

Digitale Infrastruktur

1) Etymologie: Unterbau

- Energie, Gebäude, Datenspeicher, Bandbreite, Silizium, ...
- Datensicherheit (Safety / Security)
- Personal: Wartung, Pflege, Erneuerung, Kompetenz und Bildung

2) Holistischer Ansatz braucht Vorbereitung und Durchhaltevermögen

- Technologischer Fortschritt ist exponentiell (aber flacht ab in den letzten Jahren), trotzdem, es gilt Schritt zu halten und stetig zu reevaluieren
- Manche Technologien sind flüchtig
- Kosten schwer berechenbar
 - Insbesondere bei mehr Abhängigkeit von Dienstleistern

BIZ & IT —

Why Facebook thinks Blu-ray discs are perfect for the data center

Cheap storage with a 50-year life span? Facebook spreads the gospel of Blu-ray.

JON BRODKIN - 1/31/2014, 3:00 PM

Sache des Landes?

1) Beispiele: Infrastruktur bereitgestellt vom Land

- [bwSync&Share](#)
- [BitBW](#)
- [Hochschulservicezentrum Baden-Württemberg](#)

2) Realisierbar? Kostenstellen? Bürokratie?

3) Zentrale vs. dezentrale Archivierung?

- Vertrauen? Auch in den Staat (... ich erinnere an den Pharao)?
- Vertrauen? Mehr Redundanz beinhaltet mehr Vertrauensanker.
 - Verschlüsselung löst ein Teil der Probleme

4) Welche Optionen haben wir?

Kooperationen Hochschulen I

1) Infrastruktur: **weniger ist mehr**

2) Datenspeicherung

- Achillesferse, aber handhabbar
 - „Immutabilität“
 - Idempotenz
 - Kryptographisches Konzept
 - Dokumentierte, portable und verbreitete Datenformate

3) Freiheit in der Präsentation und Benutzeroberfläche

- Kleine zielgerichtete Projekte mit kompetenten Entwicklern oder Studenten häufig ergiebiger



16 TB Festplatte: ~250€
Datenverkehr innerhalb des BelWü

News

500-Millionen-Euro-Projekt scheitert: Lidl bläst SAP-Software ab

Sieben Jahre hat Lidl mit SAP an der Einführung eines neues Warenwirtschaftssystem gearbeitet. Jetzt erklärt der Lebensmittel-Discounter das Projekt für gescheitert.

Von Andreas Weck

02.08.2018, 15:46 Uhr • 2 Min. Lesezeit

Kooperationen Hochschulen II

4) Datenspeicherung: zwei Schichten

- Datenformate, in welchem Format wird gespeichert (später mehr)
- Datenspeicherung auf Datenträgern
 - Redundanz schafft Sicherheit und erlaubt es Fehler zu begehen und zu korrigieren, indem die Daten von einer anderen Kopie geladen werden
 - Redundanz über größere Distanz erlaubt es Daten Jahrzehntereignisse zu überleben
 - Kryptographie erlaubt es Daten bei **Freunden** zu „parken“ und wiederzuholen mit der Gewissheit, dass sie nicht hineingeguckt haben und sie ebenfalls nicht verändert haben
 - Datenträger sind vergleichsweise billig
 - Bandbreite ist ebenfalls billig, insbesondere im Hochschulnetz

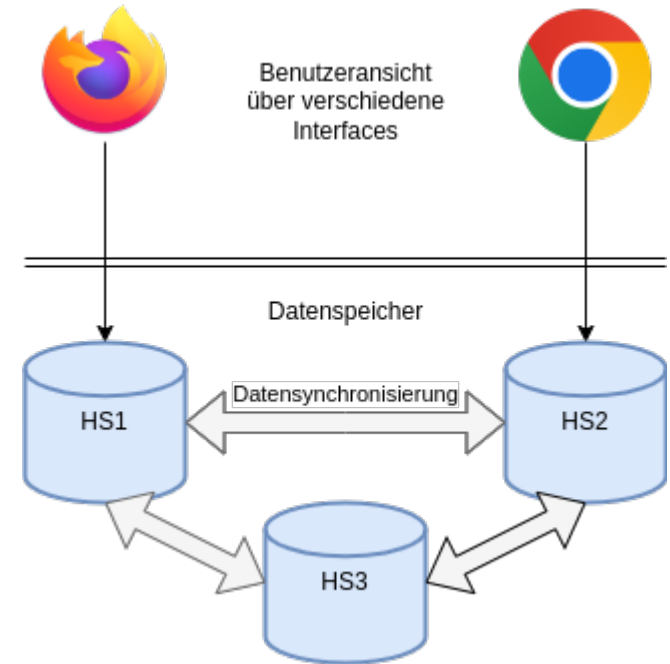
Einwurf: zentral vs. dezentral

1) Zentrale Infrastruktur

- Hochkonjunktur: Cloud Betreiber
- Abhängigkeit
- Klare Zuständigkeit, Support

2) Dezentrale Infrastruktur

- Hochkonjunktur: Twitter vs. Mastodon
- Höhere Ausfallsicherheit
- Aber: höhere Komplexität



Kooperationen Hochschulen III

5) Erst verlässliche Datenspeicher ermöglichen Experimente

- Entkoppeln der Datenspeicher von der Präsentation zum Benutzer
- Verschiedene Software kann entwickelt oder wiederverwendet werden um auf die Datenspeicher zuzugreifen
- Katastrophen bleiben aus, falls die Kopien an anderen Hochschulen nicht durch den Vorfall berührt werden
- Ermöglicht die Entwicklung verschiedener semantischer Katalogsystemen spezifisch für die Hochschulen
 - Kann standardisiert werden falls gewünscht
 - Und eine Hochschule kann sich ebenfalls wieder vom Standard verabschieden

Kooperation mit Freier Software I

1) Welche Software verwenden?

- Das Softwarekonzept muss stimmen und den Anwendungsfall konkret abdecken, sowie Spielraum lassen
- Die umgesetzten abstrakten Konzepte sind hier zu evaluieren, keine Benutzeroberfläche
- Zweitrangig die Community, in welcher die Software entwickelt wird
 - Hängt sie von einer Firma ab, gibt es diese in ein paar Jahren noch?
 - Was passiert mit den Daten, falls es die Firma nicht mehr gibt?
 - Preissteigerungen? Knebelverträge?
- Open Source und freie Software gibt hier Sicherheit
- Verifizierbarkeit der Umsetzung der abstrakten Konzepte und Architektur in tatsächlichen Code

Kooperation mit Freier Software II

2) Wert von Open-Source Software erkennen

- Entdecken von brauchbarer freier Software
 - Abgucken von bisherigen erfolgreichen Projekten
 - Ausnahmslos freie Software
 - [Internet Archive](#): entwickelte Open-Source Software
 - [Software Heritage](#): entwickelte Open-Source Software
 - [WikiData](#): entwickelte Open-Source Software
- Mit Entwicklern von freier Software reden, planen und bezahlen

Datenformate

1) Die Verwendung von offenen und dokumentierten Datenformaten ist zwingend notwendig

- Die Möglichkeit eines praktischen totalen Datenverlusts ist zu hoch
- Die Kosten bei Datenrettung und Wiederherstellung können explodieren
- Die Möglichkeiten mit offenen und dokumentierten Datenformaten zu arbeiten ist größer, häufig einfacher den Benutzern des Archivs bereitzustellen, ohne proprietäre Software beim Benutzer zu erzwingen

Ausblick: Semantisches Web

1) Semantische Verordnung von Entitäten

- Viele Beschreibungssprachen ermöglichen das semantische verlinken von Entitäten zu anderen Entitäten, vieles aber seit dem Hype nicht mehr im Gebrauch
- Deutsche Nationalbibliothek
 - Stellt die Gemeinsame Normdatei (GND) bereit
 - Formal gesehen die richtige Grundlage
 - Mitarbeit mühselig
- WikiData
 - Wikipedia als Grundlage
 - Höherer Mehrwert für Benutzer durch Tools und die Einbindung in die Wikipedia
 - Einfache Mitarbeit
- Einfach Wikipedia unterstützen? Alle Karten auf ein Pferd setzen?

Semantische Domains anderswo

1) Youtube, Instagram, TikTok

- Haben alle eigene „semantische Domänen“ erschlossen, teils öffentlich ersichtlich, teils nicht
 - Öffentlich: Kommentare, Likes, Antworten, Reaktionen, Verlinkung
 - Nicht öffentlich: Dislikes (YouTube), algorithmische Bewertung
- Kunst passiert auf diesen Plattformen. Was tun?
 - YouTube etc. erschwert die Archivierung aus ihrem Videofundus immer mehr
 - z.B. DMCA Take-Down von youtube-dl
 - Viele andere Plattformen handeln ähnlich bzw. sind komplett geschlossen

2) Kooperationen mit Sozialen Medien?

- Sollen Archive in Soziale Medien ihre Inhalte pushen?

Case Studies

1) Github - Artic Code Vault

- Siehe Bild
- Bereits Wassereinbruch

2) M-DISC (Blu-ray)

- Langlebige read-only Blu-ray Disc

3) Git-annex

- Verteiltes Datensynchronisierungswerkzeug mit Revisionssystem
- Wird bereits von Archiven verwendet

4) syncthing

- Dezentrale Datensynchronisierung

5) sqlite

- Beispiel für eine kompakte kleine Datenbank in einer Datei mit offenem Datenformat
- Skaliert bis zu vielen Terabytes

6) BitTorrent

7) NNCP



Software und Archive

Ich freue mich sehr über Fragen oder hoffe, dass eine Diskussion zustande kommt.