

# Creating new Dataset schemas

## Creating new Dataset schemas

To create a new dataset schema, the following steps must be taken:

1. **Create schema file** An xsd schema file must be created to define the new schema. These are generally placed in a folder named `gisSchemas` in the root directory of the defining plugin. The xsd file must define the fields of the new schema. See the `maeviz-bridgeDamage_1.0.xsd` in the `ncsa.maeviz` plugin for an example. As in the example, the field names must be prefixed with "maeviz." Also, the main part of the field name (after the maeviz prefix) must be 10 characters or less. The primitive type of the field (double, string, etc) must also be correctly specified.
2. **Add Schema extension point** The extension point `ncsa.gis.gisSchemas` must be extended to define the schema.

field	description
id	The id of the schema
name	A friendly name for the schema
version	A version identifier for the schema.
type	A short type name for the schema
description	A user-friendly description of the schema
file	The schema file defined in the previous step
format	The format of the schema. Predefined formats include: <code>shapefile</code> , <code>fragility</code> , <code>mapping</code> , <code>raster</code> , and <code>table</code> . If other formats are desired, they must be defined, which is outside the scope of this document
required Fields	The list of fields that are required as a minimum for this schema. These will be prompted to map to during ingestion
mapLayer	For shapefiles, the suggested smart layering position in the map. Should be between 1 and 100, where lower numbers appear above lower numbers. A mapLayer of -1 indicates that the dataset should be invisible by default
category	A category name for the schema. Categories determine what subfolders the schemas appear in the repository view.

3. **Add field metadata** An xml file should be created to define field-specific metadata for the schema. These files should be placed in a folder named `gisMetadata` that sits parallel to the `gisSchemas` folder. Files in here should have the same name as their corresponding schema xsd file, except with xsd renamed to xml. Each field should be defined using a `column-metadata` element with the following attributes:

attribute	description	required?
column-id	The id to match the column id from the xsd file. Should not include the <code>maeviz.</code> prefix.	Y
friendly-name	A friendly name for the field	
is-numeric	An indicator of whether this field should be displayed as a numeric value (for example, some schemas have string fields that should be interpreted as numbers)	
unit	The default units of this field. This abbreviation should come from a unit tag defined in the <code>ncsa.gis.units</code> extension point	
field-length	The length of the field. This should be as small as possible without truncating data	
importance	The importance of the field. Values are <code>user</code> (interesting to the user), <code>mainValue</code> (the one main important value of the dataset), <code>unique</code> (a unique key field), and <code>summary</code> (a field that can be used for grouping in summary views. If a field does match any of these importance levels, the importance attribute should be omitted).	
is-result	This should be defined if the field is a new result value for this schema	
agg-type	If the field is aggregated in a view or table, what type of aggregation makes the most sense for this field. Values are <code>sum</code> , <code>mean</code> , and <code>count</code>	