ABSTRACT

NEWBY, ASHLEY MEREDITH. Versioned Database Deployment in a Web GIS Environment for Land Management Agencies. (Under the direction of Dr. Stacy A. C. Nelson).

Geographic Information Systems (GIS) continue to merge with computer science technologies, yet there is limited scientific literature on utilizing versions of a relational database in a WebGIS environment, or solutions on how to manage federal data. This study was conducted to determine the feasibility of creating and managing an accurate original dataset for land management agencies, while minimizing database anomalies by the use of versioning delivered through the web. How can this framework be designed for ease of use across GIS skill levels? Can versioning be used with web services as a tool to minimize inaccuracies while disseminate data to a wide audience on the web? Different geospatial database frameworks were assessed to meet the needs of land management agencies, specifically the U.S. National Park Service (NPS). Different models were evaluated to establish a framework, then the workflow was tested by subjects to assess the ease of use to set up the system, followed by a survey. The result is a replicable model that is easy to use by those with a variety of GIS skills, as well as two web maps accessing two different versions of the same database. This system allows for concurrent multi-user access to vetted data in a centralized location, while also allowing individual users, teams, or projects to view the original data while being able to make edits to their own view or version. This prevents making automatic changes to the parent database without review and approval by a quality control analyst. However, there are many areas of future work, incorporating survey feedback into the workflow, expanding the system to include delivery on mobile devices, and exploring open web service standards developed by the Open Geospatial Consortium (OGC). This paper presents a first attempt to better inform the way data is collected, stored, and managed within land management agencies.

Versioned Database Deployment in a Web GIS Environment for
Land Management Agencies

by
Ashley Meredith Newby

A thesis submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the degree of
Master of Science

Forestry

Raleigh, North Carolina

2017

APPROVED BY:

Dr. Stacy A. C. Nelson
Committee Chair

Dr. Ranga Raju Vatsavai
Minor Member

Dr. James B. McCarter

Mr. William S. Slocumb

DEDICATION

To my family, and friends who are family.

BIOGRAPHY

Ashley Meredith Newby graduated from UNC Greensboro in 2007 with a Bachelor of Arts in Design and Technical Theatre. After spending some time working at Volvo's North American Headquarters in Greensboro, NC, she began her master's work at North Carolina State University in late 2012 to complete a Master of Geographic Information Science and Technology and continued for a second Master of Science in Forestry and Environmental Resources.

ACKNOWLEDGMENTS

I would like to say a huge thank you to Dr. Stacy Nelson, Bill Slocumb, Dr. Jim McCarter, and Dr. Raju Vatsavai for their time, expertise, and valuable feedback. A special thanks to Dr. Stacy Nelson for always believing in me, even when I did not. I would like to thank the National Park Service for funding and providing the data used in this study. Thank you to all my professors at NCSU who have shared their vast knowledge enabling me to grow, not just academically and professionally, but personally as well. I would like to thank Dr. Stacy Supak who guided me through the world of surveys. Thank you to Laura Thel Belica for the tough love that got me going when I was going to give up, and to Jason Matney who would always be willing to discuss the National Park Service and who inspired me to delve into the history of GIS. Thank you to my friends and family, especially my parents, sister, Michael, and Deeksha for all your moral support and willingness to proof my drafts or preview my presentations. Without all of you, this journey would not have been possible. Thank you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

Literature Review

The internet and spatial information have become a part of everyday life. The power of relational databases combined with the spatial capabilities of Geographic Information Systems (GIS) delivered through web interfaces provide users with tools that were unimaginable at the time the world wide web (WWW) was created in 1990 (Berners-Lee and Fischetti 1999) and have come a long way since the first web map. Al-Shahi et al. (2002) describe the internet as the most important development to communication since telephones and television. The use of GIS on the web, WebGIS, also marks a ‘significant milestone in the history of GIS’ (Fu and Sun 2010). It has greatly changed the way geospatial data is acquired and used. Additionally, the quality and quantity of spatial data, data with that have been geographically referenced, have been improved greatly in the past few decades and will continue to do so (Wang 2010). Access to the internet is inexpensive for the end-user, user friendly and has nearly universal accessibility.

GIS has always integrated technology from disparate sources, and now the internet, the Web, and the Cloud have integrated even more data sources with finer resolutions and in real time (Goodchild 2015). With the emergence of multiple technologies that collect large amounts of geospatial data such as Global Positioning Systems (GPS) and smart phones, we now have access to a wider scope and scale of data that leads to greater assessment areas for all types of studies around the world. In particular the release of large amounts of geospatial data to the public, such as NASA’s Landsat remote sensing data in 2008 that enabled much more analyses to take place due to no cost no longer restricting access to data (Woodcock et al. 2008). This sudden acquisition of geospatial data has been used in a variety of disciplines such as forestry, ecology, archaeology, emergency response, resource management, and public health (Yu et al. 2013; Laffan, Skidmore, and Franklin 2016; Bernardini et al. 2013; Manfré et al. 2012; Kearns, Kelly, and Tuxen 2003; Marx et al. 2014).

These rapid advancements need to be harnessed and managed to ensure quality data is available to internal and external users for reference and analysis. Downs (2015) suggested that digital assets require the same care and curation as physical assets so that they are not only available today, but are preserved for future use, similarly to how a library tracks books in a catalog. 'In an Internet-based world, value reaches beyond simply sharing data, and extends to judging data quality and to determining data fitness for consumption' (Cowen 2009). Data and information are everywhere and accessible at little or no cost. However, to gain insights, easily accessible, quality data is needed.

In order to be called geospatial data, it must have a spatial attribute. Geospatial data are a collection of facts or information that contain spatial elements, location characteristics that are geographic and spatial in nature (Joshi and Joshi 2013). Common spatial data types include aerial photography, maps, satellite imagery, vector data of points, lines, or polygons, continuous raster data based on cells such as temperature or elevation (digital elevation models (DEM)), triangular irregular networks (TIN), remotely sensed images, and light detection and ranging (LiDAR) point clouds, or tables with a spatial attribute such as latitude and longitude. They are usually stored in a Geographic Information System (GIS) with points, lines, and polygons representing vector data, and a grid of cells representing raster data. This gives data and added dimension by being able to visualize them on a map and providing them a relationship to the earth.

Some geospatial data have a time component, known as temporal data (Panigrahi 2009). Temporal changes are important in that they are a representation of agricultural history, habitat loss or gain, and urban sprawl. Change detection allows users to identify differences from one point in time to another. For example, environmental concerns such as oil spills can be tracked and verified using remote sensing data (Brekke and Solberg 2005). Another use of temporal data is to monitor climate change (Skøien et al. 2013; Zhang et al. 2014). Temporal data is particularly good for discovering patterns. For instance, habitats and forests can be monitored for change in vegetation health or loss of area (Zhang and Song 2006). When adding an attribute of time, change detection of the landscape can be visualized and quantified. How land cover change evolved informed about the ecological landscape. Sexton et al. (2013) analyzed 55 Landsat-5 images from 1984 to 2007 creating a time series that showed how the landscape of agriculture, forest, and urban areas increased and decreased from year to year. With the advances of unmanned aerial vehicles (UAVs) growing in prevalence and advancing technologically, data covering larger areas will become available at finer scales (Laffan et al. 2016). The importance of time and location stand out when plotted and analyzed.

There are many uses of these substantial amounts of geospatial data. Analysis types are network analysis directions, least cost path, distance, viewshed, hillshade, map algebra to detect change, point density, suitability analysis winery placement, heat maps, spatial autocorrelation, and clustering. Pattern discovery can be used with spatial statistics to determine how much of the pattern is created by chance, or is there a variable causing the pattern to be statistically significant, such as locating areas of high crime. Spatial analysis, statistics, and ground truthing are utilized in order to gain information from data that can then lead to results and decision making. Geospatial data can be used in a variety of analysis and tools such as classifying land cover, modeling solar radiation, or performing a suitability analysis to locate the best area to plant a winery. Bastin et al. (2013) confirmed accuracy of bird location data by asking experts to ground truth habitat for important bird areas (IBA) through the use of remote sensing data as a way to validate accuracy. International organizations such as WHO use geospatial data about malnutrition in Africa by mapping its prevalence and using spatial statistics to inform the public available through a web-based geoportal (Marx 2014). Geospatial data such as remote sensing and Global Positioning System (GPS) were used for risk and natural disaster management (Manfré 2012). Geospatial data can be used for multiple purposes. For example, soil data was used for wetland observations (Mathiyalagan et al. 2005), but soil data also aids in agriculture to inform farmers where their crops need fertilization. This allows for many advantages of problem solving, knowledge discovery, and finding efficient, sustainable solutions.

Other ways geospatial data are used involves parcel-based GIS for real estate location, valuing, school zoning, and house hunting with Microsoft's Birds Eye View web maps (Cowen 2009). Nelson, Cheruvelil, and Soranno (2006) confirmed that satellite imagery had potential to monitor macrophytes (aquatic plants) by using the Landsat-5 TM sensor in order to provide an alternative way to perform large scale macrophyte surveys that were less expensive and not as difficult. Satellite imagery was highly available and aided in natural disaster relief (Manfré 2012). 3-D models of the Earth to predict sea level rise and improve accuracy of flood plain mapping (Cowen09). In Sandby Borg, Sweden used remote sensing to locate an excavation site (Andrén et al. 2014). One way that utilized and expanded the web was the sharing of research (Berners-Lee and Fischetti 1999; Bera et al. 2015). Dias and Rafael (2007) created a model for a one-stop e-government portal allowing government entities to deliver services and be accessed concurrently between each other with web services, and to the consumers of public services. Pla and Lleopart (2010) incorporated GIS tools and an Oracle RDBMS via OGC standard web services to update city maps, including symbolization automation. Some applications include modeling, such as generating Digital Elevation Model (DEM) surfaces from free LiDAR data to detect archaeological remains of ancient cities (Bernardini et al. 2013). Others used geospatial data for emergency management, data management, and as a planning tool (Manfré 2012).

Geospatial applications can be delivered via desktop, web browsers, or mobile device, and can be collected by scientists or by Volunteered Geographic Information (VGI) as a potential source for more affordable mass data. Traditional approaches to data collection rely on trained scientists paid to go into the field to collect quality data. This unique and emerging way of collecting geospatial data involves public participation to collect data that is then used for analysis. In the work of Goodchild and Li (2012), VGI was free, provided large quantities of data, and had potential to provide real-time information. One study in particular collected sound recordings to create sound maps adding an auditory layer to the visual map (Lin 2015). VGI can also play an important role in disaster management and relief (Haworth and Bruce 2015). With this new widespread accessibility comes with its challenges such as loss of data quality assurance, liability, and security. With VGI data there is no methodology to collect data, it is dependent on the location and willingness of the participants.

While the current state of geospatial data has become increasingly advantageous, there are challenges as with most technology. The most prevalent issue with geospatial data is the lack of compatibility due the different formats the data are collected or stored in, and that they also originate from different sources. Some data might have been taken at different time intervals. Data comes in different formats such as raster, vector, LiDAR point clouds, basemaps, continuous live Global Positioning System (GPS) stream, radio based collars. Also, continuous GPS data are at times not reliable due to the lack of steady performance, or various obstacles such as loss of radio signals, battery life, or death of the animal wearing the tracking device (Dodge et al. 2014). Also, security may keep spatial data from being accessed and used. The data may have different projection and coordinate systems, and many times are in different scales creating challenges.

The current trend in Geospatial Information Science (GIS) continues to shift technologies to an online environment through the World Wide Web (WWW). Bera et al. (2015) state that to-date, the majority of geographical information was processed locally, and described the web as 'a natural extension of a researcher's desktop'. Since the invention of the WWW, it has become an integral part of everyday life connecting users around the world. The integration of the WWW with GIS has massively increased the availability and usage of geospatial data (Zhang et al. 2011). 'The term Web GIS is defined as the media, where GIS is available, using all the techniques, advantages and multimedia facilities of Web.' (Soomro, Zheng, and Pan 1999). Peng and Tsou (2003) described WebGIS as a GIS accessed through the web to integrate, disseminate, and communicate geographic information. At this point in time with technology and the advances of Geographic Information Systems (GIS), computer science, the internet, and the widespread availability of geospatial data, it now lends itself for the ability to aid in planning or making decisions while having a one-stop place to access data or information (Goodchild, Haining, and Wise 1992).

There are many benefits to allowing multiple users to access geospatial data on the internet through WebGIS such as granting access to users who may not have a software license, less human knowledge is needed, multiple users can access data and edit it at the same time, the data has a global reach, and in some cases data can be accessed offline to then be added back to the original data once it is collected in the field. Traditional GIS only benefited certain users who had access to expensive software and hardware, limiting their reach to the public (Karnatak et al. 2010). With the web, GIS extends to users who normally would not have access to it. WebGIS makes expensive software easily accessible and available to the internet audience.

There is no need for additional software or software knowledge, or to download data on the users' side, do not need to download or store massive amounts of data (Cowen 2009). Zhang et al. (2013) supported delivery of data over the web because of the benefit of the end user not having to download the data locally and analysis can take place on the hosting network's server. Skøien et al. (2013) used web services that shared and modeled data regarding climate change threat to an endangered bird species. This shift to online delivery of geospatial data can greatly benefit organizations and land management agencies. When you deliver items on the internet, it makes things cheaper and affordable where only need one license to publish letting users use your services and when it comes to the human factor it has an easier learning curve. WebGIS can reach a wider audience who can use GIS without having a license, thus saving costs, and it can enable users with limited GIS skills to perform analyses, reducing the cost of technical skills to set up the model. Supak et al. (2014) created an open source web mapping application enabling dissemination of marketing data as a decision support tool to multiple users at a reduced cost of financial, technical, and human resources. McLane and Yan (2009) also utilized web map services to disseminate information throughout a corporation so that employees can quickly gain access to valuable data without GIS software and substantial GIS training. A benefit of WebGIS allows planners in diff parts of the country to access the data where it is up to date and all the offices have the latest information. WebGIS is also aids in monitoring and decision support. Tziavos et al. (2016) also developed a framework, using Microsoft SQL Server as the RDBMS behind a web-based GIS system for monitoring the Black Sea.

With all the advantages of WebGIS, there are disadvantages to be considered. Depending on the amount of users accessing published services at a time along with the size of the service, performance may decrease. Security of sensitive information would have to be considered due to transmission of data over the internet, as well as prevent intentional or accidental modifications (Joshi et al. 2001). McLane and Yan (2009) used multiple ways of controlling security in their enterprise GIS: operating system and database authentication, access control to certain web services based on certain users, and encryption. Nevertheless, the benefits outweigh the cost, and WebGIS should be considered as a vital resource.

In order to meet these challenges, many studies have approached them in different ways. Zhang et al. (2011) used Flash WebGIS to increase speed and graphics rendering of WebGIS applications for better data analysis performance. Complex models can be accessed by users, such as decision makers, with little or no programming experience as in the work of Mair et al. (2014) who simplified model usage by delivering it on the web making it easily accessed and used. Best et al. (2007) limited spatial bias in their work with predicting marine mammal habitats by utilizing Open Geospatial Consortium (OGC) Web Feature Services (WFS) to store and serve related spatial information used in models as part of a scientific workflow. By having vetted data standardized to be used in models and easily accessible through a web-based spatial decision support system (SDSS), provides decision makers access to information they need, without the needing to know how to code. WebGIS should be considered as a vital resource with advantageous benefits as a wide audience, low costs, multi-user access, and users do not have to download data or need programming skills. WebGIS expands traditional desktop GIS from a small group of users to a multi-user reach by delivery of geospatial data and tools through the web for data sharing and decision making, taking away previous limitations (Krishna et al. 2015).

When dealing with web services, there are standard protocols. The Open Geospatial Consortium (OGC) is an international not for profit organization committed to making quality open standards for the global geospatial community (Welcome to the OGC 2017). Krishna et al. (2015) developed an enterprise WebGIS as a business solution using OGC standards in order to have interoperability. There are also other limitations such as extensive customized programming needed, and lack of flexibility where the service must adhere to the standard (Rautenbach, Coetzee, and Iwaniak 2013). OGC standards were considered for this project, but they were not compatible to be edited within the ArcGIS platform and takes away from the ease of use. Skøien et al. (2013) used OCG standard Web Processing Services (WPS) to allow users to model climate change on the web. Zhang et al. (2013) used OGC WFS standards when trying to improve query performance for emergency response during disasters. They found that using a parallel approach does reduce computation time of complex analyses that are needed for such emergency situations. There are also other limitations to OGC standards such as extensive customized programming needed, and lack of flexibility where the service must adhere to the standard (Rautenbach, Coetzee, and Iwaniak 2013). The work of Krishna et al. (2015) focused on harnessing the recent advancement in internet mapping technology, geospatial data management, and GIS API's for business applications in what they propose as a WebGIS framework instead of WebGIS technology since a framework amalgamates different technology architectures. Krishna et al. (2015) used web services standards such as OGC as well as Musinguzi, Bax, and Tickodri-Togboa (2004), Hussein, (2005) and Milosavljević, Stoimenov, and Đorđević-Kajan (2005), so that open standards are adhered to in order to extend the WebGIS system functionality beyond mapping and querying.

WebGIS has a fairly young history, yet it has made a huge impact in a short amount of time. WebGIS would not exist without the internet. Traditionally GIS relied mainly on client server architecture and the client (e.g. web browser or mobile device) and GIS server integration has remained generally the same. In contrast, the geospatial data, visualization, and analysis capabilities have dramatically improved. Initially when web maps first emerged they had limited capabilities. For example, the simple Xerox PARC Map Viewer created in 1993 returned query results of crude, pixelated images of maps from a geographic database (Putz 1994). About a decade later in the 2000s, GIS on the web progressed to more than just viewing geospatial data, it was a way to access and edit data in a database. One example of this was an interactive WebGIS geodatabase for Florida's wetlands based on a Microsoft Access database with viewing and editing capabilities accessed through a web browser (Mathiyalagan 2005). As the size and amount of geospatial data increased, the need for databases larger than MS Access could handle increased as well. The MS Access database has a limitation by the number of records contained in the database, hence the move to a relational database management system (RDBMS) since it could handle a greater volume of data. For example, Karnatak et al. (2010) used an Oracle RDBMS to provide spatial decision support to experts about biodiversity in India. As the size and types of data grew and advanced, improvement to GIS and database technologies continually adapted to the volume and different types of data.

The World Wide Web and the internet have been used interchangeably, however the internet came much earlier as a product of the U.S. funded Advanced Research Projects Agency (ARPA) originally developed to link scholars to one another starting in 1969 with the launch of ARPANET (Hafner & Lyon 1998). By 1973 a form of email started, then by the mid-1980s the term ‘internet’ was used, followed by domain creation of: net, gov, mil, int, org, com, and edu (Hafner & Lyon 1998). In 1989 the National States Geographic Information Council (NSGIC) convened and established a forum for coordinating GIS projects and government investments as it is one of the more active proponents of spatial data infrastructure projects (Cowen 2009), now almost every state has a state GIS coordinator. It was not until 1990 that the WWW was invented by Tim Berners-Lee as a way to share physics research in a consistent format at the European Centre for Nuclear Research (CERN) (Berners-Lee and Fischetti 1999). In 2005 Google Earth released and changed the way geospatial data is perceived (Cowen09). The robustness of relational databases and GIS delivered on the web can greatly aid in data management.

Sometimes databases are accessed through the web without using a GIS component. Das et al. (2012) disseminated clinical study data through a web-based data management system that integrated all data types from images, neurology metrics, genetic biomarkers and importance of different types of data being easily linked and downloaded directly from a web browser. Joshi et al. (2001) looked at database federation approach that integrated multiple database systems that allowed autonomous control at the individual database level while at the same time being accessible at the federation level.

With the availability and the benefits of geospatial data, the ability to contain and access data from a centralized location, one powerful tool of databases to utilize is versioning capabilities on the web. Now that data is available on the web, connected to databases, one important feature is to version the data disseminated on the web from databases. Versions are beneficial for workflow management in enterprise ArcSDE geodatabases, such as modeling different stages of a GIS project and testing what-if scenarios without affecting the original database. They provide a framework for security management and quality assurance in data editing, and they also support historical archiving.

Multiple data editors can access the database at the same time making edits without risk of corrupting the original data or data collected by another user. A quality control analyst could provide a level of security by employing the use of versioning. Versioning tracks changes to a database without automatically changing the original data, thus providing an official dataset. Versions can be beneficial for workflow management by separating projects, modeling what-if scenarios, provide security, quality assurance, and historical archiving (Law 2010). As defined by Bakalov et al. (2011), 'a version can be logically viewed as an alternative, independent, persistent view of the database that does not involve creating a copy of the actual data'. Bakalov et al. (2011) also describe versioning as 'a logical entity that represents a unique, seamless view of the database that is distinguished from other versions by the particular set of edits made to the version since it was created.'(2011).

There are many benefits of versioning: it supports concurrent users, prevents data duplication thus reducing storage requirements, it permits a disposable testing environment (Wall and Angryk 2015), provides database archiving abilities by time or a specific project, manages workflows, ensures data quality to prevent anomalies, has the ability to undo and redo edits, it allows users to explore alternative project ideas during a design phase, and supports edits over long-term, all without affecting the original published database (Bakalov et al. 2011). When multiple developers needed to perform testing on the same database, conflicts arose, or data needed to be copied, or code was needed to restore the original database after testing creating longer testing time, especially when the dataset was large; versioning solved this problem by giving a version to each developer that could be deleted after use (Chatterjee et al. 2004). Chatterjee et al. (2004) used versioning as a testing and development aid for database applications commonly accessed over the web such as financial services, but did not have a GIS component. Wall and Angryk (2015) used versioning in a cloud environment, but without a GIS component. Bakalov et al. (2011) combine versioning with GIS, but do not deliver the versioned data over the web for a wider audience to access. Versioning can be a significant savings of storage, server resources, and time (Wall and Angryk 2015). It allows multi-user access to a database without locking the data for days or weeks during the editing process as with traditional database management systems, preventing others from making edits until the first user has saved. Versioning allows users to access the database as if they had sole access to the dataset (Bakalov et al. 2011).

Each version cannot see other versions, thus giving the data another layer of protection so users of one version cannot harm other versions with unauthorized edits or deletions. This extra level of protection is one more way to ensure data quality. Anselma et al. (2013) used versioning of a relational database as a way to evaluate proposed updates to medial knowledge. Generally schema matching accessing the database through the command line is the computer science approach. When dealing with data warehouses, versioning preserves the history of the database, avoids information loss and simulates hypothetical business scenarios; traditional static data warehouses are not adaptable to ever changing real world data (Saroha and Gosain 2014). Future work is to develop a framework for data warehouse versioning that overcomes these challenges. Wall and Angryk (2015) looked at the value of using a MySQL RDBMS based versioning system in the cloud in spite of the trend towards NoSQL databases. Processing in the cloud allows for data to be changed and test online, without the user creating a copy of the database.

Along with this growth of technology and data acquisition came challenges for many domains, in particular, federal agencies. In a congressional research service report, Folger (2012) outlined a laundry list of issues facing federal government data such as duplicate data collection between local, state, and federal levels; data coming from different sources, different quality intended for specific objectives; the large volume of data at a national level; and incompatibility of the collected geospatial information. With the widespread access to open (free) data in a multitude of locations it has increased the requirement for geospatial data integrity, management, and preservation so that the data can be discovered, accessed, and utilized. However there are many challenges such as different sources, need metadata, size, different scales and spatial reference. These issues can be very challenging when approximately 80% of federal government data contained a geospatial component (Folger, 2012). Smith et al. (2015) concluded that 'Until most federal geospatial data management systems are interoperable, federal agencies could benefit from the development of tools and workflows enabling them to manage and deliver geospatial data in the most efficient manner possible'. Versioning does not prevent conflicting edits, however, it does allow for detection and resolution (Bakalov et al. 2009), showing another reason to have a Quality Assurance/Quality control (QA/QC) level of data quality. The next step is to incorporate versioning on the web.

CHAPTER II

Versioned Database Deployment in a Web GIS Environment for Land Management Agencies

1.0 Abstract

Geographic Information Systems (GIS) continue to merge with computer science technologies, yet there is limited scientific literature on utilizing versions of a database in a WebGIS environment, or solutions on how to manage federal data. This study was conducted to determine the feasibility of creating and managing an accurate original dataset for land management agencies, while minimizing database anomalies by the use of versioning delivered through the web. Different geospatial database frameworks were assessed to meet the needs of land management agencies, specifically the U.S. National Park Service (NPS). Different models were evaluated to establish a framework, then the workflow was tested by GIS users to assess the ease of use to set up the system guided by a tutorial, and then filled out a survey regarding the process. The result is a replicable framework that is easy to use by users with a variety of GIS skills, as well as two web maps accessing two different versions of the same database. This system allows for concurrent multi-user access to vetted data in a centralized location allowing individual users, teams, or projects to view the original data while being able to make edits to their own version. This prevents automatic changes to the parent database without review and approval by a quality control analyst. However, there are many areas of future work such as incorporating survey feedback into the framework tutorial, expanding the system to include delivery on mobile devices, and exploring open web

mapping service standards developed by the Open Geospatial Consortium (OGC). This paper presents a first attempt to better inform the way data is collected, stored, and managed within land management agencies.

Keywords: data quality, geospatial, land management agencies, RDMS, versioning, WebGIS

2.0 Introduction

The internet and spatial information have become a part of everyday life. The power of relational databases combined with the spatial capabilities of Geographic Information Systems (GIS) delivered through web interfaces provide users with tools that were unimaginable at the time the world wide web (WWW) was created in 1990 (Berners-Lee and Fischetti 1999), and have come a long way since the first web map. Al-Shahi et al. (2002) describe the internet as the most important development to communication since telephones and television (2002). The internet is inexpensive for the end-user, and user friendly with nearly universal accessibility. Also, the use of GIS on the web, WebGIS, marks a 'significant milestone in the history of GIS' (Fu and Sun 2010). It has greatly changed the way geospatial data is acquired and used. Additionally, the quality and quantity of spatial data, data with that have been geographically referenced, have been improved greatly in the past few decades, and will continue to do so (Wang 2015). This rapid acquisition of geospatial data has been used in a variety of disciplines such as forestry, ecology, archaeology, emergency response, and resource management, and public health (Yu et al. 2013; Nelson, Cheruvelil, and Soranno 2006; Laffan, Skidmore, and Franklin 2016; Bernardini et al. 2013; Manfré et al. 2012; Kearns, Kelly, and Tuxen 2003; Marx et al. 2014). The growth of the internet and the use of GIS is shifting to an online environment allowing users to access GIS information from anywhere in the world with an internet connection. Downs (2015) suggests that digital assets require the same care and curation as physical assets so that they are not only available today, but are preserved for future use. These rapid advancements need to be harnessed and managed to ensure quality data is available to internal and external users for reference and analysis.

The web continually progresses with new ways to collect and share information. The client server integration has remained generally the same. In contrast, the geospatial data, visualization, and analysis capabilities have dramatically improved. Initially when web maps first emerged they had limited capabilities. For example, the simple Xerox PARC Map Viewer created in 1993 returned query results of crude, pixelated images of maps from a geographic database (Putz 1994). About a decade later in the early 2000s, GIS on the web progressed to more than just viewing geospatial data, it was a way to access and edit data in a database. One example of this was an interactive WebGIS geodatabase for Florida's wetlands based on a Microsoft (MS) Access database with viewing and editing capabilities available through a web browser (Mathiyalagan 2005). Online access to data and maps in one location allowed for customization, reduced costs, maximized funding resources by avoiding redundant projects, enabled continuous documentation of observations, supported decision making, increased ease of use for users with limited GIS knowledge, and was independent of platforms and operating systems (Mathiyalagan 2005). However, as the size and amount of geospatial data increased, so did the need for larger databases. The MS Access database had a limitation of physical storage size in the database, hence the move to a relational database management system (RDBMS) capable of handling a greater volume of data. For example, Karnatak et al. (2010) used an Oracle RDBMS to provide spatial decision support to experts about biodiversity in India. As the size and types of data grew and advanced, improvement to GIS and database technologies continually adapted to the increased volume and different data types.

There are many terms that describe accessing geospatial datasets shared online such as Internet GIS, GIS online, Distributed Geographic Information, Web-based GIS, or WebGIS (Mathiyalagan 2005). Each term has its nuances, but to avoid confusion, this paper will utilize Peng and Tsou description of WebGIS as a GIS accessed through the web to integrate, disseminate, and communicate geographic information (2003). This shift to online delivery of geospatial data can greatly benefit land management agencies. WebGIS can reach a wider audience who can use GIS without having multiple licenses, thus saving costs, and it can enable users with limited GIS skills to perform analyses, reducing the cost of technical skills to set up the model. Supak et al. (2014) created an open source web mapping application enabling dissemination of marketing data to multiple users at a reduced cost of financial, technical, and human resources. This study utilizes these benefits of WebGIS for the NPS.

Adherence to data quality norms to store data and metadata allowed the use of Volunteered Geographic Information (VGI). This involved public participation to use mobile applications to collect data that is then used for analysis. This type of data was free, had large quantities, and provided real-time information (Goodchild and Li 2012). There was a level of skepticism regarding the use of VGI (Joshi and Joshi 2013), but just as remote sensing data can be used by archaeologists to pinpoint where to focus excavation efforts (Bernardini et al. 2013), VGI can be used as a way to pinpoint where to focus more traditional scientific methods. The NPS does not always have dedicated staff updating GIS information, so allowing VGI edits to the database, such as trail updates by maintenance staff after a trail was moved (Irwin 2016) saves time, funding, and energy by utilizing this VGI data. For data quality concerns with VGI, a quality control analyst could provide a level of security by employing the use of versioning. 'A version can be logically viewed as an alternative, independent, persistent view of the database that does not involve creating a copy of the actual data' (Bakalov et al. 2011). Versioning takes place at the individual feature class (i.e., spatial data table of the same data type) or feature dataset (i.e., a collection of related feature classes in the same coordinate system) level. Versioning tracks changes to a database without automatically altering the parent data, thus maintaining a secure and authoritative dataset. Similarly, the open source version control system, GitHub, allows developers to download parent code to modify, then saves the modifications as an edit (Brown 2014). Versions can be beneficial for workflow management by separating projects, modeling what-if scenarios, providing security, assuring quality assurance, and enabling historical archiving (Law 2010).

Recent research employed the use of parallel frameworks that divided computations out to many servers. While the work of Zhang, Zhao, and Li (2013) improved the query performance on web feature services (WFS) (e.g. geospatial data that can be edited through the web) by using parallel processes, those processes focused more on efficiency for disaster response, and less on data integrity. Since the focus of parallel frameworks was not on data integrity, a RDBMS was used for this study instead. However, Zhang, Zhao, and Li (2013) did support delivery of data over the web because of the benefit of the end user not having to download the data locally and analysis can take place on the hosting network's server. This would benefit national parks to be able to access and analyze data over the web.

Along with this growth of technology and data acquisition came some challenges. In a congressional research service report, Folger (2012) outlined a laundry list of issues facing federal government data such as duplicate data collection between local, state, and federal levels; data coming from different sources, varying in quality for specific objectives; the very large volume of data at a national level; and incompatibility of the collected geospatial information. These issues can be very challenging when approximately 80% of federal government data contains a geospatial component (Folger 2012).

One of the challenges of maintaining geospatial data is the sheer volume at a nationwide level. The U.S. National Park Service (NPS), contains 417 official park units and it would not be feasible to store all the data from all the parks in a single database. The NPS needs a central geodatabase for each park to store and edit data where it is safe from accidental deletion, inaccurate updates, or unauthorized use. There is a need for the design and development of an advanced multi-user spatial database framework for land management agencies where multiple users have the ability to access and edit geospatial data in a centralized location, through a web environment, while minimizing inaccuracies (Smith et al. 2015). This framework should allow ease of use for users with little GIS experience.

To approach these needs, this paper is modeled after the methods of Krishna et al. (2015) that developed and implemented an easy-to-use, cost-effective enterprise (i.e., multi-user) WebGIS solution for businesses by: designing and developing a model, creating a database, populating the database with geospatial data, developing the application, then perform testing and deployment. To build on these steps, the underlying RDBMS will take advantage of the benefits of versioning which are: that it supports concurrent users, prevents data duplication thus reducing storage requirements, it allows for a disposable testing environment (Wall and Angryk 2015), it provides database archiving abilities by time or a specific project, it manages workflows, and it ensures data quality to prevent anomalies (Bakalov et al. 2011). In order to meet this challenge, this study aims to approach data collection and maintenance at the individual park level within the NPS. This approach was corroborated by former NPS employee who stated that the best park data is at the individual park level (Nate Irwin, personal communication, October 13, 2016). Focusing on a local scale to collect accurate and comprehensive data at each park, in a central location, and collected using the same methods, is the first step to having a quality controlled dataset of original data for all the parks within the NPS. The need for a more centralized approach is established in the work of Smith et al. (2015) where the NPS needs park data in an enterprise geodatabase, centrally located, and populated with an accurate, official dataset deliverable through the web. Tziavos et al. (2016) also developed a similar framework, but used Microsoft SQL Server as the RDBMS behind a WebGIS system for monitoring the Black Sea, but again, the benefits of versioning were not utilized. Chatterjee et al. (2004) used versioning as a testing and development aid for database applications commonly accessed over the web such as financial services, but did not have a GIS component. This project will utilize proprietary GIS software using Esri’s platform along with an open source RDBMS to address these needs.

The literature supports the need for developing access to versioned feature services based on an enterprise geodatabase delivered in web maps. With a majority of federal data having a spatial component, the need to standardize data collection and store it in one place is very important to save money, avoid redundancy, utilize geospatial data, and reach a wider audience. The goal is to accomplish a repeatable methodology that will transfer to other land management agencies with similar needs.

3.0 Materials and Methods

3.1 Study Area

The study area was George Washington Birthplace National Monument (GEWA), a park unit of the NPS, located on the coast of Virginia south of Washington D.C. (Figure 1). The data used includes vector data shapefiles: park boundary, buildings, small scale features, roads, fences, streams, water bodies, cultural vegetation, and natural vegetation. The spatial reference of the data was NAD83 UTM Zone 18N, and was provided from the National Park Service.

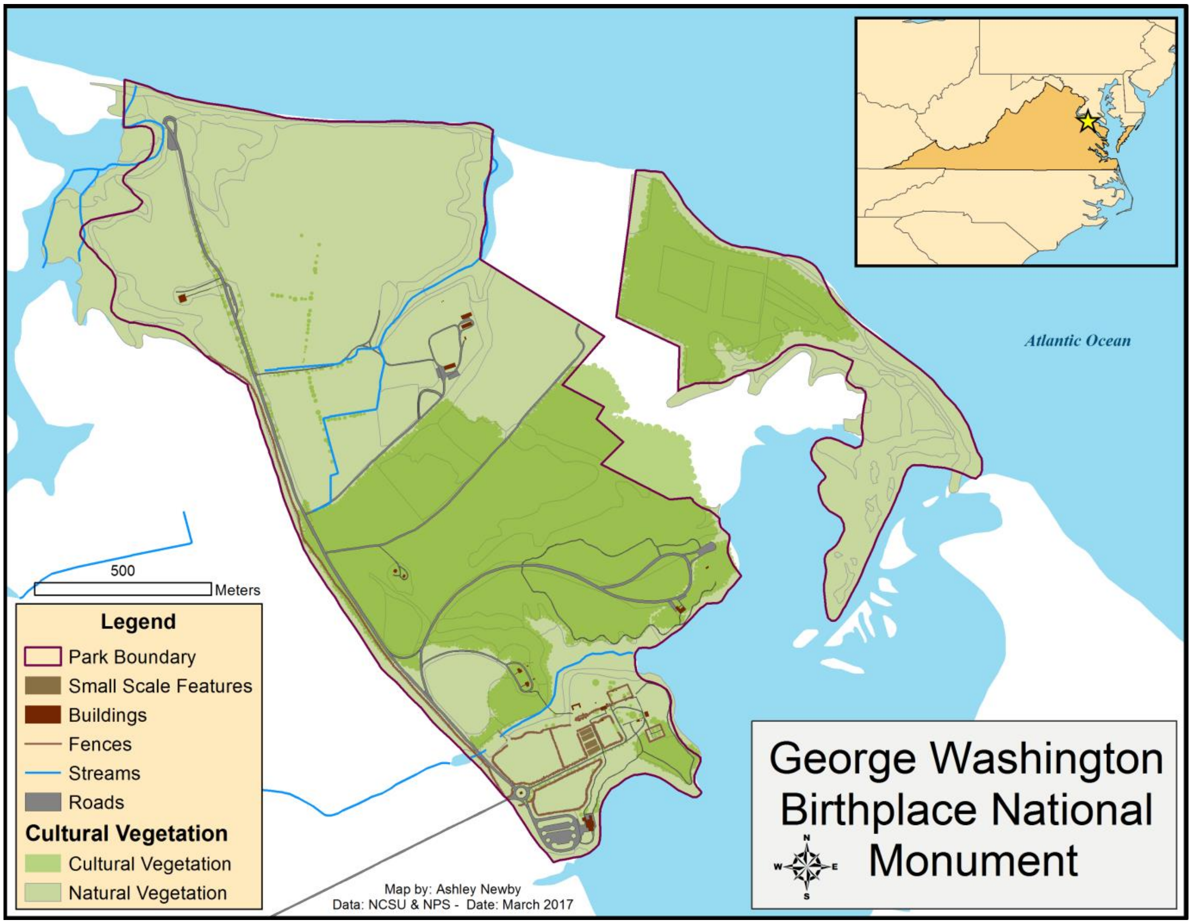


Figure 1: Map of the study area showing base data layers of GEWA

3.2 Software and Hardware Tools

For this study, proprietary software (Esri's ArcGIS for Desktop 10.4.1, ArcGIS for Server 10.4.1, ArcGIS Web Adaptor for IIS, and ArcGIS Online for Organizations) and open source RDBMS software (PostgreSQL 9.4.5) were used. SAS 9.4 was used for the statistical analyses. Process Explorer 16.2 was used for performance testing of the hardware. The software was installed on a Windows Server 2012R2 virtual machine hosted on North Carolina State University's Virtual Computing Lab (VCL) containing 8.75 GB of memory with six 2.0 GHz processors.

3.3 Framework

A workflow of how to set up the proposed framework with end products of two web maps connected to two separate versions for editing was developed. Figure 2 presents three potential models of incorporating versioning. Each model contained a parent version. For Esri enterprise geodatabases, the parent version is called 'DEFAULT' version. As versions were created, only one dataset was maintained in the database since separate tables, also known as delta tables, track the changes made to each version, leaving the parent data unchanged (Law 2010). This saves space in the geodatabase since data is not copied.

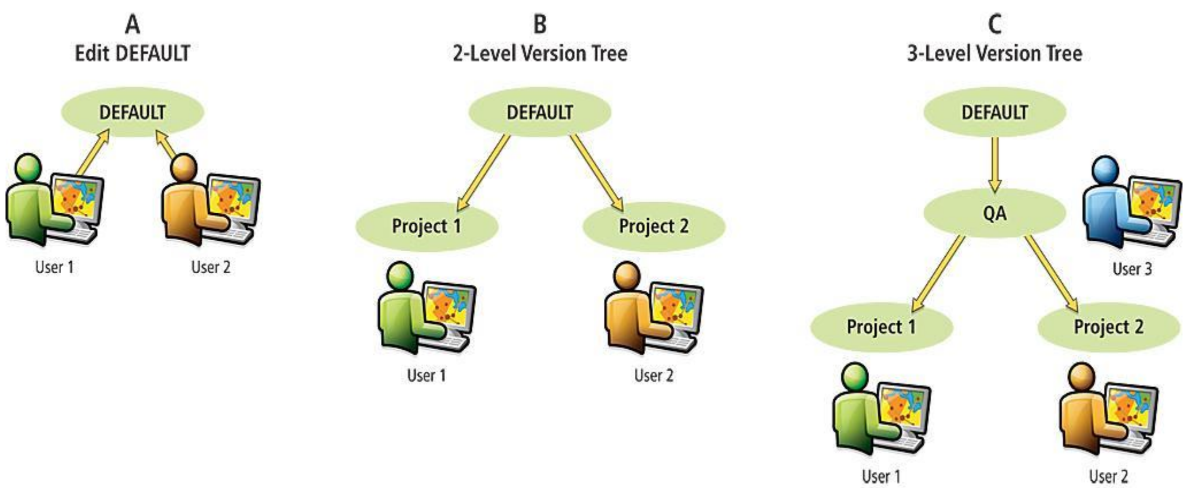


Figure 2: Potential versioning models (Law 2010)

Each model had advantages and disadvantages. Model A (Figure 3) was the simplest workflow allowing multiple users to edit the same default database concurrently. The advantage of this model was that edits were seen instantaneously, however, editing was hindered depending on when individual users save their edits since it created locks on the database. When one user starts to commit edits to the database, all other users are prevented from completing their task. This model also does not prevent unauthorized edits or allow for project data to remain private to each team, making Model A not very useful for land management agencies.

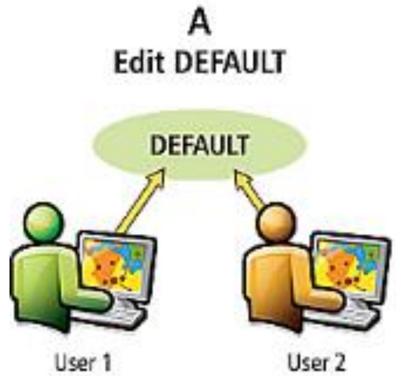


Figure 3: Model A using concurrent multi-user access (Law 2010)

Model B (Figure 4) resolved the issue of locks on the database by creating child versions that allowed editing over a long period of time with multi-user accessibility. Also, edits to each version could be kept private until the project was complete.

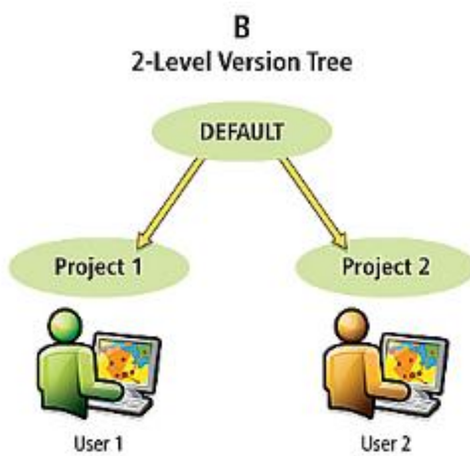


Figure 4: Model B utilizing versions (Law 2010)

The disadvantage was that there was still a lack of quality control and users could add and delete data in the parent database without any supervision. Also, adding data to the parent version with the default user without an administrator role could potentially corrupt the database by adding data to the same folder (schema) where the system tables are stored. An alternative way to protect the data was to make the parent version have protected access. However, there would still require a user to add the edits to the parent version anyway, hence the design of Model C (Figure 5).

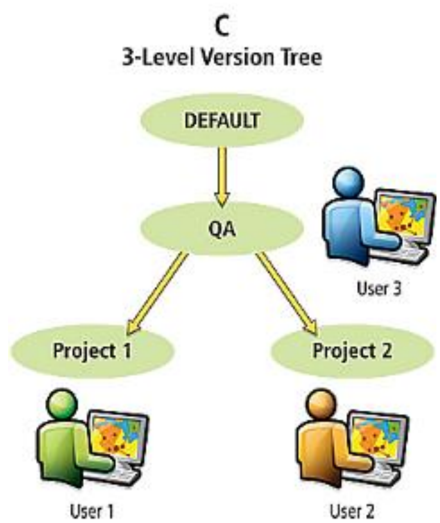


Figure 5: Model C utilizing versions with a QA/QC level of security and quality control (Law 2010)

Model C supports complex projects, and is an extension of Model B by creating child versions of the child versions (grandchildren nodes). This model is good for large projects, different teams, or numerous units of work. Versions can be divided by geographic location (city, state, country), can have multiple functions (collecting fish data vs. odonate data), can track different phases (seasons, months), or it can track the history of changes by creating an archive version of how the production database looked at a certain point in time. This model employs the use of a Quality Assurance/Quality Control (QA/QC) analyst to review edits before they are added to the parent geodatabase. The QA/QC analyst role does not require park personnel to have an expert level of GIS experience. They might need to attend a training session, but the system was designed so that users with limited GIS experience could set up a versioning framework.

The disadvantages of this model were that query performance may be slower due to the addition of delta tables in the geodatabase, edits are not seen immediately since a QA/QC analyst has to review them, and edits have to be added in the order of the version tree (i.e., grandchild versions had to save edits to the child version, then the child version could save to the parent version). In spite of the disadvantages, Model C allows for a data quality analyst to regulate all the edits added to the default data, thus minimizing data anomalies and enforcing quality control, making Model C the preliminary basis for this study.

Replication was explored as a way to edit data offline, but it was not pursued further since replication is designed to work between separate geodatabases to make their data identical to each other. Also, replication does not support a parent copy of the data, the data is not stored in a centralized location, and it does not have a QA/QC level of quality control. Replication is more suited for scenarios where a company with offices on opposite coasts of the United States need local access to the database so the data does not have to be accessed over the network. It also uses more storage since the database is copied.

Once the best framework was decided, a tutorial (Appendix A) was created to distribute to GIS users at North Carolina State University with a range of GIS skills to test the ease of setting up the framework. The assumptions were that PostgreSQL, ArcGIS for Desktop, ArcGIS for Server, and ArcGIS Web Adaptor for IIS were already installed on a virtual machine. PostgreSQL was used for the RDBMS since it adhered to the Open Geospatial Consortium (OGC) standards, free, and was a compatible option to work with ArcGIS for Server.

3.4 Survey

Ten GIS users with various levels of expertise were provided the tutorial, and then asked to complete a survey (Appendix B). This anonymous survey was administered online to test the ease of use of the workflow to set up the versioning framework. The survey went through the Institutional Review Board for the Protection of Human Subjects in Research (IRB) approval process at North Carolina State University and was deemed exempt due to the anonymity of the survey. A link to the online survey was supplied to GIS users to complete after finishing the tutorial. These users had at least a beginner GIS level of expertise, up to advanced users. The survey questions gathered information about the users' GIS background, their impressions of the tutorial using a five level Likert scale, and free text boxes for feedback and comments. After the survey was completed, the results were analyzed statistically using the chi-square test for independence to test if ease of use was independent from skill level. However, due to the small sample size, the initial analysis was followed up with Fisher's exact test.

3.5 Methodology

The multi-user geodatabase framework was comprised of the following steps using spatial data from GEWA, and expanded on Krishna et al.'s (2015) WebGIS result by adding published services from different versions of the same enterprise geodatabase:

1. Select and design a multi-user versioned data model.
2. Build an enterprise level spatial Relational Database Management System (RDBMS) for multiple users.
3. Populate the geodatabase with spatial data.
4. Check PostgreSQL and ArcGIS processes and CPU usage for performance.
5. Publish ESRI Feature Services from the data in the database connected to different versions.
6. Check PostgreSQL and ArcGIS processes again for changes in performance after publishing services from versions.
7. Create web maps from versions.

The detailed workflow for creating the versioned WebGIS framework is shown in Figure 6. Machine performance was also assessed with Process Explorer software to see if versioned feature services would use a significantly higher amount of resources on the virtual machine compared to publishing non-versioned feature services.

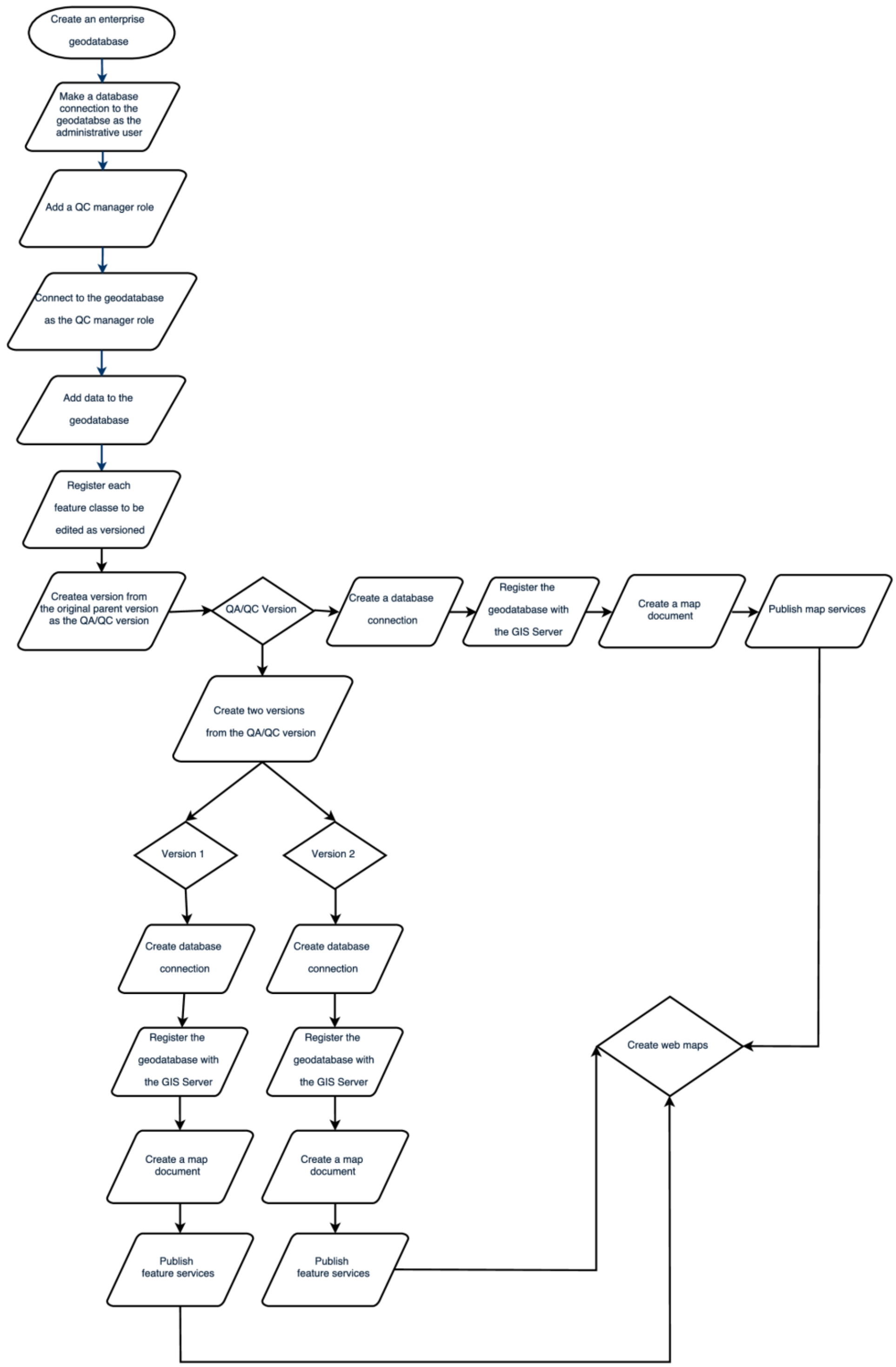


Figure 6: Workflow to create web maps from versioned data

4.0 Results and Discussion

This study created and tested a database development framework for ease of use and efficiency with the following three results: a concurrent multi-user database framework that incorporates the use of database versioning, two web maps containing viewable only layers (i.e., map services) or editable layers (i.e., feature services) published from a server and disseminated in ArcGIS Online, and a comparative analysis of efficiency and ease of use. This best available standardized framework protocol will allow GEWA, and other individual parks, to manage their data more efficiently.

4.1 Multi-User Database Framework

The best spatial relational database management system (SRDBMS) framework for users with limited GIS experience, while also minimizing data anomalies by restricting users' editing capabilities was created (Figure 7). The framework consists of an ArcGIS enterprise geodatabase using PostgreSQL for the RDBMS, a GIS server on a virtual machine, a Quality Assurance/Quality Control (QA/QC) analyst role, and ArcGIS Online for Organizations web maps.

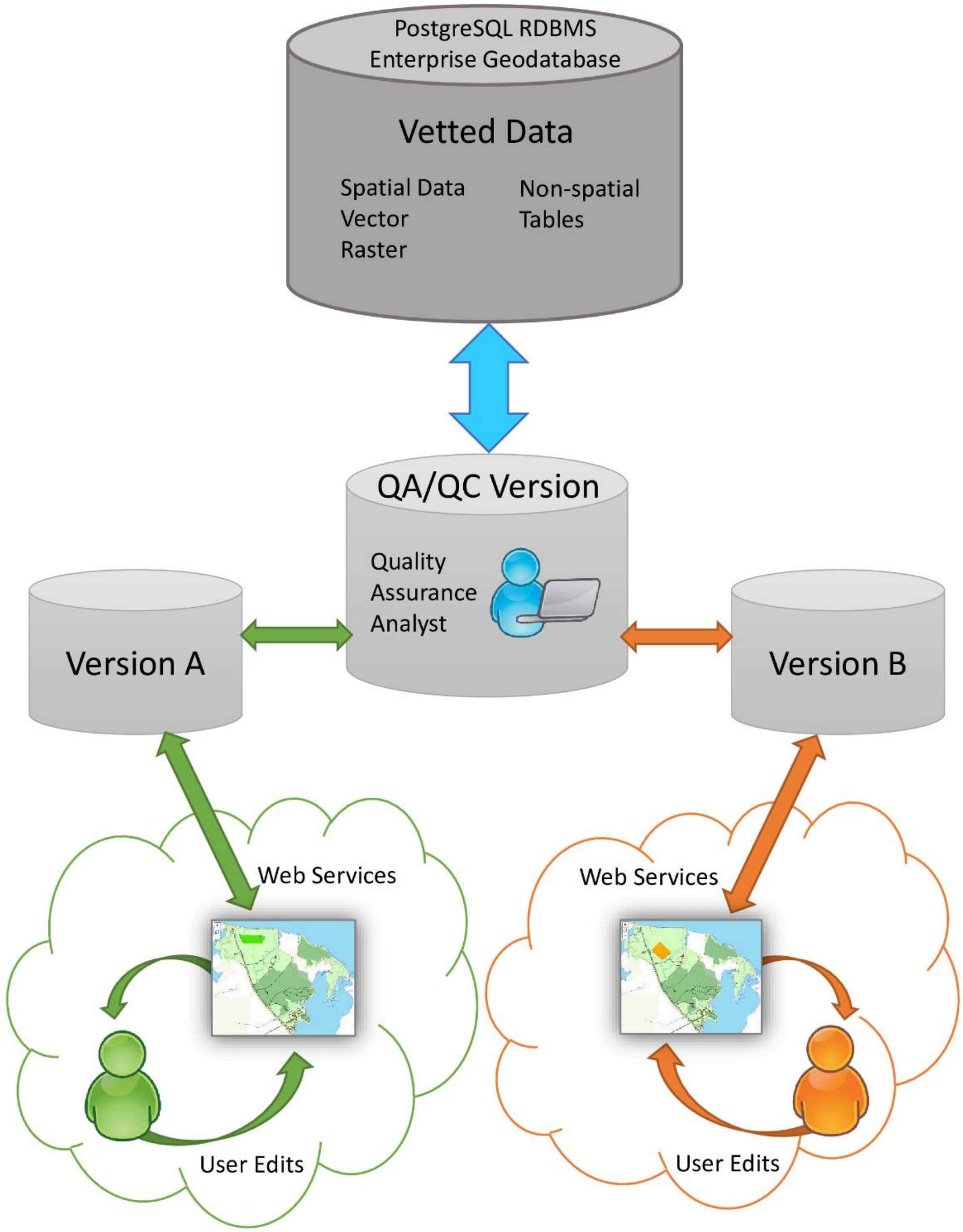


Figure 7: The database framework for versioned data delivered through the web

Using ArcMap, an enterprise (i.e. multi-user) geodatabase was created. The QA/QC analyst role was generated and was then used to import geospatial data into the geodatabase. The QA/QC analyst registered the buildings layer to be versioned. Versions were created of the buildings layer in protected mode. Versions could also be marked as private, however, only the user that created the version could view or edit that version. The base data layers were published as map services, and the versioned buildings layer was published as a feature service through ArcGIS for Server as Representational State Transfer (REST) protocol making them accessible on the web. The data were viewed and edited from a web browser presented in web maps using ArcGIS Online. It worked best to base this framework from Model C with a QA/QC version since it helped prevent corruption of the database or unauthorized edits. Open Geospatial Consortium (OGC) compliant web services such as web mapping services (WMS) were explored, but publishing a Web Map Service (WMS) or Web Feature Service (WFS) with OGC standards had a limitation in that it was not editable in ArcGIS web viewers, and certain customization abilities such as pop-ups did not function as expected. Other limitations of OGC standards contributed to this decision as well such as extensive customized programming requirement and lack of flexibility where the service must adhere to the standard (Rautenbach, Coetzee, and Iwaniak 2013).

The QA/QC analyst can view both versions to rectify discrepancies between the two versions where a researcher may need to go out and ground truth the data points, or aerial imagery can be consulted for digitizing. A training course may be necessary for each QA/QC analyst within the NPS, or each region could have a QA/QC analyst who can build the framework for each park unit within the region to allow all park staff members to contribute data if there were limited GIS personnel at a park unit. Future work would examine ways to automate the QA/QC process by employing methods of resolution between users based on area of expertise, by creating topology rules within the geodatabase to enforce topological relationships, or have different levels of acceptance where non-conflicting edits of certain data types are added directly to the parent database.

4.2 Web Maps

Two web maps were created in ArcGIS Online, both containing the basic park data for GEWA in the form of map services. The Version 1 map (Figure 8) contains an editable buildings feature class that is versioned for User 1. The Version 2 map (Figure 9) contains an editable buildings feature class that is versioned for User 2.

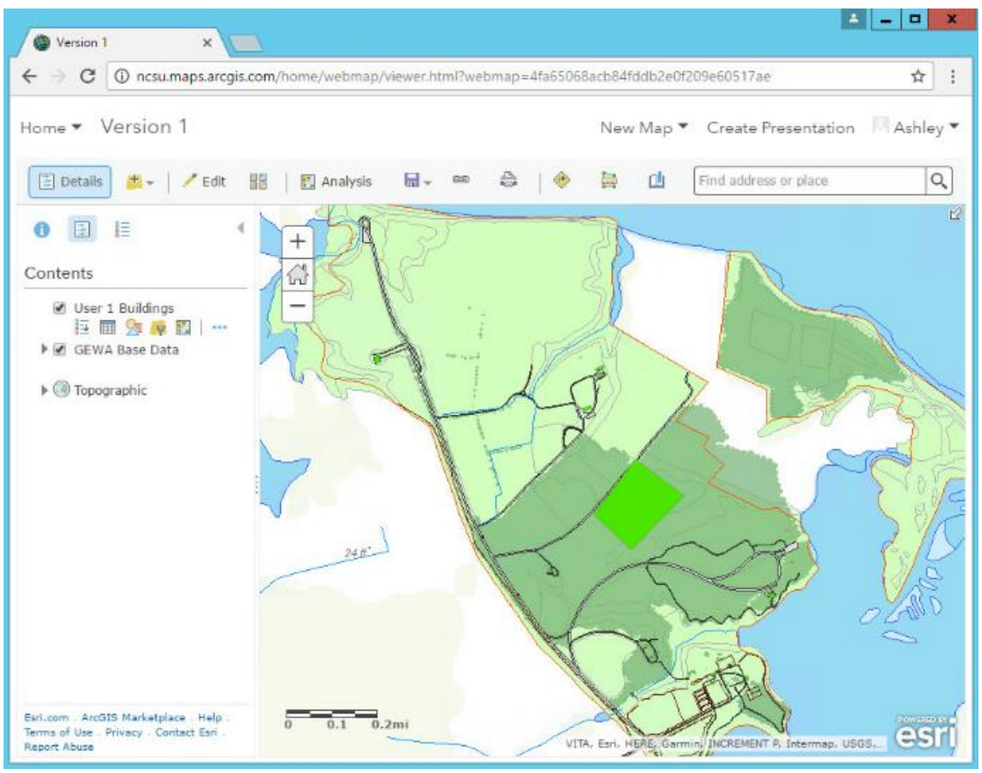


Figure 8: Web browser view of web map displaying the buildings feature class of version 1

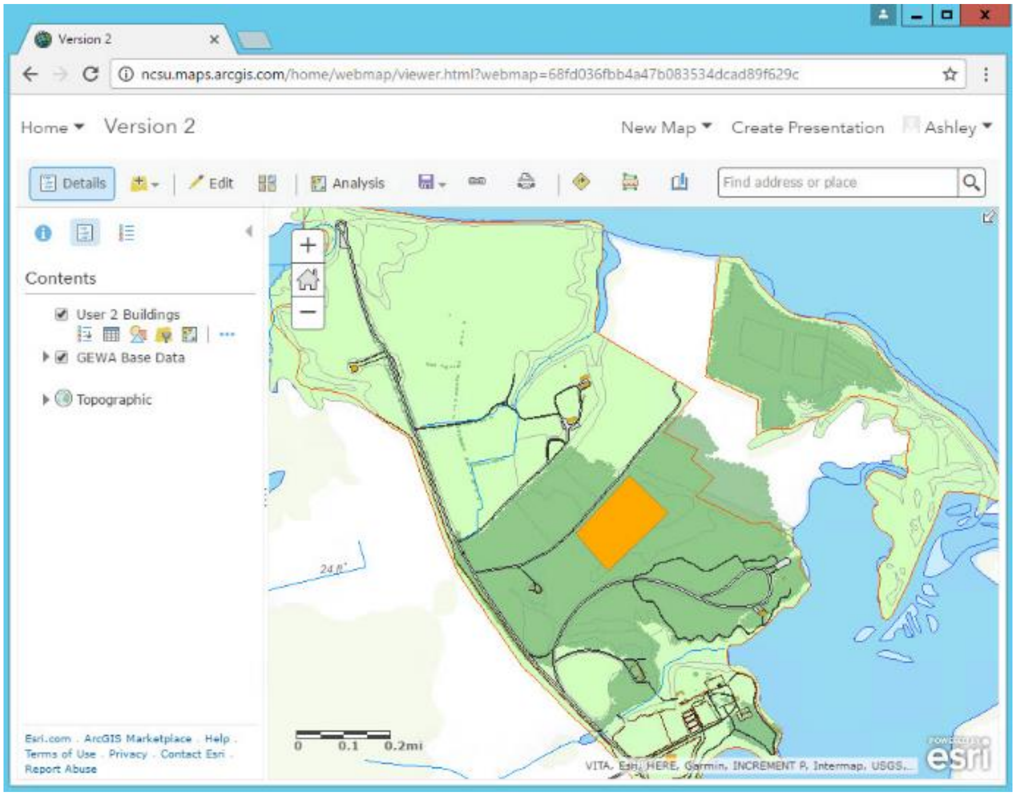


Figure 9: Web browser view of web map displaying the buildings feature class of version 2

The QA/QC analyst can view all the versions and see conflicting edits (Figure 10). They could then assign a researcher go out to the site and ground truth the data points, or aerial imagery can be consulted for verification. Utilizing versioned feature services that connect to two different versions of a geodatabase layer provides access to base data for multiple users concurrently, without them interfering with other users' edits.

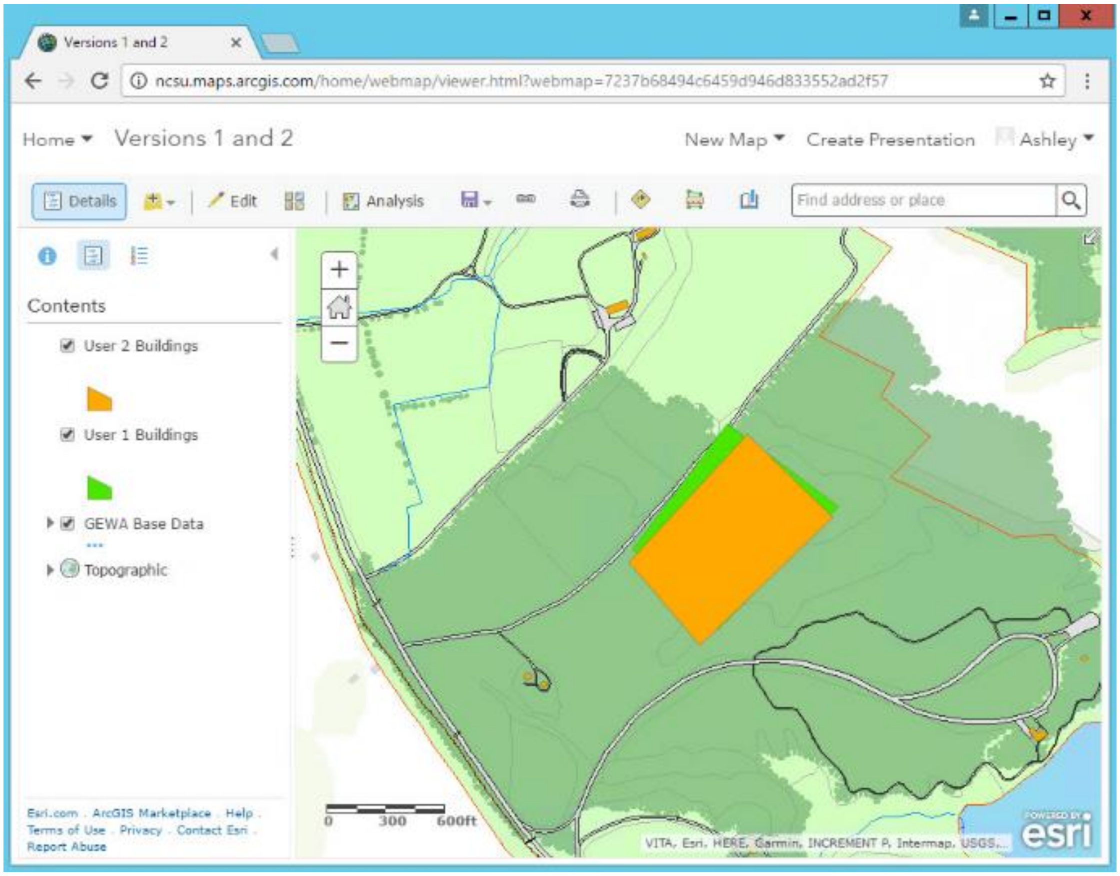


Figure 10: The QA/QC analyst view of conflicting edtis in a web browswer

A basic web map was used in the tutorial in order to start with a simple workflow that could then be built upon depending on each park units' GIS capabilities. At a minimum, having an editable feature service as a place to collect data into a common geodatabase for the park proved to be a first step into getting the park data collected into a centralized location and common format. Future work would include steps to prevent data entry error by providing a method for enforcing data integrity, such as pre-defined drop-down lists known as domains and subtypes in ArcMap.

4.3 Comparative Analysis

The survey respondents were asked to complete an online survey after finishing the framework tutorial. The survey asked the respondents which category of expertise level they fit in: light, moderate, advanced. The survey respondents consisted of mainly users with moderate GIS experience (Figure 11), however all three categories were represented. The respondents were also asked how many years they have been using GIS, the answers ranged from two to nineteen years (Figure 12).

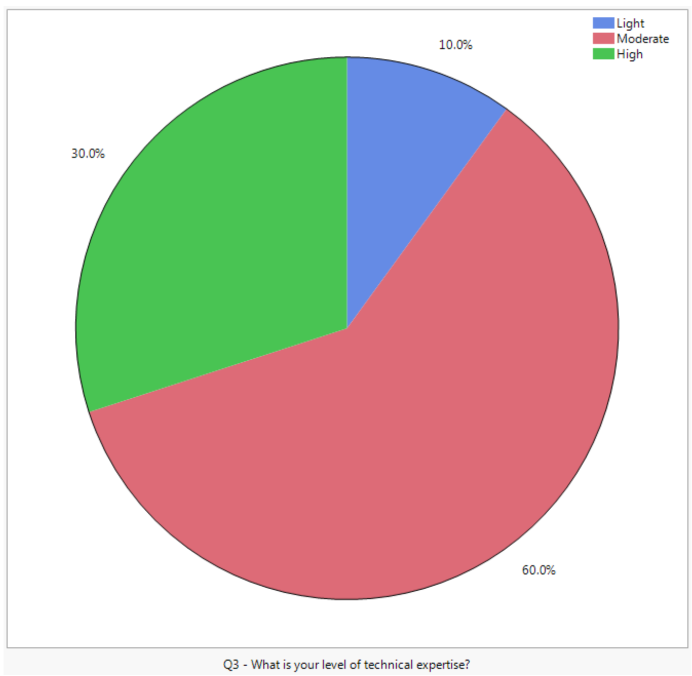


Figure 11: The respondents' expertise level: Light, Moderate, and High

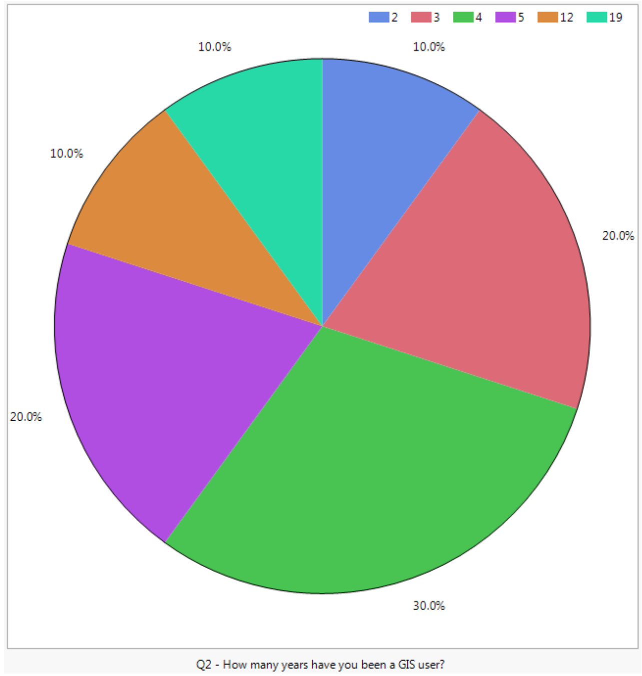


Figure 12: Respondents' GIS year of experience by percentages

The questions asked focused on evaluating the workflow of the tutorial. To improve the survey for future work, the questions can be refined to tease out the nuances between responses to the actual workflow, and which responses evaluate the organization and quality of the tutorial. The study group will also expand to a larger sample size with more beginner level of GIS experience represented.

4.3.1 Ease of Use

For a successful framework, ease of use was an important factor. Therefore, one of the survey questions asked if the tutorial was easy to use. The chi-squared test for independence was used to see if there was a relationship between skill level and ease of use at a 0.05 alpha. The chi-squared value did not fall within the rejection region with a p-value of 0.43. Due to the limited sample size of ten, Fisher's exact test was used to further validate the results. The p-value from Fisher's test was 0.60, agreeing with the chi-square value that there was no statistically significant relationship between ease of use and expertise level. Due to both the chi-square and Fisher's p-values were not below the alpha at a 0.05 level, the null hypothesis that the ease of use of the tutorial and expertise level were independent failed to be rejected. Table 1 displays the p-values for each test. The higher N increases, the results will be more reliable. This was an exploratory analysis to compose the survey and tutorial for future expansion to a larger sample size. The same tests were performed on the potential for a user with limited GIS experience to use the tutorial in the future and skill level.

Table 1: The resulting p-values for the chi-square and Fisher's Exact tests

	Chi-square P-value	**Fisher's Exact P-value**
Ease of use by skill level	0.43	0.60
Potential for future use by skill level	0.43	0.71

Likert responses were grouped by expertise level and displayed in a bar chart (Figure 13). All of the responses were in agreement to some degree. The respondents were also asked about their perceptions on future use of the tutorial by users with limited GIS experience. All the respondents either strongly agreed or somewhat agreed that a person with limited GIS experience would be able to successfully complete the tutorial (Figure 14). This was necessary if national parks with varying sized staff with varied GIS skills were going to be able to set up the framework.

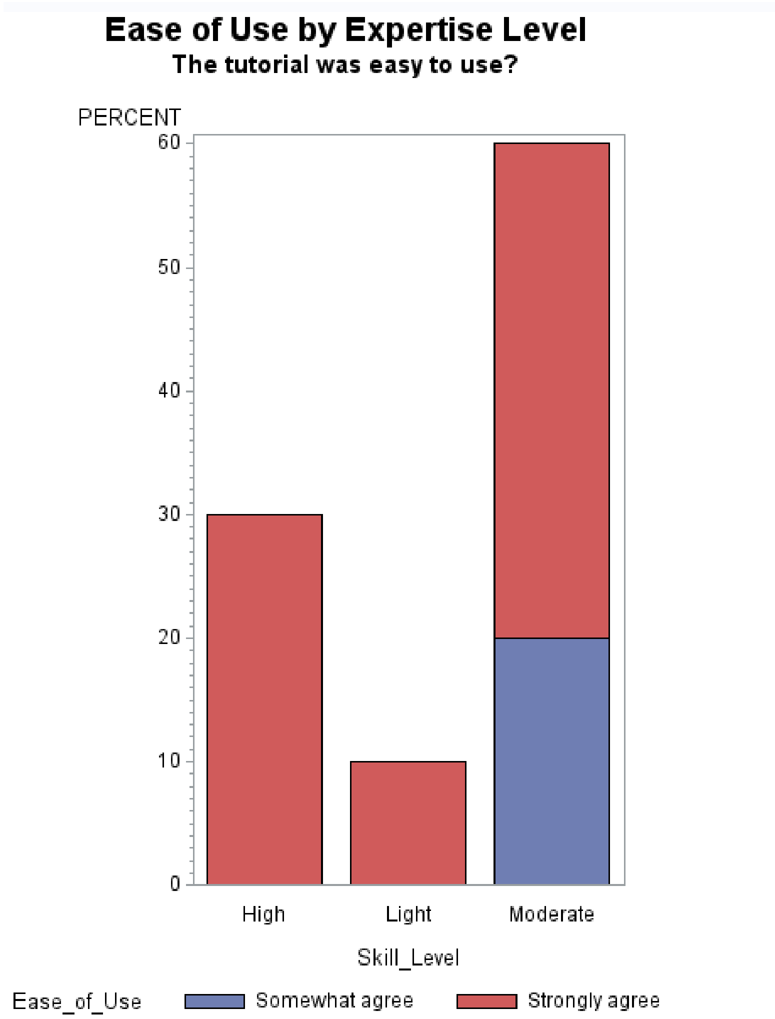


Figure 13: Bar chart comparing ease of use by expertise level. Strongly agree is shown in red, somewhat agree is shown in blue

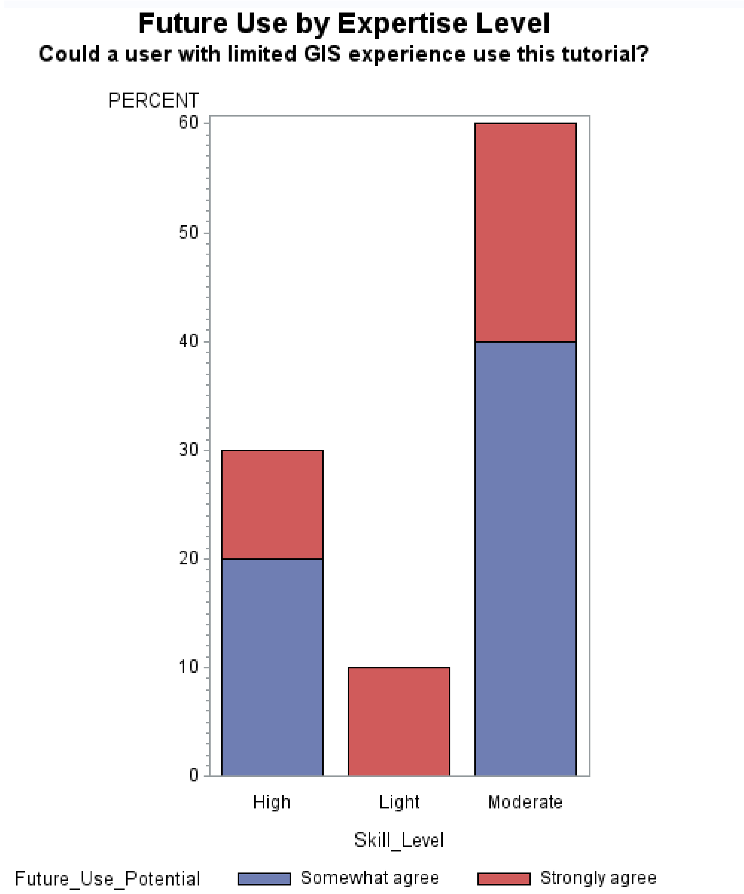


Figure 14: Bar chart comparing future use by expertise level. Strongly agree is shown in red, somewhat agree is shown in blue

The tutorial needs slight revisions to increase more of the 'Somewhat Agree' responses, to 'Strongly Agree' in future testing. The free text responses provided insightful feedback to increase ease of use for future utilization of the tutorial such as including background information of versioning and the different model scenarios, as well as adding succinct high level explanations of why the user needs to complete each tutorial step in order to aid in troubleshooting. Numbering the steps to match the initial workflow was recommended to provide clarity. Examples of how versioning was used by administrators was suggested, as well as information on how to add edits from the versions back to the original enterprise geodatabase. Lastly, links to online help documentation for troubleshooting errors were considered important to incorporate.

The overall survey provided valuable feedback on how plausible it would be to expect a wide range of experience levels set up the system described in this paper acting as the QA/QC analyst. The tutorial was easy to use and had potential to be utilized by users with limited GIS experience, and will become better with a few improvements.

4.3.2 Performance

Efficiency of the server was analyzed to see if there is a noticeable difference in the use of resources between publishing regular feature services and publishing versioned feature services. Process Explorer was used to take thirty samples of statistics of the amount of resources the machine used such as number of processes, percent of memory used, and percent commit charge (i.e. RAM plus space on hard disk used). The virtual machine was restarted before recording statistics. When web mapping services are published, they create processes to run the services: postgres.exe and arcsoc.exe. These processes were counted after publishing various scenarios (Table 2). RAM and commit charge were averaged from the thirty samples. The reason to look at an average of resource usage was because background processes could potentially skew a single observation. The results shows the cost of resources did not increase greatly with versioned feature services, suggesting that the framework was scalable as layers were published, and do not greatly increase usage of machine resources. This could be due to the way versioning does not make a copy of the data, it only tracks changes to the feature class in two delta tables, one table to track added features and one table to track deleted features.

Table 2: Computer resource usage performance evaluation

	Number of postgres.exe and arcsoc.exe Processes Running	Memory (RAM)	Commit Charge
Geodatabase with 9 base data layers with no layers published	9	26.68%	33.53%
After publishing 8 layers as map services in one map document	12	27.95%	35.15%
With 8 map services + 1 feature service published	15	28.00%	35.34%
With 8 layers published + 1 versioned feature service published (after deleting the 1 non-versioned feature service)	15	28.74%	38.78%
With 8 layers published + 5 versioned feature services published	27	35.05%	45.68%

In this study, the work of Krishna et al. (2015) was expanded by incorporating database versioning. Also the needs laid out in Smith et al.'s (2015) paper were addressed by adding park data to an enterprise geodatabase, in a central location on the VM, to create an authoritative dataset that can be accessed by the web, and controlled by the QA/QC analyst. The result of this study produced a repeatable model that was easy to use by those with a variety of GIS skills. An easy and repeatable model will allow individual parks, such as

GEWA, to accommodate a range of expertise levels to be able to set up the versioned database WebGIS framework.

5.0 Conclusion

This paper describes a successful framework for collecting and storing data for the National Park Service and land management agencies. The single user GIS model was expanded to allow multiple users to access and edit data at the same time. Using a RDBMS allows for utilization of database capabilities such as versioning and storing non-spatial data, while being accessible through the web. Versioning did work in a WebGIS environment to aid in the control of how data is collected and distributed. Many benefits of versioning include reducing storage requirements, managing workflows, ensuring data quality, and the ease of use was not dependent on GIS expertise level. However publishing map services such as OGC standards Web Map Service (WMS) or Web Feature Service (WFS) were not used due to their limitations with ArcGIS viewers. Looking at OGC standards, as well as mobile access for data collection are areas of future work. This paper introduced a new way of how data can be collected and handled within the NPS by designing an easy way to collect data into a more reliable storage system of a geodatabase through an accessible medium of the web. This method will also satisfy the requirements of individual parks, such as GEWA, by allowing them to maintain control of their individual data. Next steps include incorporating suggestions from the surveys into the tutorial, and then test again on a larger sample size before approaching individual park units. This paper demonstrates a first attempt to better inform the way data is collected, stored, and managed within land management agencies.

6.0 References

Al-Shahi, R., Sadler, M., Rees, G., & Bateman, D. (2002). The internet. *Journal of Neurology, Neurosurgery & Psychiatry*, 73(6), 619-628.

Bakalov, P., Hoel, E., Heng, W. L., Menon, S., & Tsotras, V. J. (2011). Editing and versioning for high performance network models in a multiuser environment. *GeoInformatica*, 15(4), 769-803.

Bernardini, F., Sgambati, A., Kokelj, M. M., Zaccaria, C., Micheli, R., Fragiacomo, A., ... & De Min, A. (2013). Airborne LiDAR application to karstic areas: the example of Trieste province (north-eastern Italy) from prehistoric sites to Roman forts. *Journal of Archaeological Science*, 40(4), 2152-2160.

Berners-Lee, T., Fischetti, M., & Foreword By-Dertouzos, M. L. (1999). *Weaving the Web: The original design and ultimate destiny of the World Wide Web by its inventor*. HarperInformation.

Brown, K. (2014, February 4). What Is GitHub, and What Is It Used For? Retrieved from https://www.howtogeek.com/180167/htg-explains-what-is-github-and-what-do-geeks-use-it-for/

Chatterjee, R., Arun, G., Agarwal, S., Speckhard, B., & Vasudevan, R. (2004, May). Using data versioning in database application development. In *Software Engineering, 2004. ICSE 2004. Proceedings. 26th International Conference on* (pp. 315-325). IEEE.

Cowen, D. (2009). The changing geospatial landscape: A report of the National Geospatial Advisory Committee. *Photogrammetric engineering and remote sensing*, *75*(7), 747-750.

Downs, R. R. (2015). Management, curation, and preservation of geospatial data: Introductory perspectives. *Journal of Map & Geography Libraries*, 11(2), 117-122.

Irwin, N. (2016). *How Do We Save GIS?* Center for Geospatial Analytics, North Carolina State University, Raleigh NC. https://cnr.ncsu.edu/geospatial/about/geoforum/

Folger, P. (2012, April). Issues and challenges for federal geospatial information. Congressional Research Service, Library of Congress.

Fu, P., & Sun, J. (2010). *Web GIS: principles and applications*. Esri Press.

Goodchild, M. F., & Li, L. (2012). Assuring the quality of volunteered geographic information. *Spatial statistics*, *1*, 110-120.

Joshi, P. K., & Joshi, A. (2013). Certification of geospatial data. *Current Science*, *105*(6), 759.

Karnatak, H., Saran, S., Bhatia, K., & Roy, P. S. (2010). Geospatial database organization and spatial decision analysis for biodiversity databases in Web GIS environment. *Geocarto International*, 25(1), 3-23.

Kearns, F. R., Kelly, M., & Tuxen, K. A. (2003). Everything happens somewhere: using webGIS as a tool for sustainable natural resource management. *Frontiers in Ecology and the Environment*, 1(10), 541-548.

Krishna, G. M., Ratnam, K. N., Ravi, P., & Sridhar, P. (2015). A Holistic Approach for the Development and Implementation of Robust and Cost-Effective Enterprise WebGIS Business Solutions. *Journal of Geographic Information System*, *7*(05), 478.

Laffan, S. W., Skidmore, A. K., & Franklin, J. (2016). Space, time, connectivity and conflict in biological landscapes: the fourth special issue on spatial ecology. *International Journal of Geographical Information Science*,30(1), 1-4.

Law, D. (2010, Winter). Versioning 101: Essential Information about ArcSDE Geodatabases. Retrieved March 18, 2017, from https://www.esri.com/~/media/Files/Pdfs/news/arcuser/0110/files/arcuser48.pdf

Manfré, L. A., Hirata, E., Silva, J. B., Shinohara, E. J., Giannotti, M. A., Larocca, A. P. C., & Quintanilha, J. A. (2012). An analysis of geospatial technologies for risk and natural disaster management. *ISPRS International Journal of Geo-Information*, 1(2), 166-185.

Marx, S., Phalkey, R., Aranda-Jan, C. B., Profe, J., Sauerborn, R., & Höfle, B. (2014). Geographic information analysis and web-based geoportals to explore malnutrition in Sub-Saharan Africa: a systematic review of approaches. *BMC public health*, *14*(1), 1189.

Mathiyalagan, V., Grunwald, S., Reddy, K. R., & Bloom, S. A. (2005). A WebGIS and geodatabase for Florida's wetlands. *Computers and electronics in agriculture*, *47*(1), 69-75.

Nelson, S. A., Cheruvelil, K. S., & Soranno, P. A. (2006). Satellite remote sensing of freshwater macrophytes and the influence of water clarity. *Aquatic Botany*, *85*(4), 289-298.

Peng, Z. R., & Tsou, M. H. (2003). *Internet GIS: distributed geographic information services for the internet and wireless networks*. John Wiley & Sons.

Putz, S. (1994). Interactive information services using World-Wide Web hypertext. *Computer Networks and ISDN Systems*, *27*(2), 273-280.

Rautenbach, V., Coetzee, S., & Iwaniak, A. (2013). Orchestrating OGC web services to produce thematic maps in a spatial information infrastructure. *Computers, Environment and Urban Systems*, 37, 107-120.

Smith, J. W., Slocumb, W. S., Smith, C., & Matney, J. (2015). A Needs-Assessment Process for Designing Geospatial Data Management Systems within Federal Agencies. *Journal of Map & Geography Libraries*, 11(2), 226-244.

Supak, S. K., Devine, H. A., Brothers, G. L., Rozier Rich, S., & Shen, W. (2014). An Open Source Web-Mapping System for Tourism Planning and Marketing. *Journal of Travel & Tourism Marketing*, 31(7), 835-853.

Tziavos, I. N., Alexandridis, T. K., Aleksandrov, B., Andrianopoulos, A., Doukas, I. D., Grigoras, I., ... & Teodorof, L. (2016). Development of a Web-based GIS monitoring and environmental assessment system for the Black Sea: application in the Danube Delta area. *Environmental monitoring and assessment*, 188(8), 492.

Wall, B., & Angryk, R. (2015). Storing Long-Lived Concurrent Schema and Data Versions in Relational Databases. In *New Trends in Database and Information Systems II* (pp. 97-108). Springer International Publishing.

Wang, S. (2010). A CyberGIS framework for the synthesis of cyberinfrastructure, GIS, and spatial analysis. *Annals of the Association of American Geographers*, *100*(3), 535-557.

Woodcock, C. E., Allen, R., Anderson, M., Belward, A., Bindschadler, R., Cohen, W. B., Gao, F., Goward, S. N., Helder, D., Helmer, E., Nemani, R., Oreopoulos, L., Schott, j., Thenkabail, P. S., Vermote, E. F., Vogelmann, j., Wulder. M.A., & Wynne, R. (2008). Free access to Landsat imagery. *Science*, 320, 1011, doi: l0.1126isdence.320.5879. lOll a.

Yu, X. W., Liu, H. Y., Yang, Y. C., Zhang, X., & Li, Y. W. (2013, February). Geoserver based forestry spatial data sharing and integration. In *Applied Mechanics and Materials* (Vol. 295, pp. 2394-2398). Trans Tech Publications.

Zhang, C., Zhao, T., & Li, W. (2013). Towards improving query performance of web feature services (WFS) for disaster response. *ISPRS International Journal of Geo-Information*, 2(1), 67-81.

CHAPTER III

1.0 General Conclusion and Discussion

This thesis employed the use of versioning with GIS web services in order to minimize data errors, provide security, and collect data for land management agencies, such as the National Park Service. It was an initial attempt, which will hopefully be researched further, into changing how the National Park Service (NPS) and other land management agencies in similar situations manage their geospatial data. There is a lot more work to be done, and this was one small step in heading the NPS in a new direction for quality data management, and it starts at the individual park level.

Next steps include expanding the tutorial to include survey feedback before testing again on a larger test group before approaching NPS staff. Expansion of the tutorial to include security levels is needed, as well as enhancing the web maps by creating web mapping applications, and implement suggested changes from the survey results. The free text responses from the survey provided insightful feedback to increase ease of use for future utilization of the tutorial such as including background information of versioning and the different model scenarios, as well as adding succinct high level explanations of why the user needs to complete each tutorial step in order to aid in troubleshooting. Numbering the steps to match the initial workflow was recommended to provide clarity. Examples of how versioning was used by administrators was suggested, as well as information on how to add edits from the versions back to the original enterprise geodatabase. Verify each step is consistent (e.g. all the ‘Click OK’ instructions are present). Lastly, links to online help documentation for troubleshooting errors were considered important to incorporate. Overall organizational adjustments would be beneficial as well.

No enterprise GIS can be considered a complete success without communication and training of the end users. The Bechtel Corporate GIS developed a body of literature called its GIS Desktop Procedures, to document standardized processes and 'best practices' for performing a variety of GIS-related tasks (McLane and Yan 2009). These documented workflows have been assembled to provide GIS users with step-by-step examples of how to perform any task from modeling subsurface bathymetry data to validating cut-fill calculations for engineering design (McLane and Yan 2009). This idea for training and documentation would be beneficial to apply to the versioned framework presented in this paper.

Some things did not work as expected during this study, and there are many things to consider moving forward such as the use of OGC compliant web services. OGC standards were not used for the web services due to their incompatibility with editing in ArcGIS Online web viewers. An approach to overcome this is to use an Esri enterprise geodatabase, but publish OGC layers to be consumed in another web map since there are many limitations with ArcMap viewers such as the OGC layer and the basemap have to be in the same coordinate system; group layers are not supported; some web mapping service (WMS) layers do not support pop-ups; and custom parameters are not supported (OGC 2017). Organization of spatial data is important to establish because if they are separated into feature datasets, the whole dataset gets registered as versioned. This is something to consider when setting up this versioned database framework. OGC standards and the work of Krishna et al. (2015) focused on harnessing the recent advancement in internet mapping technology, geospatial data management, and GIS API's for business applications. Although the benefits of online mapping reduce the number of licenses needed for multiple users to access the services, there is still a cost when using the ArcGIS platform when compared to free open source software. However the cost of the licensing can be preferable over the cost of hiring a developer.

This study moved away from collecting data in non-spatial tables that can be full of errors. A couple files from the original data provided by the NPS was collected in non-spatial tables that can be full of errors. Using the RDBMS allows for later integration of non-spatial or tabular data into the geodatabase to update the NPS dataset and get the data in one central location. Should need arise for data to aggregate data into regional or national model, the data is already in the format needed. Joshi et al. (2001) looked at database federation approach that integrated multiple database systems that allowed autonomous control at the individual database level while at the same time being accessible to others. This approach could fit very well for the NPS once they start to collect data in a standardize way at the part level. Since individual parks are the best source of the data, this federated idea allows them to keep their autonomy and control of their data, while enabling communication between parks. Security will be another important area to research further. Data encryption was used by McLane and Yan (2009) to prevent unauthorized access, they also enabled web application level authentication for security. Both of these methods of security can be taken into consideration for future work.

There are a lot of areas for future work including mobile application development, offline access to the application, looking further into OGC standards and open source technologies to see if there is a simple way to use them without lots of programming, look into federated databases, security, and research further into ontologies for normalization. 3D visualization on the web could also be another area to research further (Das et al. 2012). Installing PostgreSQL and ArcGIS for Server on separate machines could help improve performance since both programs are memory intensive.

Mobile application development collecting and storing data in a feature service published from a geodatabase allows for easy transition into mobile collection. Offline access to the application would be a fitting next step. Replication was explored as a way to edit data offline, but it was not pursued further since replication is designed to work between separate geodatabases to make their data identical to each other. It does not support a parent copy of the data, or have a QA/QC level of quality control. Replication is more suited for scenarios where a company with offices on opposite coasts of the United States need local access to the database so the data does not have to be accessed over the network. It also uses more storage since the database is copied.

This thesis combined versioning, GIS, and the Web in order to provide land management agencies with a method to enhance data collection, and ultimately park efficiency, which is good for all land management agencies due to the cost savings of time, money, and ability to own quality data. Lastly, this framework should be replicable to many agencies, and even other organizations such as non-profit ones, to be able to replicate this work and take GIS to the next intermingling with computer science and make versioning available on the web.

2.0 References

Ahmed, W., & Zimányi, E. (2015, September). Querying Multiversion Data Warehouses. In *East European Conference on Advances in Databases and Information Systems* (pp. 346-357). Springer International Publishing.

Al-Shahi, R., Sadler, M., Rees, G., & Bateman, D. (2002). The internet. *Journal of Neurology, Neurosurgery & Psychiatry*, 73(6), 619-628.

Andrén, A., Viberg, A., Lidén, K., Victor, H., & Fischer, S. (2014). The ringfort by the sea: Archaeological geophysical prospection and excavations at Sandby borg (Öland). *Archäologisches Korrespondenzblatt*, *44*(3), 413-428.

Anselma, L., Bottrighi, A., Montani, S., & Terenziani, P. (2013). Managing proposals and evaluations of updates to medical knowledge: Theory and applications. *Journal of biomedical informatics*, *46*(2), 363-376.

Bakalov, P., Hoel, E., Menon, S., & Tsotras, V. J. (2009, July). Versioning of network models in a multiuser environment. In *International Symposium on Spatial and Temporal Databases* (pp. 6-24). Springer Berlin Heidelberg.

Bakalov, P., Hoel, E., Heng, W. L., Menon, S., & Tsotras, V. J. (2011). Editing and versioning for high performance network models in a multiuser environment. *GeoInformatica*, 15(4), 769-803.

Bastin, L., Buchanan, G., Beresford, A., Pekel, J. F., & Dubois, G. (2013). Open-source mapping and services for Web-based land-cover validation. *Ecological Informatics*, *14*, 9-16.

Bera, R., Squividant, H., Henaff, G. L., Pichelin, P., Ruiz, L., Launay, J., ... & Cudennec, C. (2015). GéoSAS: a modular and interoperable open source spatial data infrastructure for research. *Proceedings of the International Association of Hydrological Sciences*, *368*, 9-14.

Bernardini, F., Sgambati, A., Kokelj, M. M., Zaccaria, C., Micheli, R., Fragiacomo, A., ... & De Min, A. (2013). Airborne LiDAR application to karstic areas: the example of Trieste province (north-eastern Italy) from prehistoric sites to Roman forts. *Journal of Archaeological Science*, *40*(4), 2152-2160.

Berners-Lee, T., Fischetti, M., & Foreword By-Dertouzos, M. L. (1999). *Weaving the Web: The original design and ultimate destiny of the World Wide Web by its inventor*. HarperInformation.

Brekke, C., & Solberg, A. H. (2005). Oil spill detection by satellite remote sensing. *Remote sensing of environment*, *95*(1), 1-13.

Best, B. D., Halpin, P. N., Fujioka, E., Read, A. J., Qian, S. S., Hazen, L. J., & Schick, R. S. (2007). Geospatial web services within a scientific workflow: Predicting marine mammal habitats in a dynamic environment. *Ecological Informatics*, *2*(3), 210-223.

Chatterjee, R., Arun, G., Agarwal, S., Speckhard, B., & Vasudevan, R. (2004, May). Using data versioning in database application development. In *Software Engineering, 2004. ICSE 2004. Proceedings. 26th International Conference on* (pp. 315-325). IEEE.

Coppock, J. T., & Rhind, D. W. (1991). The history of GIS. *Geographical information systems: Principles and applications*, 1(1), 21-43.

Cowen, D. (2009). The changing geospatial landscape: A report of the National Geospatial Advisory Committee. *Photogrammetric engineering and remote sensing*, *75*(7), 747-750.

Das, S., Zijdenbos, A. P., Vins, D., Harlap, J., & Evans, A. C. (2012). LORIS: a web-based data management system for multi-center studies. *Frontiers in neuroinformatics*, *5*, 37.

Dias, G. P., & Rafael, J. A. (2007). A simple model and a distributed architecture for realizing one-stop e-government. *Electronic Commerce Research and Applications*, 6(1), 81-90.

Dodge, S., Bohrer, G., Bildstein, K., Davidson, S. C., Weinzierl, R., Bechard, M. J., ... & Wikelski, M. (2014). Environmental drivers of variability in the movement ecology of turkey vultures (Cathartes aura) in North and South America. *Phil. Trans. R. Soc. B*, *369*(1643), 20130195.

Downs, R. R. (2015). Management, curation, and preservation of geospatial data: Introductory perspectives. *Journal of Map & Geography Libraries*, 11(2), 117-122.

ESRI. (2017). OGC. Retrieved from https://doc.arcgis.com/en/arcgis-online/reference/ogc.htm

Folger, P. (2012, April). Issues and challenges for federal geospatial information. Congressional Research Service, Library of Congress.

Fu, P., & Sun, J. (2010). *Web GIS: principles and applications*. Esri Press.

Goodchild, M., Haining, R., & Wise, S. (1992). Integrating GIS and spatial data analysis: problems and possibilities. *International journal of geographical information systems*, 6(5), 407-423.

Goodchild, M. F. (2015). Two decades on: Critical GIScience since 1993. *The Canadian Geographer/Le Géographe canadien*, *59*(1), 3-11.

Goodchild, M. F., & Li, L. (2012). Assuring the quality of volunteered geographic information. *Spatial statistics*, *1*, 110-120.

Hafner, K., & Lyon, M. (1998). *Where wizards stay up late: The origins of the Internet*. Simon and Schuster.

Haworth, B., & Bruce, E. (2015). A review of volunteered geographic information for disaster management. *Geography Compass*, *9*(5), 237-250.

Joshi, J., Ghafoor, A., Aref, W. G., & Spafford, E. H. (2001). Digital government security infrastructure design challenges. *Computer*, *34*(2), 66-72.

Joshi, P. K., & Joshi, A. (2013). Certification of geospatial data. *Current Science*, *105*(6), 759.

Karnatak, H., Saran, S., Bhatia, K., & Roy, P. S. (2010). Geospatial database organization and spatial decision analysis for biodiversity databases in Web GIS environment. *Geocarto International*, 25(1), 3-23.

Kearns, F. R., Kelly, M., & Tuxen, K. A. (2003). Everything happens somewhere: using webGIS as a tool for sustainable natural resource management. *Frontiers in Ecology and the Environment*, 1(10), 541-548.

Krishna, G. M., Ratnam, K. N., Ravi, P., & Sridhar, P. (2015). A Holistic Approach for the Development and Implementation of Robust and Cost-Effective Enterprise WebGIS Business Solutions. *Journal of Geographic Information System*, *7*(5), 478.

Laffan, S. W., Skidmore, A. K., & Franklin, J. (2016). Space, time, connectivity and conflict in biological landscapes: the fourth special issue on spatial ecology. *International Journal of Geographical Information Science*, *30*(1), 1-4.

Law, D. (2010, Winter). Versioning 101: Essential Information about ArcSDE Geodatabases. Retrieved March 18, 2017, from https://www.esri.com/~/media/Files/Pdfs/news/arcuser/0110/files/arcuser48.pdf

Lin, W. (2015). The hearing, the mapping, and the Web: Investigating emerging online sound mapping practices. *Landscape and Urban Planning*,*142*, 187-197.

Mair, M., Mikovits, C., Sengthaler, M., Schöpf, M., Kinzel, H., Urich, C., ... & Rauch, W. (2014). The application of a Web-geographic information system for improving urban water cycle modelling. *Water Science and Technology*, 70(11), 1838-1846.

Manfré, L. A., Hirata, E., Silva, J. B., Shinohara, E. J., Giannotti, M. A., Larocca, A. P. C., & Quintanilha, J. A. (2012). An analysis of geospatial technologies for risk and natural disaster management. *ISPRS International Journal of Geo-Information*, *1*(2), 166-185.

Mathiyalagan, V., Grunwald, S., Reddy, K. R., & Bloom, S. A. (2005). A WebGIS and geodatabase for Florida's wetlands. *Computers and electronics in agriculture*, *47*(1), 69-75.

Marx, S., Phalkey, R., Aranda-Jan, C. B., Profe, J., Sauerborn, R., & Höfle, B. (2014). Geographic information analysis and web-based geoportals to explore malnutrition in Sub-Saharan Africa: a systematic review of approaches. *BMC public health*, *14*(1), 1189.

McLane, T. J., & Yan, Y. (2009, August). Enterprise spatial data deployment strategies in a global engineering environment. In *2009 17th International Conference on Geoinformatics* (pp. 1-5). IEEE.

Milosavljević, A., Stoimenov, L., & Đorđević-Kajan, S. (2005, May). An architecture for open and scalable WebGIS. In *8th AGILE Conference on GIScience, Estoril, Portugal* (pp. 629-634).

Musinguzi, M., Bax, G., & Tickodri-Togboa, S. S. (2004). Opportunities and Challenges for SDI Development in Developing Countries-A Case Study of Uganda. *Geoinformatics*, 789-796.

Nelson, S. A., Cheruvelil, K. S., & Soranno, P. A. (2006). Satellite remote sensing of freshwater macrophytes and the influence of water clarity. *Aquatic Botany*, *85*(4), 289-298.

Panigrahi, N. (2009). *Geographical information science*. Universities Press.

Peng, Z. R., & Tsou, M. H. (2003). *Internet GIS: distributed geographic information services for the internet and wireless networks*. John Wiley & Sons.

Pla, M., & Lleopart, A. (2010). Updating of vector databases at the Institut Cartografic de Catalunya. *Catalonia*, *1*(5), 000.

Putz, S. (1994). Interactive information services using World-Wide Web hypertext. *Computer Networks and ISDN Systems*, *27*(2), 273-280.

Rautenbach, V., Coetzee, S., & Iwaniak, A. (2013). Orchestrating OGC web services to produce thematic maps in a spatial information infrastructure. *Computers, Environment and Urban Systems*, *37*, 107-120.

Saroha, K., & Gosain, A. (2014, September). Multi-version data warehouse: A survey. In *Confluence The Next Generation Information Technology Summit (Confluence), 2014 5th International Conference-* (pp. 40-45). IEEE.

Skøien, J. O., Schulz, M., Dubois, G., Fisher, I., Balman, M., May, I., & Tuama, É. Ó. (2013). A Model Web approach to modelling climate change in biomes of cowns. *Ecological informatics*, *14*, 38-43.

Sexton, J. O., Urban, D. L., Donohue, M. J., & Song, C. (2013). Long-term land cover dynamics by multi-temporal classification across the Landsat-5 record. *Remote Sensing of Environment*, *128*, 246-258.

Smith, J. W., Slocumb, W. S., Smith, C., & Matney, J. (2015). A Needs-Assessment Process for Designing Geospatial Data Management Systems within Federal Agencies. *Journal of Map & Geography Libraries*, 11(2), 226-244.

Soomro, T. R., Zheng, K., & Pan, Y. (1999). HTML and multimedia Web GIS. In *Computational Intelligence and Multimedia Applications, 1999. ICCIMA'99. Proceedings. Third International Conference on* (pp. 371-382). IEEE.

Shubik, M. (1997). Where wizards stay up late. *Complexity*, *3*(2), 48-49.

Supak, S. K., Devine, H. A., Brothers, G. L., Rozier Rich, S., & Shen, W. (2014). An open source web-mapping system for tourism planning and marketing. *Journal of Travel & Tourism Marketing*, *31*(7), 835-853.

Tziavos, I. N., Alexandridis, T. K., Aleksandrov, B., Andrianopoulos, A., Doukas, I. D., Grigoras, I., ... & Teodorof, L. (2016). Development of a Web-based GIS monitoring and environmental assessment system for the Black Sea: application in the Danube Delta area. *Environmental monitoring and assessment*, *188*(8), 492.

Woodcock, C. E., Allen, R., Anderson, M., Belward, A., Bindschadler, R., Cohen, W. B., Gao, F., Goward, S. N., Helder, D., Helmer, E., Nemani, R., Oreopoulos, L., Schott, j., Thenkabail, P. S., Vermote, E. F., Vogelmann, j., Wulder. M.A., & Wynne, R. (2008).

Free access to Landsat imagery. *Science*, 320, 1011, doi: l0.1126isdence.320.5879. lOll a.

Wang, S. (2010). A CyberGIS framework for the synthesis of cyberinfrastructure, GIS, and spatial analysis. *Annals of the Association of American Geographers*, *100*(3), 535-557.

Welcome to the OGC. (2017). Retrieved March 22, 2017, from http://www.opengeospatial.org/

Yu, X. W., Liu, H. Y., Yang, Y. C., Zhang, X., & Li, Y. W. (2013, February). Geoserver based forestry spatial data sharing and integration. In *Applied Mechanics and Materials* (Vol. 295, pp. 2394-2398). Trans Tech Publications.

Zhang, Y., & Song, C. (2006). Impacts of afforestation, deforestation, and reforestation on forest cover in China from 1949 to 2003. *Journal of Forestry*,*104*(7), 383-387.

Zhang, J., Zhu, Y., Wang, J., Sun, J., & Xu, Y. (2011). Flash based WebGIS System and its Application in Monitoring and Evaluating China's Regional Development. *International Journal of Digital Content Technology and Its Applications*, *5*, 285-295.

Zhang, Y., Song, C., Zhang, K., Cheng, X., Band, L. E., & Zhang, Q. (2014). Effects of land use/land cover and climate changes on terrestrial net primary productivity in the Yangtze River Basin, China, from 2001 to 2010. *Journal of Geophysical Research: Biogeosciences, 119*(6), 1092-1109.

APPENDICES

Appendix A. Versioning Framework Tutorial

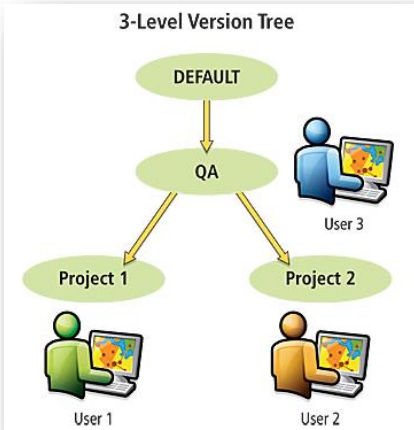
Introduction

This tutorial walks through the steps of creating versions in ArcMap, publishing services from the created versions, and then using the published services in web maps through ArcGIS Online. The only data required to complete this tutorial are shapefiles. Shapefiles of the George Washington Birthplace National Monument Park are provided on the course Moodle site.

Workflow

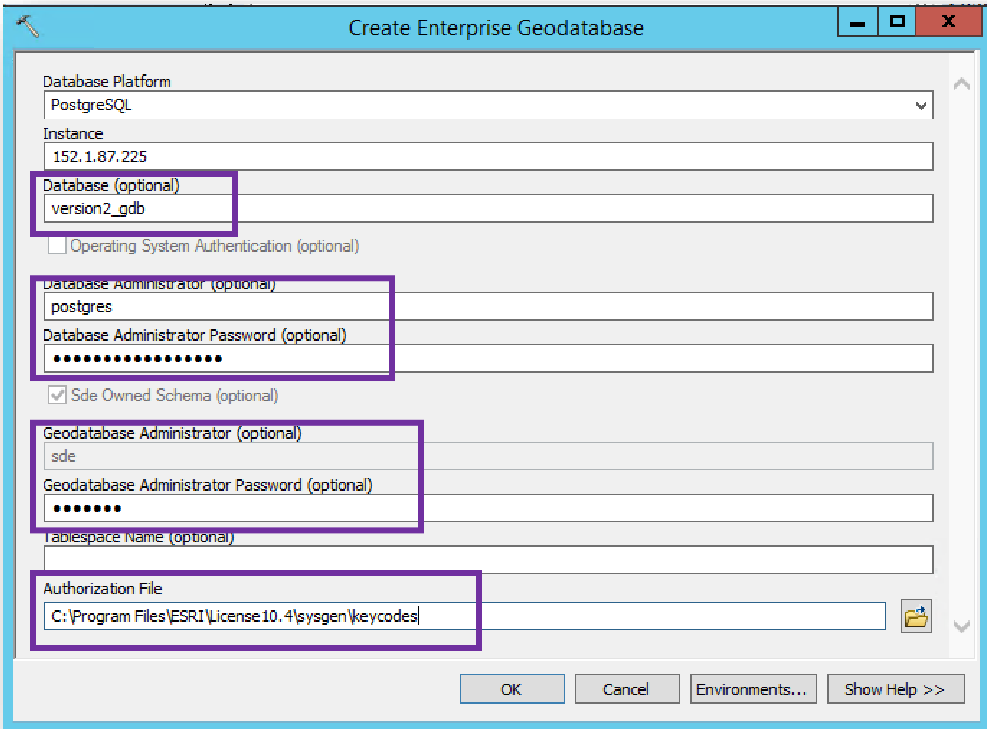
1. Create an enterprise geodatabase
2. Connect to the newly created geodatabase as the sde user
3. Add User to be the QA admin of the geodatabase
4. Connect to the geodatabase as the admin user
5. Add data to the geodatabase
6. Register the feature classes to be edited as versioned
7. Create a version from the sde DEFAULT version as the quality control version: admin.
8. From the admin version, create 2 project versions
9. Create database connections to all 3 versions
10. Register the database of each version (e.g. admin, project1, project2)
11. Publish Feature Services from the two versions
12. Publish base park data as Map Services
13. Create web maps!

Section 1 – Create Versions from an Enterprise Geodatabase using a 3-Level Version Tree Model

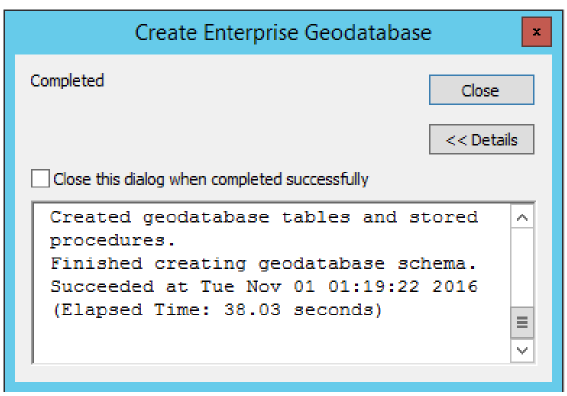


Using this model, we will create the sde.DEFAULT database, a child version for the Quality Assurance Analyst, and two project versions: Project 1 and Project 2.

1. Under the Data Management toolbox open the Create Enterprise Geodatabase tool.
   - For the Instance, enter your IP address
   - For the database name, name it "version2_gdb"
   - Use the postgres user password that should already be created
   - Use the sde user password that should already exist. If not, create a password for the sde user at this time
   - For the Authorization File, the key codes should be in this location: C:\Program Files\ESRI\License10.4\sysgen\keycodes

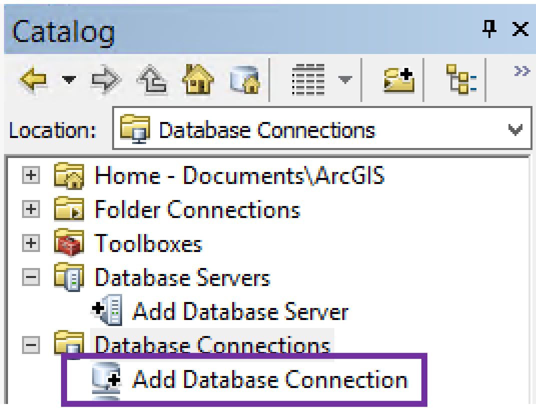


- A dialog box showing complete pops up > click Close

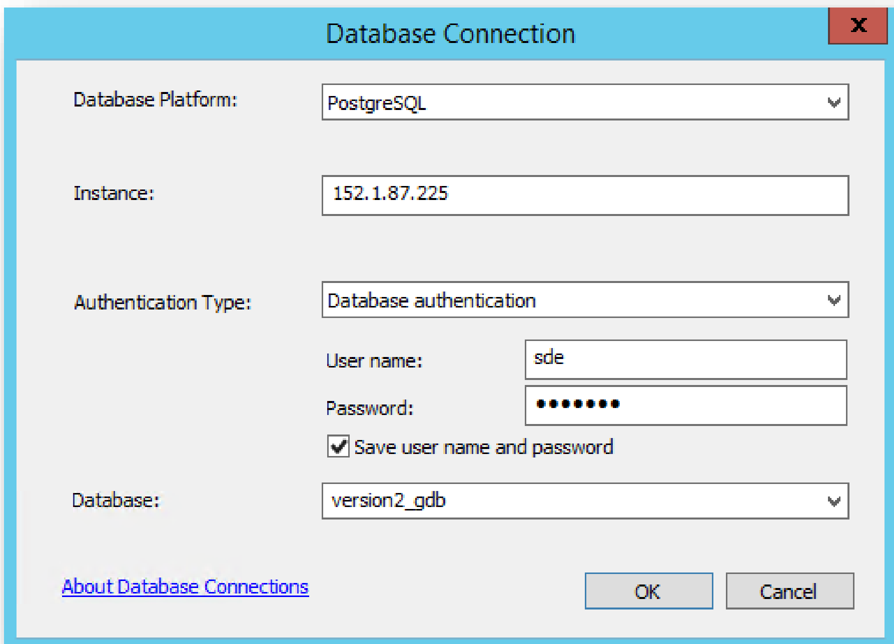


2. Connect to the database at the sde user in order to create an admin role to act as a Quality Assurance (QA) analyst

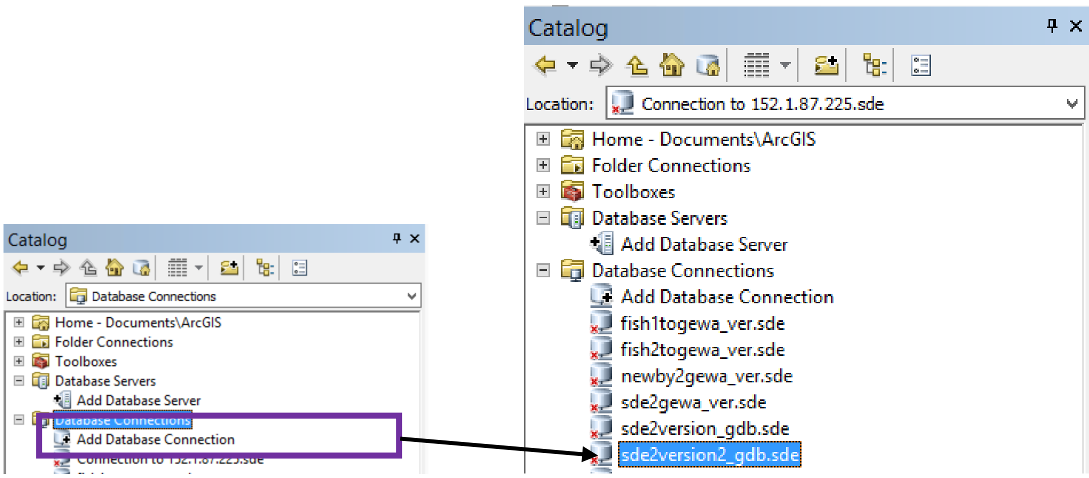
- Go to ArcCatalog > double click on Add Database Connection



- Connect to your new version2_gdb geodatabase with the sde user credentials:

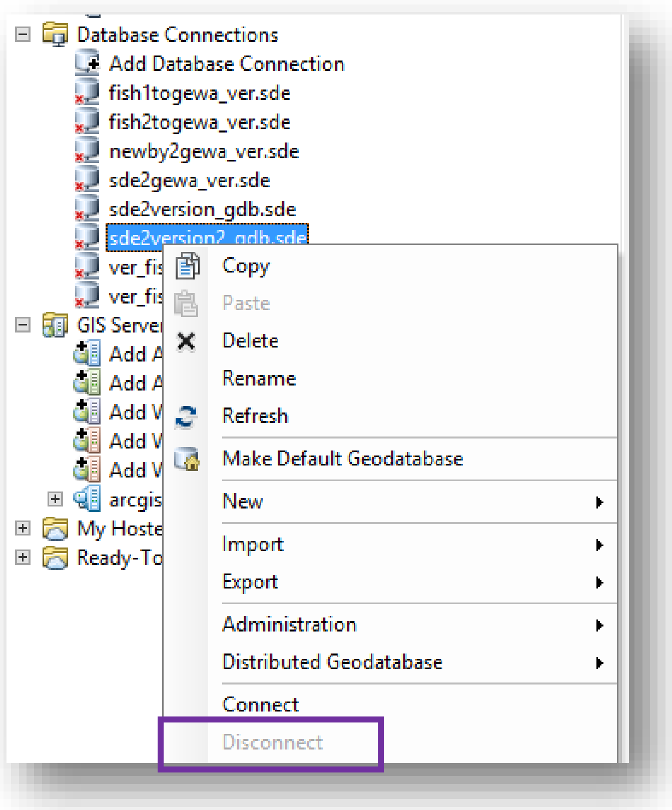


3. Rename the connection to make it easier to know which user is connecting to which database:

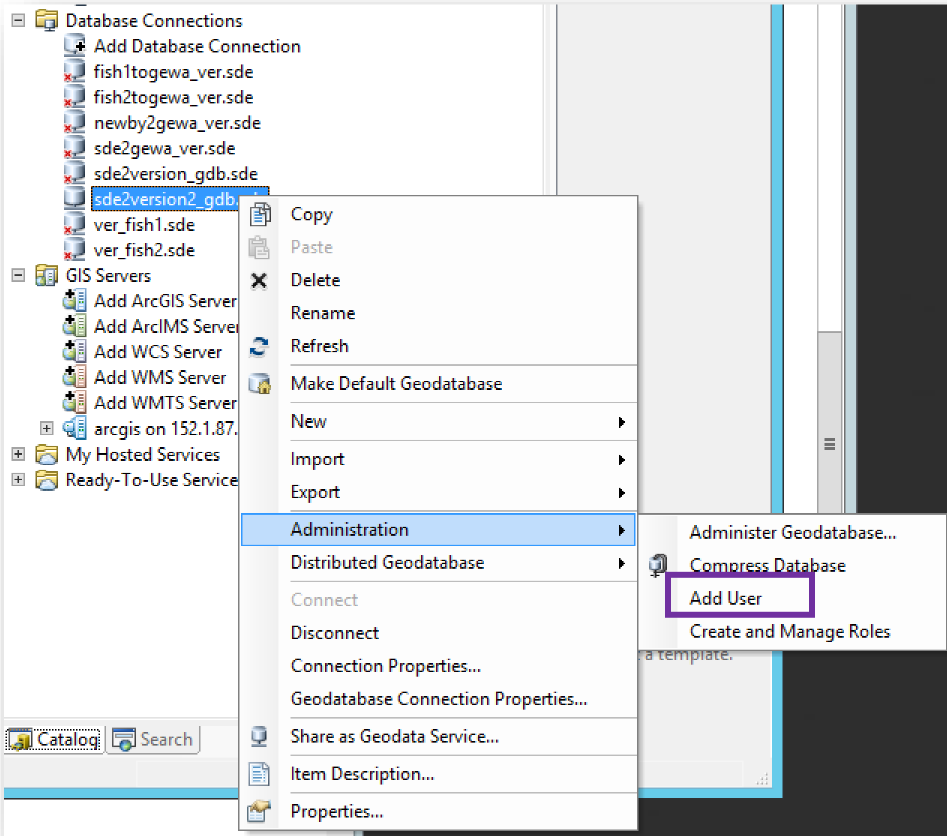


4. Now to add a QA user from the 3-level Version Tree model

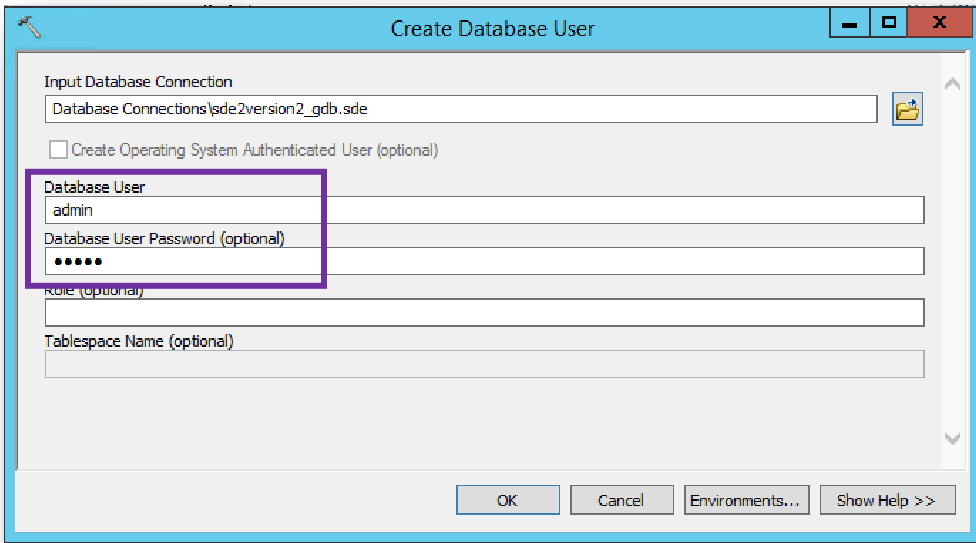
- Right click the sde geodatabase connection to version2_gdb and click Connect



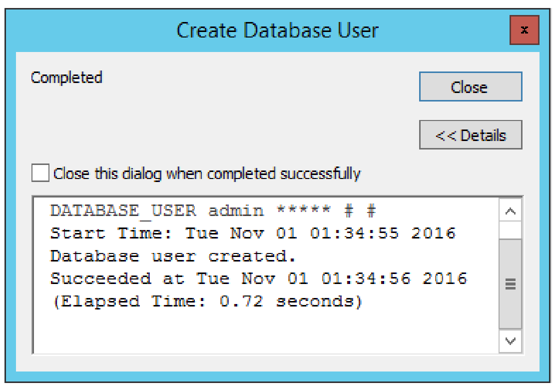
- Right click the sde geodatabase connection to version2_gdb again and click Administration > Add User



- Fill out as User: admin password: admin ← This will be the QA analyst user

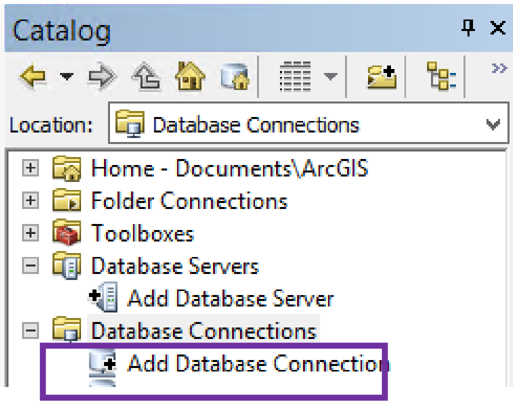


- The Complete dialog box will pop up > click Close

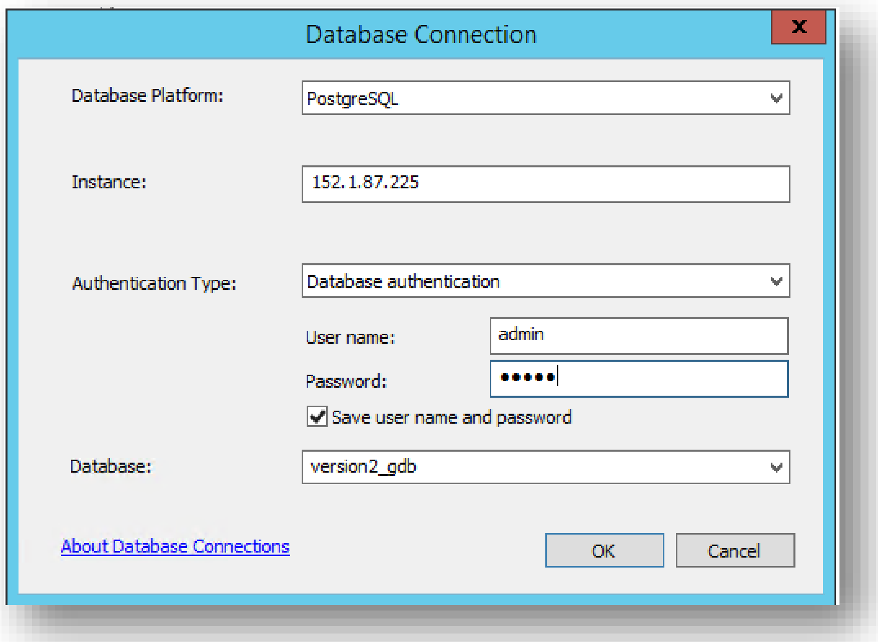


5. Next create a new database connection with the admin user

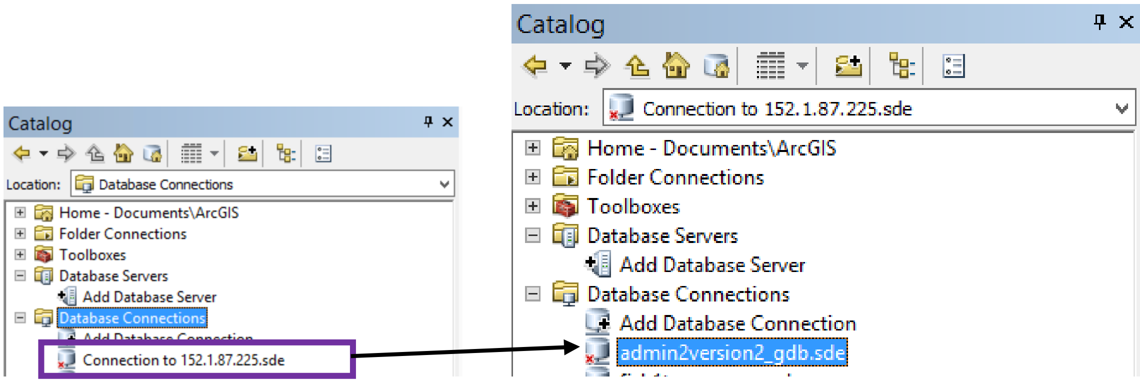
- Go to ArcCatalog > double click on Add Database Connection



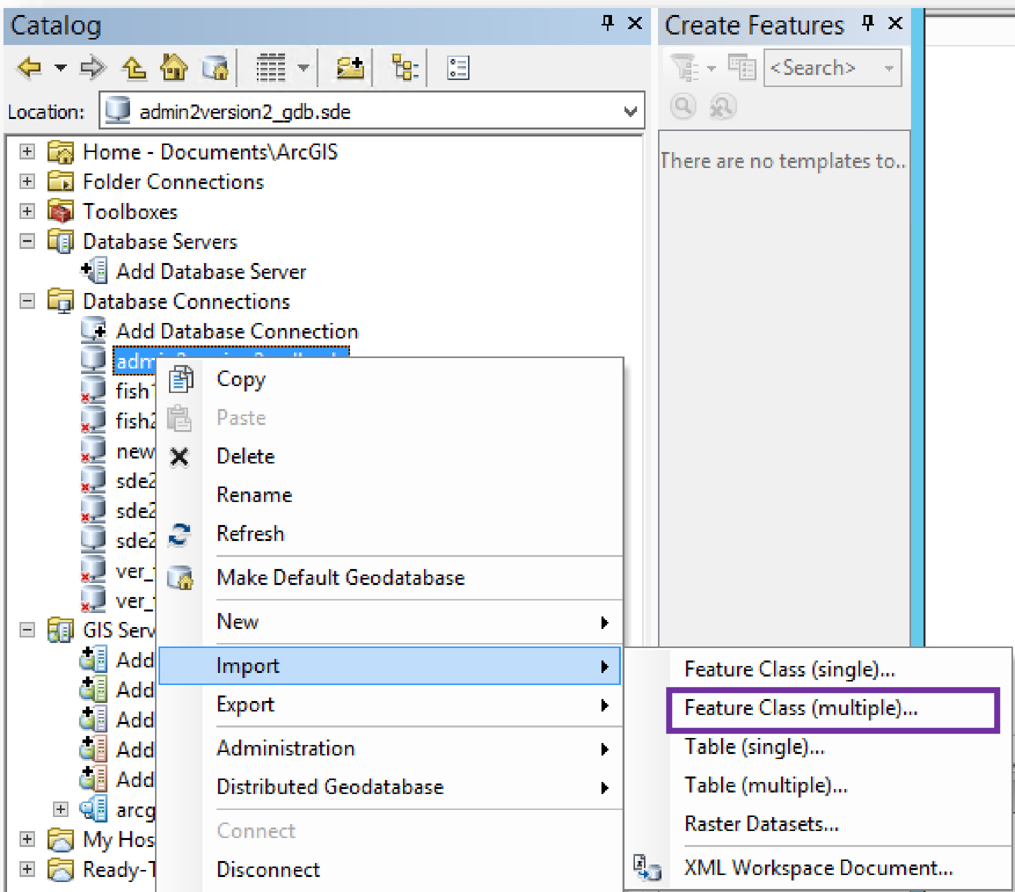
- Connect as the new admin user as below:



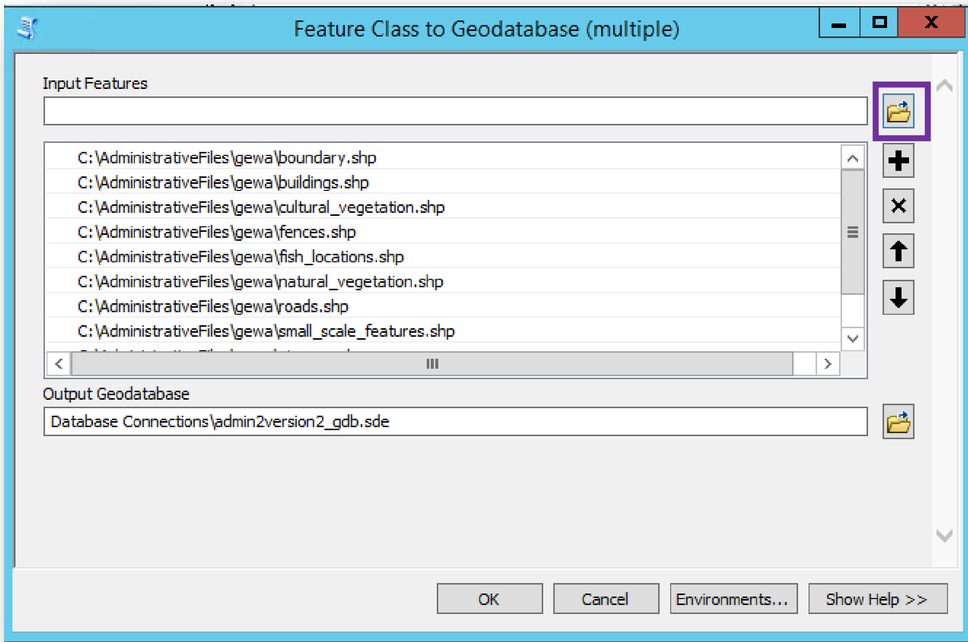
- Again, rename the connection



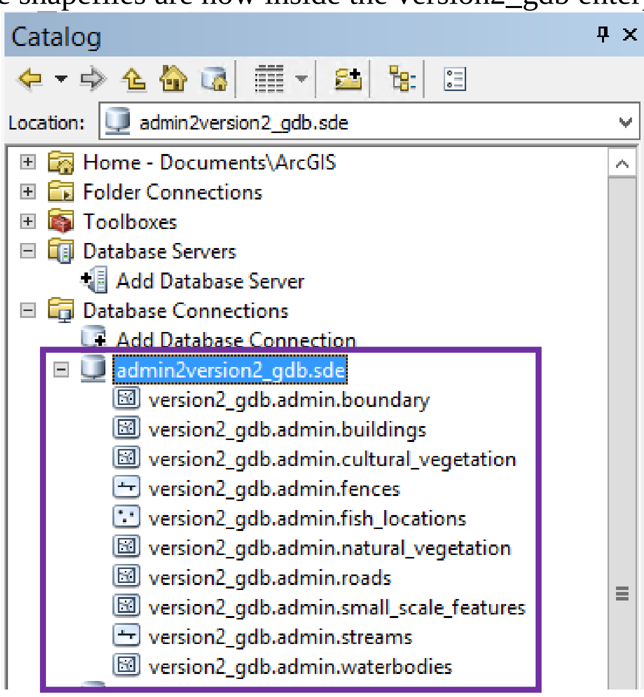
6. Now we're ready to add data. Double click on the admin2version2_gdb connection to activate it > right click on the connection > Import > Feature Class multiple



- Navigate to the shapefiles downloaded from the course Moodle page and add all the shapefiles

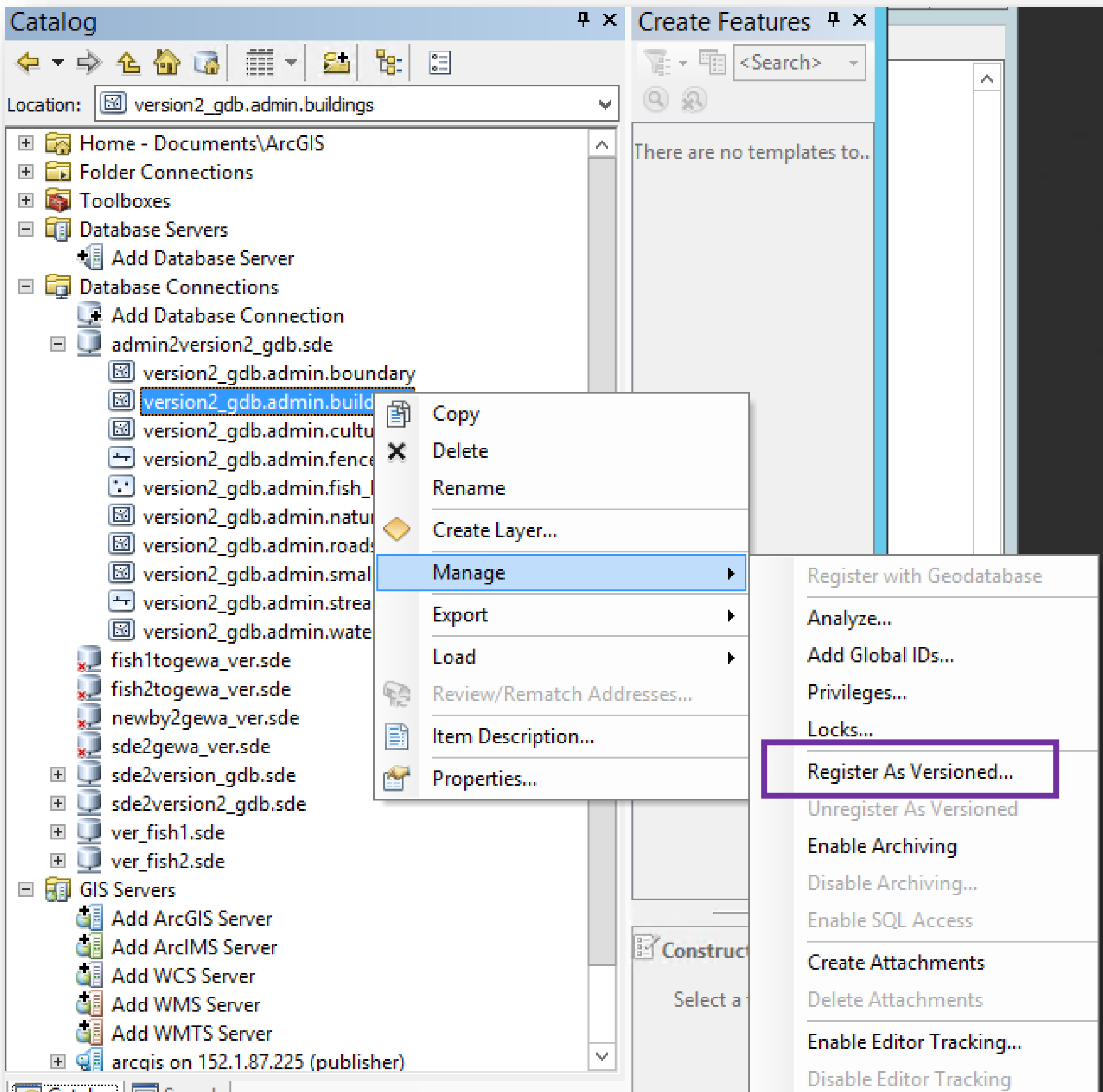


- The shapefiles are now inside the version2_gdb enterprise geodatabase:

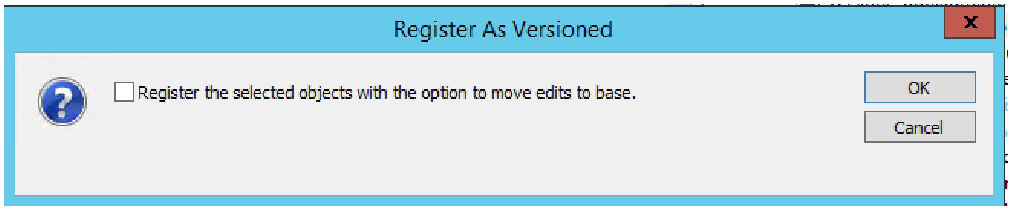


7. In order for a feature class to be edited within a version, it must be Registered As Versioned

- Right click on the feature to be edited (for this tutorial we will use Buildings) > Manage > Register As Versioned

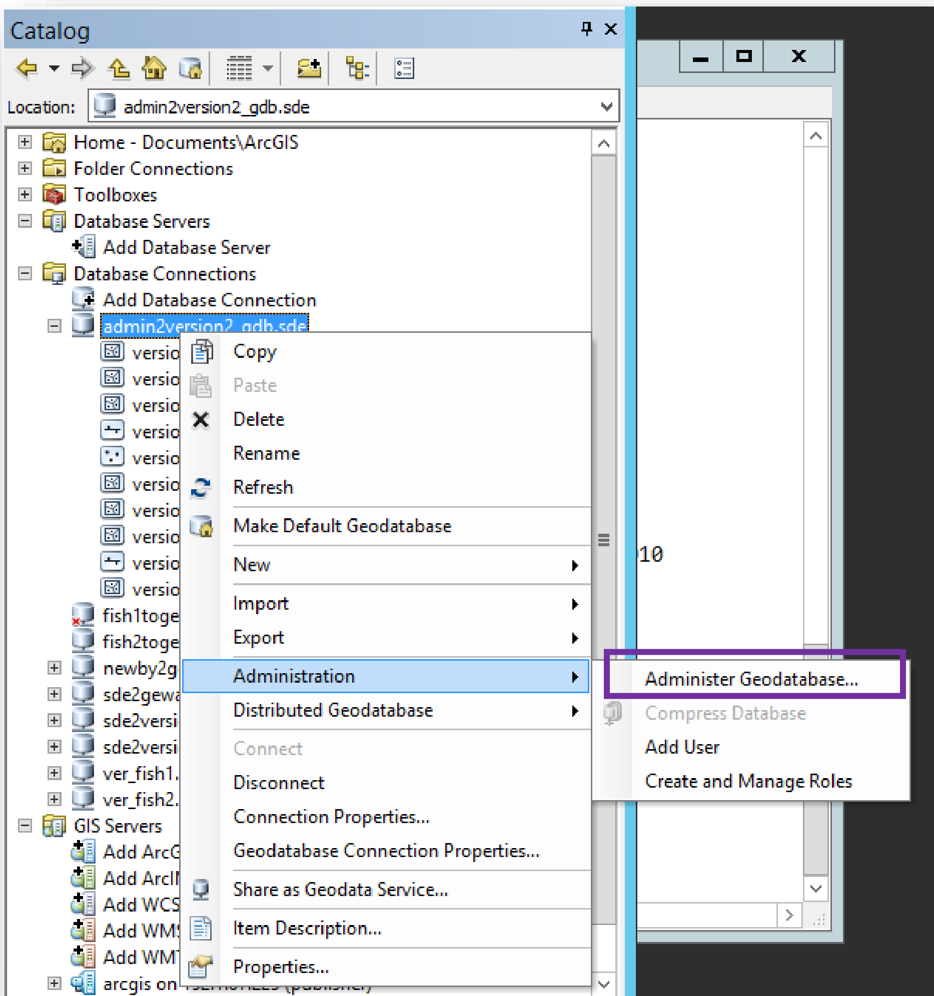


- This dialog box will pop up, do NOT check the check box > click OK

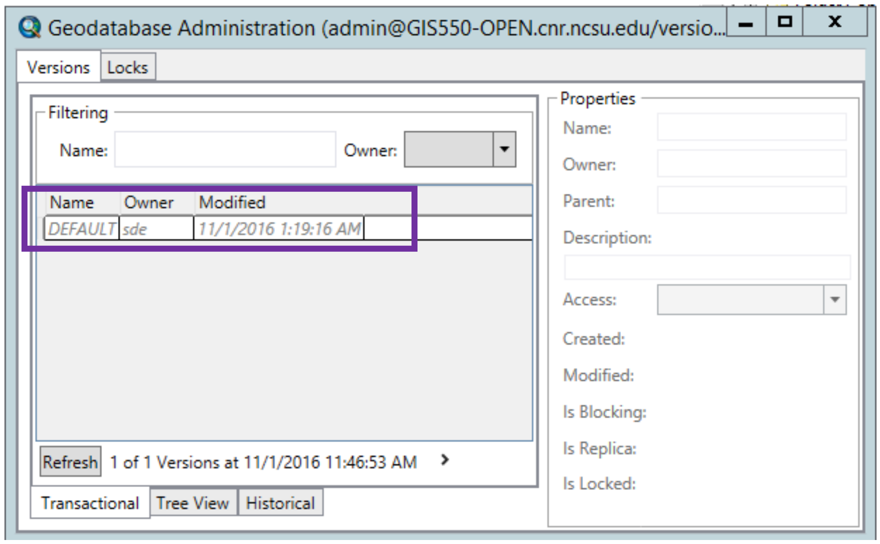


8. The next step is to create versions of the parent geodatabase (the DEFAULT version)

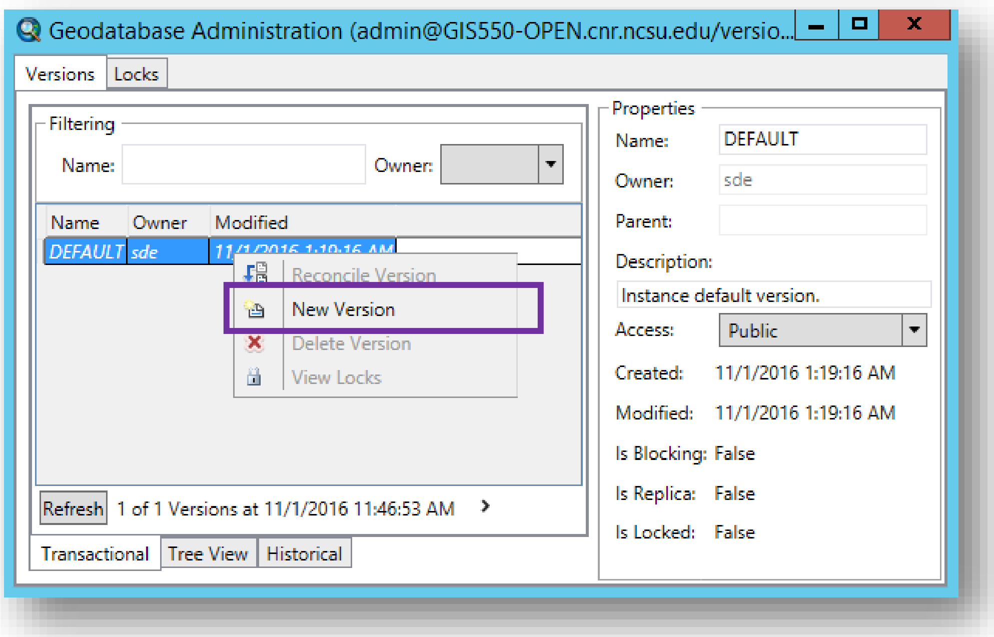
- Right click on the admin2version2_gdb connection > Administration > Administer Geodatabase



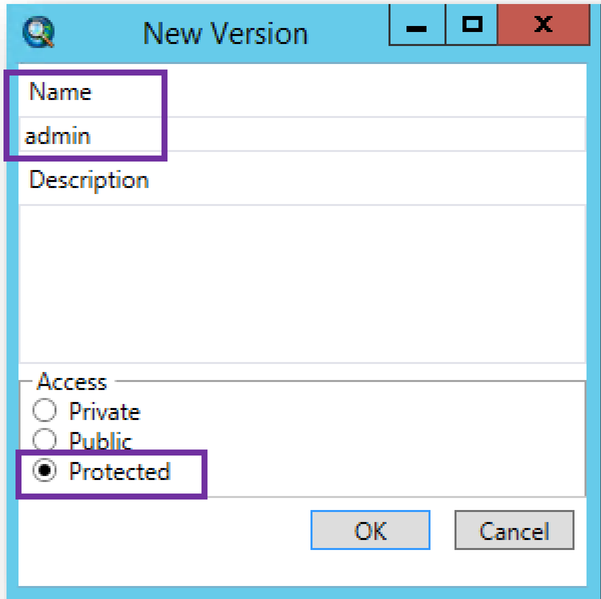
- This opens the Version Manager. The only version at this time is the DEFAULT version that gets created when an enterprise geodatabase is created.



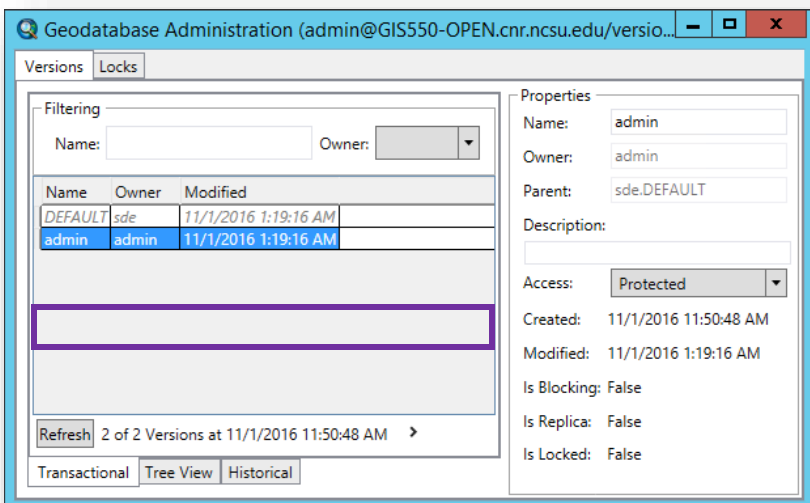
- Right click on the DEFAULT database > click New Version



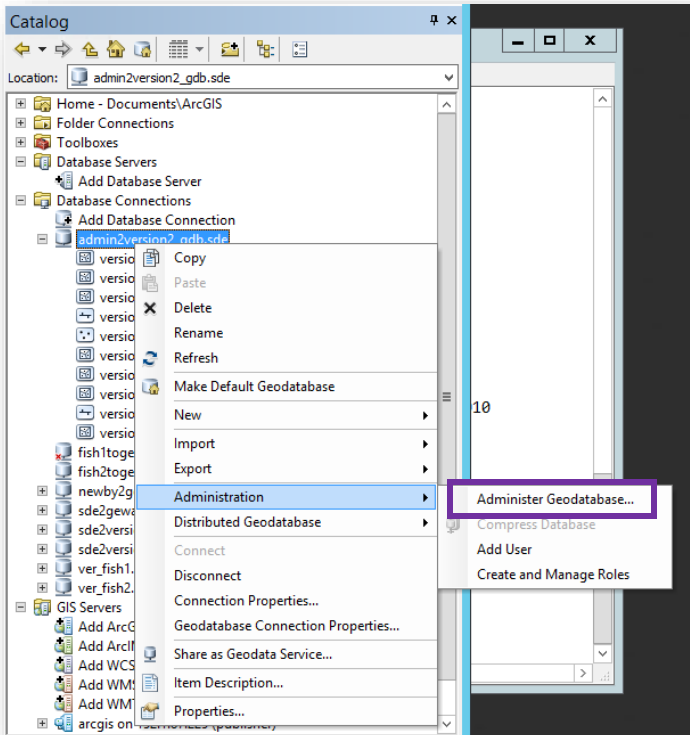
- Call the new version “admin”, add a description if desired, and check “protected” > click ok



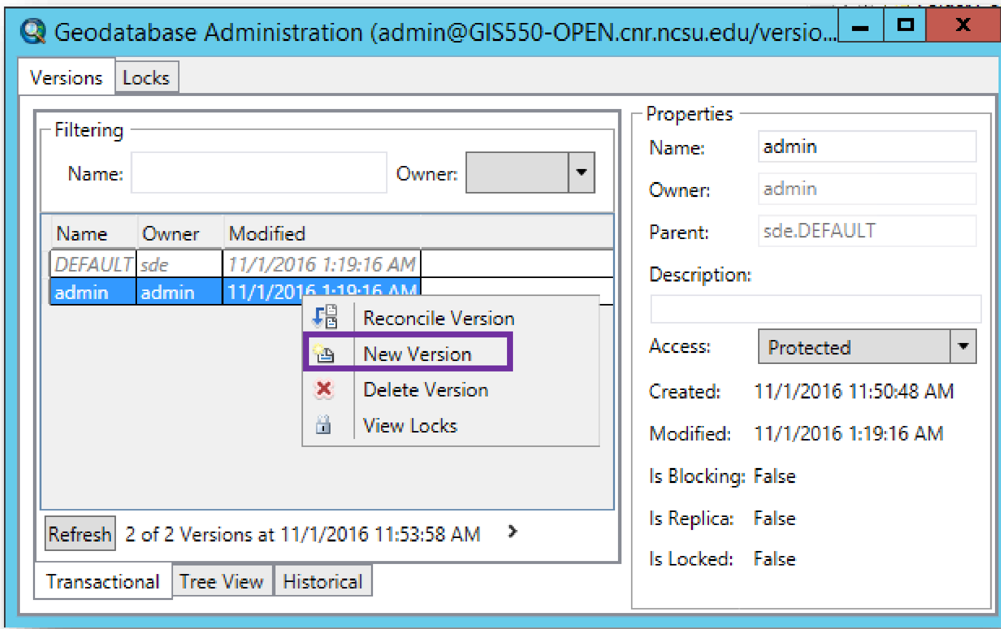
- Now there is a new version called “admin” in the Version Manager. Close this window for the changes to update.



- Go back to ArcCatalog, Right click on the admin2version2_gdb connection > Administration > Administer Geodatabase

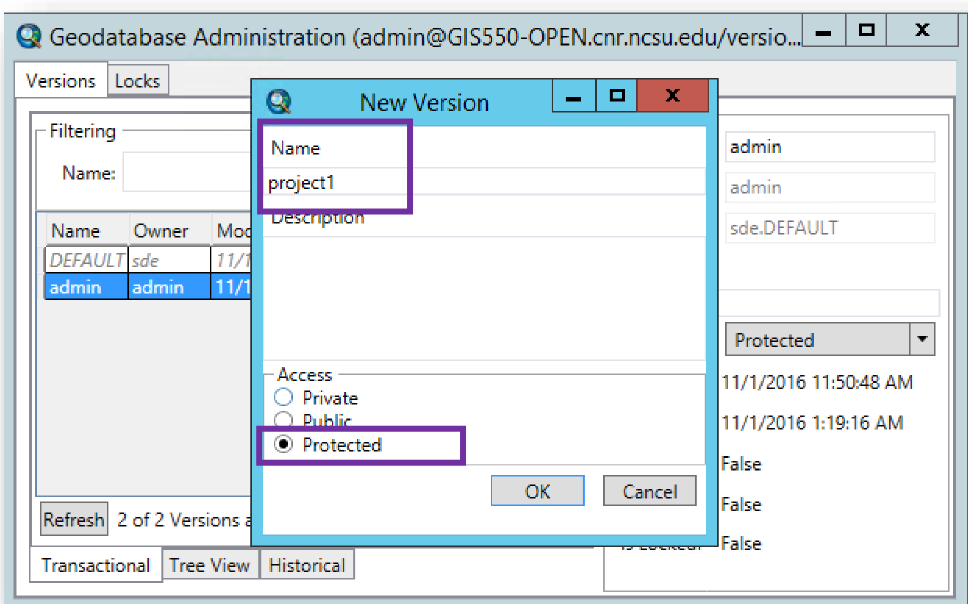


- From the admin version, create the two project versions: Right click on admin > New Version

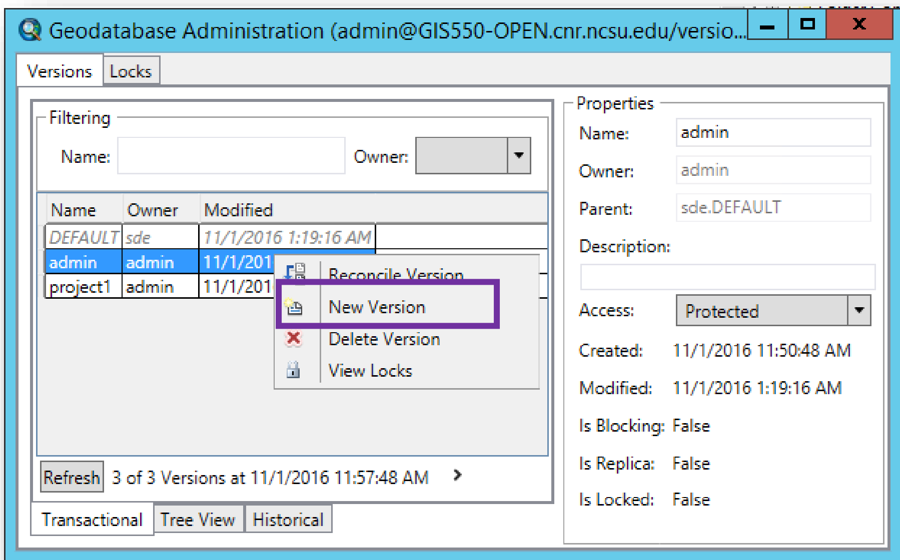


*notice how all of the options are now available: reconcile version, delete version, and view locks are now options. This is because the admin version was created with the admin user.

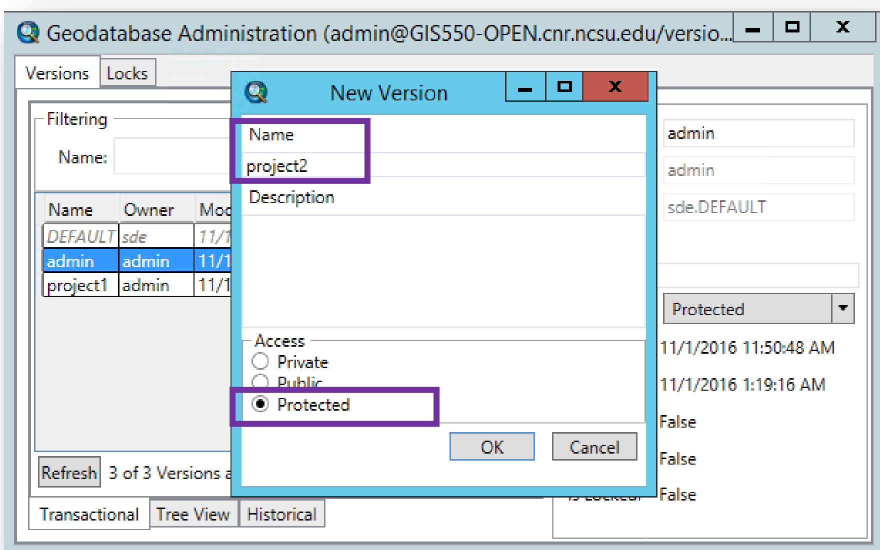
- Name this child version “project1” and access is Protected > click OK



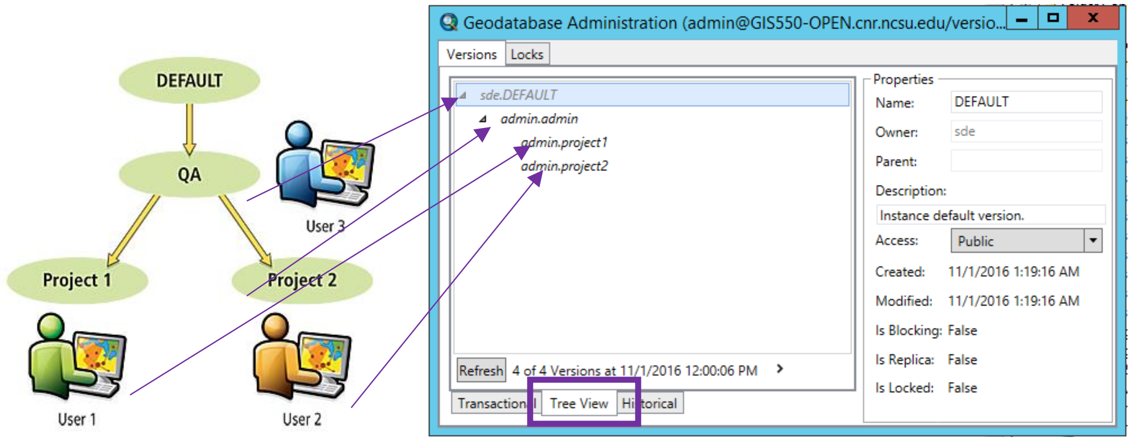
- Create the 2nd child version. Right click on the admin version > New Version



- Name the new child version "project2" and access is Protected > click ok



We have successfully created versions! Now we can view the lineage in the "Tree View Tab" of the Version Manager. Below we can see how the 3 – Level Version tree model is implemented in ArcMap

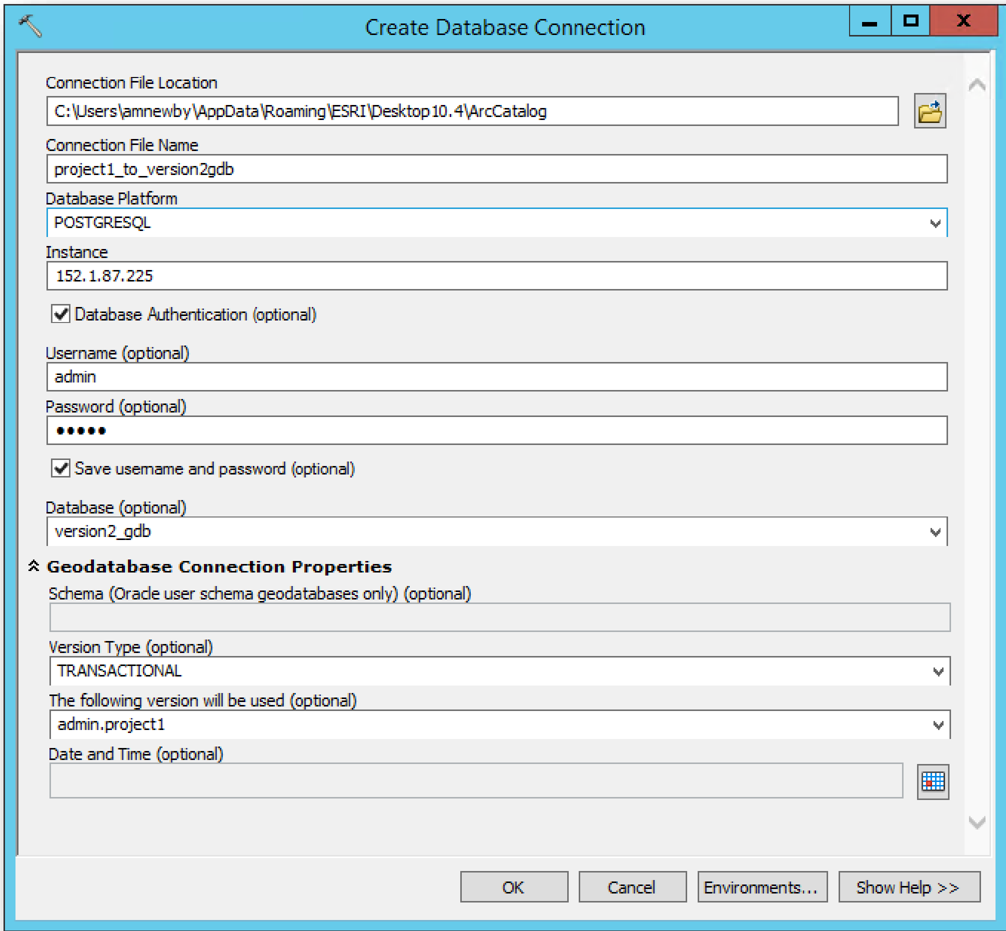


Now that the versions are created, we can allow different users to access and edit the buildings feature class through each version in a web map. But first, services need to be published from each Project version.

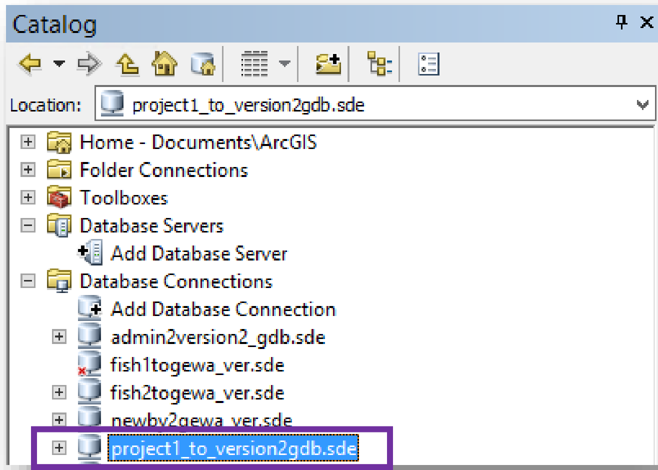
Section 2 – Publish services from two different versions

1. First, we must register the enterprise geodatabase in order to publish Feature Services:
   - Navigate to the "Create Database Connection" tool in the Data Management toolbox
     - For connection file location use: C:\Users**YourUnityID**\AppData\Roaming\ESRI\Desktop10.4\ArcCatalog
     - Use your IP address for the Instance
     - Fill out everything else as below
     - Make sure to expand "**Geodatabase Connection Properties**" to be able to select a version

*Note – ANYTIME YOU HAVE TO CONNECT TO A VERSION OTHER THAN DEFAULT, USE THE "Create Database Connection" TOOL. Right clicking the Database Connection in ArcCatalog does not work.

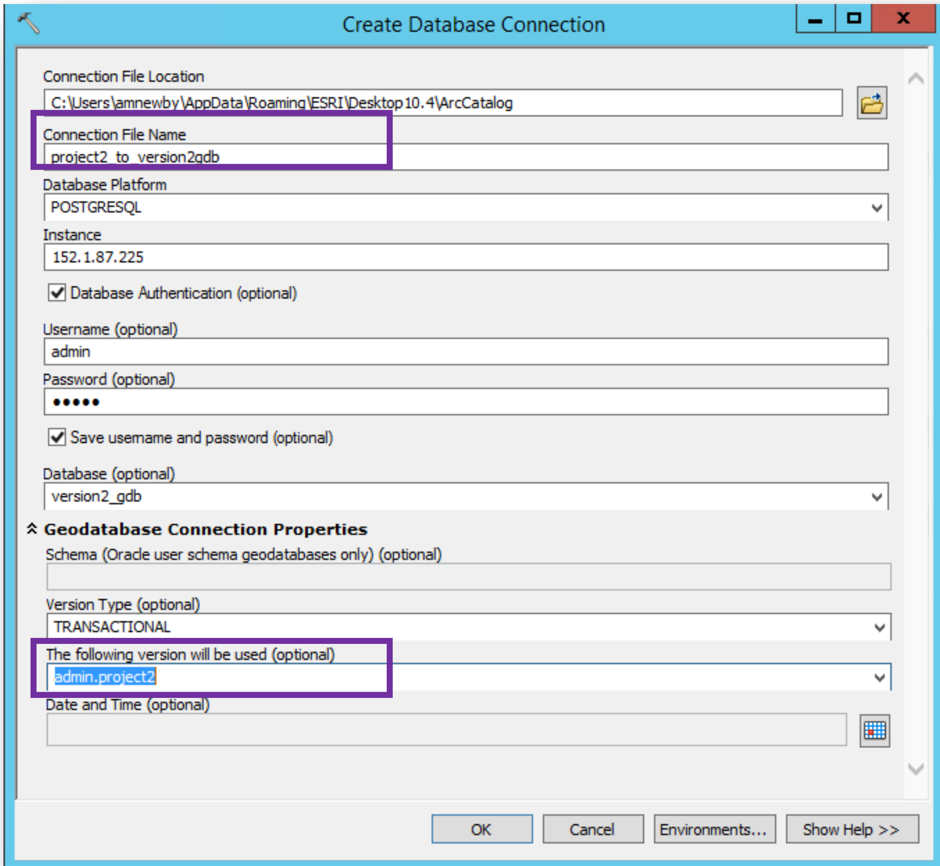


- And now we have a database connection to a SPECIFIC version (You may need to refresh the Database Connections for it to show up)

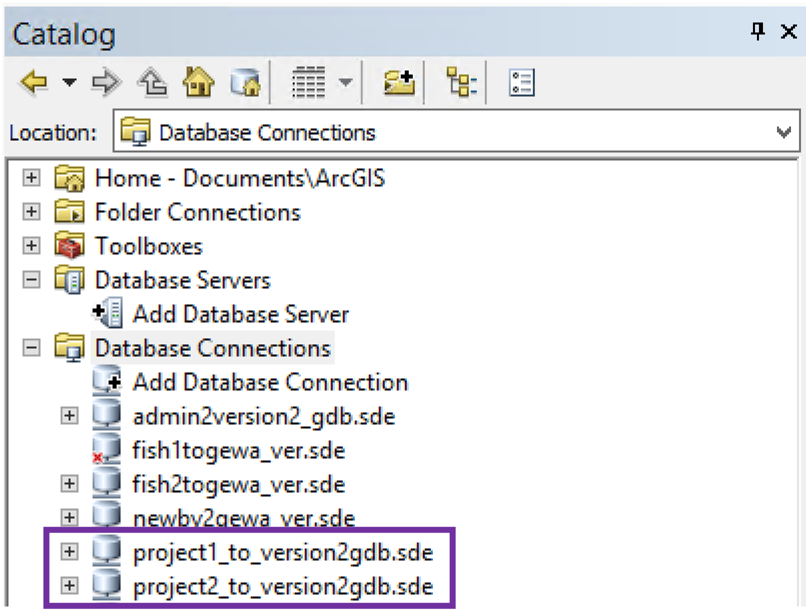


2. Repeat step 1 for the project2 version

- Navigate to the "Create Database Connection" tool in the Data Management toolbox again
    - Fill out as below, only this time the connection file name is "project2_to_version2gdb" and the version to be used is "admin.project2". Use your IP address for the Instance

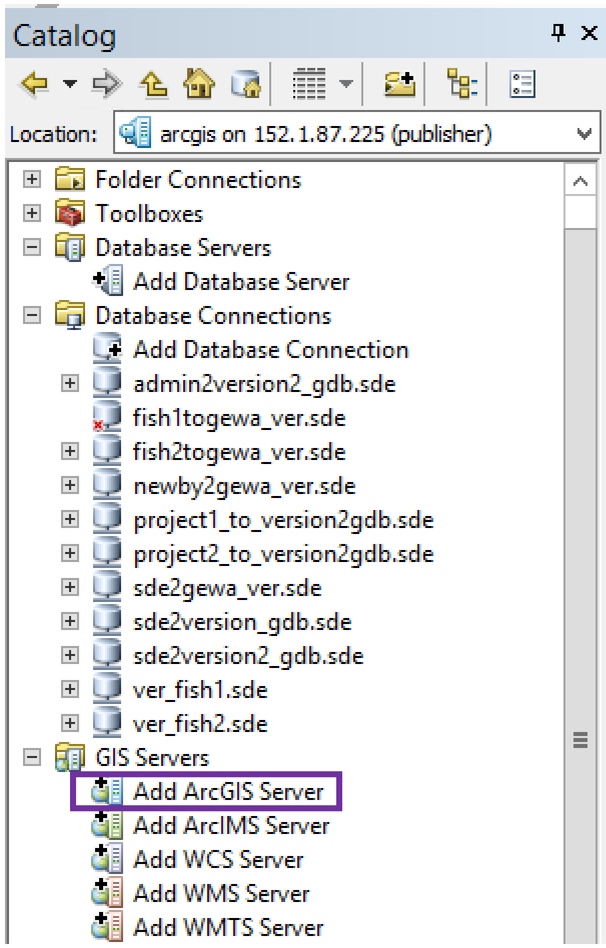


Now we have two geodatabase connections (to versions) that can be registered.

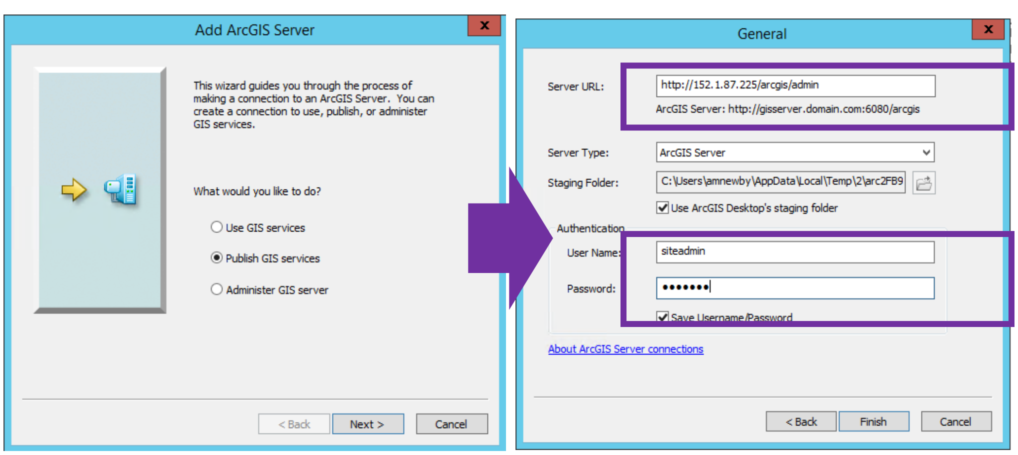


3. Create a GIS Server connection if one is not already made

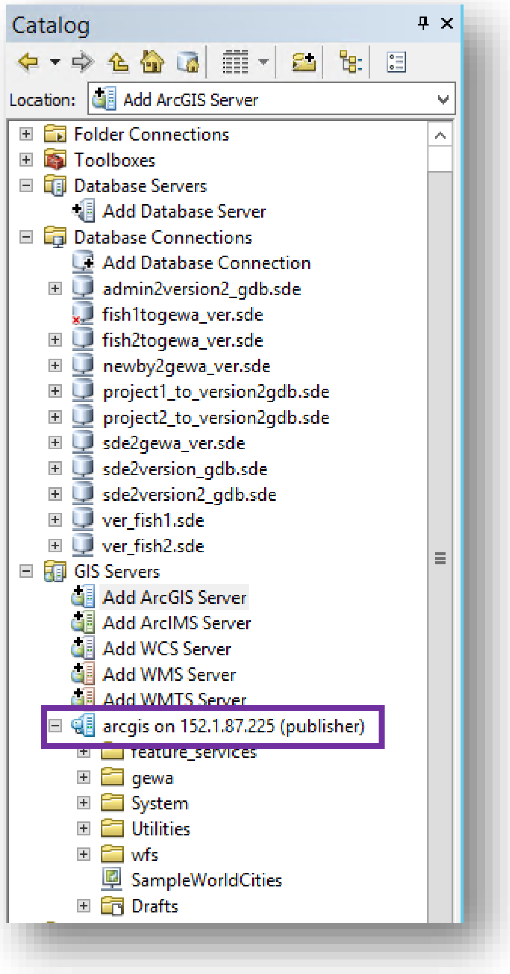
- Double click on “Add ArcGIS Server”



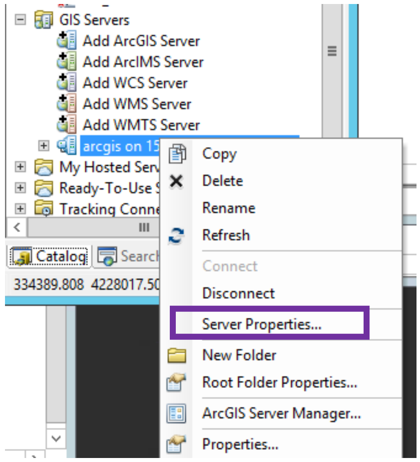
- Select "Publish GIS Services" and click Next > Replace the IP address with your own in the Server URL > enter the ArcGIS Server Manager credentials for Authentication > click Finish



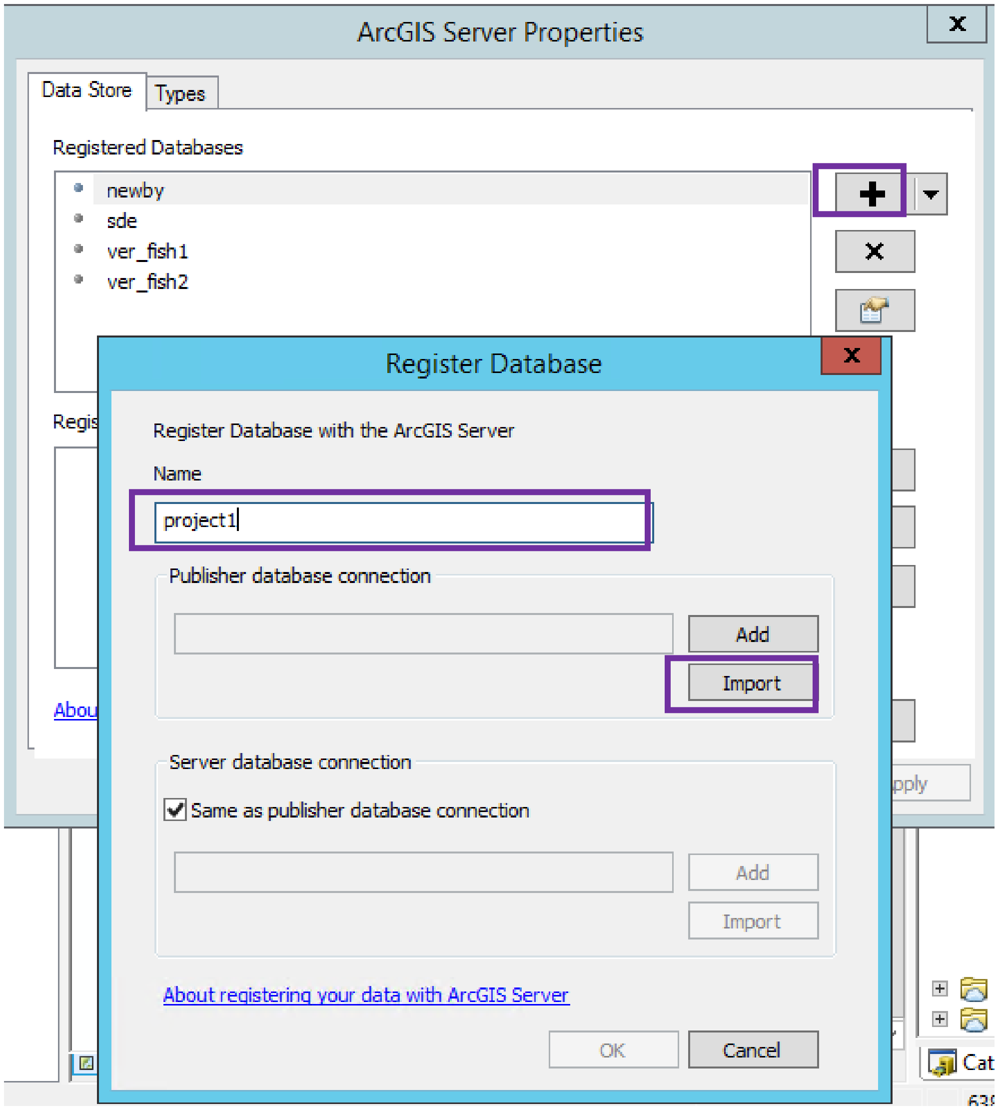
- Now the GIS Server connection shows up



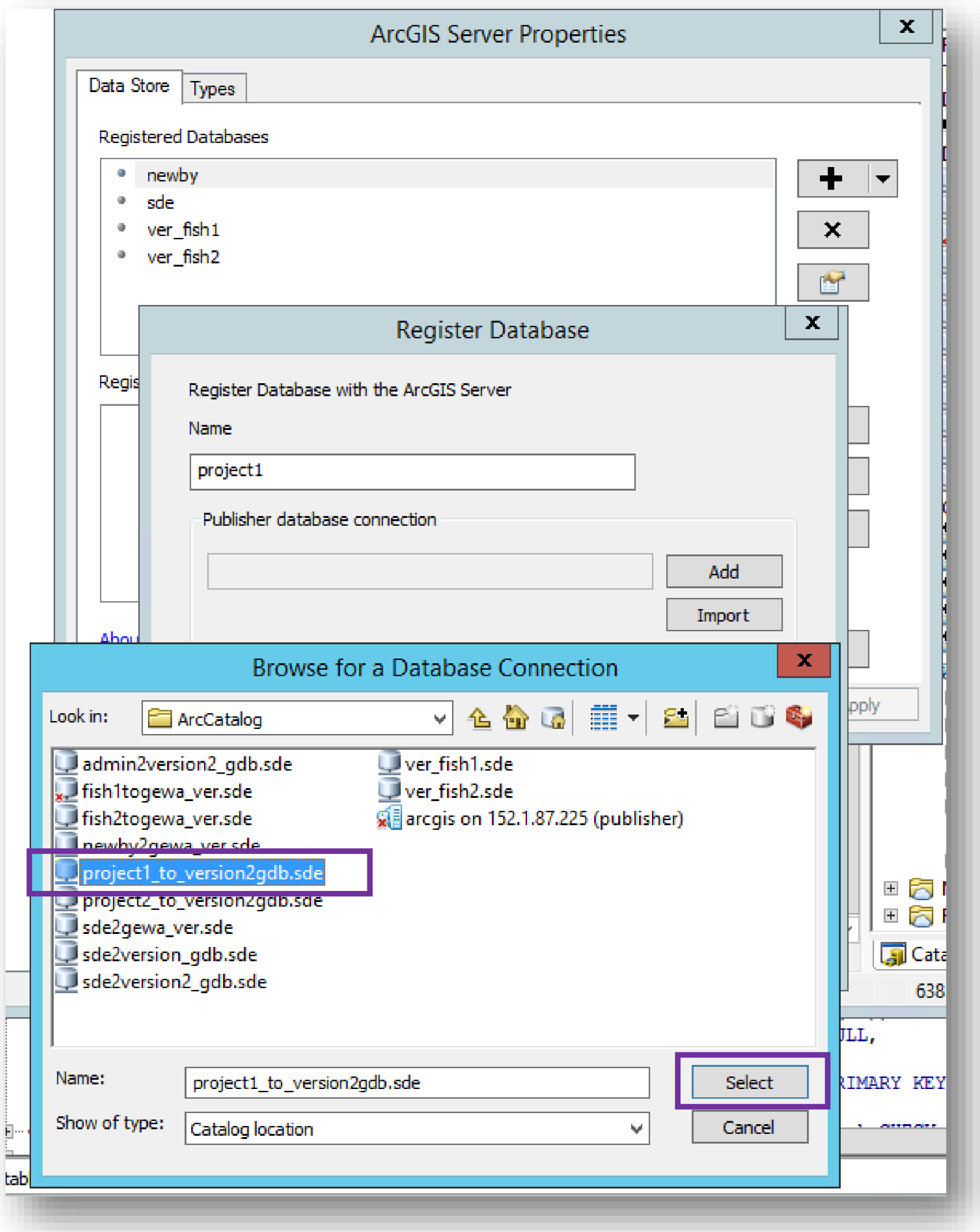
4. Register the versioned databases
    - Right click on the publisher connection > click Server Properties



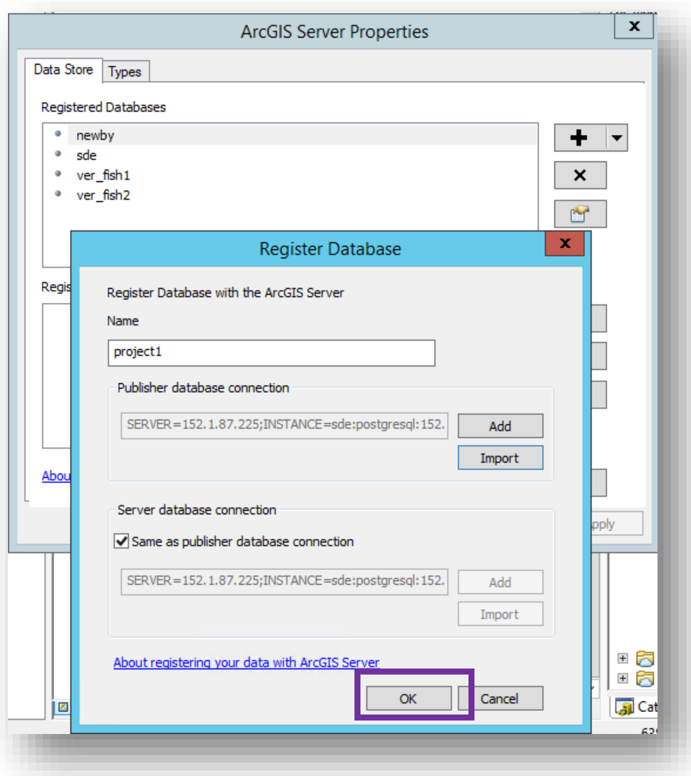
- Click on the Plus sign > Name it "project1" > Click Import



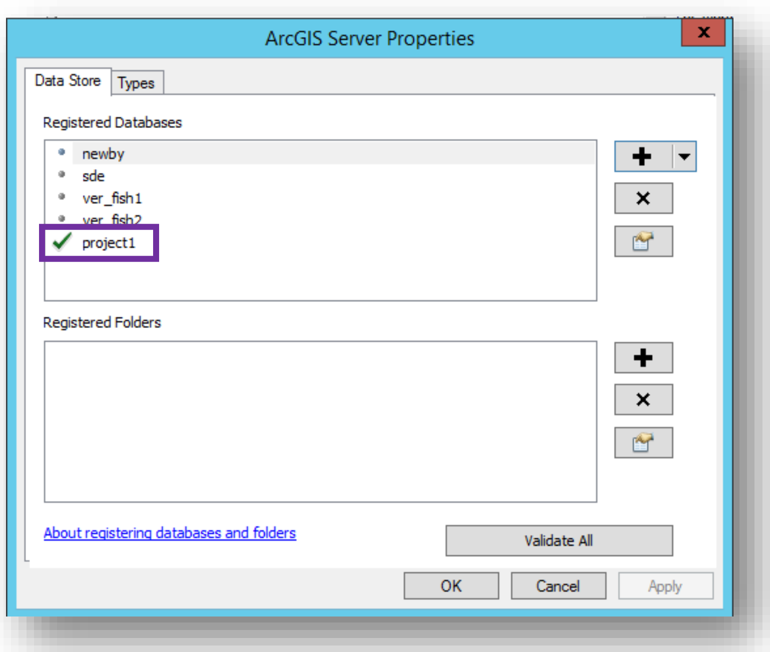
- Select the connection to the project1 version > click select


ArcGIS Server Properties
Data Store
Types
Registered Databases
newby
sde
ver_fish1
ver_fish2
Register Database
Register Database with the ArcGIS Server
Name
project1
Publisher database connection
Add
Import
Browse for a Database Connection
Look in:
ArcCatalog
admin2version2_gdb.sde
fish1togewa_ver.sde
fish2togewa_ver.sde
project1_to_version2gdb.sde
sde2gewa_ver.sde
sde2version_gdb.sde
sde2version2_gdb.sde
ver_fish1.sde
ver_fish2.sde
arcgis on 152.1.87.225 (publisher)
Name:
project1_to_version2gdb.sde
Show of type:
Catalog location
Select
Cancel

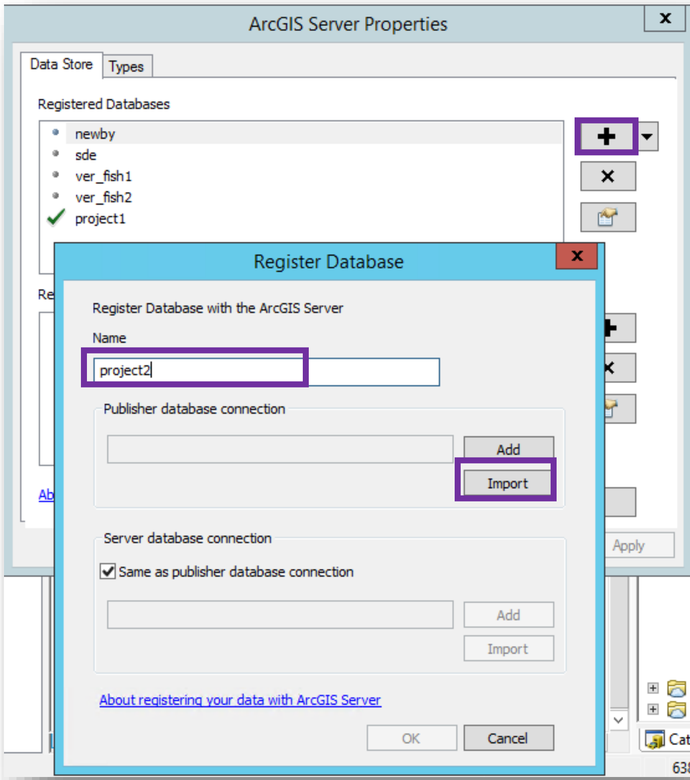
- Click OK



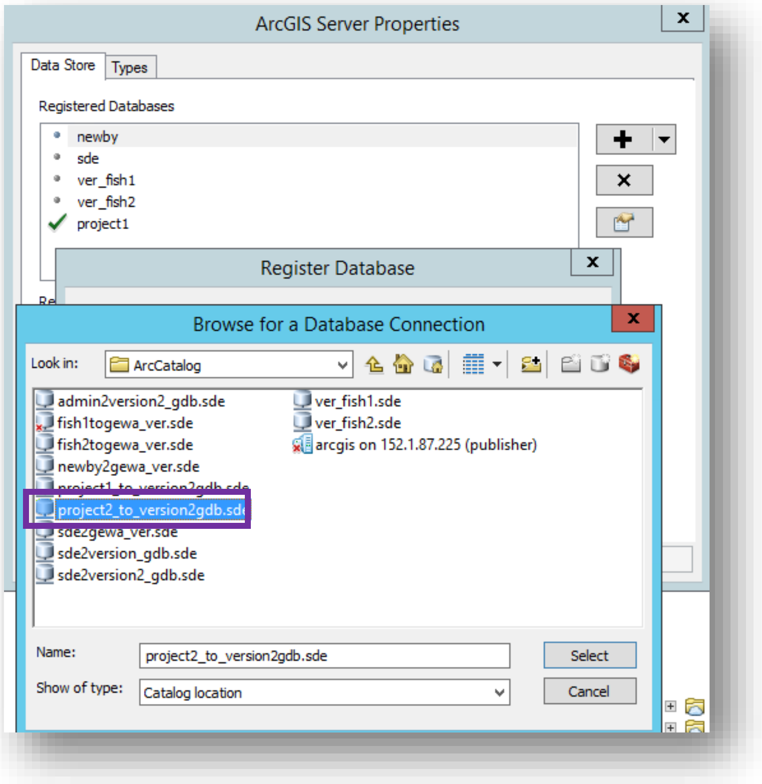
- Now the project1 version database connection is successfully registered



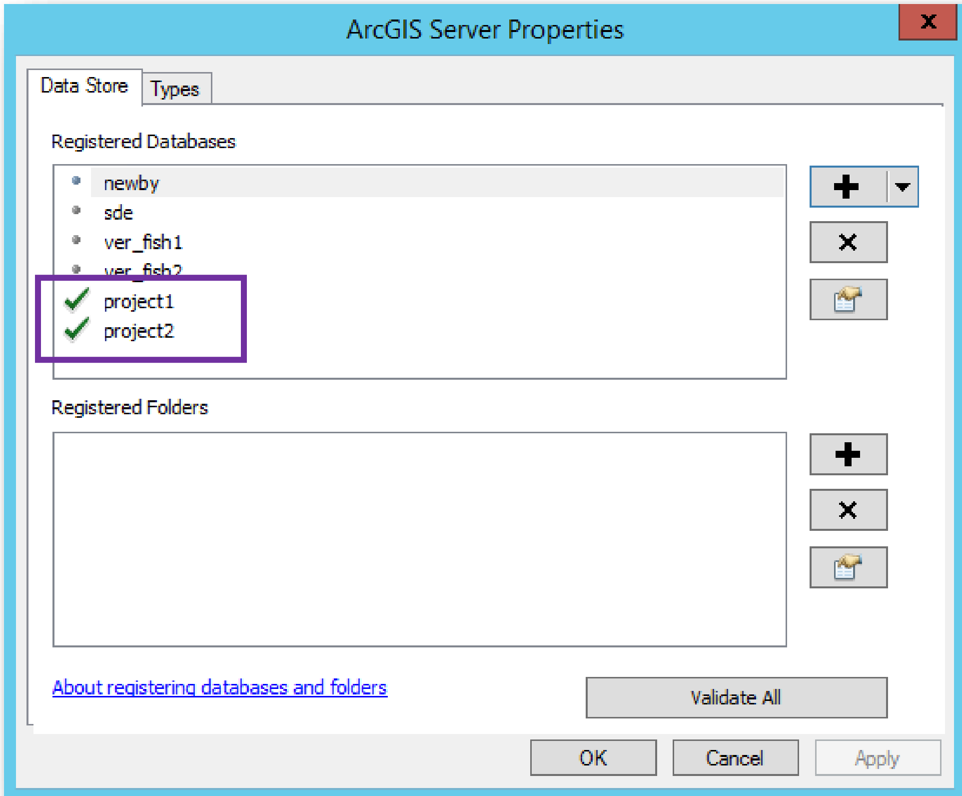
- To register the project2 version, click the plus sign again > name it “project2“ > click Import



- Select the project2_to_version2gdb.sde database connection > then click OK

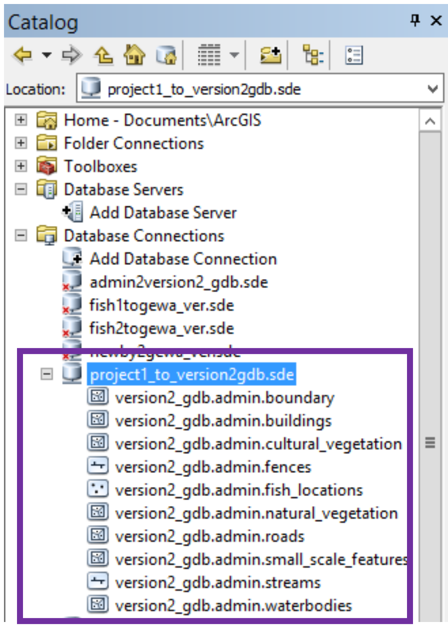


- Now both versions of the database are registered > Click OK to return to the main ArcMap page

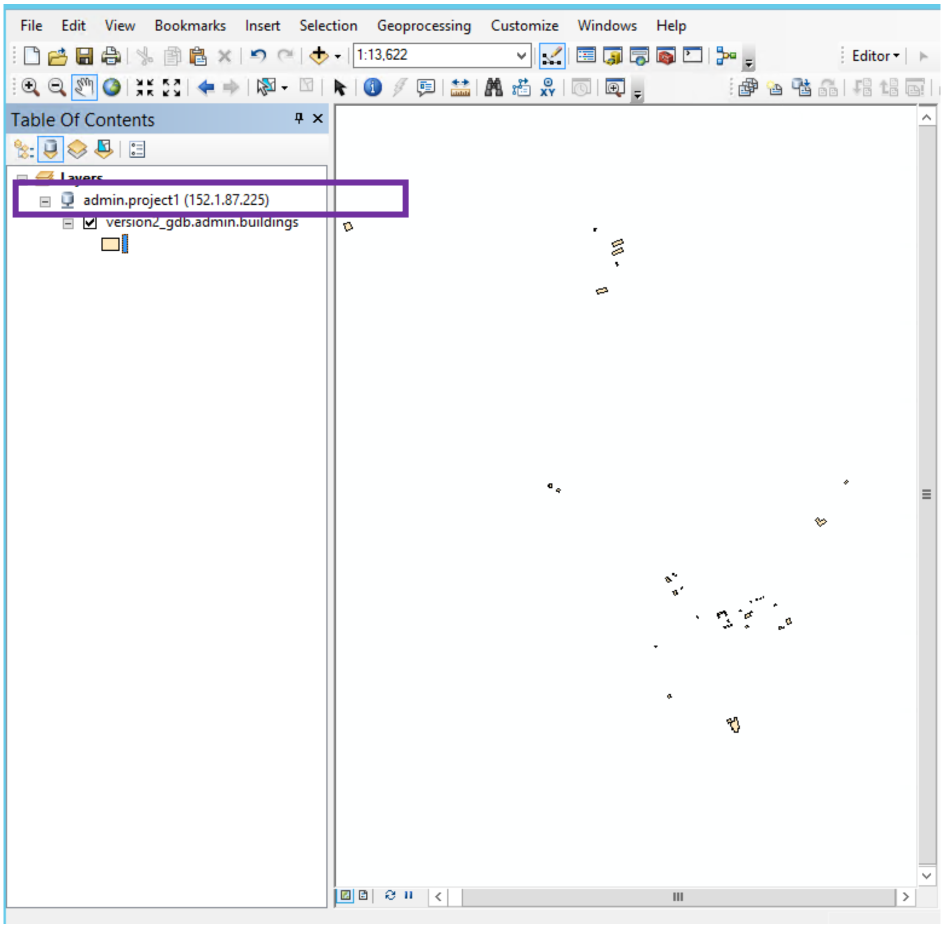


5. Publish Feature Services from each version

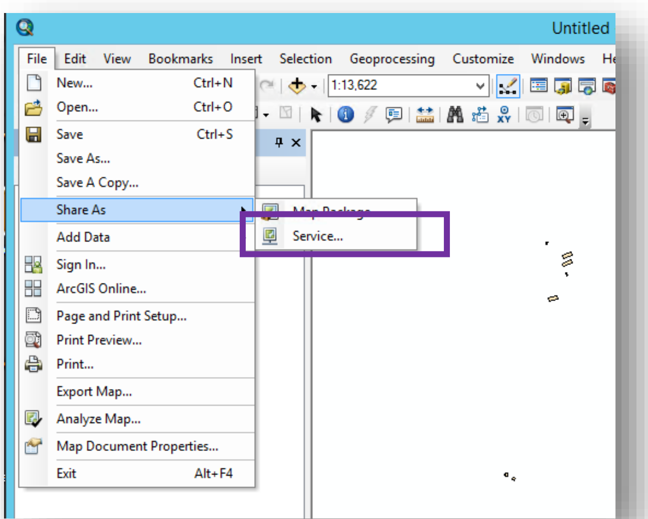
- Double click the project1_to_version2gdb.sde connection to activate it > click the "-" symbol to expand the layers



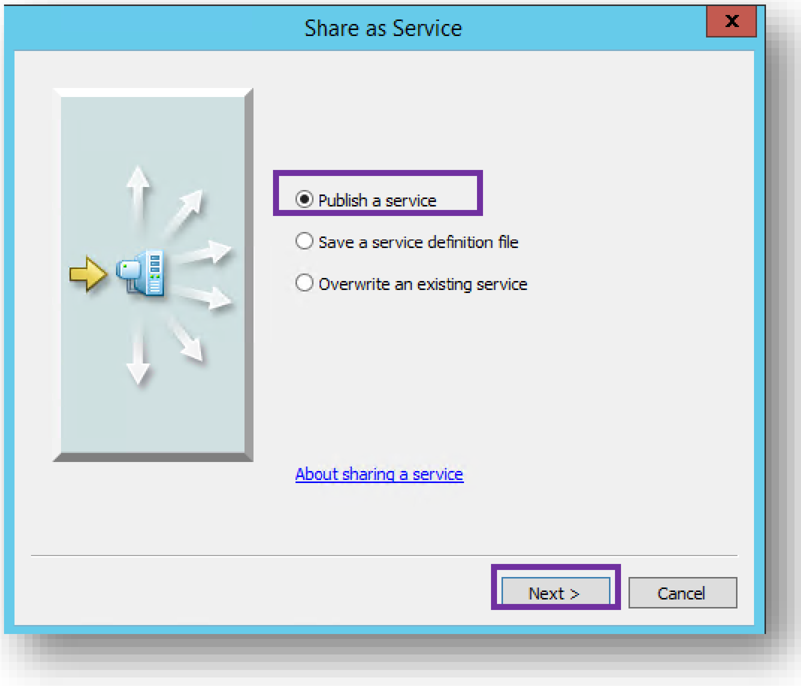
- Add the buildings feature class to ArcMap > symbolize as desired
    - *Note the data source shows the buildings are coming from the admin.project1 (152.1.87.225) database connection. Also, if you have trouble adding layers to ArcMap from a database connection, close ArcMap and reopen it.



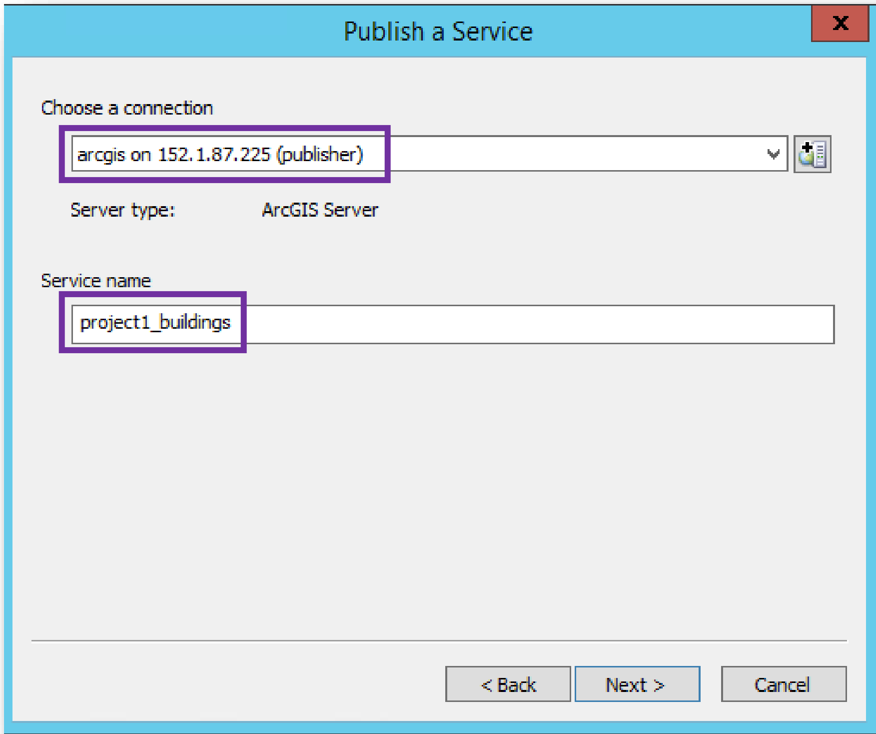
- Go to File > Share As > Service



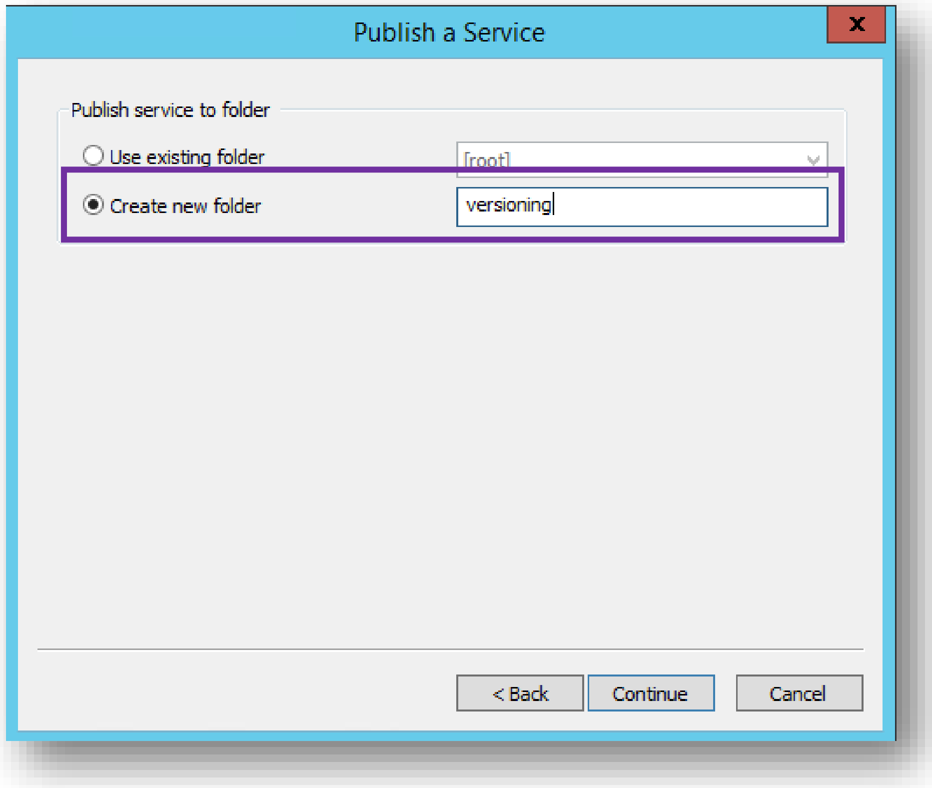
- This dialog box pops up > Choose Publish a Service > click Next



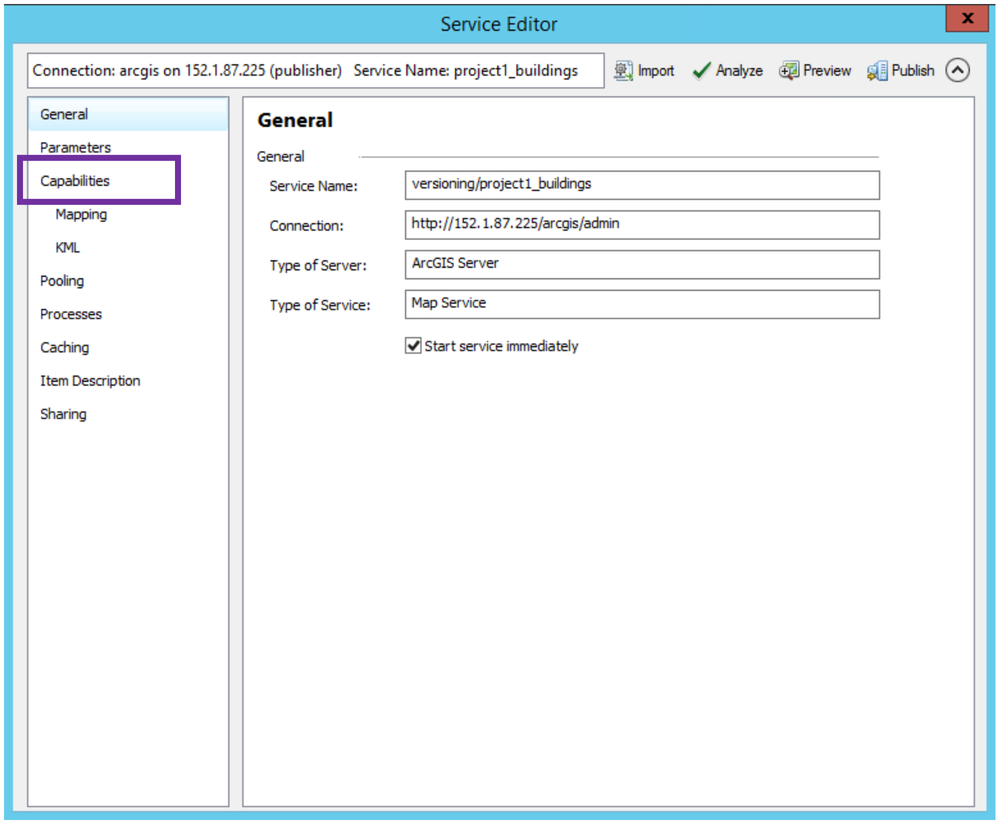
- For the connection, use the GIS Server Connection and name it "project1_buildings" > click Next



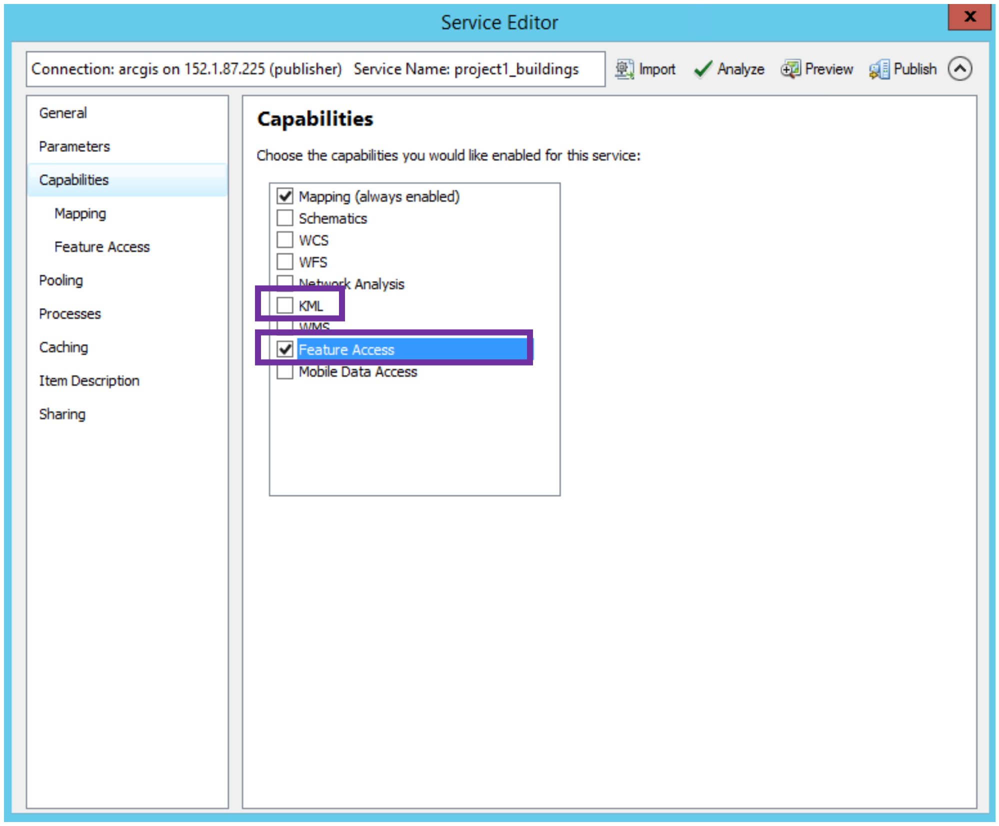
- Create a new folder called versioning > click Continue



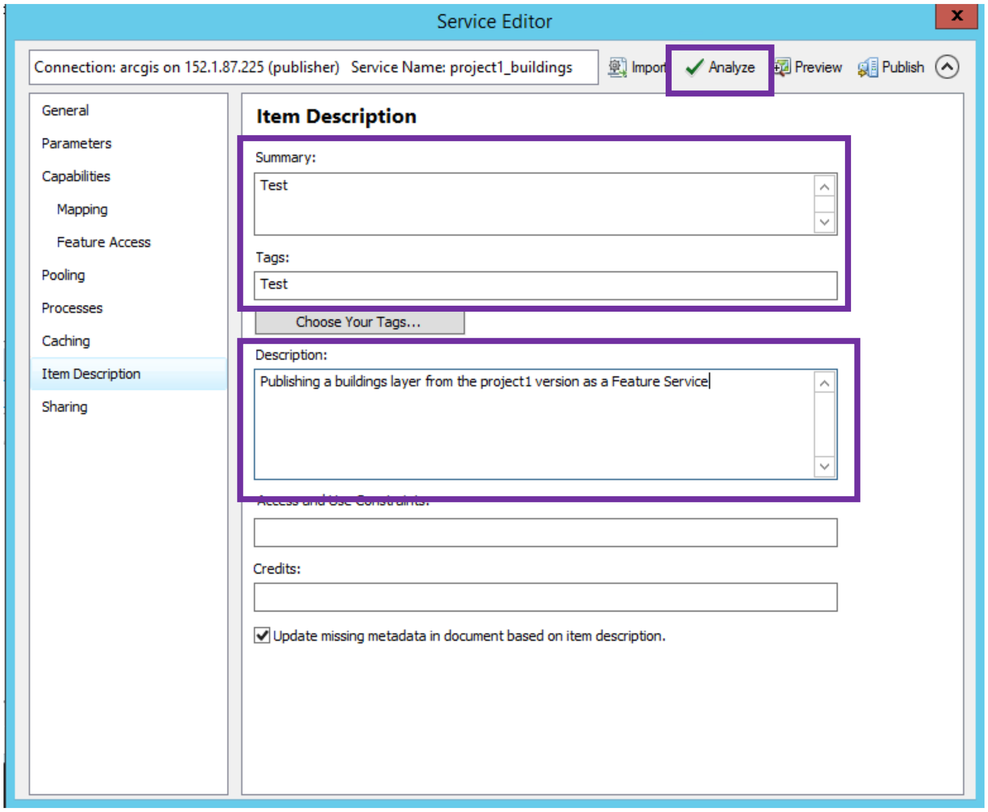
- Then the Service Editor > click on Capabilities



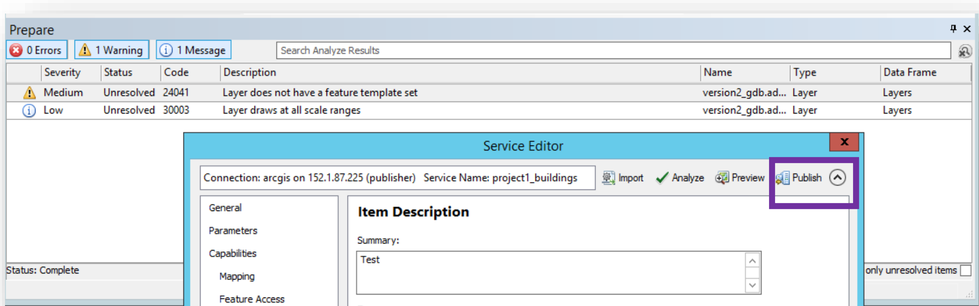
- Uncheck KML and check Feature Access



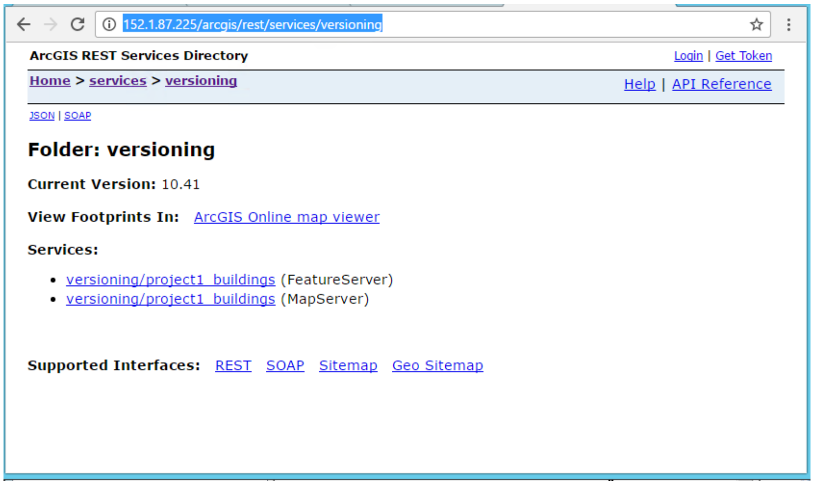
- Then skip down to the Item Description and fill out the summary, tags, and description > click Analyze in the top right corner



- If you do not see any Errors and there are no Warnings that the data will be copied to the Server > click Publish

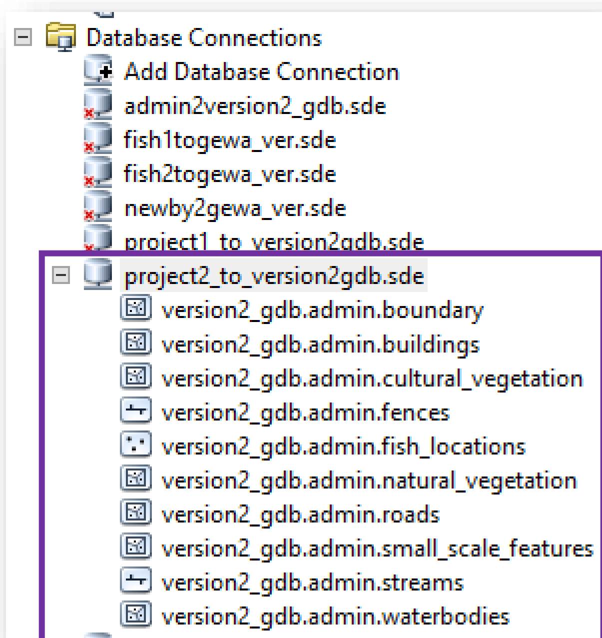


- In a web browser, navigate to the REST services interface (should be something like http://YourIPaddressHere/arcgis/rest/services/versioning) You now have a REST service published from a version!

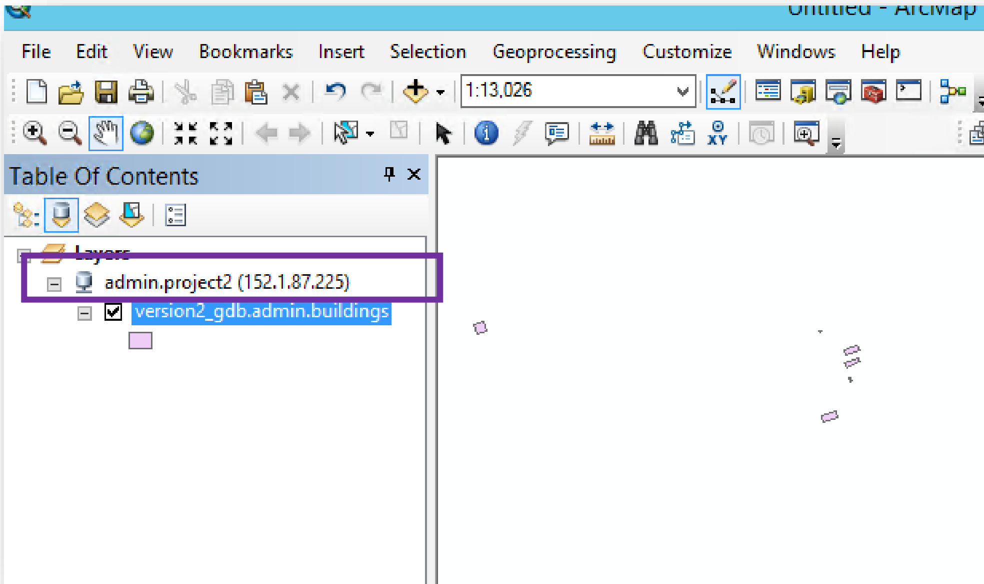


6. Now repeat the process for project2

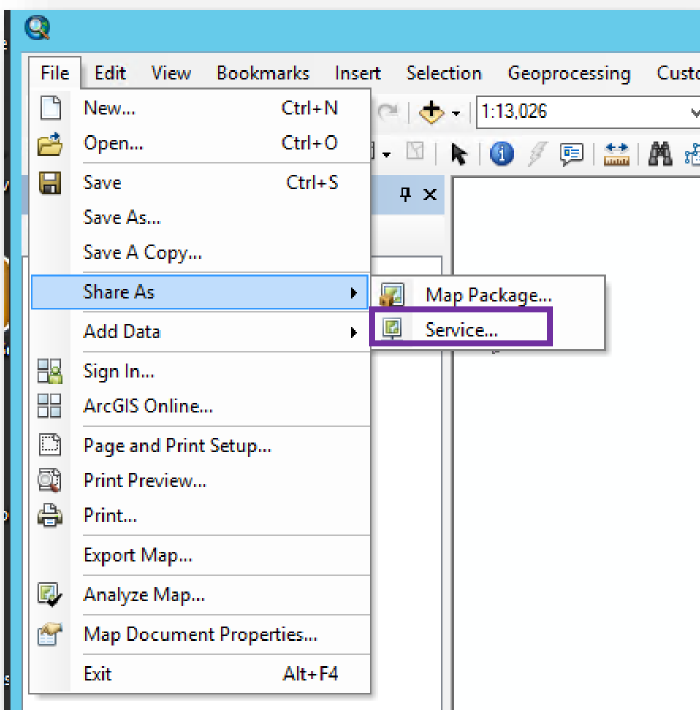
- Open a new blank ArcMap > Double click the project2_to_version2gdb.sde connection to activate it



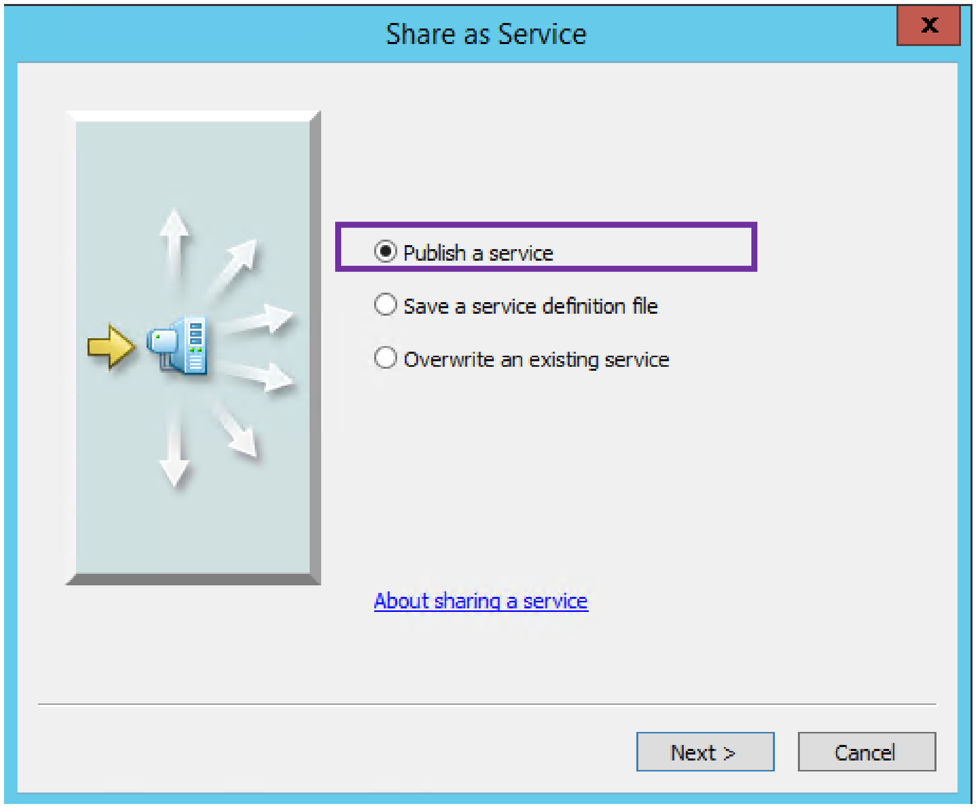
- Add the buildings to ArcMap > symbolize as desired
- *Note the layer is coming from the admin.project2 version this time



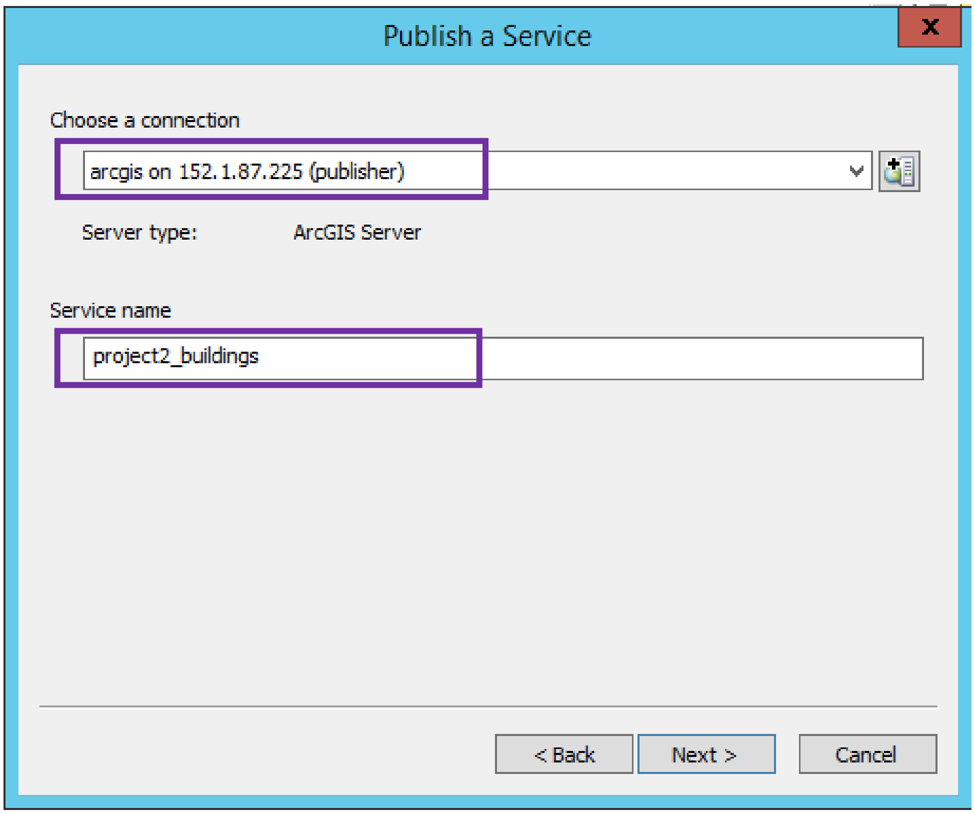
- Go to File > Share As > Service



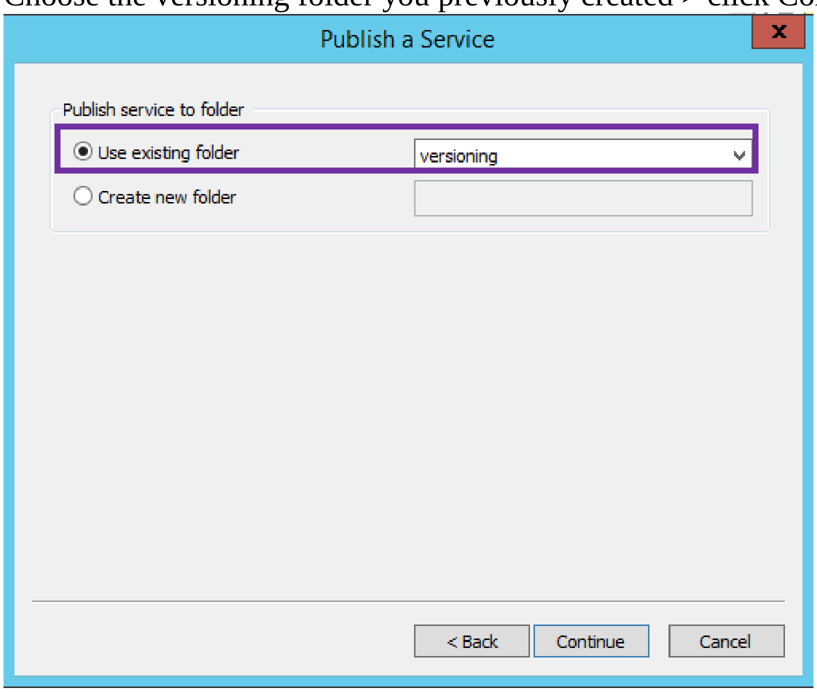
- Choose Publish a Service > click Next



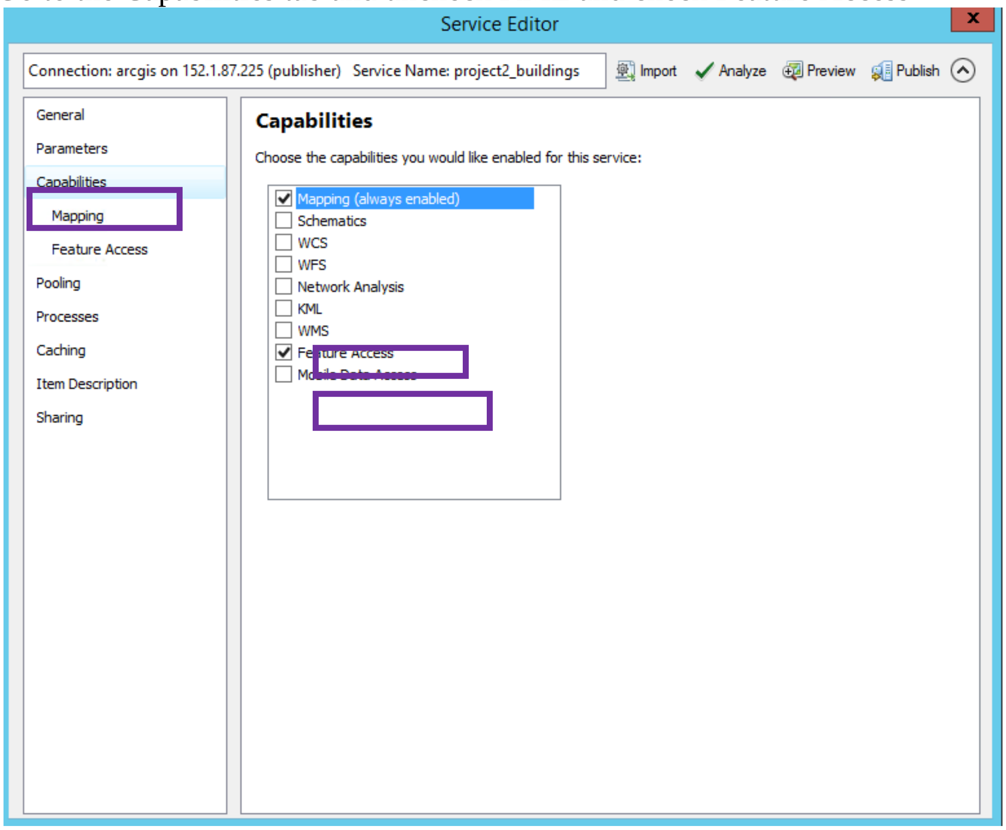
- Use the GIS Server Connection and name the service “project2_buildings” > click Next



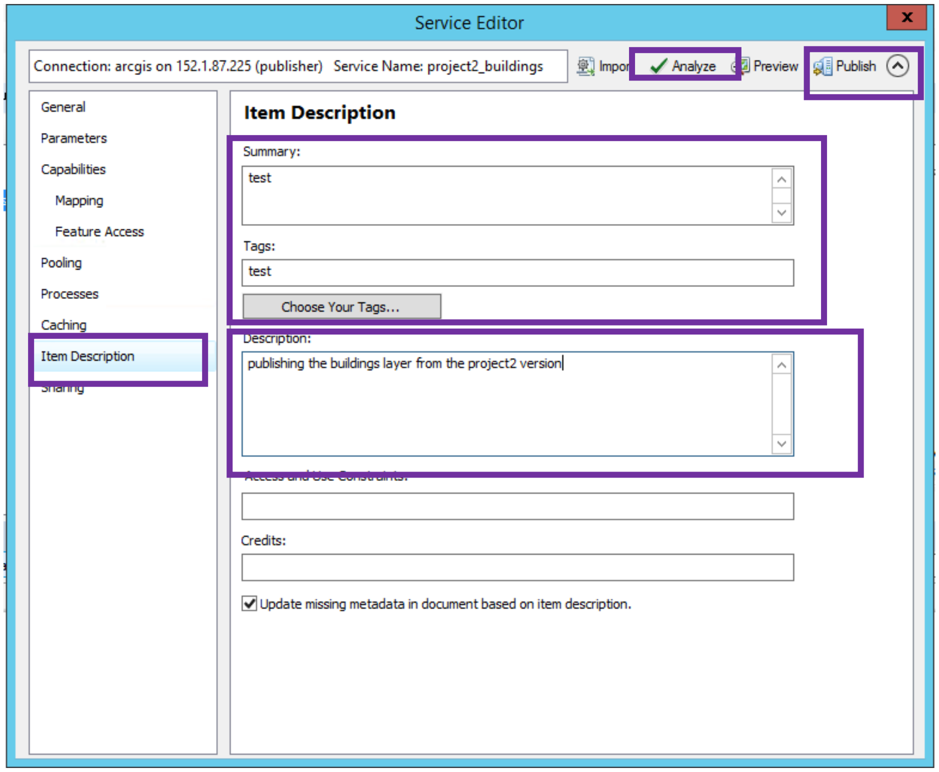
- Choose the versioning folder you previously created > click Continue



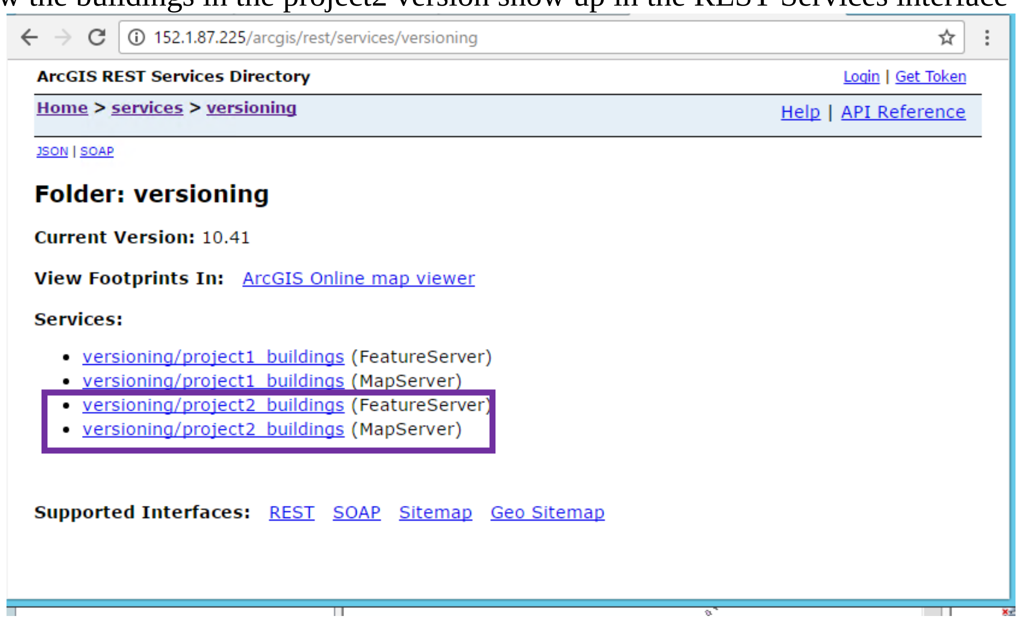
- Go to the Capabilities tab and uncheck KML and check Feature Access



- Skip down to item description, fill out the summary, tags, and description > Click Analyze
    - If you do not see any Errors and there are no Warnings that the data will be copied to the Server > click Publish

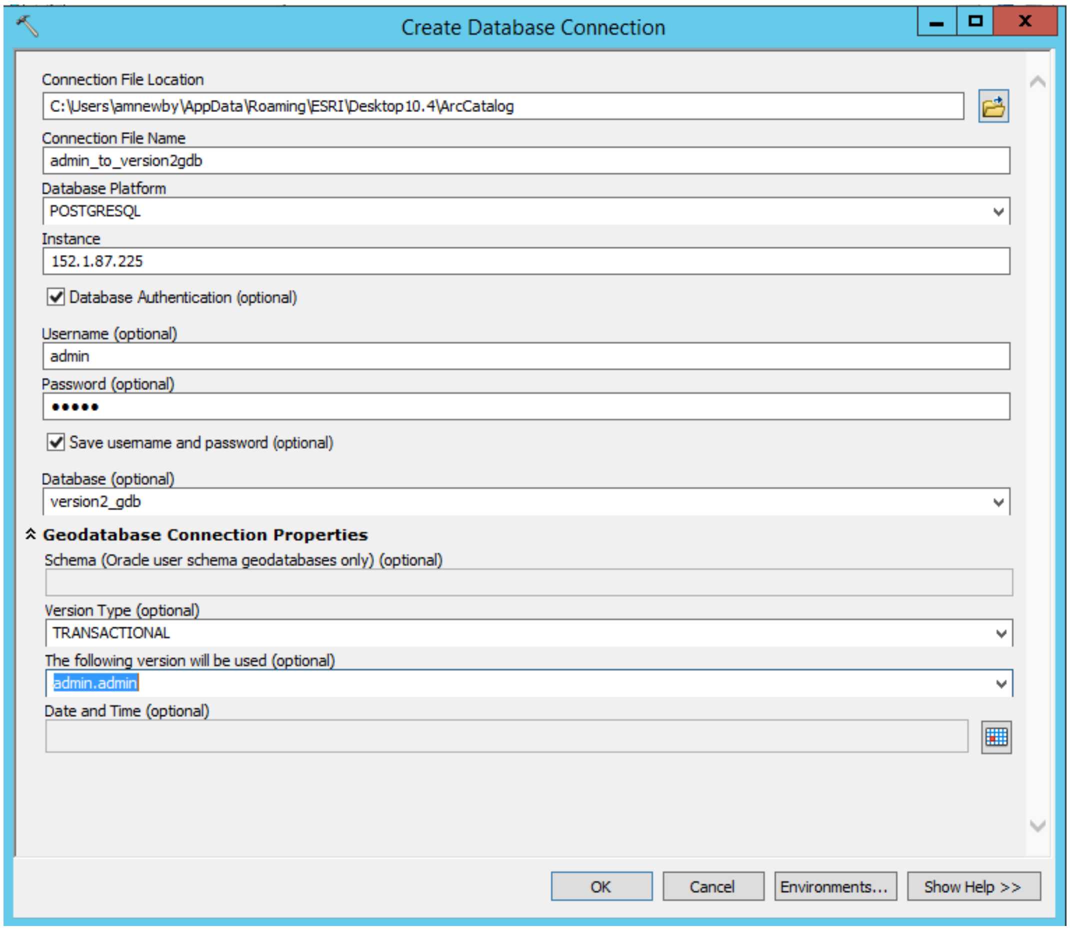


- Now the buildings in the project2 version show up in the REST Services interface

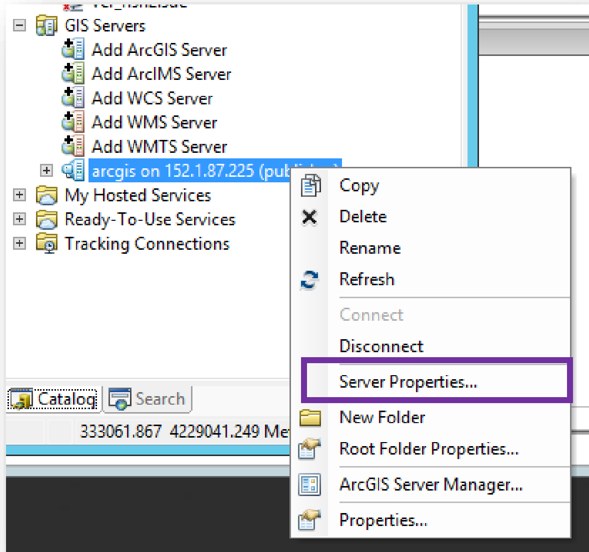


7. Publish the base GEWA data residing in the admin version as a Map Service
    - First created a database connection. Navigate to the "Create Database Connection" tool in the Data Management toolbox

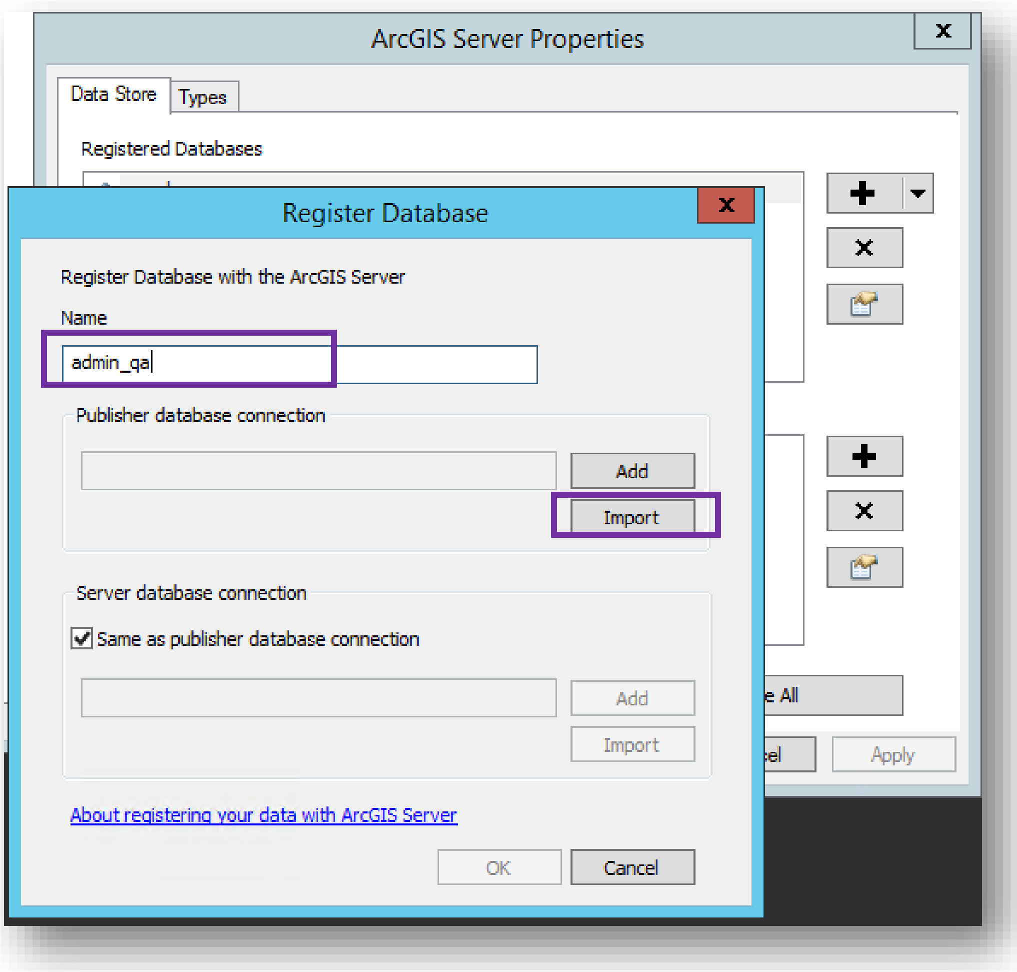
- For connection file location use: C:\Users**YourUnityID**\AppData\Roaming\ESRI\Desktop10.4\ArcCatalog
- Use your IP address for the Instance
- Fill out everything else as below



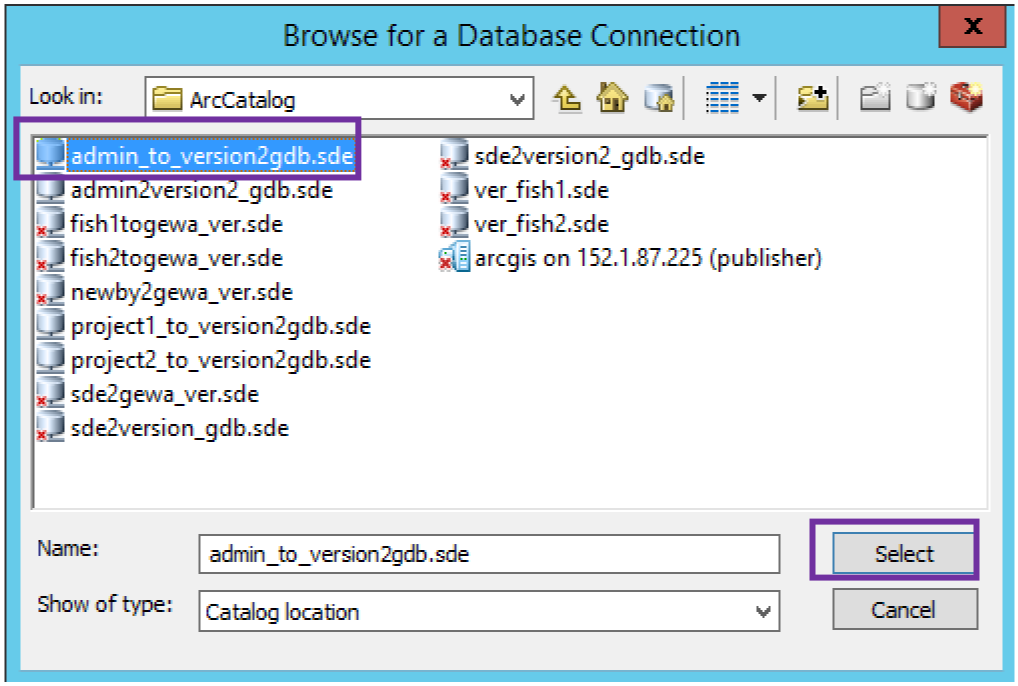
- Now register the admin version of the database with the Server Connection: Right click on the GIS Server connection > click on Server Properties



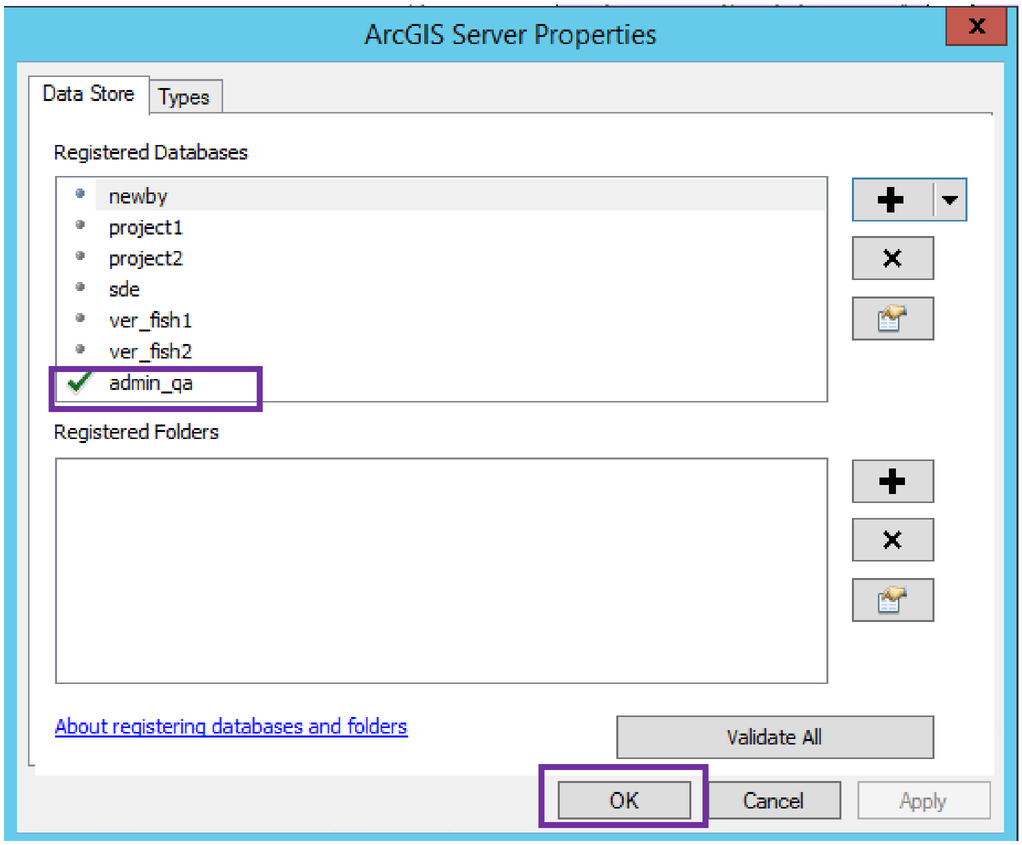
- For the Name use: admin_qa > click Import



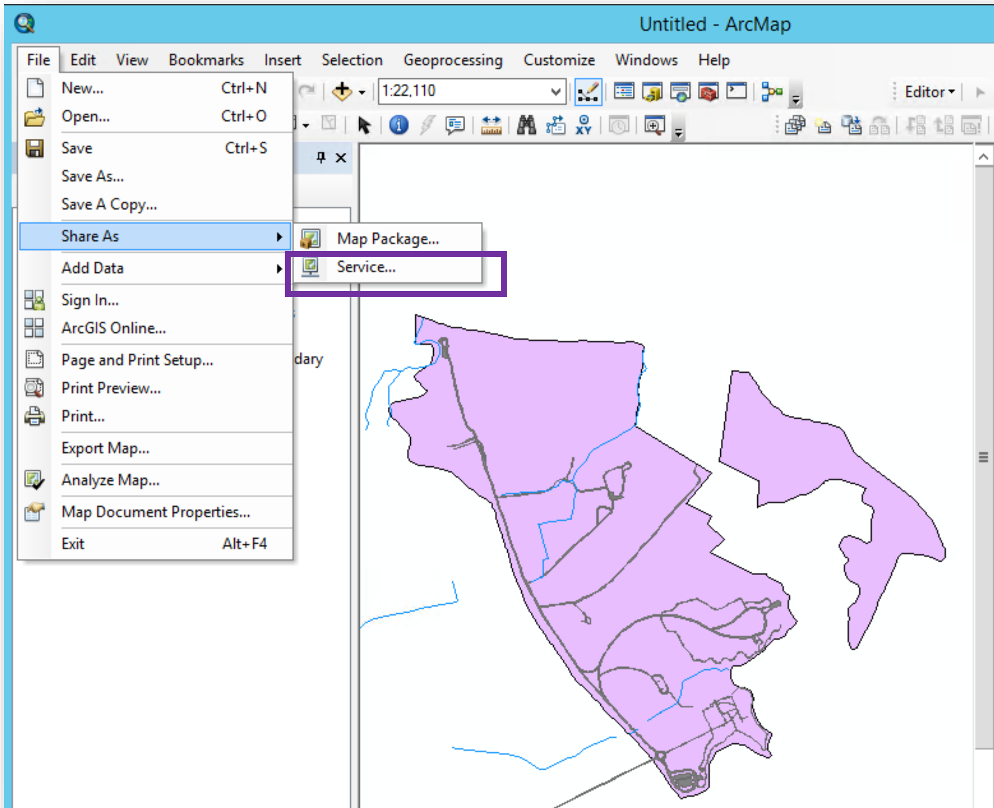
- Select the admin_to_version2gdb.sde database connection > click ok



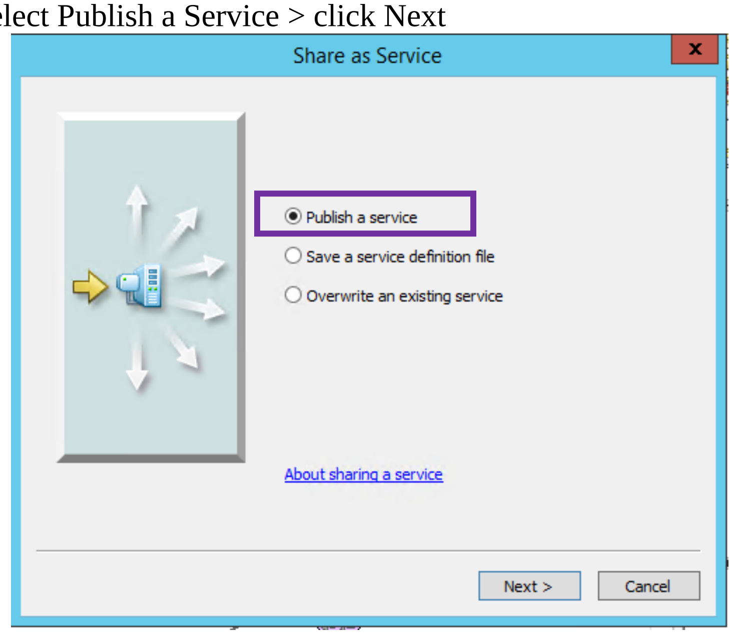
- Now the admin version is registered > click OK



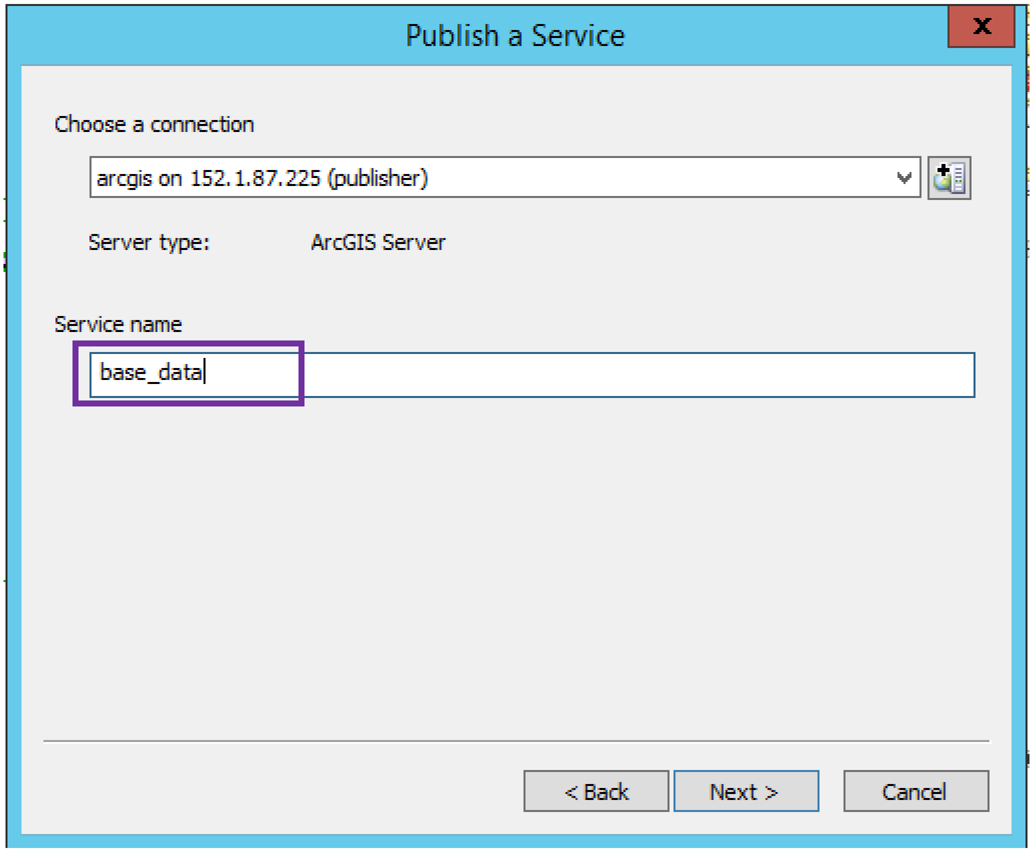
- Open a new ArcMap document > double click the admin2version2_gdb.sde database connection > Add the base data for the GEWA park to be published as a Map Service and symbolize as desired
- Click File > Share As > Service



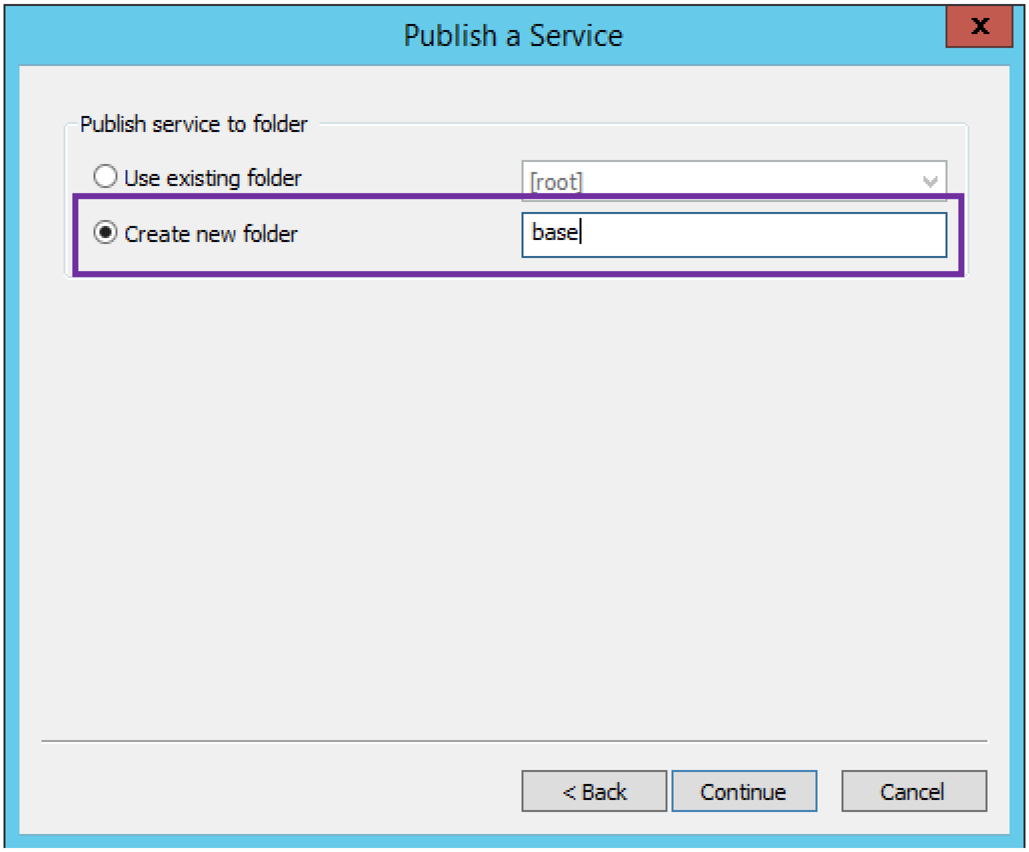
- Select Publish a Service > click Next



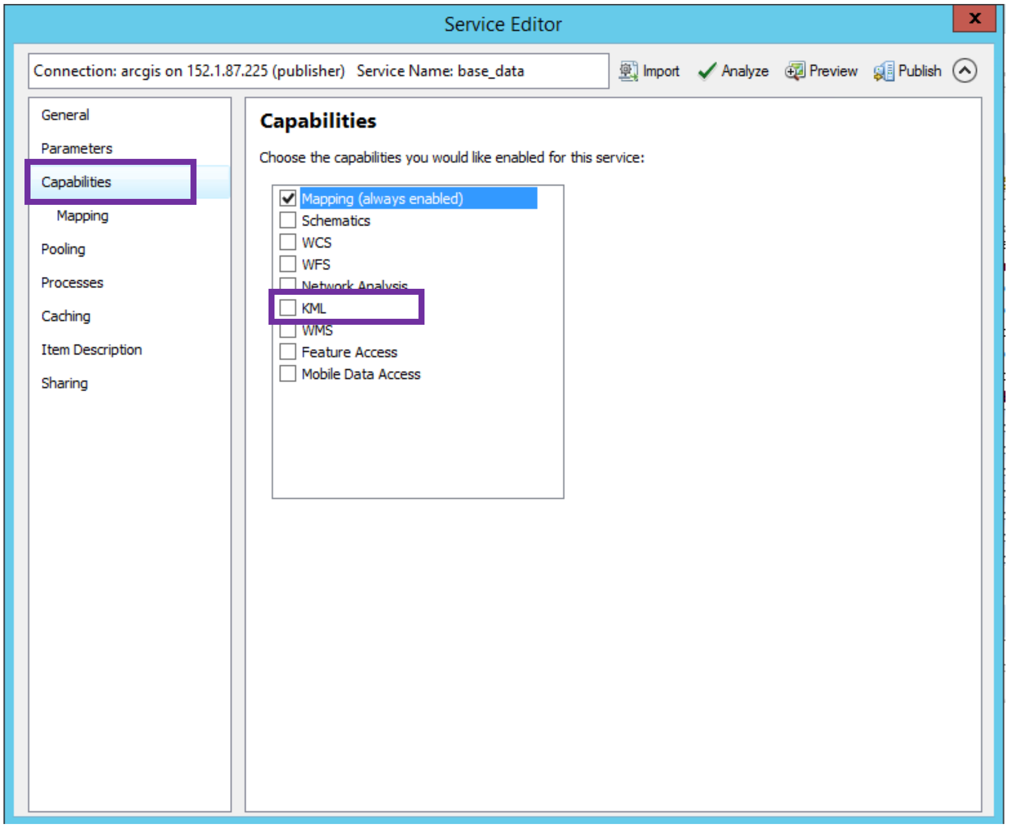
- Use the same Publisher connection > name the service “base_data” > click Next



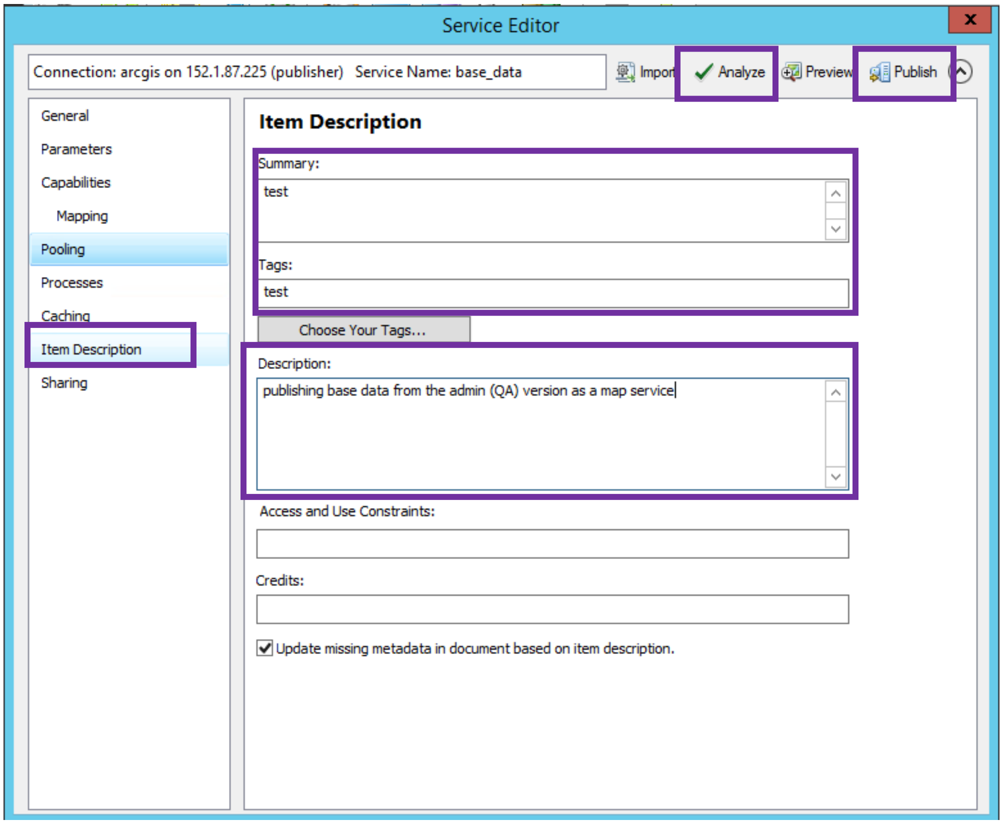
- Create a new folder for the Map Service to be placed and name it “base”



- In the Service Editor, click on Capabilities > uncheck KML



- Skip to Item Description and fill out Summary, Tags, and Description.
- Click Analyze > if there are no errors > click Publish

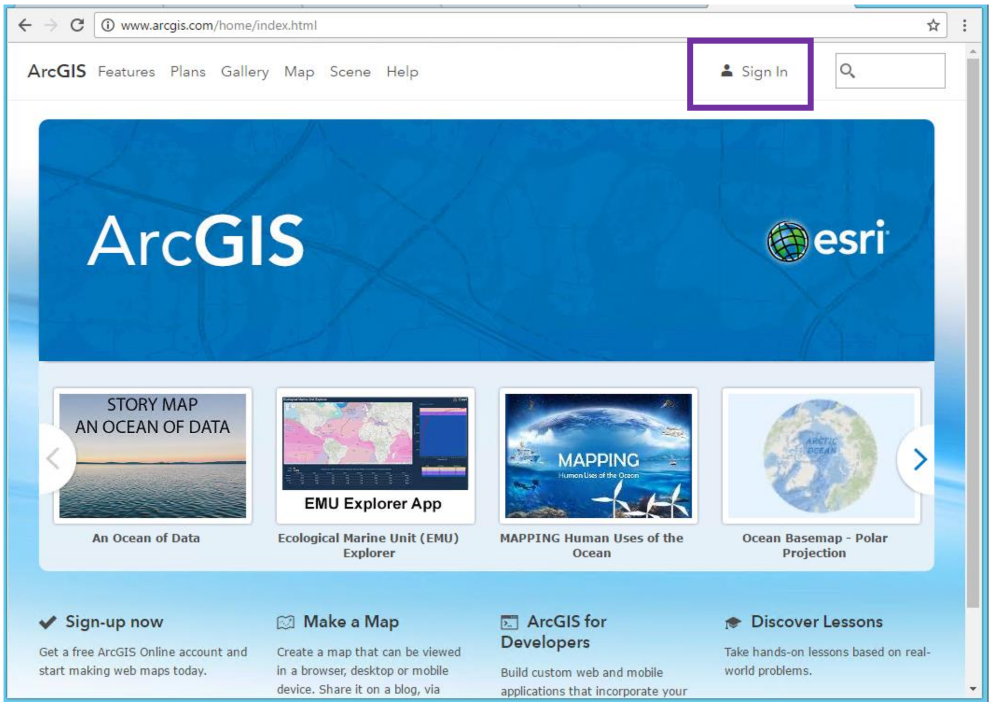


- Now the Map Service of base data shows in the REST services interface

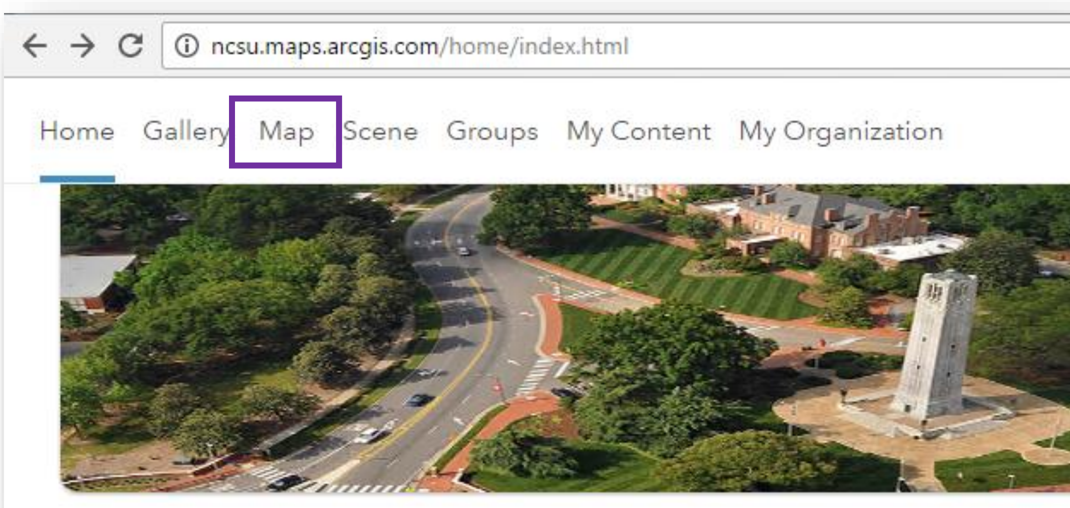
Section 3 – Making Web Maps from Versioned Feature Services

Now that you have Map Services and Feature Services, we will create two ArcGIS Online web maps, each accessing a different version.

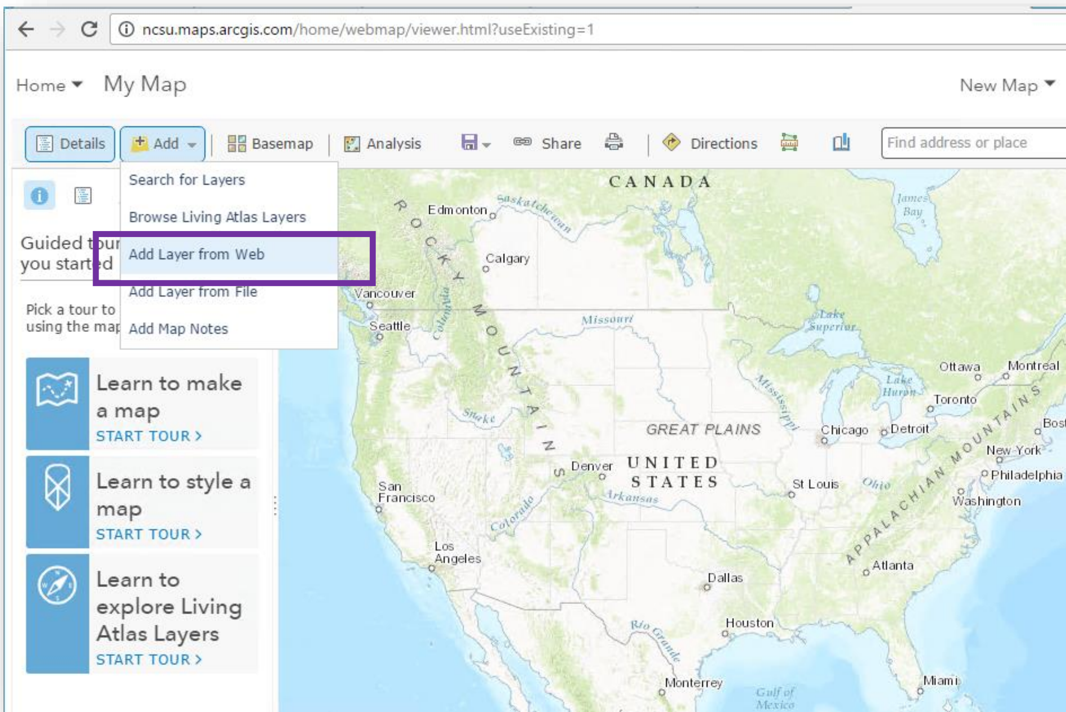
1. Navigate to http://www.arcgis.com and sign in using your ArcGIS Online for Organizations account



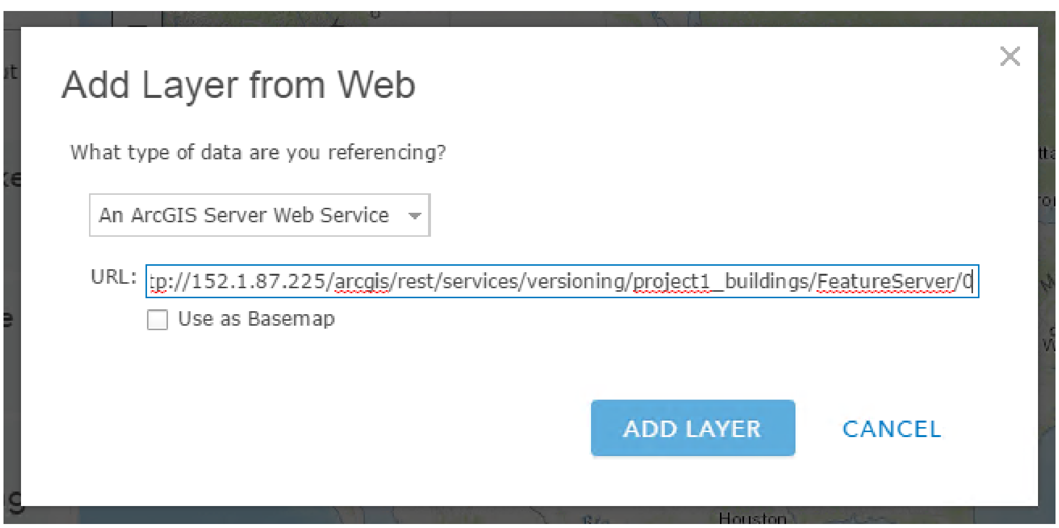
- Once logged in, click on Map to start a new map



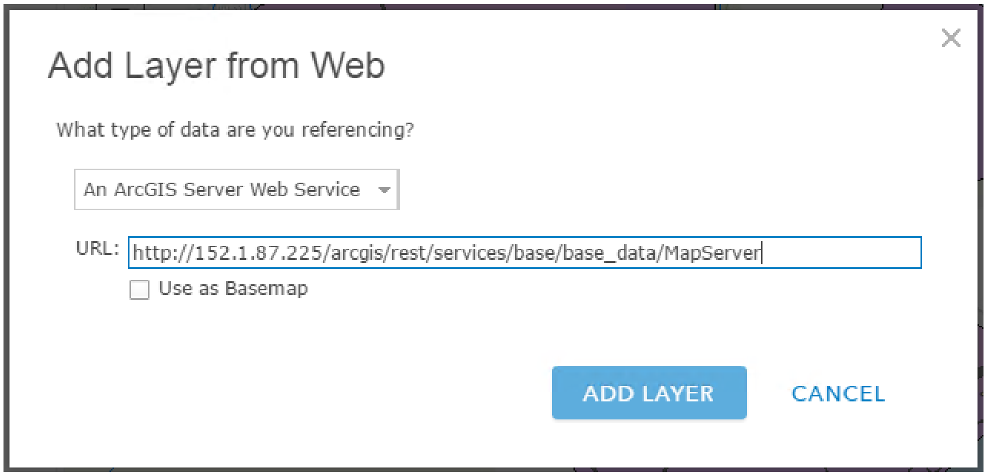
- Click Add > Add Layer from Web



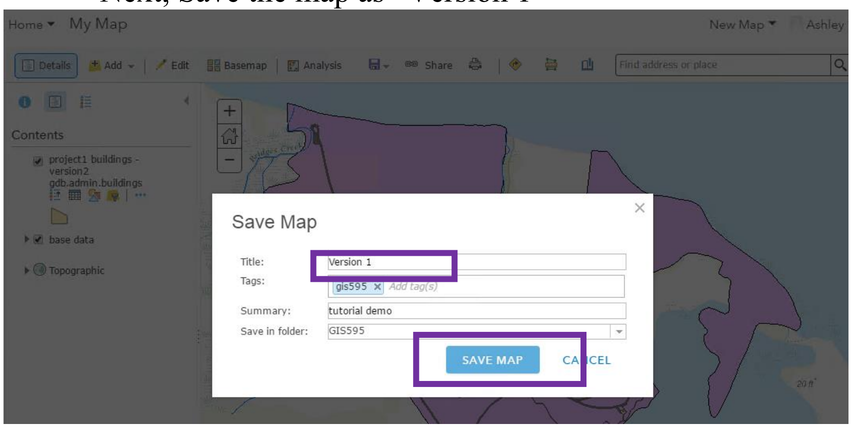
- In the URL box, enter the URL to the REST Service Endpoint of the Buildings layer coming from the Project1 version > click Add Layer



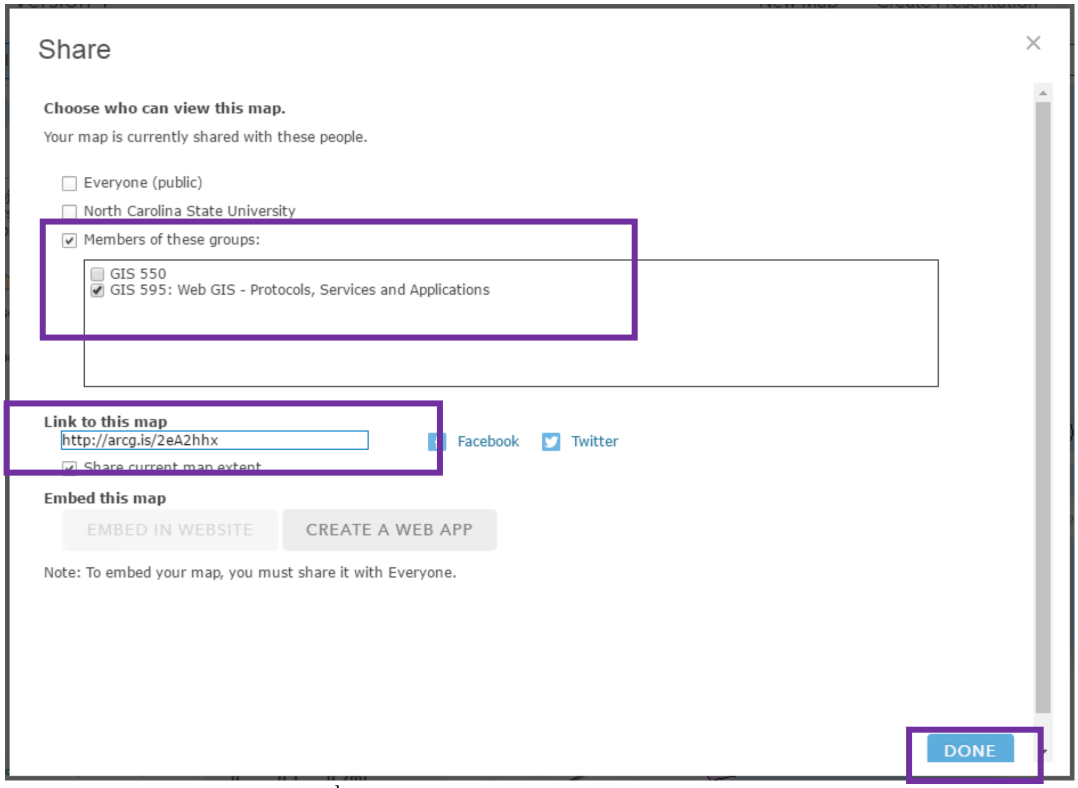
- Click Add > Add Layer from Web again, but this time add the base GEWA park Map Service data



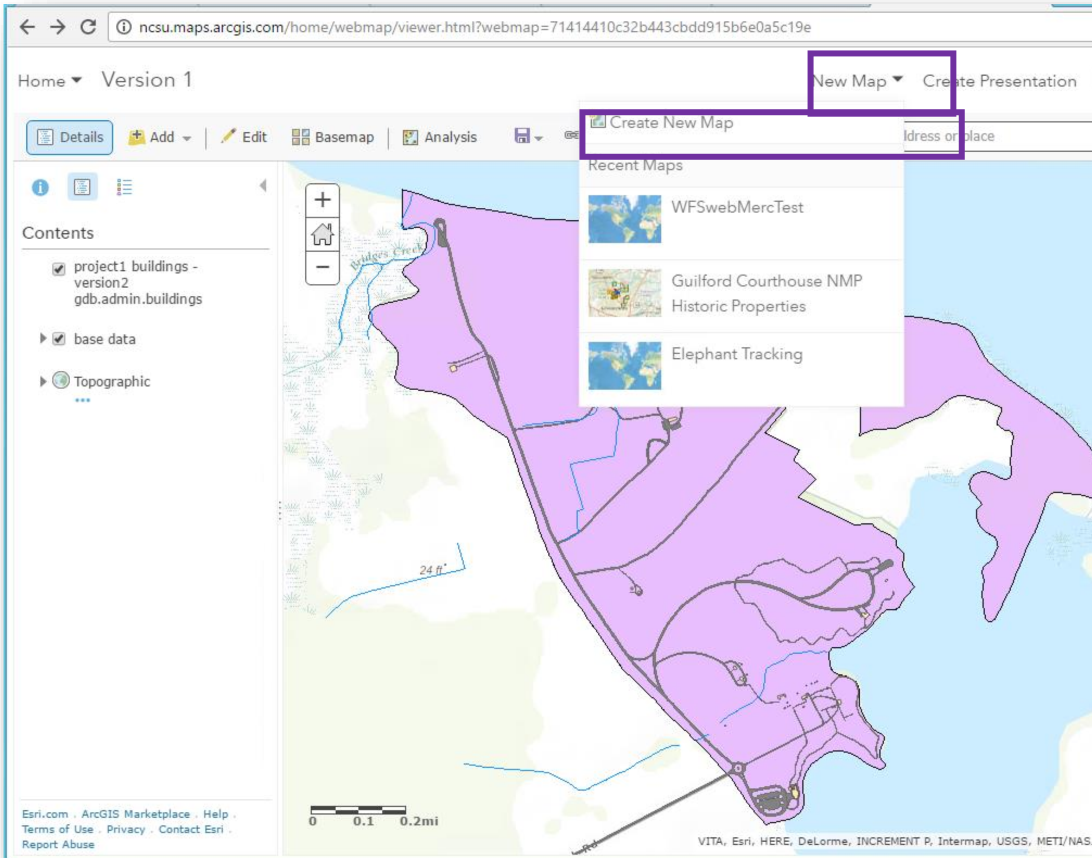
- Next, Save the map as "Version 1"



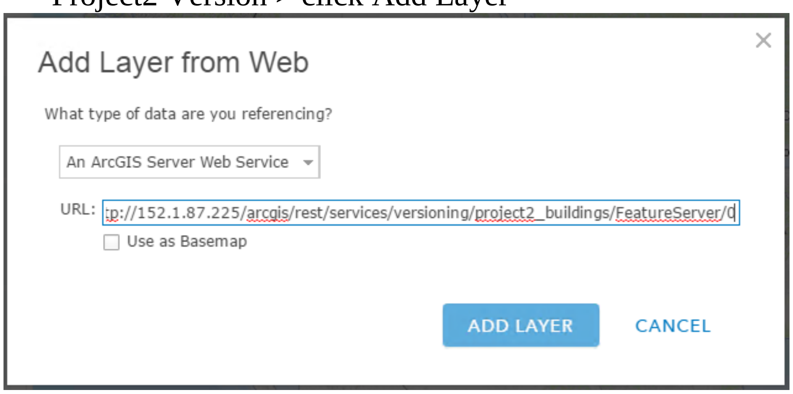
- Share the map to the GIS 595 group (the link to the map will allow anyone in the GIS 595 group access this map – hint you will need this for deliverables) > Click DONE. Now the first map is finished!



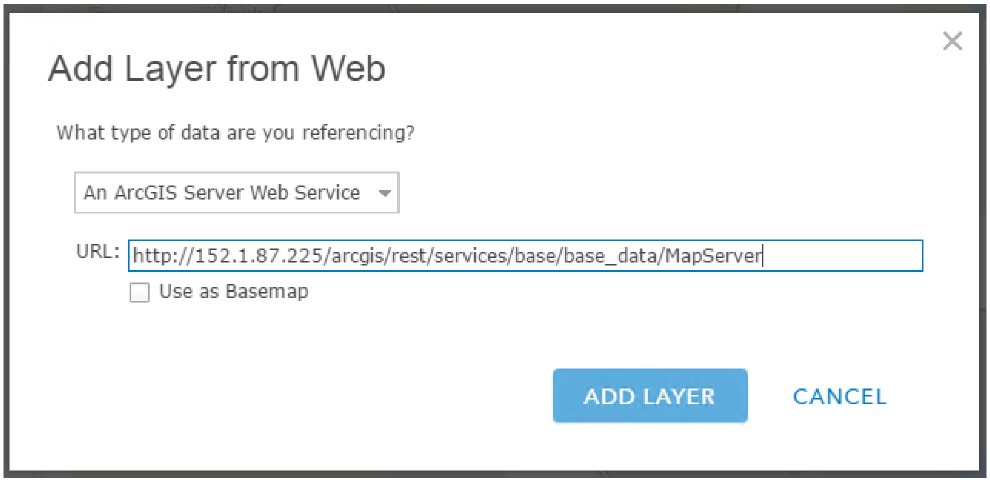
- Next create the 2nd map using the Project2 Buildings layer. In the top right corner, click New Map > Create New Map.



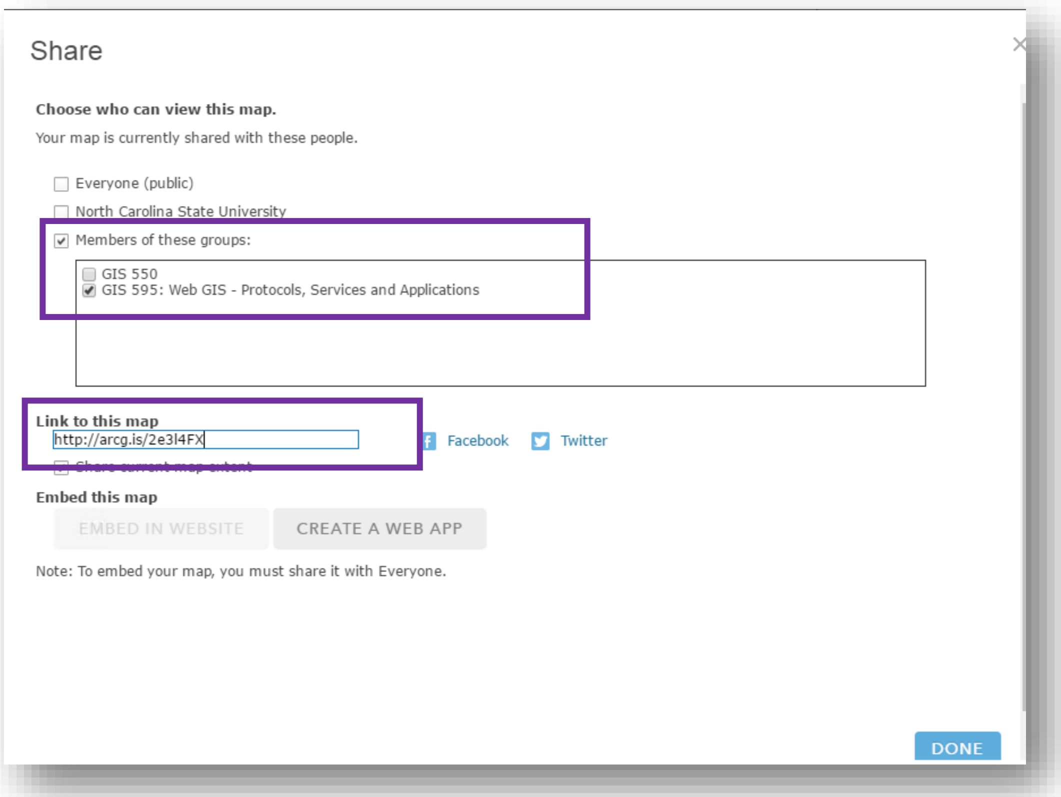
- Like before, click Add > Add Layer from Web
    - This time use the Feature Service REST URL to the buildings layer from the Project2 Version > click Add Layer



- Add the base GEWA park data as before. Add > Add Layer from Web



- Save the map as “Version 2”> Share the map the GIS 595 Group and make note of the link to share the map



CONGRATULATIONS! You have now created versions of an enterprise geodatabase, published services from the versions, and then viewed and saved them in web maps to be accessible through the internet.

Appendix B. Survey Questions

Background Experience

How many years have you been a GIS user?

What is your level of technical expertise?

- Light
- Moderate
- High

Have you ever been solely responsible for managing data as either databases or tables?

- Yes
- No

Impressions of Tutorial & Future Use

Using a five-point scale, please rate the following items:

	Strongly disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Strongly agree
1. I learned about versioning by going through this tutorial	O	O	O	O	O
2. The tutorial was easy to use	O	O	O	O	O
3. The tutorial provided enough instruction to effectively complete the task without any external help (e.g. external references, experts, message boards)	O	O	O	O	O
4. There was internal consistency in this tutorial	O	O	O	O	O
5. I felt very confident using the tutorial	O	O	O	O	O
6. Do you think a person with limited GIS experience could use this tutorial effectively?	O	O	O	O	O
7. I will likely use this tutorial in the future	O	O	O	O	O

How could the tutorial be improved?

Please provide any additional comments or suggestions on the process of completing the tutorial:

Powered by Qualtrics