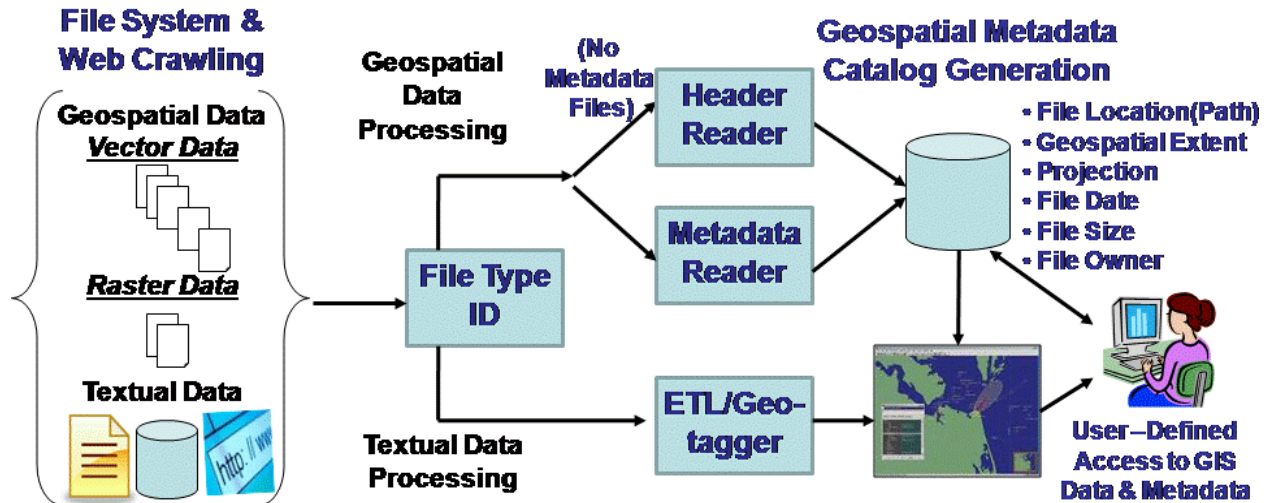


Geospatial Data Discovery

Problem: Geospatial analysts and GIS users require the ability to handle the increasing proliferation of geospatial data types and file formats; manage the massive volume of historical geospatial data accumulated over multiple decades; maintain an up-to-date catalog of geospatial data holdings; and extract location data from unstructured text for geotagging and correlation.

Solution: *Twister Data Integrator* software provides a high performance, cost-effective solution for discovering, cataloging, extracting, and correlating geospatial data.

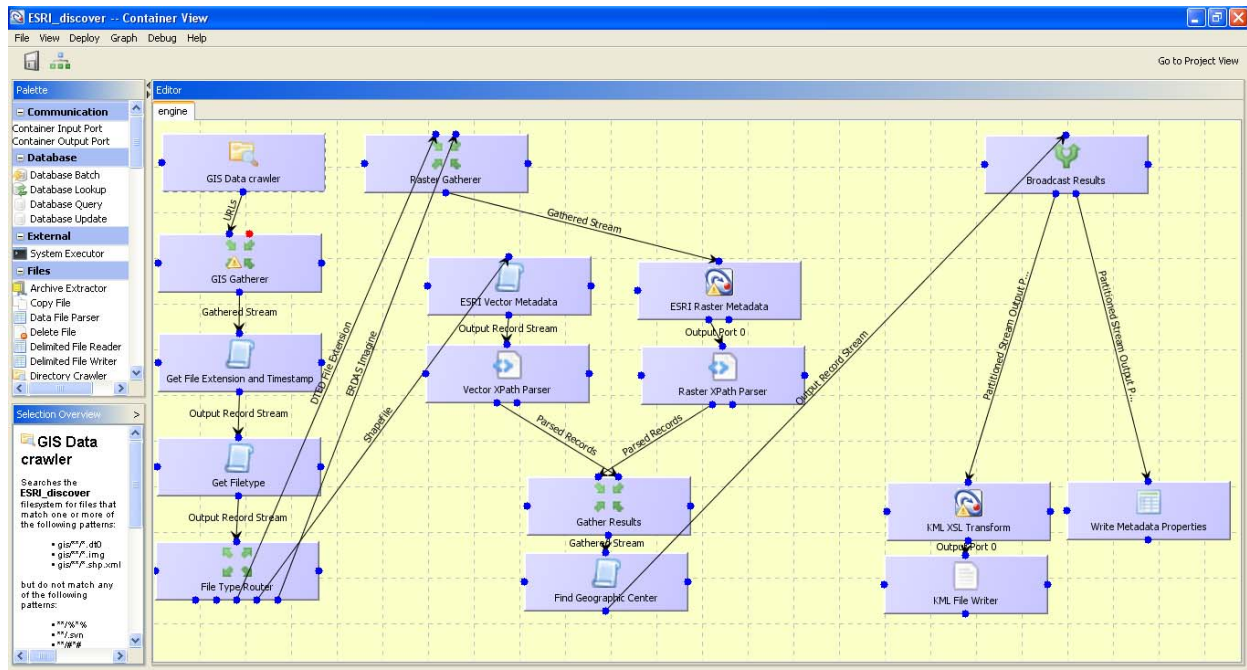


Benefits: Eliminate “geospatial data buildup” from your organization’s file systems so users can quickly locate and share geospatial data needed for analysis. Using *Twister Data Integrator*, mining and discovering geospatial data is an automated process that runs in parallel, crawling file systems or other sources of geospatial data, to identify geospatial data files, extract metadata from those files, and generate a catalog of the geospatial data holdings in a user-defined format.

- Flexibility
 - Dynamic data flow & flexible scheduling to ingest data via FTP, HTTP, CIFS; query data via ODBC/JDBC or search engine APIs; and file system/web crawling
 - Data delivery and output to any file format via templates and XSLT (e.g., KML) or to any JDBC database (e.g., ESRI Geodatabase)
- Scalability
 - Distribute processing load over hundreds of nodes
- Rapid Implementation
 - Initial data flow configuration within 2 weeks
 - Catalog decades of data within days or hours
- Time & Cost Savings
 - Reduces time spent by users searching for geospatial data
 - Leverages existing investment in geospatial data holdings by avoiding the need to reacquire or recreate data
 - Allows users to share geospatial data across an organization

How it Works: *Twister Data Integrator* operates on large data sets by executing a data flow consisting of data sources, transformations, routers, and data destinations. Each processing step within the data flow is executed by a component. The resulting data pipeline includes all of the data processing capabilities needed to ingest, transform, and deliver data, eliminating the need for custom software development.

Application Brief – Twister Data Integrator



FAQs:

Q: What types of geospatial files can be discovered with *Twister Data Integrator*?

A: Various vector and raster formatted files can be mined using *Twister*. For example, vector data such as ESRI shapefiles and raster data such as DTED and ERDAS can be identified and metadata for these geospatial data files can be extracted. SMSi understands that the processing needs of each Customer will be unique to their data holdings. Therefore, SMSi offers services to assist the Customer with custom component development and integration for additional geospatial data file types and formats.

Q: What happens when geospatial data files are deleted or updated and new files are created?

A: *Twister* can be scheduled to periodically check the file system that was crawled to generate the geospatial metadata catalog to identify new, updated, or deleted files. The metadata catalog can be updated by *Twister* if files have been added, modified, or deleted. By utilizing *Twister's* delivery features, users can be notified of new data availability via watch lists and alerts.

Q: How does geotagging work?

A: *Twister* integrates many open source and commercial ETL tools that extract entities from textual data, including place names and locations, and eliminates the need for custom software development to utilize these ETL tools. By cross referencing these extracted locations with geospatial coordinate data through third party software such as *Group 1*, or by delivering these place names directly to your GIS software, these referenced locations can be geocoded with latitude and longitude coordinates.

Q: What if I want to display the metadata in a geospatial view using my favorite GIS software?

A: *Twister* can deliver the metadata in any format, either to a delimited file, relational database, XML document, or specific format for visualization. For example, *Twister* can generate KML for visualization of the metadata in ESRI ArcGlobe or Google Earth, providing a geospatial view of data holdings based on geospatial extent.

Q: Can the generated metadata catalog be shared with users in other organizations?

A: Yes, depending on the organizations' IT infrastructure and security environment, *Twister* can be used to provide shared access to the generated metadata. If there is a common data model or format that is used to enable data exchange between organizations, then the *Twister Data Integrator* can be used to deliver the generated metadata in this format. If the organizations share access to a centralized geospatial metadata repository, then *Twister Data Integrator* can deliver the generated metadata directly to the shared repository. If no common data model or centralized repository exists, then the generated metadata catalog can be shared with other users via *Twister Data Server's* federated query capability. *Twister* mediates access to the geospatial metadata catalog, translates the user queries, and collects the query results for delivery to the user. *Twister Data Server* can be configured to share data across multiple networks operating in multiple security domains.