

Save 100 Hours on Your Next Project

ScribeKey's GIS Data Profiler for ArcGIS 9,10

Create structured, concise, and detailed data profiles of your geospatial datasets for more efficient:

- Dataset Revisions
- Application Development
- Integration, Migration, and ETL
- Data QA/QC and Cleansing
- Schema Matching
- Metadata & Data Dictionary Development

Highly beneficial for:

- Developers and System Integrators
- Data Brokers and Providers
- Data Evaluators
- Data Analysts
- Project Managers
- End Users

Easy to use:

- Built with ArcObjects™ and .NET
- Industrial strength batch command line
- Wide variety of configurable settings
- Also includes tools to import/flatten XML and generate FGDC CSDGM metadata
- Comprehensive user guide includes Quick Start and Tutorial using US Census data.

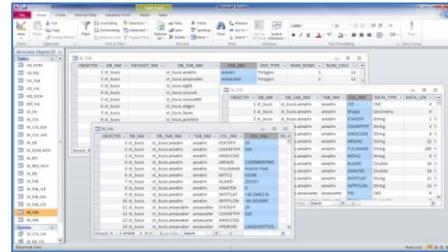
Try it out:

- Download your evaluation copy or request a demonstration at www.scribekey.com
- Single user license: \$395 USD (discount for .edu)
- All HTML, Metalayer, and XML/Metadata tools are fully functional in the evaluation version.
- Webinar or on-site training, project implementation assistance, and custom profiler development available.

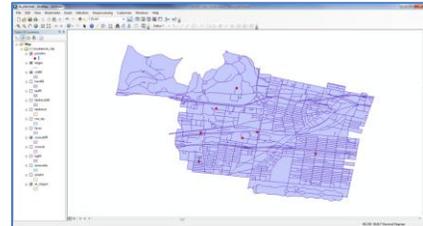
Get the Details on Your Data



www.scribekey.com



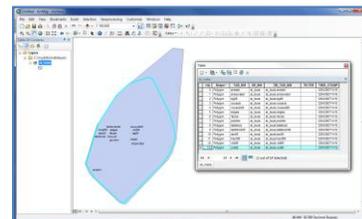
PROFILE: Capture the structure, contents, and meaning of Shapefile or Geodatabase (mdb, gdb, sde, sdc) feature classes, attributes, domains, metadata, indexes, relationships, and more, in an easy-to-use/share MS Access database.



CLIP: Easily create smaller subsets of large volume feature classes (and related tables) or predefined locale specific subsets of large multi-layer datasets.



DESCRIBE: Generate HTML data dictionaries and tables describing your datasets with the look and feel of your organization's website.



METALAYERS: Create polygons showing the spatial extents of your feature classes, joined with profile database tables in ArcMap™ for enhanced metadata.

What is data profiling? Data profiling is systematically exploring and capturing the essential information describing the structure, meaning, and contents of a database. Data profiling is an important tool in mainstream IT data warehousing, business intelligence, and data-centric decision support systems. If you think of your database as a large book with lots of details, the data profiler can be thought of as a tool to generate a table of contents and an index, or if you think of your database as a large building with lots of floors and rooms, the data profiler is a tool that can generate a set of architectural blueprints. This information is highly useful for:

- Describing the data for end users, developers, and data analysts
- Determining if the data is suitable for a given purpose
- Determining the level of effort for data cleansing, integration, or ETL operations
- Determining if metadata is present and accurate
- Creating enhanced data description materials, including vocabularies and keywords, to facilitate data querying

Who benefits from data profiling? Application developers, data evaluators, data brokers, data providers, data analysts, project managers, and end users all benefit from the structured and detailed information generated through data profiling. Data profiling can save hundreds of hours on many types of projects.

What types of projects does data profiling help with?

- Dataset revision tracking
- Application development
- Data evaluation and analysis
- Data cleansing and QA/QC
- Data merging, migration, and ETL
- Data dictionary and metadata development

How does ScribeKey's profiler work? Built with ESRI's ArcObjects and .NET, ScribeKey's Profiler is an easy-to-use command line, batch processing application which connects to an ESRI data source and then creates and populates an MS Access database with data profiling information and metadata. From this MS Access database, profiling information can be loaded into other databases, saved as HTML, MS Excel spreadsheets or XML, and used to generate charts and graphs for easily understood data description presentations. Additionally, the MS Access database can be used as a front end by linking to tables in a more industrial data store such as SQL Server, Oracle, or MySQL.

What information is captured with ScribeKey's profiler?

- **Feature Classes/Tables:** Name, Database, Geometry Type, Spatial Extent, Coordinate System, Number of Records, Number of Fields, Indexes, and more.
- **Data Fields:** Name, Data Type, Data Length, Sample Values, Number of Nulls, Percentage Complete, Data Value Patterns, Min/Max Values and Lengths, and more.
- **Data Values:** Lists of data values, frequencies, and percentages.
- **FGDC XML Formatted Metadata** (when available): Title, Publication Date, Origin, Abstract, Purpose, Entity Definition, Attribute Definitions, and Domain Value Lists.
- **ESRI Geodatabases:** Datasets, Domains, Subtypes, and Relationships.
- **Relationships:** Checks cross-column value matches between 2 columns, all columns in 2 tables, or all columns in the database.
- **Domain Matching:** Differences and similarities between the value sets in 2 columns.
- **Regular Expression Matching:** Matches and mismatches between the set of values in a column and a regular expression.

- **XML Documents:** Flattened XML elements and attributes loaded into a database table for flexible query and use.

What types of data sources does ScribeKey's profiler work with? ScribeKey's profiler works with ArcGIS 9 & 10, and supports all major ESRI (TM) data stores including Shapefiles, Personal Geodatabases, File Geodatabases, SDE, and SDC.

What other tools are available with ScribeKey's data profiler?

- **Clipping Tools:** For creating subsets of larger databases using a polygon or polygons from an existing feature class. Additionally, copies the subsets of rows in non-geometric tables related to the clipped features.
- **HTML Tools:** For generating sets of interlinked HTML pages for presenting the contents of the profile database as an interactive data dictionary. These HTML formatted data dictionaries can be used by federal, state, and local GIS data brokers for providing detailed dataset descriptions to their end users. These data dictionaries can also be used by data providers and application development teams for easily sharing data description information. HTML templates can be used, to ensure the resulting output matches the look and feel of an existing website.
- **Metalayer Tools:** For generating polygon layers describing the spatial extent of a given feature class (or set of feature classes) with either simple bounding boxes, or more detailed convex hulls. These metalayer polygon features can be linked with the tables in the profile results database and used from within ArcMap, or a web based GIS application, to provide a single view of fully integrated data and metadata.
- **XML/Metadata Tools:** The profiler also includes tools for exporting XML metadata out of Geodatabase blobs, generating the Entity/Attribute/Domain XML sections for FGDC CSDGM metadata, and importing any XML document into a 'flattened' database table, allowing easy access, query, and manipulation of the data contents.

How do ScribeKey's configurable profiler settings help with large datasets?

ScribeKey's profiler provides 3 different operating modes including:

- **Schema Mode:** Only data structure information and any available metadata information is loaded into the profile database. This mode works very quickly on even very large geospatial datasets.
- **Sample Mode:** Data structure, metadata, and sample values are loaded into the profile database, using a maximum rows setting. This mode can be used very effectively with large datasets to gain an initial overview of data contents and values, and help to identify primary keys and domain values. Additionally use or ignore lists of tables and columns can be configured to focus only on data elements of interest.
- **Full Mode:** Data structure, metadata, sample values, null counts, percentage complete, and exhaustive lists of unique values for specifically designated fields are generated. This more time consuming and processor intensive mode is used after Schema and Sample mode profiles have been run, to allow data profilers to produce fully descriptive profiles in the most time efficient manner possible.

What is the difference between the evaluation and licensed versions of ScribeKey's profiler? The HTML, metalayer, and metadata/XML tools are all fully functional in both the evaluation and licensed versions. With the evaluation version, the data profiling tools are limited to processing 10 feature classes or tables, 10 columns for each feature class or table, 100 column values for each column, and 1000 sample rows. For clipping, the evaluation version is limited to processing 10 feature classes or tables.