An ESRI® White Paper • May 2009

ArcGIS® 9.3 Geocoding Technology

ESRI 380 New York St., Redlands, CA 92373-8100 USA
TEL 909-793-2853 • FAX 909-793-5953 • E-MAIL info@esri.com • WEB www.esri.com

ArcGIS 9.3 Geocoding Technology

An ESRI White Paper

Contents **Page**

The Geocoding Engine 1

Supported Environments, Inputs, and Outputs 2

Typical Deployment Scenarios and Performance 2

Customization Options 3

Data Sources, Functionality, and Quality 3

ArcGIS 9.3 Geocoding Technology

ESRI's ArcGIS® 9.3 geocoding engine is a proven, flexible, cross-platform technology common to ArcGIS Desktop (including ArcGIS Engine), ArcGIS Server, and ArcGIS℠ Online offerings. It supports geocoding within any ArcGIS deployment scenario, including ArcReader™, ArcMap™, and ArcCatalog™, and custom applications built with the integrated development environment accessing ESRI's developer components.

The ArcGIS 9.3 geocoding engine supports a rich set of geocoding styles that are immediately usable for many addressing systems worldwide. In addition, it is fully customizable, allowing you to build, publish, and share effective and focused address locators that reflect custom rules, address types, and data sources. The ArcGIS 9.3 geocoding engine also allows you to build locators and gazetteers for nonaddress descriptors—such as feature identifiers, codes, or names—a process that is beyond the capability of traditional geocoders.

ESRI is continually adding new geocoding styles and data sources to its products and Web offerings, and existing styles are continuously refined according to user feedback and requirements. At ArcGIS 9.3.1, ArcGIS Server developers may also access Microsoft's Virtual Earth™ geocoding environment, providing extra geographic coverage for international geocoders.

The Geocoding Engine

The ArcGIS 9.3 geocoding engine is a general purpose record-matching processor. It is spatially aware and able to be tuned at build and run times to test agreement and discriminate differences between input and reference data according to probabilistic rules you define.

The principal elements of interest for the geocoding engine are the

- Address locator
- Reference data, which is typically embedded within an address locator

The address locator is an index structure that supports rapid execution of the matching process your rule base supports and includes the underlying reference data—such as street centerlines; address points; or any other point, line, or polygon feature class.

Reference data determines the spatial precision of geocodes. Point reference data supports exact locations while linear reference data supports interpolated locations. Tolerances and parameters within the geocoding process permit control over sensitivity to spelling or alternate naming, or number agreement in address elements, thus allowing the user to handle issues like local variability in addresses while using the same reference data. By default, geocoded locations share the coordinate precision of the underlying reference data, but this precision may be rounded if desired.

An address locator style has no dependency on a source of reference data; ESRI, its business partners, and users around the world can use geocoding reference data or fully

built locators from a variety of sources. These include user-created data, commercial products from suppliers like Tele Atlas® and NAVTEQ®, and free reference data downloaded from the Census TIGER 2008 site.

Address locators encapsulate one or more geocoding rule bases and reference datasets, making them portable and distinct from the reference data itself, which when embedded cannot be seen within a locator. Locators built with ArcGIS Desktop may be stored in any ArcGIS workspace, file system, or geodatabase and managed with ArcGIS 9.3 standard administration tools.

Address locators may be simple (single step) or composite (multiple step), providing a cascading hierarchy of geocode precision within the one ArcGIS process. Match preferences may be set at run time, and match quality is recorded both as a flag for the locator that provided the best geocode result and as a normalized score out of 100 for that locator.

Supported Environments, Inputs, and Outputs

Because geocoding is a core function of ArcGIS, the supported environments for address locators are the same as any other data accessible by ArcGIS. ESRI's support site contains detailed information on supported operating systems and database versions for current and previous releases of ArcGIS.

Geocoding is normally performed on input tables, but GIS feature classes may also be used as inputs. Input tabular data may be external to an ArcGIS workspace (for example, from an OLE DB connection) or natively recognized by ArcGIS, such as CSV, dBASE and Excel® files, or tables in any ArcGIS workspace format: ESRI's file geodatabase; personal geodatabases (Microsoft® Access®); and personal, workgroup, and enterprise geodatabases using a DBMS platform. ArcGIS Server additionally supports serialized transmission of lightweight representations of tables and features when geocoding record sets transported via the Web.

Geocoding processes normally output point feature classes, but optionally a standardized representation of an input table may be written with address fields formatted according to the rule base. Use cases for this include separating address components into a schema useful for address labels.

The ArcGIS geocoding engine supports reverse geocoding, the process of finding an address from a given x,y location. This is exposed in the user interface in ArcMap and is available programmatically for desktop and Web APIs.

Typical Deployment Scenarios and Performance

ArcGIS 9.3 address locators support the same author-publish-use paradigm as data and map products in ArcGIS. The ArcGIS Desktop applications, ArcMap and ArcCatalog, have menus and geoprocessing tools for creating locators in any workspace type. Locators may then be copied or published as workspace-based resources for desktop clients or published as ArcGIS Server geocoding services for consumption by Web applications.

Geocoding may be an interactive or batch operation. Both deployment scenarios are supported by ArcGIS Desktop 9.3 and Web client applications—an ArcGIS Desktop 9.3 application may geocode interactively or in batch mode using a workspace-based locator or an ArcGIS Server geocoding service. An ArcGIS Server 9.3 Web application may send individual geocoding requests or the equivalent of whole tables as record sets for batch processing.

J-9790

Geocoding performance is highly data dependent and with many processing variables, but for well-designed implementations of standard locator styles, users can expect to achieve performance with the following orders of magnitude:

Geocoding Performance

Architecture	Interactive	Batch
ArcGIS Desktop	Subsecond	2,250,000/hour
ArcGIS Server—Web client	< 2 seconds	15,000/hour per process
ArcGIS Server—Local client	Subsecond	2,000,000/hour per process

Customization Options

ArcGIS 9.3 provides a set of platform-independent software components called ArcObjects™, written in C++, with a wide variety of development environments available. Technical resources you need to develop using the ArcGIS 9.3 platform are now located within the ArcGIS Resource Centers; however, the ESRI® Developer Network (EDN℠) site provides a useful entry point into technical Resource Center sites for our APIs here, cross-referenced by application and development technology. ArcGIS Desktop, ArcGIS Engine, and ArcGIS Server application areas are relevant to Software Developer Kit (SDK) environments for geocoding, and ArcGIS Mobile can also access ArcGIS Server geocoding functions executed server side.

Data Sources, Functionality, and Quality

ArcGIS 9.3 geocoding derives x,y coordinates from inputs using reference data. Data errors are possible in the input and reference data; these can cause an undesirable result, such as failure to geocode or geocoding successfully to the wrong location. Because users can often determine these data issues by inspection **after geocoding,** a rematch dialog is provided for geocoded feature classes that may be used anytime to select and repair unmatched or falsely matched feature selections individually or in bulk.

ESRI provides reference data from commercial suppliers such as NAVTEQ and Tele Atlas. Commercial suppliers offer data with varying positional accuracy, for example, postal code centroid, place location, or rooftop position or interpolated along a street centerline. ESRI's address locators honor the precision and geometry of the supplier's reference data, optionally with an offset for address parity (side of street) or intersection (back along the street from an intersection), which may be desired for cartographic purposes when using centerline reference sources.

Reference data represents a physical delivery point, and geocoding against that data constitutes **address validation** of the input data. Address validation means that input addresses are properly formed and the address actually exists.

Because geocoding results also contain the reference data values for address components, ESRI geocoding supports **address correction** of misspellings, transpositions, house number best approximations, and so on, understood by the probabilistic engine rules.

ESRI is able to supply geocoding engine rule bases that are not packaged with any of our products; examples of these include single-line input rule bases for the United States, Canada, and Europe, enabling the parsing and geocoding of address sentences rather than address fields, much like the user experience of a search engine.

About ESRI

For four decades, ESRI has been helping people make better decisions through management and analysis of geographic information. Our culturally diverse staff work with our business partners and hundreds of thousands of people who use GIS to make a difference in our world.

A full-service GIS company, ESRI offers support for implementing GIS technology from the desktop to enterprise-wide servers, online services, and mobile devices. GIS solutions are flexible and customizable to meet the needs of all our users.

Our Focus

At ESRI, we focus on promoting the value of GIS and its applications throughout the world and pay close attention to our users' needs. Our software development and services respond to our customers with products that are easy to use, flexible, and integrated. Our technology is multidisciplinary, productive, and valuable to our users.

We have a strong commitment to educating our customers through ESRI's various training programs. ESRI is a socially conscious business and invests heavily in issues regarding education, conservation, sustainable development, and humanitarian affairs.

Contact ESRI

1-800-GIS-XPRT (1-800-447-9778)
Phone: 909-793-2853
Fax: 909-793-5953
info@esri.com
www.esri.com

Offices worldwide
www.esri.com/locations

ESRI
380 New York Street
Redlands, California
92373-8100 USA

GSA Contract Holder
Contract GS-35F-5086H

Printed in USA