AN ESRI
WHITE PAPER

JUNE 2018

Data in ArcGIS: User Managed and ArcGIS Managed

380 New York Street
Redlands, California 92373-8100 USA
909 793 2853
info@esri.com
esri.com

esri | THE SCIENCE OF WHERE™

Table of Contents

Data in ArcGIS: User Managed and ArcGIS Managed

Introduction

Your data storage and management infrastructure is a vital aspect of your ArcGIS deployment. It shapes how your organization will access, manage, contribute and edit data, and provides the foundation for how your data interacts with your larger GIS implementation and beyond. Where to store and maintain data is an important choice for ArcGIS administrators, analysts and publishers, and these choices are often reviewed and revised as new options are made available.

In this Esri whitepaper, we will outline two ways data can be handled within the ArcGIS platform. At a high level, these two ways are referred to as *user managed* and *ArcGIS managed*. These are new terms that will be explained further throughout this document. Both approaches, user managed and ArcGIS managed, can be leveraged together in your deployment based on different use cases. As every deployment is unique, it is ultimately your choice in how, and to what extent, you use each of these options.

After reading this whitepaper, you will be equipped with the knowledge you need to determine your approach to data storage in ArcGIS. For users of ArcGIS Enterprise, we will also have explained how the ArcGIS Data Store is **not** a replacement for existing enterprise geodatabases.

Categories of data within ArcGIS

There are two overarching ways to use data with ArcGIS today. One is the approach where data owners and administrators are responsible for *providing and managing* the underlying storage infrastructure such as file servers and database servers. This can be referred to as the *user managed* approach, where the underlying infrastructure and databases must be managed entirely by the administrators of the system.

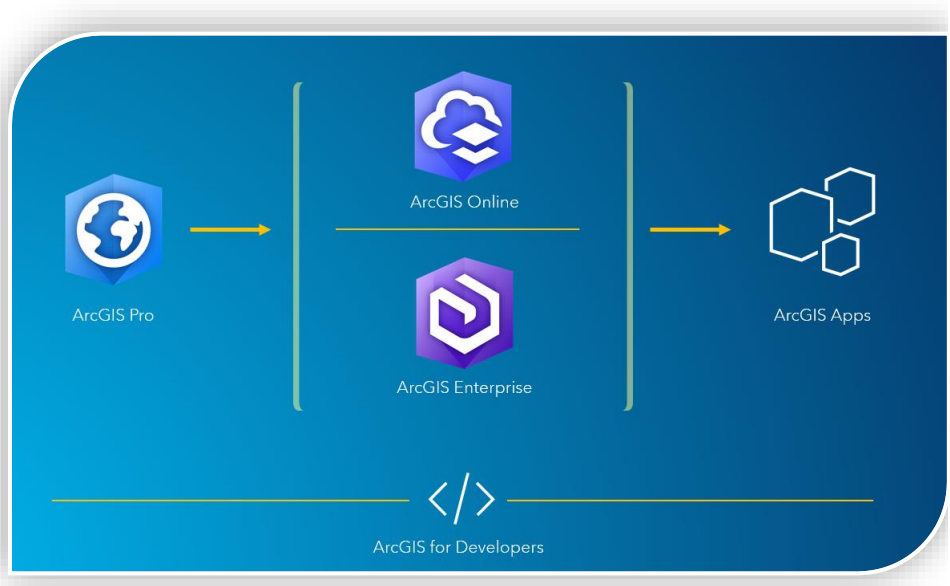
The other is what can be considered the *ArcGIS managed* approach. This approach introduces the concept of hosted data because the data is stored within (hosted by) the ArcGIS system and exposed solely through REST based web services that are shown to end users as layer items within their portal. In this data model, there is no direct access to underlying files or databases by users or administrators. Therefore, this approach is considered ArcGIS managed, since the underlying database is not exposed for direct management by an administrator.

This characteristic–whether or not the administrator has direct access to the underlying data source(s)–is unique between the two approaches.

The modern ArcGIS platform

To understand how data is accessed and used in ArcGIS, it is helpful to first understand the modern ArcGIS platform.

In 2018, ArcGIS is made up of several products that can be used independently, but are designed to be used together, where they provide their full value. The three primary offerings are ArcGIS Desktop (which includes ArcMap and ArcGIS Pro), ArcGIS Online, which is a software-as-a-service cloud-based mapping offering, and ArcGIS Enterprise, which provides a complete mapping and analysis platform running on your own infrastructure on-premises or in the cloud. Supporting these core products is a plethora of applications that run on devices of all types–desktops, smartphones and tablets. The ArcGIS platform also has a rich developer ecosystem, via APIs and SDKs to extend and further develop core products.



The modern ArcGIS platform

Consuming data in ArcGIS

With this understanding in place, let's take a deeper look at how data can be used in the platform and among products.

Within ArcGIS, there are many types of applications, including Esri-built applications, partner applications, and customer-built applications. These applications consume data in different ways depending on whether they are Windows desktop applications, native applications running on a mobile device such as a smartphone or tablet, or a web application running in a browser.

The applications of choice for many GIS analysts are the ArcGIS Desktop applications: ArcMap and ArcGIS Pro. Because they are desktop applications running on a Windows environment, ArcMap and ArcGIS Pro can read data from file-based sources like file geodatabases and shapefiles. They can also directly connect to databases and enterprise geodatabases like those built on Oracle, SQL Server, and PostgreSQL.

Mobile devices and web applications, however, cannot connect directly to these data sources. Instead, these applications must connect through a web service. With today's technology, these web services are made available through a standard referred to as REST (Representational State Transfer), where a server application acts as an intermediary between the underlying data source–the files or databases containing the data–and the client application running in a mobile device or browser. The two offerings that provide the ability to publish web services within the ArcGIS platform are ArcGIS Enterprise and ArcGIS Online.

History of data in ArcGIS

The use of data within ArcGIS has expanded, as its product offerings have, over many years. For decades, the common approach was to use file-based data sources sitting on disk on individual analysts' workstations or on network file shares. These file-based sources have taken many forms over the years including coverages, shapefiles, or more comprehensive formats like the file geodatabase. Many of these formats are still viable data sources today.

As organizations grew, and centralized data management became more of a focus, many found they needed better methods of collaborating on data maintenance tasks, data schemas and definitions. This led a move toward centralized repositories of data rather than files scattered across workstations and multitudes of network shares. These centralized repositories, still prevalent today, are most frequently built on top of relational database systems like Oracle, Microsoft SQL Server, and PostgreSQL that include native spatial data types. Within ArcGIS, these databases are enriched by Esri's enterprise geodatabase technology to provide features and functionality like versioning and topologies. Enterprise geodatabases are sometimes referred to as SDE or ArcSDE databases because of the names from previous generations of the Esri technology that powers this functionality.

Whether file-based or in the form of an enterprise geodatabase, what is common to this approach to data is that management of the data–the underlying location and lifetime of the data–is the responsibility of the data creators and administrators of the system, therefore considered *user managed.*

User managed data and referenced data
Many of these datasets, originating in file-based format or in enterprise geodatabases, are exposed to web and browser-based applications through web services via ArcGIS Enterprise. When data is published to ArcGIS Enterprise, the publisher can leave the data in its source location and to reference it via the web service. Using this option, the web service run by ArcGIS Server (map service, feature service) connects to the original source and no copy is made. If the service, referencing the data, is deleted, the data itself is not deleted. It remains in the source database until it is explicitly deleted in the source.

ArcGIS managed data and hosted data
Now, let's fast forward. With the introduction of ArcGIS Online, a new data paradigm was introduced: the concept of *ArcGIS managed* and *hosted* data. Since ArcGIS Online is a software-as-a-service offering, the data inherently needs to be managed by ArcGIS, not the end user, for it to be available in ArcGIS Online. This means that there is no database that a GIS analyst directly connects to create datasets and web services. Instead, the data *is* the web service–there is no distinction between the two. When the web service is deleted, the underlying data is also deleted automatically.

The idea of ArcGIS managed data sources and hosted data originated in ArcGIS Online as a technical necessity, but also powers another important capability available with both ArcGIS Online and ArcGIS Enterprise: the ability for GIS analysts and non-GIS users alike to publish data without needing to interact with the backend database infrastructure. For all their power and functionality, database systems require the resources to maintain it. Because of this, data stored in these systems are often the most important datasets that require the most advanced capabilities and levels of control.

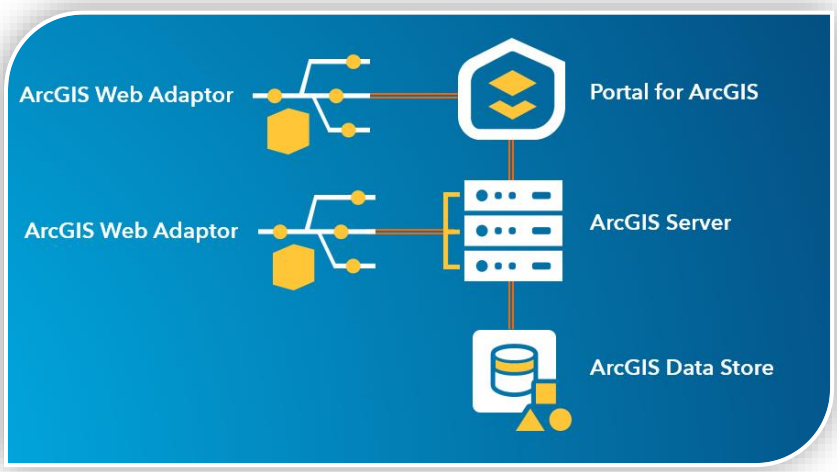
This leaves a gap: how can users quickly and easily take the many datasets available to them–spreadsheets, .CSVs, etc.–and make use of them in modern web-based applications or simply make them available to others in their organization without requesting the support of a database administrator to help deploy the data to an enterprise geodatabase? Thus, began the need for hosted layers in ArcGIS Enterprise.

ArcGIS Data Store

With this background, let's look at ArcGIS Enterprise specifically and the role of the ArcGIS Data Store component.

To support hosted data in ArcGIS Enterprise, there is a backend component that must be installed to store the data in an ArcGIS managed approach. This is the relational data store type* of the ArcGIS Data Store component.

Pictured below is an example of how the ArcGIS Data Store fits into the logical architecture of the base ArcGIS Enterprise deployment.



Here, the relational ArcGIS Data Store has been registered with an ArcGIS GIS Server, and is set as the managed database for the hosting server. The hosting server is federated with the Portal for ArcGIS component. This makes up an ArcGIS Enterprise base deployment.

*Note that the **relational ArcGIS Data Store** is only one of three types of ArcGIS Data Stores in ArcGIS Enterprise. The other two types of the ArcGIS Data Store are the **spatiotemporal ArcGIS Data Store**, designed for storing large amounts of data often with time components, and the **tile cache ArcGIS Data Store**, for storing 3D (scene) layers. The spatiotemporal ArcGIS Data Store is only needed if you are leveraging GeoEvent Server and/or GeoAnalytics Server in your ArcGIS Enterprise deployment.*

Hosted layers in ArcGIS Enterprise

Many common workflows and their subsequent outputs in ArcGIS Enterprise depend on the ability to create hosted layers. Thus, hosted layers are not only created as a direct action from uploading a dataset and explicitly choosing to publish it as a new layer. Hosted layers are also inherently created as the output of many actions within the Enterprise portal, such as running analysis tools and as part of distributed collaboration workflows where feature layers are copied.

Earlier we mentioned one of the most fundamental workflows for hosted layers: the ability to take existing data and quickly visualize and analyze it on a map absent of having to interact with a larger database system. This is what we refer to as self-service mapping, and can be appealing for users with any level of GIS and analysis experience. It enables users to turn their existing data into hosted layers, thus unlocking different features on the layer such as the ability to enable editing, syncing, exporting, deleting and sharing capabilities. It also means that this data creates a web service and can be used in apps and analyzed using a simple web-based interface.

ArcGIS managed or user managed?

It depends on your use cases and intentions for your data. Note it does not have to be either/or, as both options can be used in your deployment together.

As mentioned, some datasets will inherently be created as hosted layers as part of a workflow that generates the data. For example, using the analysis tools available in the Enterprise portal Map Viewer will always store the output results as hosted feature layers. Publishing spreadsheets or .csv files will inherently result in a hosted feature layer.

In other cases, you have a choice. This is the case when working with data in ArcMap or ArcGIS Pro and deciding how to publish your map and associated layers by either copying or referencing data.

Hosted layers are lightweight and scalable. A regular ArcGIS Enterprise deployment can support thousands (even tens of thousands) of hosted layers with little overhead beyond the storage requirements for the data itself. On the other hand, services published with references to data have a higher server cost associated with them. An ArcGIS Server site within an ArcGIS Enterprise deployment can typically serve out hundreds of services at most because of higher memory (RAM) requirements for this type of service.

Thus, some general trends emerge from how to best choose between ArcGIS managed data published as hosted layers and user managed data published as referenced layers. Layers intended for self-service mapping, often created from CSVs, spreadsheets and shapefiles coming from individual contributors' local machines, may not require storage in an enterprise geodatabase. Sometimes these datasets are short lived, used for temporary analysis or as a draft service, and therefore do not need to be incorporated into a larger enterprise geodatabase. These are good candidates for hosted layers, where the data will be ArcGIS managed.

In some ways, it may be easier to flip the question around: assume that the default should be to store data as ArcGIS managed and consider what exceptions exist that would require a user managed data store. The answer is typically driven by a need for specific functionality that is only available with services published from referenced datasets:

- Advanced enterprise geodatabase functionality like versioning, archiving, and topologies.
- Need for direct SQL access and control not available via the REST API. This may include integration with other systems that require SQL-level access.
- Need for a dynamically rendered map service (server-side rendering as opposed to client-side rendering). Some high end cartographic functionality is only supported in map services. One example is use of Maplex labeling for fine-grained control over label placement.

Often there is a tendency for authoritative information that is maintained by an organization or team to be stored in an enterprise geodatabase, whereas the less rigorously maintained data coming from a single individual is an excellent candidate for publishing as a hosted layer.

The below table includes some of the differences between ArcGIS managed and user managed data:

	User managed (enterprise geodatabase)	ArcGIS managed (hosted data in the ArcGIS Data Store)
Database Technology	Option of RDBMS (Oracle, SQL Server, PostgreSQL, Informix, Db2, SAP HANA, etc.)	The database is chosen by ArcGIS; not a bring-your-own-database nor a general-purpose database.
User Access	Multiuser, with both direct connect from desktop applications and via REST-based map and feature services.	Multiuser, solely via REST-based map and feature services.
Rendering	Map services can render the output server-side using map image layers. Both map and feature services can render client-side via feature layers. Map image layers support advanced renderers and cartographic options not available in feature layers.	Hosted feature layers only support client-side rendering with out-of-the-box symbol set and cartographic options.
Versioning and archiving	Supports both traditional versioning and branch versioning. Archiving historical snapshots is supported.	Versioned editing is not supported; the last edit submitted is stored for the feature layer. No archiving.
Topology	Topology rules can be created and enforced.	Topology rules not supported.
Scalability	The underlying RDBMS can be scaled to support a large number of users, editors and data using native RDBMS features and scaling of the server hosting the database. Individual map and feature services that reference user managed data sources can have a substantial memory footprint on the server. Typical configurations can support in the range of hundreds of services.	The relational data store type of ArcGIS Data Store can be scaled vertically; adding more capacity & resources to a single machine. A secondary machine can be added as passive backup and a failover host. Individual hosted feature layers have very low memory footprint on the server. Hundreds to thousands of hosted feature layers can be supported on a standard configuration.
Use as managed database for hosting server	Not supported, however the geodatabase can be registered as an additional data store, or the managed database on a non-hosting server.	Supported – only the relational ArcGIS Data Store can be registered as the managed database for the hosting server site.

Summary and conclusion

When comparing ArcGIS managed and user managed data sources, the core distinction is that in the ArcGIS managed approach, the system manages both the lifetime of the data and the under-the-hood management of the database. In the user managed approach, the lifetime of the data and management of the database is managed by individuals, most often the database administrator. This can be a shift in how many organizations are used to managing their data and databases, and because of this, organizations can and should choose to leverage the benefits of both approaches, depending on the intended use of their specific data.

The second distinction between the two approaches is the behavioral rules and regulations that can be imposed on your data. As mentioned, settings on ArcGIS managed (hosted) items can be imposed to control whether or not the item can be viewed by specific members, deleted, edited, synced, shared, exported and more.

However, a benefit of user managed data in enterprise geodatabases is that administrators can impose organizational-wide rules for managing spatial integrity, quality, accuracy, data relationships and advanced versioning and data editing workflows. In this regard, enterprise geodatabases are often a preferred approach for data that requires systematic vetting processes before exposing as a service. Enterprise geodatabases also allow low-level and fine-grained control over database details such as database indexes and statistics exactly because of the user managed nature of the database. For larger datasets that are the subject of heavy and performance-sensitive queries, many database administrators may prefer to keep them in a user managed database in order to maintain tight control over these kind of details.

When considering how and where to store your data, it is recommended to understand the format, characteristics of your data, and how you want to interact with it. For example, if versioning, archiving, direct SQL access and advanced query support are key functionalities needed for your data, you will almost certainly decide to store and manage your data in an enterprise geodatabase. The enterprise geodatabase also supports certain specific data types: multipatch datasets and mosaic datasets, for example.

However, if you have a dataset that was downloaded from an open data site, for example, and you need to visualize it on a map, hosted data could be a more appropriate approach for that dataset.

Note that your storage choices for your datasets can always change–you can choose to create hosted data out of data that was formerly stored in your enterprise geodatabase, or you can export your hosted data and incorporate it into your enterprise geodatabase. The choice does not have to be permanent, and can be adjusted based on your ongoing organizational needs.

Hosted data can and should be integrated into your GIS, but it is not necessary to use it exclusively. Your organization should continue to take advantage of both the self-service benefits of hosted data and the robust data regulation and management capabilities of a user managed approach, such as the enterprise geodatabase. Again, the options do not need to be used exclusively and it is perfectly appropriate to leverage both in your GIS, based on the specific use cases for your users and your data.

esri | THE SCIENCE OF WHERE™

Esri, the global market leader in geographic information system (GIS) software, offers the most powerful mapping and spatial analytics technology available.

Since 1969, Esri has helped customers unlock the full potential of data to improve operational and business results. Today, Esri software is deployed in more than 350,000 organizations including the world's largest cities, most national governments, 75 percent of Fortune 500 companies, and more than 7,000 colleges and universities. Esri engineers the most advanced solutions for digital transformation, the Internet of Things (IoT), and location analytics to inform the most authoritative maps in the world.

Visit us at esri.com.

Contact Esri

380 New York Street
Redlands, California 92373-8100 USA

1 800 447 9778
T 909 793 2853
F 909 793 5953
info@esri.com
esri.com

Offices worldwide
esri.com/locations

For more information, visit **esri.com/enterprise**.