



Introducing ArcScan™ for ArcGIS®

An ESRI® White Paper • August 2003

Copyright © 2003 ESRI
All rights reserved.
Printed in the United States of America.

The information contained in this document is the exclusive property of ESRI. This work is protected under United States copyright law and other international copyright treaties and conventions. No part of this work may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying and recording, or by any information storage or retrieval system, except as expressly permitted in writing by ESRI. All requests should be sent to Attention: Contracts Manager, ESRI, 380 New York Street, Redlands, CA 92373-8100, USA.

The information contained in this document is subject to change without notice.

U.S. GOVERNMENT RESTRICTED/LIMITED RIGHTS

Any software, documentation, and/or data delivered hereunder is subject to the terms of the License Agreement. In no event shall the U.S. Government acquire greater than RESTRICTED/LIMITED RIGHTS. At a minimum, use, duplication, or disclosure by the U.S. Government is subject to restrictions as set forth in FAR §52.227-14 Alternates I, II, and III (JUN 1987); FAR §52.227-19 (JUN 1987) and/or FAR §12.211/12.212 (Commercial Technical Data/Computer Software); and DFARS §252.227-7015 (NOV 1995) (Technical Data) and/or DFARS §227.7202 (Computer Software), as applicable. Contractor/Manufacturer is ESRI, 380 New York Street, Redlands, CA 92373-8100, USA.

ESRI, the ESRI globe logo, ArcScan, ArcInfo, ArcEditor, ArcMap, ArcGIS, ArcView, www.esri.com, and @esri.com are trademarks, registered trademarks, or service marks of ESRI in the United States, the European Community, or certain other jurisdictions. Other companies and products mentioned herein are trademarks or registered trademarks of their respective trademark owners.

Introducing ArcScan for ArcGIS

An ESRI White Paper

Contents	Page
Introduction	1
What Can You Do With the ArcScan Extension?.....	1
Batch Vectorization.....	1
Interactive Vectorization.....	2
Raster Snapping.....	2
Raster Tracing.....	2
Raster Selection.....	3
Simple Raster Editing.....	3
ArcScan Basics.....	3
Understanding a Raster Data Set.....	3
The Cell.....	3
Rows and Columns.....	4
Values.....	4
Coordinate Space and the Raster Data Set.....	5
Georeferencing a Raster Data Set.....	6
An Overview of the ArcScan Extension.....	6
Who Should Use ArcScan?.....	6
How Does ArcScan Work in the ArcGIS Environment?.....	6
Getting Started With ArcScan.....	6
Symbolizing Raster Layers for ArcScan.....	7
Selecting Connected Cells.....	7
Cleaning Up the Raster Data.....	7
Using the Raster Snapping Tools.....	7
Tracing Raster Cells.....	7
Batch Vectorization.....	8
More Information	8

Introducing ArcScan for ArcGIS

Introduction

ESRI® ArcScan™ for ArcGIS®, the raster-to-vector conversion extension to ArcInfo™, ArcEditor™, and ArcView®, allows you to scan images of vector maps and convert them into vector-based feature layers such as shapefiles and geodatabase feature classes. Vectorization can be performed manually by interactively tracing raster cells or automatically using the batch mode.

The process of converting raster data into vector features relies on user-defined settings. These settings allow you to influence the geometric composition of the output vector features. Once you have determined the optimal vectorization settings for your data, they can be readily saved and reused.

The ArcScan extension also provides tools that allow you to perform simple raster editing to prepare your raster layers for vectorization. This practice, known as raster preprocessing, can help you eliminate unwanted raster elements that are not in the scope of your vectorization projects.

What Can You Do With the ArcScan Extension?

ArcScan provides tools that allow you to convert scanned images into vector-based feature layers. ArcScan also supports the ability to select and edit raster cells to help you refine the image you are working with. When combined, these tools extend the ArcMap™ application within the ArcGIS Desktop products by providing new techniques for integrating features derived from raster imagery into your geographic information system (GIS).



A typical scanned parcel map from an engineering department

Batch Vectorization

One of the key features of the ArcScan extension is its ability to automatically convert raster data into vector features. This process, known as batch vectorization, can significantly reduce the time it takes to vectorize scanned images.

ArcScan supports two types of vectorization methods: centerline and outline. Depending on your requirements and the type of scanned images you are working with, the vectorization method you employ will vary.

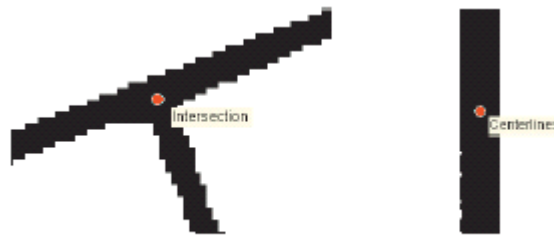
Batch vectorization requires settings that influence how the output vector features are generated. These settings, also known as styles, can be saved and reused with raster images that possess similar characteristics.

Interactive Vectorization

Along with batch vectorization, you can also generate features manually. This process is known as interactive vectorization and is similar to the existing techniques used to create features with the ArcMap Editor. Interactive vectorization consists of two components: raster snapping and raster tracing.

Raster Snapping

The ArcScan extension supports the ability to snap to raster cells. Although not required for raster tracing, raster snapping can help ensure that you create features accurately. You can snap to raster centerlines, intersections, corners, ends, and solids.

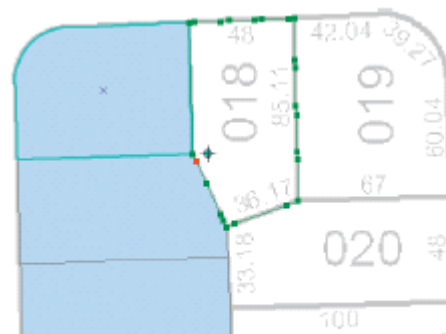


Snapping to raster intersections; snapping to raster centerlines

Raster Tracing

Raster tracing is useful in cases in which you need to have more control over the vectorization process or need to vectorize a small portion of an image. The Vectorization Trace tool allows you to manually trace raster cells and generate features for raster data that you wish to vectorize.

With the Vectorization Trace tool, you simply point the cursor in the direction you wish to vectorize and click. With each click, features will be generated at the centerline of the raster cells. The current vectorization settings will influence the output vector geometry. You have the option to generate line or polygon features.



Raster snapping and tracing used to create polygon features

Raster Selection

ArcScan supports tools for selecting raster cells. You can create raster selections interactively by clicking a series of connected cells or by executing an expression-based query. These selection tools can help you define the scope of your vectorization. The

Select connected cells dialog box allows you to perform complex cell selection based on pixel area and envelope extents. This can help you filter which cells you wish to omit from or include in the vectorization.

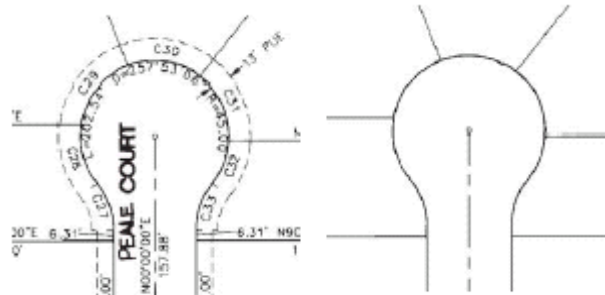


Image prior to and after cleanup

Simple Raster Editing

ArcScan also supports tools for editing raster images. You can draw, fill, and erase raster cells all within an ArcMap edit session. These steps, known as raster cleanup, allow you to eliminate raster cells that are not in the scope of the vectorization. In addition, you can export the modified raster to a new file in case you need to preserve the original copy.

ArcScan Basics

ArcScan is designed as an extension to ArcInfo, ArcEditor, and ArcView for the purpose of converting raster data into vector-based feature layers. In addition, you can leverage the tools and commands available in the Editor toolbar to further refine this newly created data. ArcScan goes beyond the traditional methods of manual vectorization by introducing new tools to automate the raster selection and cleanup process. Whether you choose to trace raster cells interactively or batch vectorize an entire raster image, ArcScan provides an efficient way to integrate raster data into your GIS.

This section provides an overview of the raster data sets as well as an introduction on how to get started with ArcScan. It also describes how ArcScan works in conjunction with the Editor toolbar. Topics covered include enabling the ArcScan extension, adding the ArcScan toolbar, symbolizing raster data, and setting the vectorization options.

Understanding a Raster Data Set

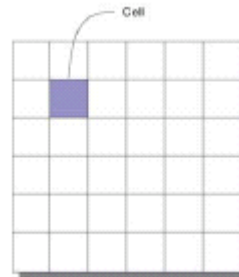
Vector data—such as coverages and shapefiles—represents geographic features with lines, points, and polygons. Rasters—such as images and grids—represent geographic features by dividing the data into discrete squares called cells. A raster can represent the following:

- Thematic data such as land use, temperature, and elevation
- Spectral data such as satellite images and aerial photographs
- Pictures such as satellite images and aerial photographs

In ArcScan, you will work primarily with rasters that represent scanned maps and drawings. This section will provide a brief overview of raster data.

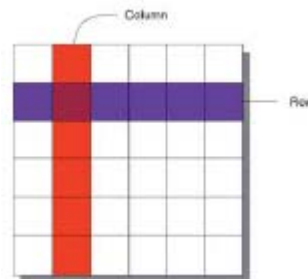
The Cell

A raster data set is made up of cells. Each cell, or pixel, is a square that represents a specific portion of an area. All cells in a raster must be the same size. The cells in a raster data set can be any size that you desire, but they should be small enough to accomplish the most detailed analysis. A cell can represent a square kilometer, a square meter, or even a square centimeter.



Rows and Columns

Cells are arranged in rows and columns, an arrangement that produces a Cartesian matrix. The rows of the matrix are parallel to the *x*-axis of the Cartesian plane, and the columns are parallel to the *y*-axis. Each cell has a unique row and column address. All locations in a study site are covered by the matrix.

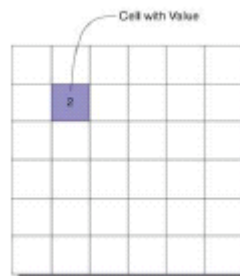


Values

Each cell is assigned a specific value to identify or describe the class, category, or group that the cell belongs to or the magnitude or quantity of the phenomenon that the raster describes. The characteristics the values represent include soil type, soil texture, land use class, water body type, road class, and housing type.

A value can also represent the magnitude, distance, or relationship of the cell on a continuous surface. Elevation, slope, aspect, noise pollution from an airport, and pH concentration from a bog are all examples of continuous surfaces.

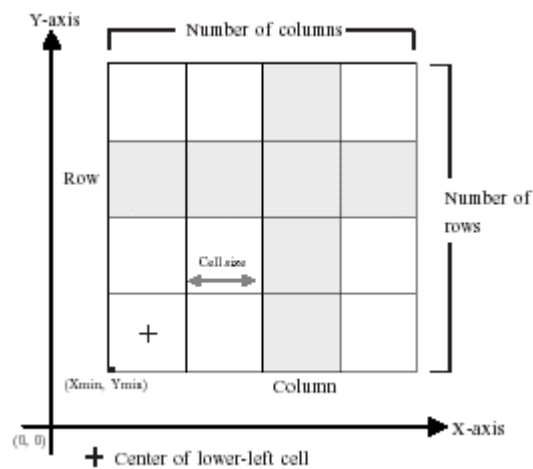
For rasters representing images and photographs, the values can represent colors or spectral reflectance.



Coordinate Space and the Raster Data Set

Coordinate space defines the spatial relationship between the locations in a raster data set. All raster data sets are in some coordinate space. This coordinate space may be a real-world coordinate system or image space. Since almost all raster data sets represent some real-world location, it is best to have that data set in the real-world coordinate system that best represents it. Converting a raster data set from a nonreal-world coordinate system—image space—to a real-world coordinate system is called georeferencing.

For a raster data set, the orientation of the cells is determined by the x - and y -axes of the coordinate system. Cell boundaries are parallel to the x - and y -axis, and the cells are square in map coordinates. Cells are always referenced by an x,y location in map coordinate space and never by specifying a row-column location.



The x,y Cartesian coordinate system associated with a raster data set that is in a real-world coordinate space is defined with respect to a map projection. Map projections transform the three-dimensional surface of the earth to allow the raster to be displayed and stored as a two-dimensional map.

The process of rectifying a raster data set to map coordinates or converting a raster data set from one projection to another is referred to as geometric transformation.

Georeferencing a Raster Data Set

To georeference a raster data set from image space to a real-world coordinate system, you need to know the location of recognizable features in both coordinate spaces. These locations are used to create control points. The control points are used to build a polynomial transformation that will warp the image from one coordinate space to another. This can be done with the Georeferencing toolbar.

Control points are locations that can be accurately identified on the raster data set and in real-world coordinates. These identifiable locations may be road and stream intersections, building corners, bridges, the mouth of a stream, rock outcrops, and identifiable points on geometric landscape features such as the end of a jetty of land, the corner of an established field, or the intersection of two hedgerows.

For each control point selected on the input raster data set, the output location may be specified either by graphically selecting a point that is already in the desired output coordinate system or by typing in the known output coordinates. The relationship between the control points chosen in the raster data set and the output coordinate space is then determined.

Using this relationship and a polynomial transformation, the raster data set is converted from nonreal-world space to real-world space.

An Overview of the ArcScan Extension

The ArcScan extension provides tools and commands that support the conversion of raster data to vector features. This process, referred to as vectorization, can be performed interactively or in an automated fashion.

The interactive vectorization experience, referred to as raster tracing, requires that you trace the raster cells in the map to create vector features. The automated vectorization experience, referred to as batch vectorization, requires that you generate features for the entire raster based on settings that you specify.

ArcScan also supports tools to help with the preparation of raster data for vectorization. These include tools for selecting and editing raster cells to help you define the scope of the raster to vector conversion.

Who Should Use ArcScan?

Organizations that need to convert raster images into vector-based feature layers are the primary candidates for using the ArcScan extension. Since a large amount of geographic information still exists in the form of hard-copy maps, having a tool to integrate these documents into a GIS is crucial. These legacy documents may be derived from engineering, survey, and cartographic professionals. ArcScan provides an efficient way to streamline this integration when compared to traditional techniques such as digitizing.

How Does ArcScan Work in the ArcGIS Environment?

The ArcScan tool set is an add-on component of the ArcGIS Desktop suite. Licensed as a separate extension, ArcScan works within the ArcMap environment and relies on its own user interface that supports the tools and commands used for the vectorization process. The ArcScan extension is designed to work with the ArcView, ArcEditor, or ArcInfo licenses.

Getting Started With ArcScan

As with other ArcGIS extensions, you must first enable the ArcScan extension in ArcMap before you can use it. The ArcGIS Desktop Administrator will manage the licenses and will notify you if one is not available. You must also add the ArcScan toolbar to your map to access the tools and commands that support the vectorization

work flow. Any map document that is saved with the ArcScan toolbar present will automatically retain the toolbar the next time the document is opened.

Since ArcScan is designed to work with the Editor, you must start an edit session to activate the toolbar. This means that all of the editor tools and commands can be used in conjunction with the ArcScan tools and commands. ArcScan will respect editor settings such as the snapping environment, current edit tasks, and target feature layers.

Symbolizing Raster Layers for ArcScan

ArcScan can vectorize any raster format supported by ArcGIS Desktop as long as it is represented as a bi-level image. This requires that you symbolize raster layers with two unique colors. You can use the Unique Values or Classify rendering option to separate a raster into two colors. Most scanned documents tend to consist of two colors that delineate the foreground and background values. Typically, the foreground is represented as a dark color such as black, and the background is represented as a light color such as white. However, these colors can be reversed or represented by different values. As long as the two colors possess unique values, ArcScan will support vectorization for the current foreground raster cells.

Selecting Connected Cells

ArcScan supports the ability to select connected raster cells. Connected raster cells are raster pixels that share contiguous borders. This can be either in a side-by-side or diagonal arrangement. This functionality allows you to select portions of the raster for various reasons such as vectorization, export, and removal.

Raster selections can help you focus on the important parts of the raster data while allowing you to isolate parts of the data you are not interested in. Raster selections can be performed interactively by clicking a series of raster cells or by a Structured Query Language expression-based query using the Select Connected Cells dialog box. Both methods allow you to create a new selection set, add the current selection set, or remove from the current selection set.

Cleaning Up the Raster Data

ArcScan supports tools to perform simple edits on raster layers. These include a tool called Raster Cleanup, which supports its own edit session. Raster cleanup is sometimes essential for raster layers that contain a lot of unwanted cells, commonly referred to as noise. You can also use the Raster Cleanup tools to add new raster cells to the image.

Once you have performed the edits on the raster, you can save them to the current file, export the edited raster to a new file, or perform the vectorization and discard all changes. Used in combination with the raster selection tools, Raster Cleanup is a powerful function that can make your vectorization experience more efficient and, therefore, increase productivity.

Using the Raster Snapping Tools

The ability to snap to raster cells has been introduced with the ArcScan extension. This functionality allows you to accurately trace raster data to create vector features, which can enhance the interactive vectorization experience. The supported raster snap agents include centerlines, corners, intersections, ends, and solids.

ArcScan relies on the current Editor snap tolerance settings for raster snapping tolerance. Other settings include showing snap tips, which can help you distinguish which part of a raster's connected cell you are snapped to.

Tracing Raster Cells

Interactive vectorization is characterized as the process of manually tracing raster cells. This can be accomplished using the standard Editor sketch tools or the ArcScan

Vectorization Trace tool. Used in conjunction with raster snapping, raster tracing can be an effective and accurate way to convert raster data to vector features.

Depending on the current edit target layer, you have the option to create line or polygon features. You can also control the geometric composition of the output vector features by adjusting the vectorization settings prior to tracing. Once you have added new features to your database by tracing, you can leverage other tools, such as topology, advanced editing, and spatial adjustment, to modify the data, if necessary.

Batch Vectorization

Batch vectorization is defined as an automated technique for converting raster data into vector features. This process relies on user input to control how to perform the vectorization. Factors, such as image resolution, amount of noise in the image, and the actual content of the scanned document, all play a role in determining the outcome of the vectorization.

ArcScan supports two vectorization methods: centerline and outline. Centerline vectorization will generate vector features along the center of the raster linear elements. This method is typically used for vectorizing parcel and survey scanned maps. Outline vectorization will generate vector features at the border of the raster linear elements. This method is typically used for vectorizing land use and vegetation scanned maps.

As previously mentioned, the success of a vectorization may be determined by the state of the scanned document at the time of the conversion. It is sometimes necessary to modify the image prior to the generation of features. This process, referred to as preprocessing, can help you clean up certain portions of the raster that will help define the scope of the vectorization. The Raster Cleanup tool provides the means to perform these steps. In addition, raster selections can be used in combination or independently of raster cleanup to isolate the raster cells you wish to vectorize.

Besides manipulating the original raster, the most influential factor in batch vectorization is the settings. These settings control which cells are vectorized as well as the amount of generalization and smoothing that is applied to the output vector data. You can modify the settings and preview them immediately in the map to see how they will impact the vectorization. Once you have determined the appropriate settings, you can vectorize the entire raster layer or a defined area of it. After data creation, you can use other Editor tools, such as topology, advanced editing, and spatial adjustment, to further refine the data, if necessary.

More Information

For more information on ArcScan for ArcGIS, please visit our Web site at www.esri.com/arcscan.