embodiments, the additional dimension identified by the electronic processor 300 may also include an identified duplication or discrepancy, such as when two nodes are associated with two different components or stages with different identifiers but, based on similarities between the identifiers, are likely referring to the same component or stage. For example, when data records associated with an entity are provided by two different organizations, the organizations may create their own unique identifiers or may change the format of an identifier. Thus, these records may not be accurately represented in the original data graph because the identifiers are not identical. To identify these similarities, the electronic processor 300 may be configured to apply a similarity algorithm to the original data graph (or the databases used to generate the original data graph) to detect similarities. For example, when the entity is a product, the electronic processor 300 may be configured to apply a similarity algorithm to databases provided by two organizations involved in the manufacturing or distribution of the product. Applying the