

Digital Intermediate  
Deployment Overview  
Company Confidential

Prepared by:



## Table of Contents

Executive Summary .....	3
Business Challenges.....	3
1. Projects Consuming The Provider’s Storage Resources For Extended Periods .....	3
Video Deliverables.....	3
Extended Quality Control Periods.....	3
2. Data Security .....	3
3. Inefficiency In Data Production Line Reduces The Provider’s Overall Production Capacity .....	4
4. Operational Efficiency Challenges .....	4
5. Separation of “Church and State” .....	4
The Provider’s Data Management Goals / Benefits .....	5
1. Generate Incremental Revenue.....	5
2. Worldwide Standardization Of Workflow For Projects .....	5
3. Improved Customer Communication .....	5
4. Reduce Storage Consumption.....	5
5. Reduce Operator Error & Required Skill Set & Training .....	5
6. Reduce Operational Costs Via Automation .....	5
Sample Digital Intermediate Workflow .....	6
DataFrameworks Initial Deployment Outline .....	7
1. Phase 1.....	7
a) Simplify The Process Of Storing, Finding, and Reporting On Data.....	7
b) Standardize Data Organization & Naming.....	12
c) Production Facing Reports And Views .....	13
d) Introduce Degree Of Automation .....	14
e) Simplify Common Data Management Tasks .....	15
f) Synergies With The Provider’s Existing Autodesk Infrastructure.....	15
DataFrameworks Subsequent Deployment Possibilities.....	16
Phase 2 .....	16
a) Tighten Security And File System Forensics .....	16
b) Possible Integration With Other Infrastructure Components .....	16
Phase 3 .....	17
a) Possible Integration Points With Autodesk Infrastructure.....	17

## Table of Figures

Figure 1. High Level Overview Of A DI Pipeline .....	6
Figure 2. Overview Of Location Dependent User Access .....	8
Figure 3. Overview Of Location Independent User Access.....	9
Figure 4. Overview Of Data Management Challenges With Autodesk “Soft Import”.....	10
Figure 5. Overview Autodesk “Soft Import” With Location Independence.....	11
Figure 6. Sample View Of Project Data .....	13
Figure 7. Sample View Of Project Data Storage Costs .....	14

## **Executive Summary**

Efficient data management can help a Digital Intermediate Service Provider (The Provider) address their customer concerns and overall business challenges as images services transition from analog to digital workflows. The purpose of this document is to outline an understanding of the existing business challenges that DataFrameworks can help to resolve and to establish a plan for the pilot deployment of DataFrameworks in terms of:

- Stages Of DataFrameworks Deployment
- Expected DataFrameworks Deliverables At Each Stage
- Milestones In Terms Of Business Objectives Achieved
- Reversion Plans

## **Business Challenges**

### **1. Projects Consuming The Provider's Storage Resources For Extended Periods**

Currently there is no financial incentive for customers to have their project data removed from The Provider's online storage resources. Two factors often contribute to a reduction in The Provider's production capacity due to excess / extending consumption of storage resources:

#### **Video Deliverables**

One might present the argument that the only incentive for the customer is to actually have their project data stay online for as long as possible. For example, a customer might incur charges to restore their project data to online storage, if they subsequently needed another video deliverable format. However, if the customer's project data still resides on The Provider's online storage resources, a charge to restore their project data would be avoided. Therefore, there is basically an incentive for customer is to keep their project data on The Provider's storage resources for as long as conceivably possible.

#### **Extended Quality Control Periods**

The Provider's customers ultimately perform the quality control process on images services provided by The Provider. The customer's project data will remain on The Provider's online storage resources until the customer has completed their quality control process and "accepted" the work. Obviously, delays or lack of a "sense of urgency" in the customer's quality control process will ultimately result in a reduction in The Provider's overall production capacity, or an increase in storage related capital expenditures to meet demand.

### **2. Data Security**

Many of the major motion picture studios have focused on the value of their content and are mandating that both vendors and internal divisions develop and implement formal security policies.

Some of the following factors contribute to the increased content security concerns:

- Sharp increase in piracy for theatrical release content and new production for DVD release
- Existing library material destined for DVD release
- Evolving digital workflows put elements and finished content at risk
- Physical asset security procedures cannot protect electronic based content

In addition, changing technology has increased content security risks:

- File based networks
- Assets and images are easily transported across a network or across the world
- Traceability of file / data actions is difficult
- Electronic file based operations can be performed faster than real time
- Assets are generated easily and often
- Ability to move content quickly across networks without formality associated with physical media
- Ability for others to penetrate electronic workspaces from the outside
- Transportation of content across both secure and unsecured networks

**3. Inefficiency In Data Production Line Reduces The Provider’s Overall Production Capacity**

The raw quantity of data can quickly consume The Provider’s overall production line capacity. Efficient management of the data flow is key to maximizing overall job throughput on limited resources. The level of management complexity is increased with multiple services / workflows offered by The Provider to customers. All The Provider’s jobs ultimately share the same fixed storage resources, these workflows might include:

Feature Film Digital Intermediate  
 Restoration Of Feature Films  
 Television Digital Intermediate  
 Visual Effects  
 Feature Marketing  
 Data Coordination Services Of Clients VFX Project Data

The competitive environment requires The Provider to increase production line flexibility and adaptability to meet customer requirements.

**4. Operational Efficiency Challenges**

Increased volume coupled with shorter timeframes of projects increases the operational complexities associated with data management. Careful coordination of the data and shared resources is required among multiple jobs and the creative staff to ensure maximum efficiency in the data production line. Automation of routine tasks can: a) reduce The Provider’s operational costs, or b) increase overall production capacity without a corresponding increase in operational costs.

**5. Separation of “Church and State”**

Studios are increasingly demanding their service providers physically separate “production networks” from the general “Information Technology network,” due to piracy concerns. The requirement introduces additional constraints for The Provider’s personnel not directly involved in the creative aspects of the project. For example, Customer Service Representatives need the ability to effectively communicate project status to customers, but don’t necessarily need physical access to actual image data. A mechanism is needed to provide The Provider’s personnel with the information required to perform job function, without necessarily increasing piracy risks or exposure points.

## The Provider's Data Management Goals / Benefits

### 1. **Generate Incremental Revenue**

The revenue / service charge for the active Digital Intermediate component (Scan, Grade, Conform) can be estimated and accounted for with a reasonable degree of accuracy. However, two components of the process are difficult to estimate and generate revenue:

- Video Deliverables
- Extended Customer Quality Control Process

In the analog world, customers were charged for every tape consumed / generated. Current DI proposals typically have the storage costs well understood and accounted for in the up front processes (scan, conform, grade). Difficulties arise in accounting for time delays in video deliverables, and the customer's quality control process both of which can require project data to be resident on storage resources for extended periods of time. DataFrameworks can introduce automated data management standards and processes which can, account for and report on all the data associated with any given project.

### 2. **Worldwide Standardization Of Workflow For Projects**

This approach will help The Provider implement a repeatable workflow that can be 'franchised' to locations around the world, creating a uniform quality experience world-wide for their customers. The implementation of DataFrameworks into Autodesk applications will allow for a uniform view and organizational structure of storage across all applications, and clearly defined end-to-end workflow. Consistent, predictable, and repeatable data organization produces a more approachable environment, improves the ability to locate data, and reduces storage waste.

### 3. **Improved Customer Communication**

Production facing reports and views will enable Producers and Client Representatives the ability to quickly see all the elements in one view, even though the view on the underlying file systems may be disjointed and distributed. Such reports and views provide the ability to inform clients and vendors what data has been received and is currently online.

### 4. **Reduce Storage Consumption**

DataFrameworks can coordinate more efficient use of high performance, high cost, resources. It can motivate users to complete and offline project data, freeing up valuable resources for more revenue generating projects. A mechanism is required to provide visibility, cost, and trends of storage resource consumption. DataFrameworks provides this functionality out of the box.

### 5. **Reduce Operator Error & Required Skill Set & Training**

There is often a need for Operators & Data Managers to have Unix System Administrator skill sets. DataFrameworks can provide easy to use, secure and automated processes to help reduce error through DataFrameworks managed data tasks

### 6. **Reduce Operational Costs Via Automation**

Automation reduces operational errors and their associated costs. Standardization combined with automation sets the foundation for geographically distributed workflows.

Note: The DataFrameworks Initial Deployment Outline section will articulate when and how the above goals are addressed.



## DataFrameworks Initial Deployment Outline

The following provides a high level overview of the phases and goals of an initial DataFrameworks pilot deployment.

### 1. Phase 1

#### a) Simplify The Process Of Storing, Finding, and Reporting On Data

One of DataFrameworks strengths is in allowing users to know what sorts of assets are consuming storage within a meaningful business-oriented hierarchy at the present time and historically over time. The software provides many ways to keep assets organized and has a large amount of useful reporting. A common namespace traversable from a single point allows users to need only think of data in terms of a meaningful business hierarchy, instead of focusing on the underlying storage bins.

#### **Deliverables**

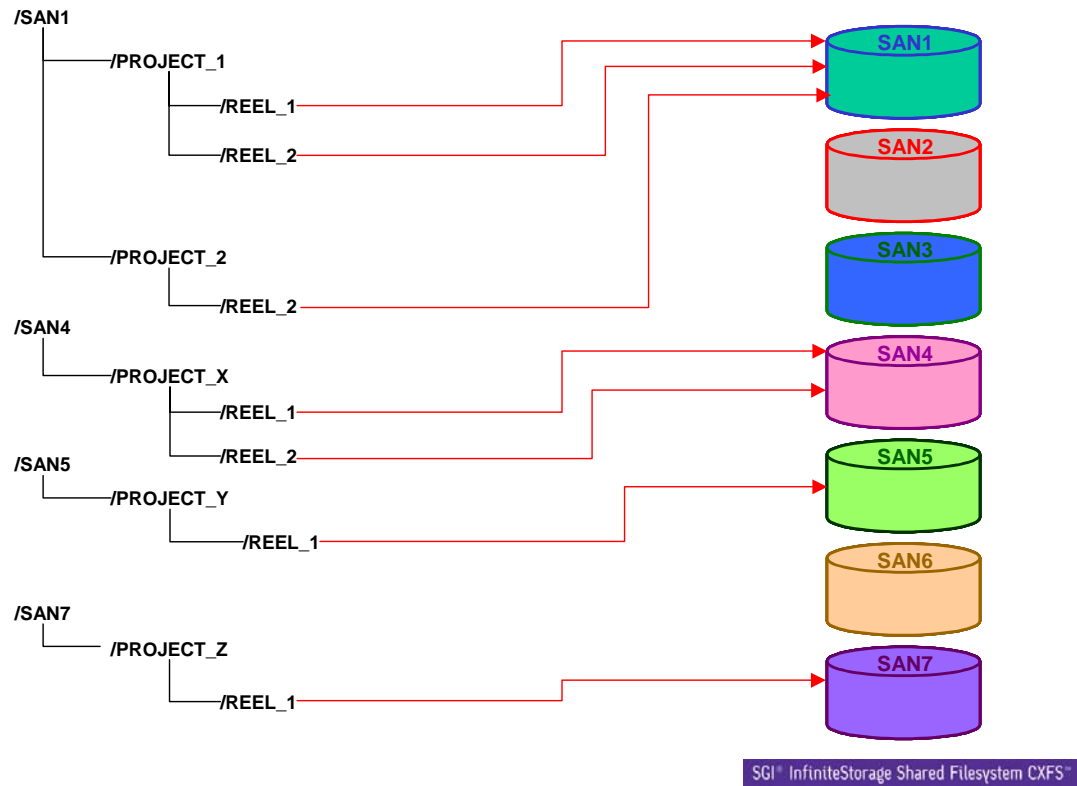
- The DataFrameworks team will assist in the initial deployment of the software. The software will try to preserve any established operational pipeline and folder hierarchies in place at The Provider, whenever possible. This will enable The Provider to continue to operate with their existing operational practices while testing the possible benefits and risks of an abstracted namespace. DataFrameworks will train the The Provider's Scanning department on creating a project and initiating a scan via the DataFrameworks GUI, ensuring both file systems structures (logical and target SAN) are created from the onset of a project.
- Reversion Plan: Parallel files system access structures will be established to ensure a reversion plan for The Provider, at this point the DataFrameworks logical data path will be maintained in parallel to the target SAN file system path. DataFrameworks will initially target each project to a specific SAN to map directly to The Provider's existing data management processes. The initial targeting of a project to a specific file system SAN will enable The Provider's existing data management scripts and processes to continue in tact.
- The DataFrameworks team will attempt to automate the insertion of scans into the DataFrameworks system. This will depend of the logistics of the process in place. DataFrameworks will investigate the possibility of further simplification, by providing the Scanning department with a basic script, which might prompt the operator to input any specific information required by The Provider (work order number, etc). (It is not uncommon for The Providers scanning department to manually create file system directories via existing infrastructure components, for example, the Thomson Phantom TransferEngine Software.)

#### **Abstract Data From Physical Infrastructure Complexity**

Efficient data management ultimately increases production capacity, while inefficient data management results in a loss of production capacity. Most Providers dedicate resources on a physical basis, i.e. Project X is assigned to /SAN1. This requires a show to reside entirely on a single SAN with everything encapsulated into one directory.

## User Access

Requires user knowledge of infrastructure, location, and naming conventions of all data for a given project.



Manual management of the quantity, location, and performance requirements of data

Figure 2. Overview Of Location Dependent User Access

A workflow strategy based on physically dedicating resources ultimately will:

- Reduce overall utilization rates
- Result in unnecessary data movement
- Limit production line flexibility / adaptability

The dedication of a project to single physical SAN, combined with the complexity of The Provider's hardware infrastructure, often results in some of the following operational procedures:

- Email generation as result of data movement operations with the new directory location. The new directory path location is then copied, and pasted into the next work order.
- Work Orders and email broadcasts contain the actual path to data to avoid / reduce operational errors.
- Constant Work Order generation by Producers to avoid / reduce operational errors.



Manages the quantity, location, and performance requirements of data.

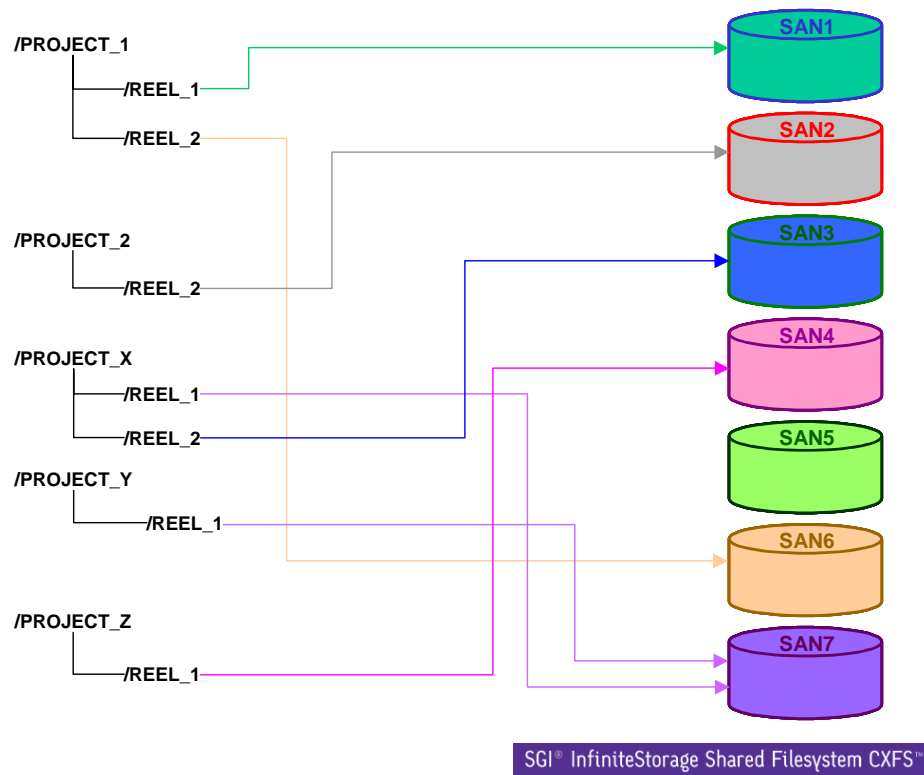


Figure 3. Overview Of Location Independent User Access

#### Deliverable

- Reversion plan: DataFrameworks provides parallel file system paths: a) logical path linked to actual target physical path, and b) actual target physical path. The existence of the actual target physical path will enable The Provider's existing data management tasks and scripts to operate in parallel
- DataFrameworks attempts to make use of and incorporate existing scripts developed by The Provider's data management staff, whenever and wherever practical.

Challenges Addressed:

Reduced Infrastructure Complexity For Creative Staff

Challenges Addressed:

Autodesk "Soft Import" And Dependency Of Project To Specific SAN Location

Logical data locations will increase flexibility to: a) adapt to changing production and scheduling changes, b) move data to ensure / meet real time performance requirements, and c) remove the dependency of a project to a specific physical storage location. It is not uncommon for The Provider's infrastructure, to have projects are tethered to their original SAN location due to physical path locations. The following operational complexities and limitations are present when projects reference a specific SAN:

- If project data needs to be physically moved due to production and scheduling constraints, the Autodesk soft import links must be updated to reflect the new physical locations. Updating the “soft import” reference can be rather complex, as it is not uncommon for the links to reference 1000 clips, where the links are not all in the same directory but spread out through directories.
- Projects must be restored back to their original physical location if they need to be brought back online.

The following outlines the current operational complexities should changing schedule and priority changes require The Provider’s data management to move data that has been “soft imported” into Autodesk applications.

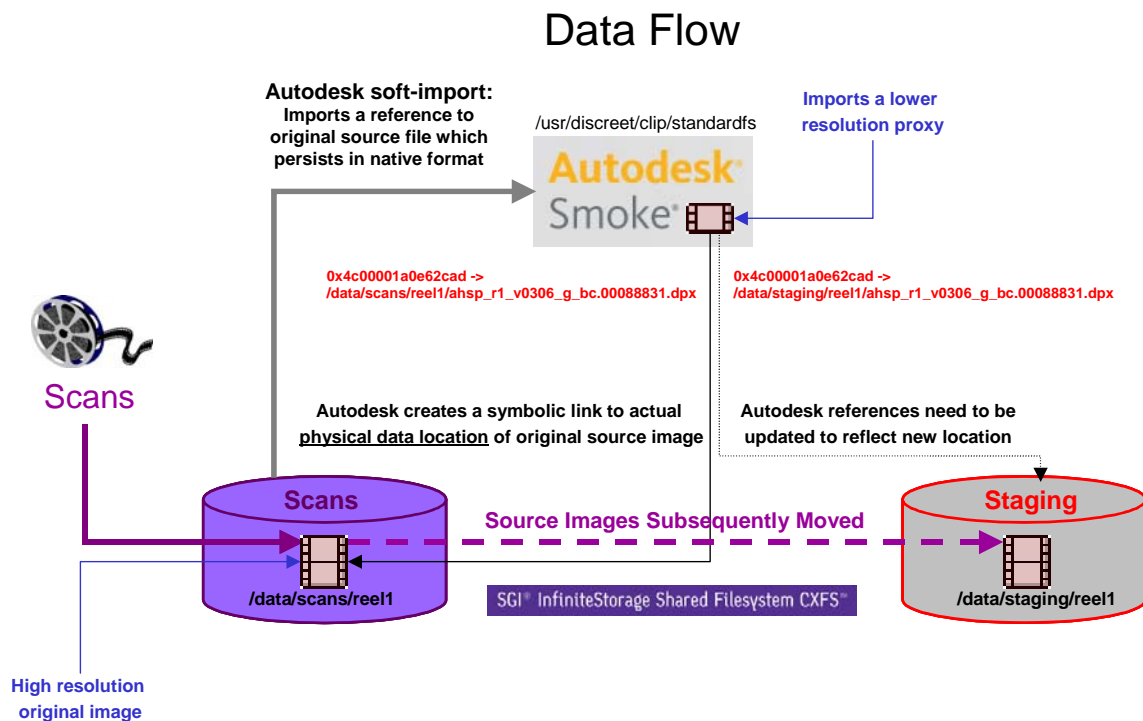


Figure 4. Overview Of Data Management Challenges With Autodesk “Soft Import”

Often the increased operational complexities make it impractical to actually move data from original SAN location, therefore reducing The Provider’s flexibility to adapt to changing schedules and priorities.

To actually move data that has been “soft imported” into the Autodesk Smoke application, The Provider’s data managers need to perform the following tasks:

- Actual movement of data from original source SAN to new destination SAN
- Manual update of Autodesk “soft import” references to reflect new location for each file. The following provides an example:
  - Original Autodesk Smoke Reference  
0x4c00001a0e62cad -> /data/scans/reel1/ahsp\_r1\_v0306\_g\_bc.00088831.dpx
  - Updated Autodesk Smoke Reference  
0x4c00001a0e62cad -> /data/staging/reel1/ahsp\_r1\_v0306\_g\_bc.00088831.dpx

Note: An update is required for each individual file which could require thousands of individual updates.

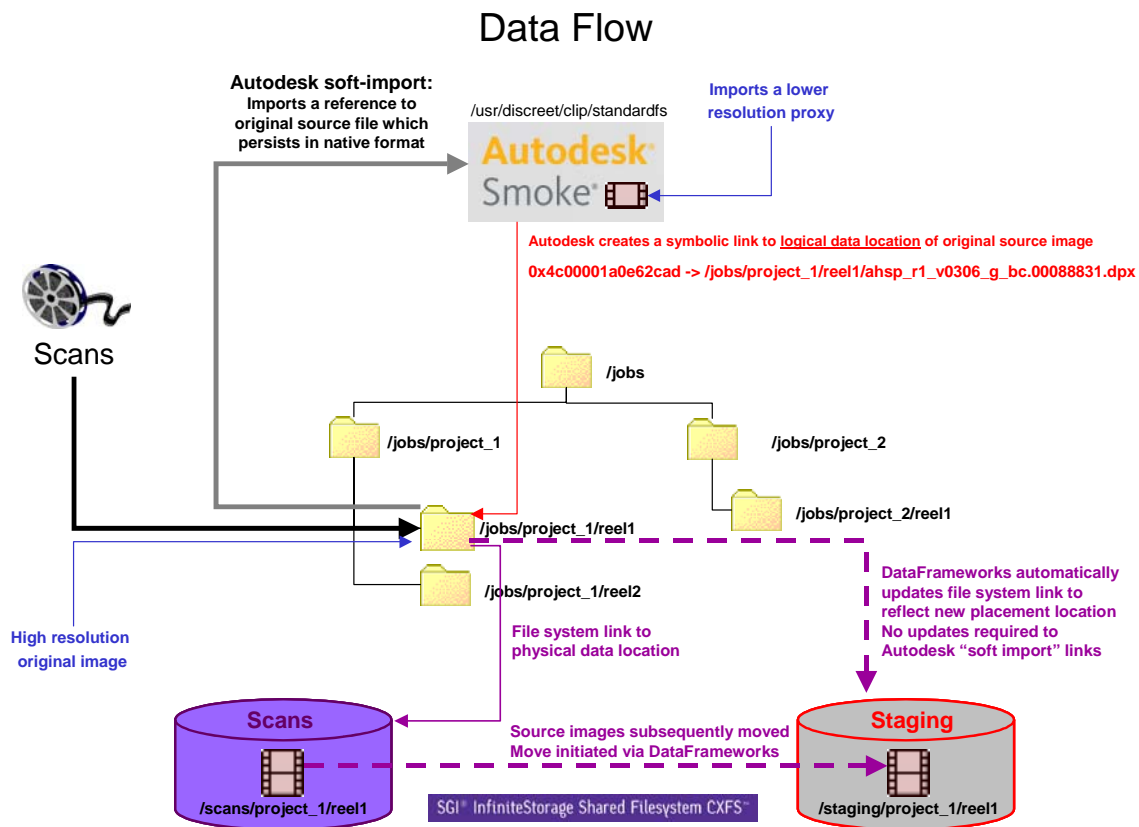


Figure 5. Overview Autodesk “Soft Import” With Location Independence

DataFrameworks removes the need to perform manual updates for each file referenced by the Autodesk “soft import.” To move data that has been “soft imported” into the Autodesk Smoke application from the logical data locations (established with DataFrameworks), The Provider’s Data Managers need to perform the following tasks:

- Data Manager invokes DataFrameworks “Supervised Move” operation
  - DataFrameworks initiates and validates actual movement of data from original source SAN to new destination SAN
  - DataFrameworks automatically updates the logical data references to reflect the new target physical location
- Autodesk references to logical data location do not need to be updated

b) Standardize Data Organization & Naming

Naming conventions provide consistent, predictable, and repeatable data organization which the essential the basic foundation for worldwide distribution of editing and production.

The Provider’s often has defined and outlined basic naming conventions, workflows, and practices but lacks an automated mechanism to lock down adherence. DataFrameworks will inherit and automate on a go forward basis the implementation and management of The Provider’s desired naming conventions. The following provides an example of naming conventions:

(scn = scans | exp = Fire Exports | rnd = Render | ooh = Out Of House)

SCANNING:	show/xx_r1/xx_scn/xx_wo27201234
SCANS FOR CLIENT:	show/xx_client/xx_scn/xx_wo27201234
EXPORTS:	show/xx_r1/xx_exp/xx_rlv123_v01/2048x1556
PICKUPS:	show/xx_r1/xx_exp/xx_rlv123_v02/2048x1556
RENDER:	show/xx_r1/xx_rnd/xx_rlv123_v01_[date]_cc1/2048x1556
XYZ RENDER:	show/xx_r1/xx_rnd/xx_rlv123_v01_[date]_xyz_cc1/2048x1556
OUT OF HOUSE:	show/xx_ooh/xx_wo[wo#]_[barcode ]/[shotname]/2048x1556

Challenges Addressed:

Reduce Confusion And Delays In Process Of Storing And Finding Data

While The Provider may have defined and published naming conventions, they inevitably lack an automated mechanism to enforce and monitor adherence. Conformance with defined naming conventions is often directly related to the number of individuals involved with the creative aspects of any given project. Delays and inefficiencies are less problematic when a project is currently active and fresh in minds of the individuals actively working on project. There is a business risk associated with understanding of project related data. Consistent, predictable, and repeatable data organization provides the foundation for automation of data management and data clean up tasks, the ability to realize automation efficiencies is seriously jeopardized when the knowledge/understanding of project data resides within the memory of key individuals. The problem set becomes exasperated when projects are subsequently revisited 4-6 months later.

### Deliverables

- DataFrameworks will work with The Provider's Producers and Data Managers to recommend, configure, and implement the desired naming conventions within the DataFrameworks "framework" configuration.

c) Production Facing Reports And Views

The following screenshot provides an example of "production facing" view where the Producer can quickly identify and view all data and corresponding storage for a particular project (or sub component of a project).

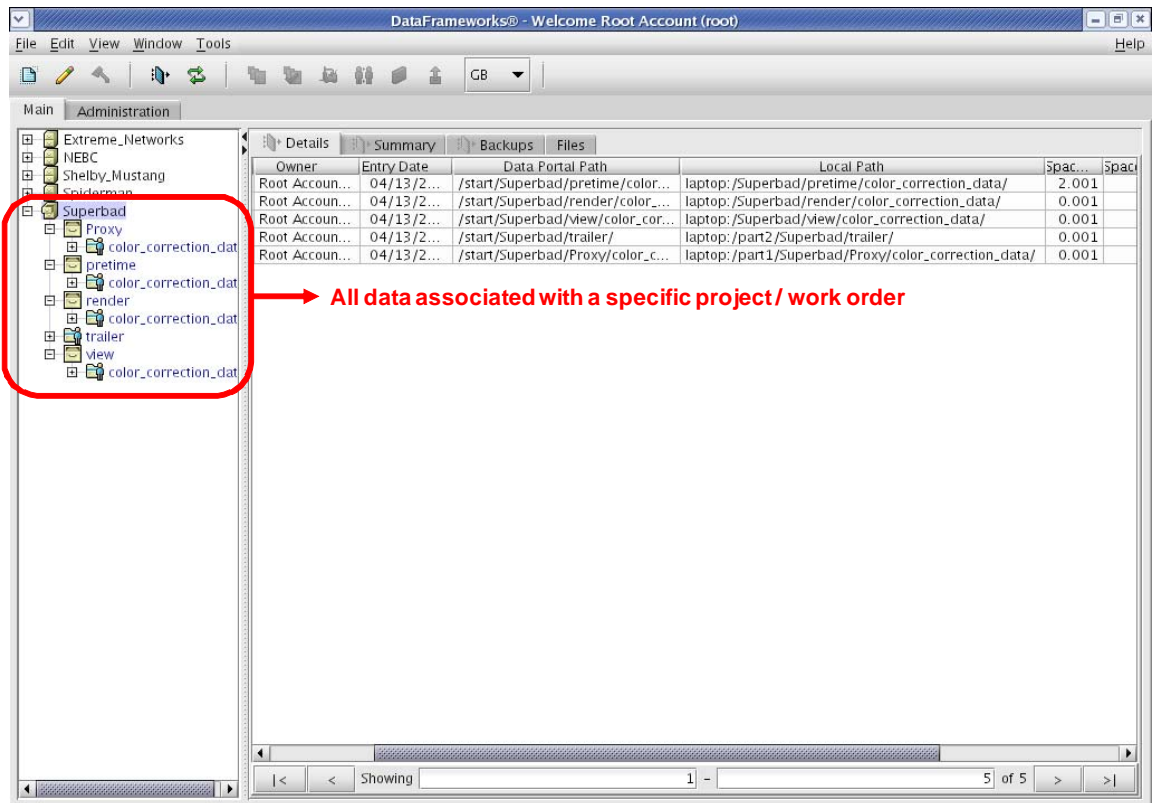


Figure 6. Sample View Of Project Data

The following screen shot provides an example of “production facing” view where storage costs can be assigned and presented for all data and corresponding storage for a particular project (or sub component of a project).

The screenshot shows the DataFrameworks application interface. On the left is a tree view of project folders including Extreme\_Networks, NEBC, Shelby\_Mustang, Spiderman, and Superbad. The main window displays a table with the following columns: Id, Status, Owner, Entry Date, Data Portal Path, Space R..., Spac..., and Cost Reserved. The Cost Reserved column is highlighted with a red box, and a red arrow points to it from the text 'Cost for storing data associated with projects'.

Id	Status	Owner	Entry Date	Data Portal Path	Space R...	Spac...	Cost Reserved
8	OK	Root Accoun...	10/21/2...	/start/NEBC/Fort_St_Johns/Char...	0.001	0.000	\$0.98
64	OK	Tom Ledoux...	10/21/2...	/start/Extreme_Networks/Deve...	0.001	0.000	\$0.98
73	OK	Root Accoun...	11/27/2...	/start/Spiderman/Production/O...	0.001	0.000	\$0.98
52	OK	Root Accoun...	10/21/2...	/start/Extreme_Networks/Sum...	0.001	0.000	\$0.98
40	OK	Root Accoun...	10/21/2...	/start/Spiderman/Library/Obje...	0.001	0.000	\$0.98
6	OK	Root Accoun...	10/21/2...	/start/NEBC/Fort_St_Johns/Char...	0.001	0.000	\$0.98
62	OK	Tom Ledoux...	10/21/2...	/start/Extreme_Networks/Deve...	0.001	0.000	\$0.98
71	OK	Root Accoun...	11/27/2...	/start/Spiderman/Production/O...	0.001	0.000	\$0.98
50	OK	Root Accoun...	10/21/2...	/start/Extreme_Networks/Sum...	0.001	0.000	\$0.98
38	OK	Root Accoun...	10/21/2...	/start/Spiderman/Library/Obje...	0.001	0.000	\$0.98
4	OK	Root Accoun...	10/21/2...	/start/NEBC/Fort_St_Johns/Char...	0.001	0.000	\$0.98
69	OK	Root Accoun...	11/27/2...	/start/Spiderman/Production/O...	0.001	0.000	\$0.98
48	OK	Root Accoun...	10/21/2...	/start/Extreme_Networks/Proje...	0.001	0.000	\$0.98
35	OK	Root Accoun...	10/21/2...	/start/Shelby_Mustang/289_Hi...	0.001	0.000	\$0.98
2	OK	Root Accoun...	10/21/2...	/start/NEBC/Fort_St_Johns/Char...	0.001	0.000	\$0.98
60	OK	Tom Ledoux...	10/21/2...	/start/Extreme_Networks/Deve...	0.001	0.000	\$0.98
57	OK	Root Accoun...	10/21/2...	/start/Extreme_Networks/Sum...	0.001	0.000	\$0.98
46	OK	Root Accoun...	10/21/2...	/start/Extreme_Networks/Proje...	0.001	0.000	\$0.98
11	OK	Root Accoun...	10/21/2...	/start/NEBC/Fort_St_Johns/Char...	0.001	0.000	\$0.98
67	OK	Tom Ledoux...	10/21/2...	/start/Extreme_Networks/Deve...	0.001	0.000	\$0.98
58	OK	Gavin Keeler...	10/21/2...	/start/Shelby_Mustang/Fastbac...	0.001	0.000	\$0.98
55	OK	Root Accoun...	10/21/2...	/start/Extreme_Networks/Sum...	0.001	0.000	\$0.98
12	OK	Root Accoun...	10/21/2...	/start/NEBC/Fort_St_Johns/Char...	0.011	0.011	\$11.19
166	OK	Root Accoun...	01/04/2...	/start/Spiderman/Library/Obje...	0.147	0.000	\$73.41
120	OK	Root Accoun...	12/05/2...	/start/Spiderman/Production/O...	0.088	0.000	\$88.28

Figure 7. Sample View Of Project Data Storage Costs

**Deliverables**

- DataFrameworks will work with The Provider’s production personnel to understand, configure, and implement a few production reports and views.

d) Introduce Degree Of Automation

**Deliverables**

- The DataFrameworks monitors will alert on various “non-conforming” events, DataFrameworks will work with The Provider’s personnel to configure alerts to proper personnel.
- The initial scanning script will automate the processes setting up and allocating file system resources in preparation of project data.

- e) Simplify Common Data Management Tasks  
Automate Conformance Validation And Identification Of Non-Conformance
- f) Synergies With The Provider's Existing Autodesk Infrastructure  
The standard DataFrameworks product will provide the following synergies with The Provider's existing Autodesk infrastructure:
- Facilitate paradigm shift from proprietary stonefs implementation to standard file systems
  - Provide a methodology / structure to facilitate data interchange among Autodesk suite of applications
  - Enable worldwide standardization for project management
  - Enable a repeatable workflow that can be duplicated in locations around the world
  - Provide ability to allocate and report on storage utilization across all file systems and facilities
  - Overall project resource consumption across all file system types

## DataFrameworks Subsequent Deployment Possibilities

### Phase 2

Phase 2 of a DataFrameworks deployment would focus on providing additional efficiencies and possible integration with other infrastructure components.

#### a) Tighten Security And File System Forensics

Many Providers enable a large population of individuals at a location, to have administrator **root** privileges in order to facilitate workflow.

Many of the common data management tasks within image processing require administrator (**root** or **sudu**) privileges on critical machines (meta data servers, etc.). Data managers are often “entry level” positions and many do not possess the level of technical expertise common with System Administrators.

Film scanners require real time access to accurately scan film without interruption. However, the SAN attached to the scanner is basically a modern day film container. Ultimately data must be moved off the SAN's used for scanning in order to make room for new scans. Methodologies like backing up to tape, or slow copy mechanisms tie up valuable scanning resources (which are typically a billable item / service to customers). The faster the scan data can be removed from the SAN, making way for new scans, the more revenue can be generated.

SGI's CXFS copy is a high-speed data copy mechanism. However, the CXFS copy utility may require root-level (administrator) privileges to execute, which can introduce potential security risks. The DataFrameworks interface can be configured to incorporate functionality to allow Data Managers the ability to move data via CXFS copy without providing each Data Manager with root-level (administrator) privileges. Furthermore, the CXFS copy option will only be displayed if the source and destination SAN are SGI CXFS sans, further guiding the user to the acceptable or most efficient options.

Challenges Addressed:

Reduce operational error and a level of security exposure with file-based workflows.

Defragmentation

The conform process typically does not layout the data in a nice, un-fragmented file system layout, required for real time playback. Therefore the files must be defragmented to get acceptable performance for the subsequent color grading / correction process. SGI's file defragmentation utility **xfstools** also requires administrator privileges. Data Managers need to carefully coordinate the timing, and use of the defragmentation utility to avoid adversely affecting the performance of the SAN at critical times. The DataFrameworks interface can be configured to incorporate defragmentation workflow functionality. The interface can be configured to incorporate “custom actions” (for example: defragment). The “custom action” invokes scripts for specific actions, which might allow Data Managers the ability to defragment files, without providing each Data Manager with root-level (administrator) privileges.

#### b) Possible Integration With Other Infrastructure Components

- Render Management
- Backup Software
- Asset Management
- Work Order Management

### **Phase 3**

Phase 3 of a DataFrameworks deployment would focus on possible integration points with Autodesk products commonly utilized at Providers. Existing and future Autodesk infrastructure components would be evaluated for synergistic opportunities.

- a) Possible Integration Points With Autodesk Infrastructure
- Hook into wiretap API to provide aggregate reporting, allocation of resources, trend usage, and enforce structure within smoke / flame.
  - Provide hook to stonefs and standard file system support in smoke / flame that will improve collaboration and project sharing between Lustre and Smoke / Flame