**Relational Database Schema Conventions for the Master Concordance File**

Schema Fields on the Concordance File:

These three fields currently on the concordance file will be used to document the RDB structure:

data\_type

cardinality

rdb\_table

Conventions:

**Cardinality:**

Code each field as either “ONE” or “MANY” representing the relationship between a single tax return and possible values of the variable. For example, each filing contains one EIN. A single filing will include multiple board members.

**RDB Table naming conventions:**

RDB tables are organized according to forms and schedules as follows:

FORM-PC-TABLE-00-DESCRIPTIVENAME

SCHED-A-TABLE-00-DESCRIPTIVENAME

Where TABLE-00 contains all one-to-one relationships from the form or schedule, and TABLE-01 or higher contain one-to-many relationships.

Alternatively if we want to organize schemas by parts instead of forms/schedules we could do this, but I don't see a good reason for that?

FORM-PC-P-01-TABLE-00-DESCRIPTIVENAME   (form 990-PC, part 01, all one-to-one variables)

SCHED-A-P-00-TABLE-01-DESCRIPTIVENAME

FORM-PF-P-02-TABLE-00-DESCRIPTIVENAME

**Data Types**

Use schemas version versus our own version? Add another field for schema\_data\_type?

Primary Keys

These variables should be included on all tables:

* EIN
* Document Locator Number
* org name
* tax year
* submission date
* what else?

Merges would be performed on?

EIN-TaxYear will not be unique if they append a filing. EIN-TaxYear-SubmissionDate should be.

Is DLN unique?