Archivierung Digitaler Kunst aus technischer Sicht

Warum herkömmliche Speichermedien für die Langzeitarchivierung digitaler Kunstwerke nicht geeignet sind.

Armin Weißer¹ · Thomas Ludwig²

Online publiziert: 25. Januar 2021
© Der/die Autor(en) 2021

Interviewpartner: Armin Weißer, iTernity GmbH Freiburg im Breisgau, Deutschland. Über Armin Weißer: Dipl. Inf. Armin Weißer ist seit über 15 Jahren als professioneller Software-Craftsman tätig und verantwortet seit 2016 als Director Software Development bei der iTernity GmbH die Entwicklung der Speicherlösungen iCAS und iCAS FS. Für technologische und technisch konzeptionelle Entscheidun-

gen bedient er sich dabei dem bekannten Designleitsatz „Form folgt Funktion“, den er um „Funktion folgt Operabilität“ erweitert, um dem Betriebsaspekt der langlebigen Softwareprodukte von iTernity ein angemessenes Gewicht zu verleihen. Über die iTernity GmbH: iTernity entwickelt Softwarelösungen für den Schutz und die langfristige Integrität sensibler Daten. Der software-basierte Ansatz ist die Grundlage für zukunftssichere Datenspeicherung und rechtskonforme Archivierung und ermöglicht Hardwareunabhängigkeit, Flexibilität sowie Integritätsschutz bei geringen Gesamtkosten. Über 1200 Organisationen aus verschiedensten Branchen setzen bei der langfristigen Speicherung ihrer Daten auf iTernity Lösungen

**Interview mit Dipl. Inf. Armin Weißer,
Director Software Development bei der
iTerntity GmbH**

Zusammenfassung

Eine zentrale Aufgabe eines Museums ist die Aufbewahrung von Kunstwerken für nachfolgende Generationen. Neben der Konvertierung und Digitalisierung ursprünglich analoger Werke steigt auch die Anzahl an sogenannten born-digital Werken, wie z. B. digitalen Videos und Fotos.

Die Frage nach adäquaten Speichermedien für die langfristige Archivierung dieser Werke stellt IT-Abteilungen in Kunsthäusern und Museen vor komplexe Herausforderungen. Neben der Unveränderbarkeit und Integrität der Kunstwerke dürfen auch steigende Datenmengen, Kostenaspekte und Investitionssicherheit nicht außer Acht gelassen werden.

Dipl. Inf. Armin Weißer, Director Software Development bei iTernity, beleuchtet im Interview die technologischen Herausforderungen bei der Archivierung digitaler und di-

Armin Weißer
armin.weisser@iTerntity.com

✉ Thomas Ludwig
ludwig@dkrz.de

¹ iTernity GmbH, Freiburg, Deutschland

² Deutsches Klimarechenzentrum, Hamburg, Deutschland

gitalisierter Kunstwerke. Wie kann ein software-basierter Ansatz Zukunftssicherheit schaffen? Mit welchen Schwierigkeiten sind herkömmliche Speichermedien verbunden? Wie kann eine langfristig sichere Archivierung von digitalen Kunstwerken in der Praxis umgesetzt werden?

Interview: Digitale Kunstwerke langfristig sicher archivieren – aber wie?

Herr Weißer, die Medienkunst hat sich in den letzten Jahren schnell weiterentwickelt. Die Anzahl an digitalen Werken nimmt stark zu. Welche Herausforderungen bringt die Archivierung von digitalen Kunstwerken aus technischer Sicht mit sich?

Die erste technische Herausforderung ist es, eine wachsende Datenmenge effizient zu verarbeiten und trotzdem jedes Objekt mit der größten Sorgfalt zu behandeln.

Die zweite Herausforderung ist es, eine hohe Verfügbarkeit, Integrität und modulare Erweiterbarkeit des Archivsystems sicherzustellen und gleichzeitig den Verwaltungsaufwand sowohl für die Anwender als auch für die IT zu minimieren.

Und zuletzt die Aufgabe, digitale und digitalisierte Kunstwerke im Ganzen zu archivieren. Die Anwender wollen insbesondere sehr große Dateien mit mehreren Terabyte archivieren, ohne sie aus rein technischen Gründen vorher zerteilen zu müssen.

Glücklicherweise sind das Herausforderungen, mit denen die Kunsthäuser nicht allein sind. Die iTernity GmbH beschäftigt sich schon seit vielen Jahren mit dem Thema Langzeitarchivierung von digitalen Inhalten. Unsere Kunden, aus unterschiedlichen Branchen, haben durchaus ähnliche Herausforderungen wie die Kunsthäuser.

Wenn man z.B. den reinen Workload betrachtet – wie viele Dateien welcher Größe müssen in welcher Zeit geschrieben und/oder gelesen werden – gibt es bei der Archivierung von digitalen Kunstwerken durchaus Ähnlichkeiten mit der Speicherung von Full-Backups. Bei beiden nimmt insbesondere die Größe einer einzelnen Datei stark zu und übersteigt schnell die Kapazität einer Festplatte.

Auch die Notwendigkeit nach langfristiger Integrität und Unveränderbarkeit der Daten teilen sich die Kunsthäuser mit Kunden anderer Branchen, wie z.B. Kliniken, Industrieunternehmen oder Banken. Wir behandeln im Grunde alle Daten unserer Kunden wie wertvolle Kunstgegenstände. Ob es sich nun um ein einseitiges PDF mit den Notizen des letzten Meetings, das Backup einer virtuellen Maschine oder um eine künstlerische Videoinstallation handelt: Es gibt viele Gründe für langfristigen Integritätsschutz, und die Angst von erpresserischen Ransomware-Angriffen ist nur einer davon.

Aufgrund der wachsenden Datenmengen ist auch die Planungssicherheit ein großer Aspekt. Die Kunsthäuser

wollen und sollen sich schließlich auf ihre Kernaufgabe konzentrieren, ohne ihre Zeit mit dem technischen Betrieb, der Erweiterung oder der Interaktion mit dem digitalen Langzeitarchiv zu vergeuden. Die Lösung sollte also modular von ein paar wenigen Terabyte- bis in den oberen Petabyte-Bereich skalieren können, ohne dass sich der Aufwand für die Fach- und IT-Abteilung erhöht.

Das Kunsthhaus Zürich setzt für die Archivierung digitaler Kunstwerke die Archivplattform iCAS FS ein. Wie läuft die Archivierung eines digitalen Kunstwerks konkret ab?

Beim Kunsthhaus Zürich liegt der sogenannte „Non-Application-Data“-Fall vor, d.h. nicht Applikationsserver, sondern Menschen interagieren direkt mit dem Langzeitarchiv. Hier sind eine nahtlose Integration in die Arbeitsumgebung der Fachabteilung und eine intuitive Handhabung des Archivs sehr wichtig [1].

Beim Kunsthhaus Zürich hat man sich für drei Speicherbereiche mit einer jeweils spezifischen Mindestaufbewahrungsdauer entschieden. Auf den Windows-Workstations der Fachabteilung werden diese über das SMB/CIFS-Protokoll mit geeigneter Benutzer-Authentifizierung als Netzlaufwerke eingebunden und sind somit verwendbar, wie eine externe Festplatte. Die digitalen Kunstwerke werden mit den bekannten Methoden der lokalen Dateiverwaltung auf einen der Speicherbereiche kopiert und schließlich als schreibgeschützt markiert. Der Schreibschutz kennzeichnet die Datei als nicht veränderbar. Das unterscheidet sich von der Handhabung bis hierhin nicht von einem herkömmlichen NAS.

Im Gegensatz zu einem herkömmlichen NAS kann man den Schreibschutz einer Datei auf einem Langzeitarchiv aber erst dann wieder entfernen, wenn die Mindestaufbewahrungsdauer abgelaufen ist. Im Fall digitaler Kunstwerke ist die Mindestaufbewahrungsdauer entsprechend lange angelegt. Mit der langfristigen Unveränderbarkeit der Objekte sind die wertvollen Inhalte somit auch vor einer Ransomware-Attacke sicher, obwohl sich diese ja vornehmlich auf eingebundenen Netzlaufwerken ausbreiten würde. Bei iCAS FS beißt Malware zum Glück auf Granit.

Für andere Branchen gibt es gesetzliche Vorschriften, was die Mindestaufbewahrungsdauer und das Löschen von Daten angeht – Stichwort DSGVO. Diese rechtlichen und regulatorischen Aspekte spielen aber bei der Archivierung von digitaler Kunst bisher keine Rolle. Hier geht es vielmehr um die Konservierung der Kunstwerke – deshalb arbeitet das Kunsthhaus auch mit einer „unbegrenzten“ Aufbewahrungsdauer für die digitalen Kunstwerke.

Wie sieht die Archivlösung für den Einsatz im Kunsthaus konkret aus?

Das Archivsystem ist ein Cluster bestehend aus mehreren baugleichen, dedizierten x64-Servern mit jeweils 24 internen Hard Disk Drives (HDD) vom Typ 7200k LFF. Auf jedem Server läuft eine software-basierte Appliance bestehend aus einem Enterprise-Linux, einem verteilten Dateisystem und weiteren Softwarekomponenten für das Monitoring und das Cluster-Management. Es gibt in einem iCAS FS Cluster keine Unterscheidung zwischen Daten- und Metadaten-Knoten und damit keinen Single Point of Failure.

Die Knoten werden in der Regel auf unterschiedliche Brandschutzzonen im Rechenzentrum des Kunden verteilt und sind untereinander mit einem dedizierten 10 Gbit/s-Ethernet verbunden. Auch eine Verteilung der Knoten auf verschiedene Geo-Lokationen ist möglich.

Die HDDs selbst werden in den Servern mit einem RAID-6 zusammengefasst, partitioniert und mit einem lokalen Dateisystem formatiert. Die virtuellen Disks stellen dann wiederum die Speicherbausteine für ein clusterweites, verteiltes Dateisystem dar. Auf dem verteilten Dateisystem werden schließlich logische Speicherbereiche angelegt und per SMB/CIFS oder NFS präsentiert. Die SMB bzw. NFS Server-Prozesse laufen implizit auf jedem Cluster-Knoten mit. Es sind hier also keine dedizierten Gateways für SMB/NFS notwendig.

Insgesamt war es uns wichtig die Gesamtarchitektur so einfach wie möglich zu halten. Für die Hochverfügbarkeit der externen Schnittstellen (SMB, NFS und HTTPS) verwaltet der Cluster virtuelle IPs, die er zu jeder Zeit einem der verfügbaren Knoten zuweist. Vom Client aus betrachtet ist der Cluster also eine zusammenhängende, ausfallsichere Ressource im IT-Stack und nicht eine Ansammlung von einzelnen Storage-Servern.

Die Speicherbereiche für die Ablage der Daten werden jeweils mit einem eigenen Replikationsfaktor und einem Read/Write-Quorum ausgestattet. Bei einem Cluster bestehend aus drei Datenknoten könnte man sich z.B. für einen Replikationsfaktor von 3 und ein Quorum von 2 entscheiden.

- Das Read/Write-Quorum von 2 wäre in diesem Szenario als Kompromiss zwischen Verfügbarkeit und Konsistenz zu verstehen.
- Solange zwei von drei Knoten im Cluster verfügbar sind, lassen sich Dateien lesen und schreiben. Damit ist gewährleistet, dass zu jeder Zeit mindestens zwei identische Kopien einer Datei vorliegen und auch ein Lesen eines Kunstwerkes nur dann erfolgreich ist, wenn es auf mindestens zwei Knoten Bit für Bit identisch ist.
- Auf der anderen Seite kann bei diesem Setup jederzeit einer von drei Servern bedenkenlos zu Wartungszwecken heruntergefahren werden, ohne dass die Verfügbarkeit

oder die Konsistenzkriterien des Archivs herabgesetzt werden müssen. Die Synchronisation mit dem dritten Server erfolgt nach dessen Neustart automatisch im Hintergrund.

Digitale Kunstwerke müssen auf ihrer langen Reise verlässlich geschützt werden. Wie lässt sich die Integrität und Unveränderbarkeit der Daten langfristig sicherstellen?

Naja, zu viel „reisen“ sollen die Daten im Optimalfall nicht. In meinem Gedankenmodell sollen es sich die Daten auf dem Archivsystem bequem machen und sie sollen vor allem gesund bleiben. Für mich sind Daten, die nicht bewegt werden müssen, gute Daten, da eine Datenbewegung auch immer Kosten und Risiken mit sich bringt. Ganz verhindern kann man die Datenbewegung allerdings nicht, da die physischen Festplatten als Speichermedium natürlich nicht verlässlich und langlebig genug sind.

Neben der physischen Speicherebene ist die Datenintegrität durch unbeabsichtigte oder böswillige Veränderung von außen gefährdet, z.B. durch einen Benutzerfehler oder eine Ransomware-Attacke. Hier greift der oben beschriebene Schreibschutz. Er verhindert, dass ein Kunstwerk vor Ablauf der Mindestaufbewahrungsdauer von außen über die SMB-Schnittstelle verändert werden kann. Diese Logik wird softwareseitig auf dem verteilten Dateisystem mit einem Chain-of-Responsibility-Pattern gelöst, bei dem jede Dateioperation durch eine Verarbeitungskette wandert und verschiedene Hürden nehmen muss, bevor sie an ein darunter liegendes Speichermedium delegiert wird. Eine verändernde Dateioperation wird in dieser Kette dann entsprechend streng geprüft und von der Software zurückgewiesen.

Die Daten werden zudem synchron auf die verschiedenen Cluster-Knoten repliziert und auf Block-Ebene verschlüsselt.

Welche Rolle spielt die redundante Datenspeicherung beim Schutz der Datenintegrität?

Um das Risiko eines Datenverlustes durch eine Beschädigung des physischen Speichermediums zu verringern wird innerhalb eines Archivknotens ein RAID eingesetzt. Eine zusätzliche clusterweite Replikation schützt zudem vor einem Katastrophenfall, bei dem ggfs. ein ganzer Archivknoten oder eine ganze Site zerstört wird.

Die clusterweite Replikation hat aber noch einen anderen Nutzen. Durch sie lässt sich sehr zuverlässig die Korruption einer einzelnen Datei nachweisen und auch wieder heilen. Den ersten Schritt bezeichnen wir als Self-Testing, den zweiten als Self-Healing.

Abb. 1 Architektur von iCAS FS mit der Anbindung verschiedener Datenquellen

Das Self-Testing operiert auf einer einzelnen Datei und vergleicht den zum Zeitpunkt der Archivierung gespeicherten Hash-Wert mit dem aktuellen Hash-Wert des Dateieinhaltes.

Das Self-Healing sucht sich eine nicht beschädigte Kopie und ersetzt die defekte Datei mit der validen Kopie.

Kann damit ein Backup der Daten ersetzt werden?

Ein Backup ist zunächst nichts anderes als eine weitere Kopie der Daten, auf die nur lesend zugegriffen werden kann. Ein Backup ergibt aber nur dann Sinn, wenn dabei auch ein Medienbruch – also die Kopie der Daten auf ein anderes Speichermedium in einer anderen Feuerzone oder sogar einer anderen Geo-Lokation – erzwungen wird.

In unserem Fall werden die Daten durch das Archivsystem selbst auf unterschiedliche Lokationen repliziert und durch den Schreibschutz bis zum Ende der Mindestaufbewahrungsdauer gegen Manipulation von außen geschützt. Ein versehentliches Löschen innerhalb der Mindestaufbewahrungsdauer ist damit ebenso ausgeschlossen wie eine mutwillige Manipulation.

Einem Datenverlust durch Hardwaredefekte oder Katastrophen wird durch den Einsatz von RAID und der clusterweiten Replikation im Zusammenhang mit dem Self-Healing Verfahren vorgebeugt, so dass in der Summe ein zusätzliches Backup für den Kunden unnötig wird.

Das Zurücksetzen des kompletten Datenbestands auf einen bestimmten Zeitpunkt in der Vergangenheit ist mit dem Verfahren allerdings (noch) nicht gewährleistet. Wir arbeiten an der Integration eines Snapshot-Features in iCAS FS, um auch diesen Aspekt abzudecken. Damit könnten

auch solche Dateien wiederhergestellt werden, die nicht mit einem Schreibschutz versehen wurden.

Ein zusätzliches Backup kann in manchen Fällen trotzdem Sinn ergeben, so wie es auch Sinn ergeben kann, ein Backup selbst zu archivieren und wie ein Kunstwerk zu konservieren. Auch für diesen Einsatzzweck lohnt sich der Einsatz eines Langzeitarchivsystems.

Auch im Kunstwesen herrscht der Trend steigender Datenmengen und immer größerer Dateien wie z.B. 4k-Videos. Wie kann dem Datenwachstum Rechnung getragen werden?

Die Größe einer einzelnen Datei liegt bei digitalen Kunstwerken schnell mal bei mehreren Terabyte. Die Tendenz steigt aufgrund von Full HD und 4K oder sogar 8k auflösenden Videos.

Es ist notwendig, dass das Archivsystem mit sehr großen Einzeldateien umgehen kann und dass sich die Gesamtkapazität modular, durch das Hinzufügen weiterer Festplatten (Scale-Up) oder durch das Hinzufügen weiterer Archivknoten (Scale-Out), erweitern lässt.

Nach einer Erweiterung des physischen Speichers können dann die logischen Speicherbereiche erweitert werden. Für die Fachabteilung im Kunsthaus läuft das alles transparent ab.

Wie lässt sich eine Vereinbarkeit von Kosten, Datenwachstum und Investitionssicherheit schaffen?

Das steigende Datenwachstum erfordert eine Archivlösung, die flexibel und wirtschaftlich erweitert werden kann. In-

Tab. 1 Vergleich: Warum herkömmliche Speichermedien für die Archivierung von digitaler Kunst nicht geeignet sind

Lösung	Nachteile
Proprietäre Hardware-Appliance	Hardware- & Vendor Lock-In Hohe Anschaffungskosten Spezielles Know-How für den Betrieb notwendig
SAN	Keine Archivierungsfunktion (WORM) Kein Schutz vor Ransomware Erfordert spezielle Infrastruktur Hohe Anschaffungs- und Wartungskosten
NAS	Keine Archivierungsfunktion (WORM) Kein Schutz vor Ransomware
Server mit internen Disks	Keine Archivierungsfunktion (WORM) Kein Schutz vor Ransomware Limitierung auf die Kapazität eines Servers, bzw. die Verteilung der Daten muss von außen gesteuert werden Keine Hochverfügbarkeit Hohes Risiko bei Hardware-Ausfällen
Cloud-Speicher	Hohe laufende Kosten pro TB Hohe Latenzen und limitierter Durchsatz durch WAN Anbindung Ggfs. proprietäre Desktop- und Web-Applikationen, die erlernt werden müssen Daten bleiben nicht im Haus
Software-basierte OpenSource-Lösungen	Beratungsintensiv Hohe Komplexität im Betrieb Kein Support Kein One-Stop-Shopping

dem sich die Gesamtkapazität durch das Hinzufügen zusätzlicher Archivknoten modular ausbauen lässt, kann das System mit steigenden Anforderungen flexibel wachsen. So lässt sich sicherstellen, dass das System dem Datenwachstum auch auf lange Sicht standhält.

Dieser steigende Speicherplatz darf aber nicht teuer sein. Ein Archivsystem, das auf Standardhardware basiert, ermöglicht im Gegensatz zu proprietären Systemen eine hohe Flexibilität bei der Speicherauswahl und langfristige Kosteneinsparungen. Die Hardwareunabhängigkeit stellt zudem sicher, dass bereits getätigte Investitionen in Speicherhardware geschützt werden. Weitere Kosten können durch einen selbstverwaltenden Betrieb des Archivsystems eingespart werden, da keine Administration, Schulung oder spezielles Anwenderwissen nötig sind.

Was sollten Museen und Kunsthäuser beim Aufbau einer langfristigen und sicheren Archivierungsstrategie beachten?

Die Kunsthäuser müssen sich sehr genau überlegen, welche Trade-Offs sie bereit sind einzugehen, wenn es sich um die Konservierung von digitalisierten und born-digital Kunstwerken handelt (s. Tab. 1). Im Entscheidungsprozess würde

ich aber folgende Punkte auf die Liste der verpflichtenden Kriterien setzen und keine Kompromisse eingehen:

- WORM-Funktion mit Retention-Management und Zugriffsschutz Schutz vor Ransomware-Attacken und Veränderung der Daten
- Flexible Replikation (synchron und asynchron) im Zusammenspiel mit Self-Healing für langfristige Datenintegrität auch im Katastrophenfall oder bei Hardware-Defekten
- Standard Dateisystem-Schnittstellen wie SMB/NFS zur einfachen Integrierbarkeit in bestehende, organisatorische Prozesse
- Flexibles und erweiterbares Speichermanagement Möglichkeit klein zu starten und nach Bedarf die Kapazität zu erweitern
- Unterstützung sehr großer Einzeldateien von mehreren Terabyte damit immer größer werdende Dateien im Ganzen archiviert werden können

Folgende Fragen sollten vor einer Entscheidung zusätzlich beantwortet werden, sind aber als Trade-Off gegen die verpflichtenden Kriterien zu halten:

- Sollen die Kunstwerke im Haus bleiben oder vertraue ich sie der Cloud an?
- Welche Expertise und welche Ressourcen muss ich als Kunde beisteuern, um die Lösung betreiben zu können?
- Wie hoch sind die Gesamtkosten des Betriebs (total cost of ownership, TCO) über die nächsten 5 oder 10 Jahre?
- Vertraue ich darauf, dass ggfs. benötigte Spezialhardware in meinem Betrachtungszeitraum ausreichend zur Verfügung steht? Oder setze ich auf eine grundsätzliche Hardwareunabhängigkeit der Lösung?

Anhang

Langfristig sichere Datenarchivierung mit der Scale-out Archivplattform iCAS FS

iCAS FS ist eine Scale-Out-Plattform für die langfristig Speicherung von sensiblen Daten, wie beispielsweise digitalen Kunstwerken. Die Appliance-ähnliche Architektur basiert auf Standardhardware und einem Linux-Betriebssystem. iCAS FS ist ab 20 TB unbegrenzt skalierbar und ermöglicht, steigende Datenmengen im Kunstumfeld flexibel zu bewältigen (s. Abb. 1). Als selbstverwaltende Lösung reduziert iCAS FS den Aufwand für die IT-Abteilung auf ein Minimum und ermöglicht eine Kostenreduktion um 53 % im Vergleich zu Public-Cloud-Speichern.

Open Access Dieser Artikel wird unter der Creative Commons Namensnennung 4.0 International Lizenz veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprüng-

lichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Artikel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.

Weitere Details zur Lizenz entnehmen Sie bitte der Lizenzinformation auf <http://creativecommons.org/licenses/by/4.0/deed.de>.

Literatur

1. Kunsthaus Zürich (2019) Anwenderbericht. https://iternity.com/files/assets/03_Downloads-PDFs/DE/Case-studies/Anwenderbericht_Kunsthaus-Zuerich.pdf. Zugegriffen: 21. Nov. 2020