

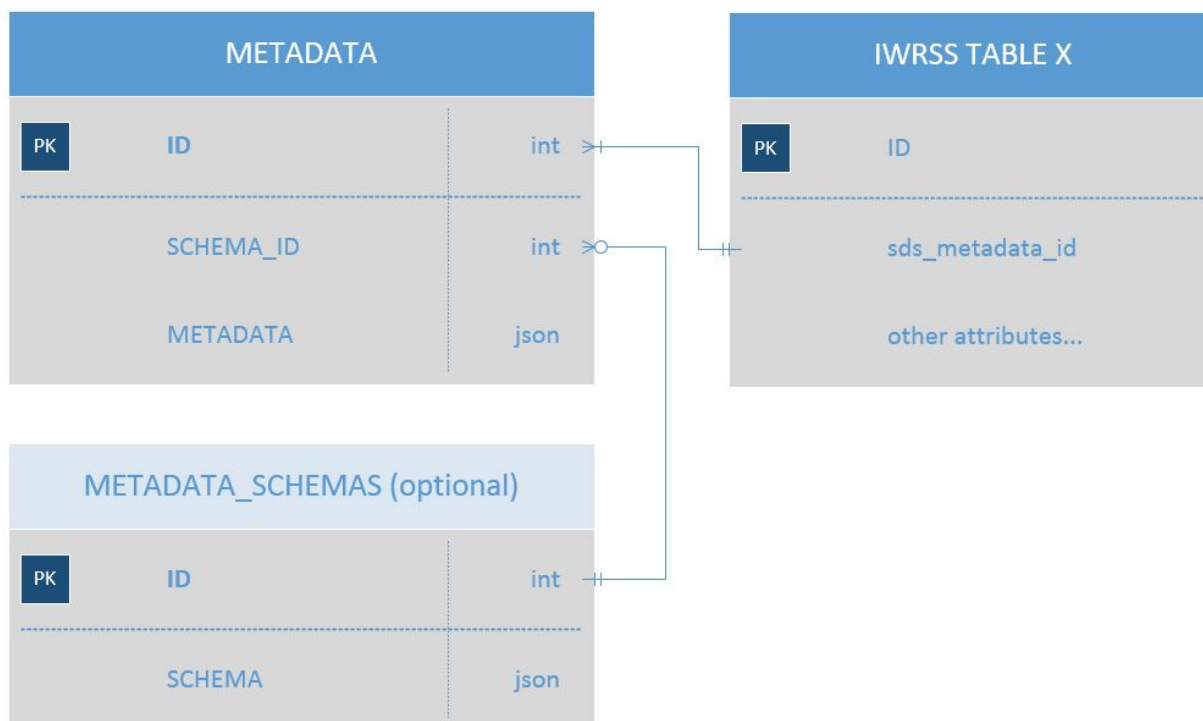
Attachment E.1: IWRSS Metadata Specification

Recommendation Summary

The following is the metadata approach we recommend for use by the IWRSS data model that will allow cross-agency indexing of available data with minimal data modeling effort up front and implementation customization down the road. PostgreSQL v > 9.2 or Oracle v > 12g are required for this implementation as it leverages the JSON data type inside of the database.

Design

Without delving specifically into any metadata fields that will be included, this recommendation will apply to any FGDC compatible XML metadata schema, as well as customizations of those schemas. The extent to which we validate the actual data structure can be implementation specific, or even metadata template specific within the implementation.



The basic structure of this design is shown above: Each table in the IWRSS data model will already be required to have the **sds_metadata_id** field that is designed to be a foreign key to a table that tracks metadata. We also leverage that field to store a foreign key reference to a metadata table. Our design differs from SDS in the fact that the metadata table doesn't just provide a reference to an XML metadata file on disc or online, but holds the metadata information itself in the table as JSON (JavaScript Object Notation) formatted data that indexed and exposed to the database for querying. Each record in the IWRSS data model will be

allowed to reference one metadata file in the metadata table, rows within a single IWRSS table may share metadata references if there are no record-specific differences in the metadata records, or may have their own unique metadata record.

One big benefit to storing the metadata as JSON in the database is the ability to add, remove, utilize or ignore the various metadata tags that can be used to track information about datasets. Some datasets will use tags that others will not, while some tags will be shared by almost all of the metadata records.

Each individual JSON metadata value can be unique but the design holds more query power when metadata objects share common attributes that can be aggregated and queried. Validation schemas can be used to make sure that rules governing the shared attributes are followed. Each metadata record can be associated with a metadata schema record. Schema records can be used to run validation routines on each row of metadata on insert or update operations. The level of validation required can be implementation, dataset or agency specific.

Data Format Conversion

FGDC compliant metadata is generally stored in XML format. This design requires a XML to JSON data type conversion that is easily accomplished as long as the XML is properly formed. Routines to convert between XML and JSON and back are widely available and customizable if needed.

Data stored in the JSON data type in the database are indexed and made available to querying for easy searching of the database based on metadata attributes.

Inter-Agency Accessibility

If all participating agencies store the metadata associated with their data using this design and expose those metadata records that apply to sharable data using a web API, the metadata can be mashed up and used to drive the content of interactive searchable web interfaces allowing users from any agency with access the ability to search the holdings of each of the participating agencies by metadata information. Paired with OGC compliant spatial data services, this approach makes the data sharing mission of the IWRSS work group easily attainable.

Other Considerations

Because the JSON data type exposes a queryable dataset to the database we should also consider migrating those fields that may end up largely being null or the same value across many rows in certain tables to the metadata spec. This way the data can still be populated and available without cluttering up the primary tables.