Is Your Data Held Hostage? Migrate It to Autodesk Vault

Robert Stein – Autodesk
Data Management Business Consultant

PL1821   Is your project data management (PDM) system holding your data hostage? Is your PDM system making it difficult to do your job? Is your PDM system stopping you from integrating into other business systems? If you answered yes to any of these, you should attend this class and have your eyes opened to how easy it could be to migrate from your current system to Autodesk Vault software. We discuss the entire process from cradle to grave. Bring new life to your data with Autodesk Vault.

Learning Objectives

At the end of this class, you will be able to:

- Describe the process of migrating to Autodesk Vault

- Analyze your data

- List The Requirements for Migrating to Autodesk Vault

- Estimate the duration of your data migration.

About the Speaker

Based in Novi, Michigan, Robert works as a Data Management Business Consultant for Autodesk focusing on data management projects. Prior to working for Autodesk, Robert worked in varying roles over the years as a designer, CAD manager and consultant at a variety of companies. Robert has worked on a number of large scale implementations and custom programming projects with Autodesk Manufacturing products such as Inventor, AutoCAD Electrical, Vault Family, and AutoCAD Mechanical. Robert is in his tenth year as a presenter at Autodesk University.
Robert.stein@autodesk.com

Introduction

Over the years I have heard customers tell me numerous times that they feel like they are held hostage to their data management system. Although that may feel to ring true, this class will hopefully shed some light on the truth that you can move to a different data management system. It will take some work, but can absolutely be done. In this ninety minute lecture we will discuss the reasons some organizations might want to move, the methods used to do so, and the entire process. We will also discuss some of the real world projects I have worked on performing these migrations.

What is Autodesk Vault?

Autodesk Vault is a software solution that helps to organize, manage, and track data creation, simulation, and documentation processes for design, engineering, and construction teams. With Autodesk Vault you can get more control over design data with revision management capabilities and quickly find and reuse design data, for easier management of your design and engineering information.

With the above description you can see that Vault truly is a system that you can migrate your data to. I have worked with customers who have ITAR controlled data in systems such as SmarTeam and ProIntralink. I have worked on projects merging multiple Autodesk Vault servers into one server. I have worked on projects where we have loaded data from local computers and network drives.

The Vault is a central repository for all of your engineering data, marketing, office and much more. It is secure using vault logins or windows authentication to log into the system.

Migration Process

For all of my migration projects I have followed a simple process. What's that you ask? A simple process for migration? Yes, it's a simple process that from a high level can be applied to almost any migration.

Data Workshop

The first steps to the migration process is to understand what you are up against. You may remember from the requirements section I mentioned a lot of things to think about. The first thing I like to do for any migration project is to start off with a data cleansing workshop. During this data cleansing workshop the main goal is to gain an understanding of the data "as-is" state and the "to-be" state. There is a lot of information to cover in this type of workshop. Depending on the size of the project I have seen them be as simple as 1 day all the way up to a few weeks.

During this workshop you will want key players from all of the departments who create or consume the data that is being migrated. They don't need to be there the entire time however they may need to be called in and out of the workshop. During this workshop you will want to cover the following topics:

- Types of Data
- Quantity of Data
- Data relationships
- Workflows of current / future system
- Security model of current system
- Properties of Current / Future System
- Users / Groups
- Duplicate Files
- Broken References
- Categories (classification of files)
- Revision methodology
- Data Analysis
- Business Logic
- Required integrations such as SAP / SharePoint
- Data Integrity
- Metadata Mapping
- And so on

This data is used to build a plan of attack on how to approach the rest of the process. It is important to get a very good understanding of the topics above so that you don't run into any surprises during the migration.

In the class files there is an excel spreadsheet that will help you collect a lot of this information. In addition to the spreadsheet there are some other resources that you will find valuable.

- Autodesk Reference Repair Utility – Available on the wikihelp.
- Autoloader – Great tool for scanning files on disk to see if they have any missing relationships. Good for inventor and AutoCAD files.
- Robocopy – Build into windows 7, this is a great tool for copying large amounts of data. If the copy fails because of network issues, you can restart it and it will skip over any files it already copied. Very useful.

It will be a good idea to build a project plan so that everyone involved knows the actions they are responsible for, dates those actions are due, and impact if they aren't completed. You can use Microsoft Project for this or I have also used Visio to give a graphical view of this. I like using visio because you have a quick visual and you can put details in each day for who is responsible for what.

Monday	Tuesday	Wednesday	Thursday	Friday
	11/12/2013 Workshop	11/13/2013 Workshop	11/14/2013 Workshop	11/15/2013 RS – Travel Day
Client – Continue to work on Fixing Duplicate Files, References, and deciding which wins between duplicates across Vaults. Complete Configuration of Dev Vault (Transitions, property mapping)				Client – Run Exports
11/18/2013	11/19/2013	11/20/2013	11/21/2013	11/22/2013
RS-Work on Manipulation of XML Import Files				
	Test Import 1	Test Import 2	Thanksgiving Day 11/28/2013	11/29/2013
11/25/2013	11/26/2013	11/27/2013		
RS – Manipulate Second Export XML's and Prepare for Test Import to Dev Server				
Client Testing of Imported Data				
12/2/2013	12/3/2013	12/4/2013	12/5/2013	12/6/2013
Run Ref Repair Utility Run Export from Production Server	XML Manipulation		Import clean data into Production Vault	
12/9/2013	12/10/2013	12/11/2013	12/12/2013	12/13/2013
Import Validation	Go Live Training / Half Day Mornings			
12/16/2013	12/17/2013	12/18/2013	12/19/2013	12/20/2013
	Go Live Support / Half Day PM			
12/23/2013	12/24/2013	12/25/2013	12/26/2013	12/27/2013
12/30/2013	12/31/2013	1/1/2014	1/2/2014	1/3/2014

NON Working Days
Autodesk University Week

Extract

After the Data Workshop stage, then the ETL process starts. The ETL Process stands for Extract, Transform and Load. This is where the heavy lifting and time consuming tasks come in. But it is also where you have a large opportunity to fix metadata.

Extract

Transform

Load

During the extract stage of the process we use different tools to pull the data out of the current system. This can be done many different ways such as the system might have a method built in to export the data. Other systems do not. Here are a few ways that I have done this with projects I have worked on.

- Autodesk Proprietary Tools - Work with Autodesk Consulting to find out details.
    - o Vault Data Transfer Utility
    - o Vault ETL
- Network Drives – If the data already exists on a network drive then the data is already exported. You would want to create a duplicate copy so you can utilize that in the next stage of the process.
- Custom Developed Program – most of the projects I have worked on have had a custom developed tool for the export. For example one project we were exporting from SmarTeam which we partnered with a company who already had tools to export from that system. On another project we were exporting from ProIntralink which a different partner had some tools to export this data.

During the export process we typically get the data into what we consider as a "package" file. This package contains the physical files and the metadata / relationships in some type of a file. This file could be csv file, access database, xml file or really any format. This metadata file will be used during the rest of the process.

Transform

During the transform phase this is where we manipulate and massage the data to get it ready to import into Vault. Most clients take this opportunity to clean up the data. Some of the things that I typically do during this phase are:

- Update Properties – Utilize the mapping spreadsheet to find old property names in the metadata file and update them to the new property names.
- Usernames - Utilize the mapping spreadsheet to find old usernames in the metadata file and update them to the new usernames.
- Other updates to the metadata
- Fix Duplicate Files – We have to understand the business logic behind this.

This phase can be done using any tools you like. I use a combination of tools depending on the task. For simple XML editing I use:

- XMLSpy by Altova – This tool is expensive but has a grid view that really helps to visualize the data while you are editing.
- Notepad ++ - This is a great free tool. However on some of my larger projects this tool doesn't work as well.

If you have exported your data into some type of a database you can of course use database manipulation to make these "transform" changes. Some of these tools could be:

- SQL
- Oracle
- MS Access

One more option would be to write a program to manipulate the data for you based on a set of business logic rules. This would typically be done in something like Visual Basic or C#.

- Visual Studio .NET 2010
- Powershell

Load

During the load phase you will use one of many tool options to actually load the data into Vault. But before doing this, you need to think of a few things here. For example:

- Are all the properties defined in Vault
    - Properties mapped for file write back?
- Are all the users defined in Vault
    - Are they part of the correct group?
- Are all the groups defined in Vault
    - Do they contain the correct roles?
    - Do they contain access to the correct Vault?
- Are all the workflow definitions in Vault
    - Do they have the proper transition information
    - Do they have the proper criteria and actions
    - Is the proper state based security assigned?
- Are all the category definitions in Vault
- Are all the revision scheme definitions in vault
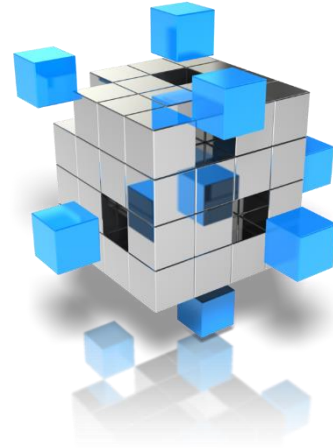
Basically does your Vault have all of the configuration to cleanly receive the data you have just spent so much time working on? If you don't have the placeholders for this data, whatever tools you are using to load the data will most likely fail. The question will be does it fail gracefully or not.

There are multiple tools available to load data into Autodesk Vault. Depending on the size of the migration, the complexity and type of data that will help you choose the tools. Most of them will require some type of a project with Autodesk Consulting to accomplish that. I have listed some of these tools below:

- Autoloader – This is a good entry level tool that is available to the public. This tool allows you scan your data, then import it directly to Vault. This tool is good for data that is sitting out on network drives, or local PC's and weren't really part of a PDM system. http://images.autodesk.com/adsk/files/best_practices_for_using_autoloader1.pdf
- Vault ETL – The Vault ETL tool has been developed by Autodesk Consulting and utilizes some of our other technology I will outline later. This tool is really good for taking data from a network drive with no revisioning, files stored on a server with revisioning (different versions in different folders) and files stored in another data management system.
- Vault Data Transfer Utility – This tool is what the Vault ETL is built on top of. This is the tool that does the loading of the data for the ETL tool as well as many of our other projects. This tool takes an XML file and the associated data and loads it into Vault.
- Add Project – For inventor files you can use Add Project directly from the software. This is good for projects that are sitting on network drives, or on local computers.
- Task Scheduler – This is a good way to migrate the physical files from one release to another release of Inventor. Here is a link to a whitepaper that discusses that. (Typically done after the data is loaded to Vault. http://images.autodesk.com/adsk/files/migrating_autodesk_inventor_data_in_autodesk_vault.pdf

Testing Data Integrity

Now that we understand the high level process of getting the data loaded we need to talk about testing the data. The testing phase of the ETL process can be just as important as any other phase in the project. This is often times an overlooked part of the project, or sometimes a part of the project that is given too much attention.

I have had projects on both ends of the spectrum of testing the data integrity. All I can do is give you some guidelines and you will have to make the decisions. Here are some of the things I have seen over various projects.

- Focus on a "Random" testing of the data. What I mean by this is don't get tied up on some key assemblies that you know already have issues. Chances are if it had "issues" in the old system, it will have the same issues in the new system.
- Choose a target and stick with it. If you are comfortable saying that I will check 10% of the files loaded, then stick to testing 10%.
- You also need to pick a success percentage. For example of the 10% of files in the Vault what percentage am I comfortable with having some type of failure. So maybe 1/2 % of the 10% is acceptable.
- Don't set unrealistic expectations. Having a data migration of 100% is a great goal but be flexible on it. There might be things that just can't come across.
    - For example if the source system supports something with the CAD system it was designed for, chances are even though Vault may have a CAD Addin for that application it may not support all the functionality.
- Develop a team of testers. Don't rely on just one person to test this data. Think about it, if you are loading 300,000 files. 10% of that is 30,000 files. For one person to test that is unrealistic in a reasonable amount of time.
- Document what you are going to test.
    - In most of our scenarios you have a list of what data is being exported then imported. You got this during the export phase. You can import that list into Excel and use some simple excel macros / formulas to randomly select whatever percentage of files. Then this can also become your to do list and documentation of success / failure on each file.
    - Build in a few extra columns for notes, status etc…
- If you hit your failure rate, stop testing.
    - If you stated that 2% failure is unacceptable, and you reach that at 3% why would you continue with the other 7%. One caveat to that is if you don't know exactly what the problem is. You may need to test some additional files to figure out what the cause is, so you can go back and fix it in the system.

Use these as guidelines not hard fast rules and you will do just fine during your data testing.

Lather, Rinse and Repeat
This process typically happens a number of times throughout the duration of the project. For example you will miss things during your transform phase, maybe your testing phase uncovered some broken links that you want to go back and fix before importing into production. This is why you should always build some testing cycles into the project plan. You should also have a test environment so that you can perform these tests without getting in the way of production.

Requirements for Migrating to Autodesk Vault

Now that we have discussed the process of migrating and analyzing your data. Let's talk about why you want to move to Vault and what it takes.

When we talk about requirements for migrating to Autodesk Vault we really need to consider a lot of things. In this section we will discuss what those topics are.

Why are you looking to migrate?
This is a great question. Why are you looking to change to a new PDM system? There are many reasons and I am sure we can all list them. But do those reasons have value to the organization? Do they fall in line with the plans of the organization for future growth? By changing will you become more streamlined? Will you be able to integrate into other systems? What are the things that are going to hit home with the executive level managers who are going to have to sign off on the costs for this. Let's think about these things for a few moments. Here are some examples I have heard from customers in the past.

- Why are you looking to change to a new PDM system?
    - Our current system is out dated.
    - It takes a team of people to keep it alive and functioning.
    - The CAD integrations with the current system hold us back.
    - We can't move to the latest version of Inventor because our current system is always one to two releases behind.
    - By changing to a new PDM system that solves these problems we can utilize the latest functionality in our CAD systems because we can upgrade.
    - By moving to a new PDM system we can integrate directly with SAP, Sharepoint or whatever other system we want.

What is your current system?

When answering this question there are a few things I typically ask as well. Once I know what system the client is looking to move off of I ask some follow up questions. For example let's say the client said we are currently on BrandXPDM. I may follow up with the following questions:

- What are the things you like most about your current system? (It's a global solution that we can replicate to our offices in China and Australia)
- What gaps do you see in your current system? (CAD Integrations are very bad)
- What are the end user pain points? (i.e. takes two minutes to log in, slow and difficult UI etc…)
- Do you have any security restrictions such as ITAR? Does the current system have certification for that restriction? (Yes we have ITAR data but our system isn't certified.)

By following up with these questions you gain a little understanding about what the Vault implementation might look like. For example in this scenario we know that we need to do a replicated vault, we know that the CAD integrations that Autodesk has to offer will really be a benefit. We know that the end users will benefit from the performance and UI. We also know that we will have some security to set up. None of these are show stoppers for moving to vault but all good things to know up front.

I have worked on migrations from a few different systems. Here are some of the systems I have migrated to Vault:

- SmarTeam
- ProIntralink
- Windows Explorer – This is honestly a pretty common scenario.
- Vault – I have done several different projects but mainly this is done when you have multiple vaults and you want to merge into one vault. For example you might have acquired a company, two different business units now want to be on the same vault.

Whatever the current system is, there is data in it and you are looking to get it out into Vault.

What types of Data is in your current system?

No matter what system you are working with the data coming into Vault will be of importance. I have developed a spreadsheet that helps you to collect this information and it will help for future use. When thinking about this you will want to consider the following things.

- Data Types – i.e. Inventor, AutoCAD, Catia, Microstation, Solidworks, Word, Excel etc…
- Relationships – Which of these files have relationships with other files. – i.e. Inventor files have relationships, AutoCAD files can have relationships etc..
- Quantities – This can help you estimate a few things. Such as how long the "import" will take.
- What are the authoring tools for your current data?

- o This will give you a good understanding of the legacy data you will be maintaining in the new system.
- What are the authoring tools for your future data?
  - o This will give you a good understanding of the new data you will be creating. A lot of times clients will move to a new PDM system when they change CAD systems.

Data Mapping

When looking to move all of your data to a new system there is a lot to think about when it comes to data mapping. Not all PDM systems are exactly the same. However there will be a lot of similarities. We need to capture both, the differences and the similarities. What do I mean by that? See the example table below.

Property Mapping Example:

Source System Property Name	Vault Property Name
Title	Description
DesignBy	Designer
Designed By	

This is a simple example of a mapping for properties. This is important because during a migration is a great time to consolidate multiple properties into one combined property. In the above example there were two properties "DesignBy" and "Designed By" which are really the same thing. So in the new Vault we can merge them together into Designer field.

This was just one example of the types of mappings we need to think about. Below is a list of all of the things I typically consider.

- Properties – Think about merging like property names into one property name. Also think about property type. For example if the old property type is a string, you probably won't be able to migrate it into a date field.
- Lifecycle States – Most systems have a concept of lifecycles. This is a great opportunity to revise your workflow. I have some customers who had 8 lifecycle states in the old system and in the new system they are at 4.
- Lifecycle Definitions – Even if your old system doesn't have this, you will need this in the new system. In Vault you can have multiple different workflows. So you will need to tell the system which workflow state goes with which workflow.
- Categories – Categories are like a classification of files. If you are coming from some system besides vault you probably don't have this but you will need to define it for the new system.
- Revision Scheme – This is a great opportunity to clean up revisions. This could mean mapping an old revision of 3 to C. This doesn't happen often but I have seen it once or twice.

- Missing Metadata – For example let's say your engineering department hasn't done such a great job filling out properties but all the information is in the ERP system. In theory you can do an export from the ERP system and use that during the migration to populate those properties into the PDM system. It takes some work but it can be done.
- Usernames – On a recent project of merging two vaults together I have had a customer ask me to map usernames. For example what we did was in Vault A we had a user naming convention that was like RobStein. In vault B we had SteinR. In the new merged Vault we might want both of those names mapped to ADSSTEINROB. All three of the users are technically the same. If we modify the fields such as created by to have all the same ADS username it helps searching later.

Test Environment

With most implementations clients will have a test environment and a production environment. This type of project isn't any different. You will need to have a test environment. In fact you will want to test the process multiple times. As you go through this process you will be taking snapshots of data and users will continue to modify data. As you fix things for the test mergers your users will be modifying other files and possibly introducing new issues into the current system.

The test environment should be the same as the production environment will be.

- Hardware – Make sure the hardware is comparable to the production environment.
- Operating System – This should be the same version, service packs and hotfixes.
- SQL – This should be the same version, service pack and cumulative updates.
- Vault ADMS – This should be the same version, service packs and hotfixes.
- Disk Space – Make sure you have enough disk space for multiple iterations if testing.

Estimating a Migration

When estimating your migration you will need to follow the basic format below. This isn't an exact science since there can be so many unknowns such as server speed, number of files, number of revisions, number of versions and so on.

Task	Estimation Rules
Data Migration Workshop	I have found for most projects a good average to estimate for this task is 3 days. You can always modify it as you get into it during the first day.
Planning	This will really depend on what you find out during the Data Migration Workshop. As a rule I would estimate 1 day minimum for planning.
Extract	This is based on a few factors. • Number of Files • "Power" of servers performing export • Software being used for export • Complexity of file relationships • Physical file size. As a rule of thumb I use about 450 files per minute when I am using our Autodesk tools for the export. With other tools I would recommend running a few test exports and timing them for an average.
Transform	This is probably one of the more unpredictable stages. It really depends on a lot of things: • Number of Files • What you will be transforming ○ Property merging ○ Workflow manipulation ○ Revision synchronization ○ Relationship building ○ Business logic • Number of different sources
Load	This is based on a few factors. • Number of Files • "Power" of servers performing export • Software being used for import • Complexity of file relationships • Physical file size. As a rule of thumb I use about 450 files per minute when I am using our Autodesk tools such as Data Transfer Utility for the import. With other tools I would recommend running a few test exports and timing them to get an average.
Testing Data Integrity	This depends on the number of people testing and the number of files. I would estimate 5 minutes per file on average for testing. This should give you time to open the file, verify links, close the file and then clean your workspace.

Conclusion

If you feel your data is held hostage in one of those other PDM systems don't worry, we can get that data into Vault. Throughout this class you have seen multiple examples of how to go about the migration process.

I will be available in the Answer Bar from 2:00pm to 6:30pm on Wednesday. Please feel free to come visit and ask more questions there.

I hope this class has helped. If you would like to find out more information please feel free to email me at Robert.stein@autodesk.com