

How can taxonomic changes be managed in a scientific collection without reorganising specimen records?

Short answer

A scientific collection database should be designed so that the specimen record stays stable even when taxonomy changes. The safest approach is to separate the specimen or occurrence record, the identification record, and the taxon or name reference and connect them through stable identifiers rather than rewriting the specimen rows whenever names change.

Detailed answer

A robust scientific collection database should treat **the specimen as the stable core object** and treat taxonomic opinions as linked, updateable information rather than as something embedded permanently in the specimen record. In practice, that means the database should separate at least three things:

1. the **specimen or occurrence entity**,
2. the **identification history**, and
3. the **taxonomic reference**

This information matters because Darwin Core distinguishes an **occurrence** from an **identification** and gives each its own identifier terms, including `occurrenceID` for the occurrence and `identificationID` for the identification.

Each specimen should therefore have a **persistent identifier** that does not change when the scientific name changes. Darwin Core explicitly recommends that `occurrenceID` be a persistent, globally unique identifier, and CETAF likewise promotes HTTP-URI-based stable identifiers for digitised biological collections so records can remain consistently referable over time.

The next layer is the **identification table**. Instead of storing only one taxon name directly in the specimen row, each determination should be stored as a separate linked record with its metadata, such as who made the identification, when it was made, according to which source, and whether it is current or superseded. Darwin Core `identificationID` is designed precisely for that identification-level body of information. This structure preserves the full determination history instead of overwriting older identifications whenever taxonomy is revised.

The taxonomic name itself should also be handled carefully. A database should avoid treating a literal name string as the only anchor. Darwin Core provides terms such as those `acceptedNameUsageID` for the currently accepted taxon concept and `taxonID` for the taxon information record. This supports a model where a specimen is linked to an identification, which in turn links to a managed taxon reference, checklist entry, or taxonomic backbone. When taxonomy changes, you usually update the linked taxon relationship or add a new identification record instead of reorganising the specimen records.

This means the database design should follow a normalised relational model rather than flat spreadsheet logic. A practical structure would include:

- a **Specimen or Occurrence table** for the physical object and its stable identifier
- an **Identification table** for one-to-many determinations over time
- a **Taxon or Name table** for accepted names, synonyms, and checklist references
- optional **Reference, Agent, and Event tables** for literature, identifiers, collectors, determiners, and dates

That design reduces repeated editing, preserves provenance, and makes future taxonomic updates far less disruptive. It also aligns better with biodiversity informatics practice than storing the current name as the main organising principle for the whole collection database.

References

1. [Darwin Core](#)
2. [Darwin Core List of Terms](#)
3. [CETAF Stable Identifiers](#)
4. [ABCD - Access to Biological Collection Data](#)

