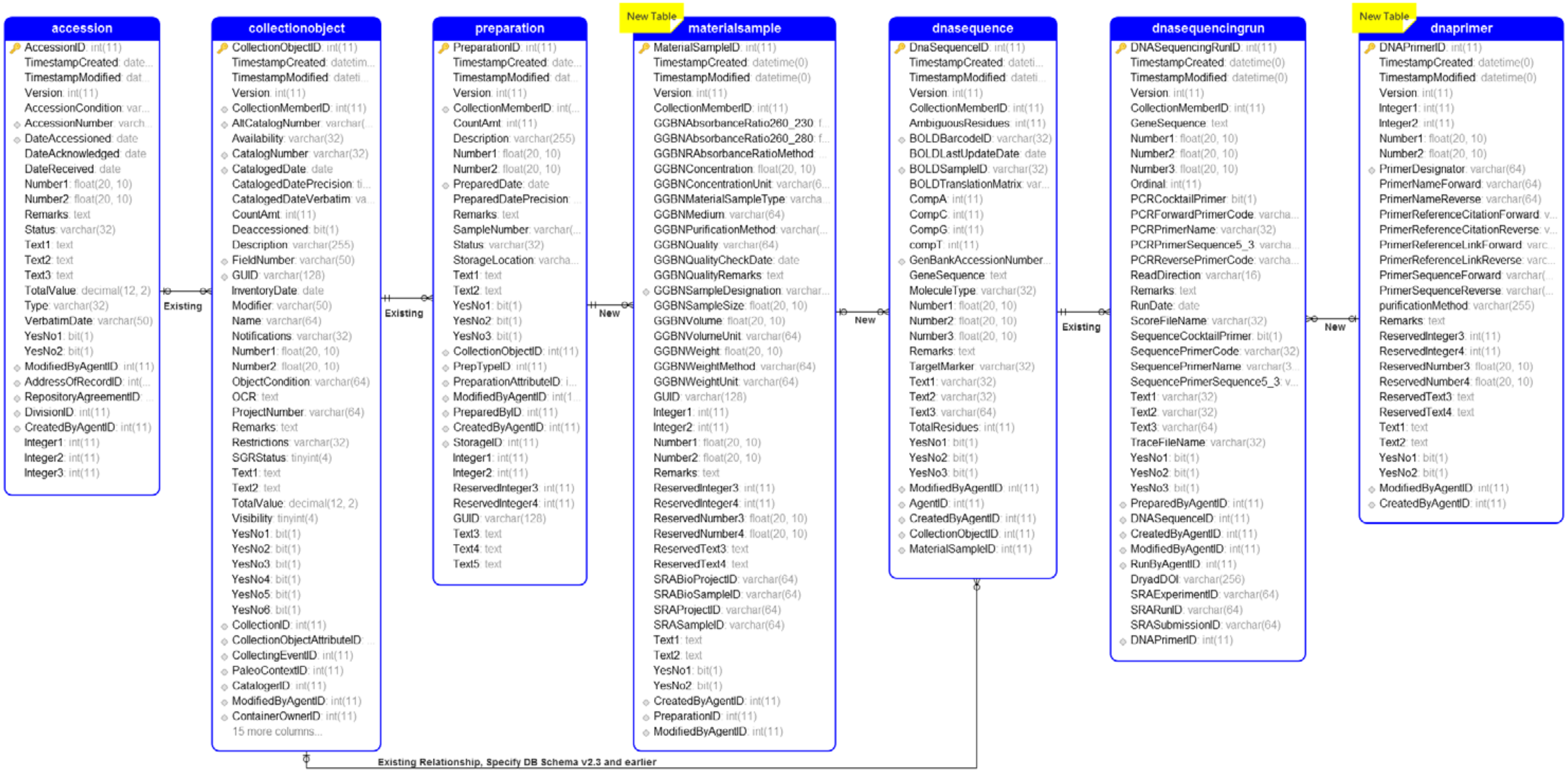
Specify Database Schema update (v2.4) for processing museum data associated with tissues, DNA extracts, sequencing runs, primers, and Sanger sequence and NGS data. These changes are effective with the release of Specify 6.6.06, and Specify 7.3.0 for the web.

Specify Software Project, 12 October 2017



Specify Software Database Schema Changes, Specify 6.6.06 and 7.3, 12 October 2017

The Specify Project has updated Specify's database schema (v2.4) to effectively manage catalog information associated with tissues, DNA extracts, and Sanger method DNA sequences, sequencing runs, and primers. Prior to Specify 6.6.06 and 7.3.0, schema v2.3 related Collection Object records directly to DNA Sequence table records. With increased use of tissues and DNA extracts as museum samples, Specify needed to accommodate additional cataloging scenarios with an additional data table (Material Sample) and two new relationships in-between the Collection Object and DNA Sequence tables, as shown.

These changes were made in collaboration with the Global Genome Biodiversity Network (GGBN) to bring Specify into compliance with the GGBN DarwinCore Material Sample extension to support standard-based digitization and communication of museum genetic material records. Specify 6.6.06 and 7.3.0 also added database support for Next Generation Sequencing (NGS) information deposited with NIH, Dryad, or other data archives and aggregators. Various fields for NIH SRA identifiers have been added to the database schema for web link-outs to records in those repositories. (In our 2017 collaboration with GGBN, we also created the capability for automatic assembly and publishing of any flavor of Darwin Core Archive files in the Specify 7 web platform.)

There are two basic options for describing and cataloging tissue samples in Specify. In the first, using the tables in the partial database schema shown, a tissue sample would be considered a Preparation of a Collection Object, with a count, description, and a Preparation Type of "Tissue" from the look-up table PrepType. Any other detailed or local metadata about the tissue itself would be handled in the Preparation Attribute table (not shown). A DNA extract sample derived from a tissue is considered a “Material Sample” according to the GGBN Material Sample Vocabulary data elements, information about the extract would be stored in the Material Sample table. This option is a best suited for collections that have some frozen tissues derived from their physical specimens and which want to use one collection database for both types of specimens/samples. The second basic approach in Specify, which may be preferred for institutions with separately managed genetic material repositories, is to catalog tissues as ‘first-order’ Collection Objects. The data tables and relationships shown would still apply. Material Sample would still be used for DNA extract information in this approach.

Any Specify collection which has used the DNA Sequence table with records related directly to Collection Object in schema v2.3 (and earlier) that wishes to use the new Material Sample and relationships will need to have their existing DNA Sequence records adjusted to the new schema by Specify Project staff. Also Specify installations that use Preparation for cataloging tissues in the first method and which want to use the Material Sample table for DNA or other extracts, will need their Preparation data forms updated in order for the Material Sample data fields to visible. Sites installing Specify for the first time with releases 6.6.06 or 7.3.0 or later, will receive the updated data entry forms.

These changes apply to both Specify 6 and 7 databases; the two platforms will continue to use the same database schema. Sites using both Specify 6 and 7 to access a database will need to update both Specify 6 on workstations, and Specify 7 server software (including the server embedded Specify 6 copy) in order to continue to use both platforms. We will maintain both relational paths between the Specify Collection Object and DNA Sequence tables, but collections should consistently use only one of these two logical pathways.

Got feedback or questions? Let us know at: support@specifysoftware.org.