



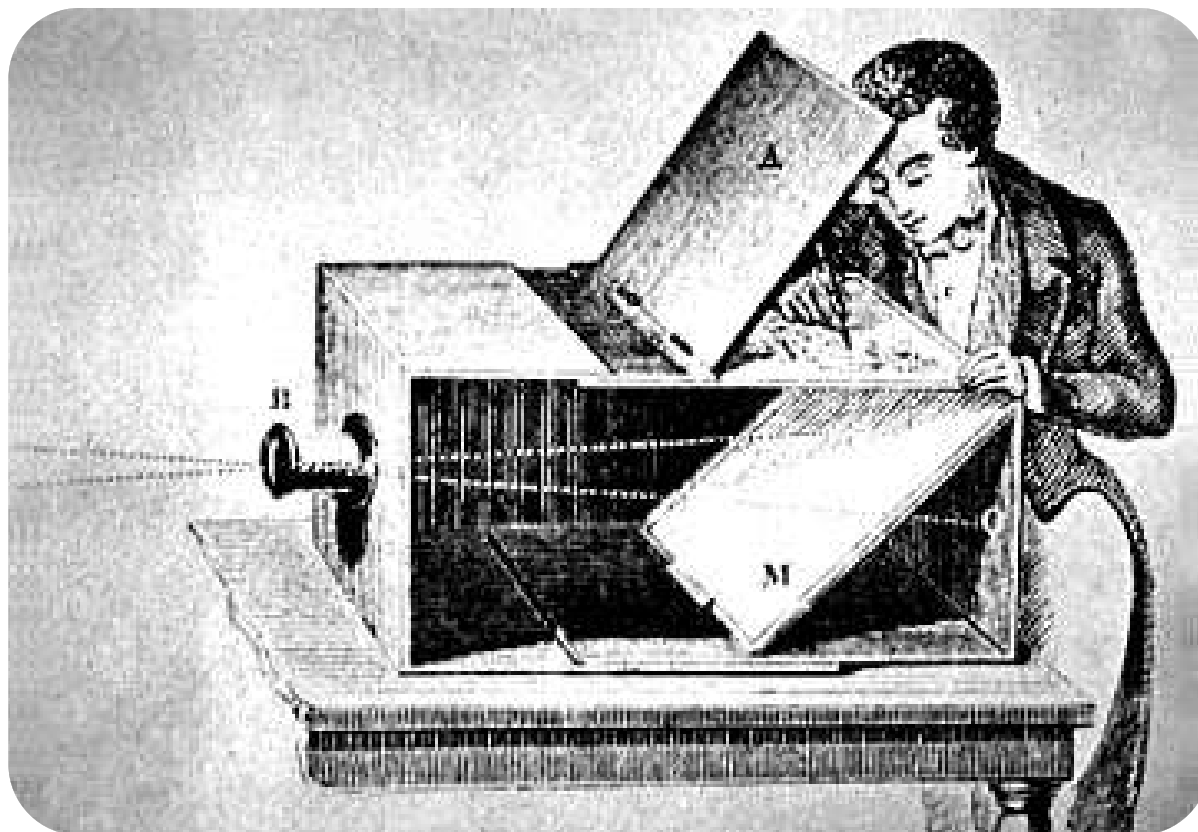
- “Digitaalse ainese ja informatsiooni säilitamine ning kasutamine on nii tuleviku majanduse arengu nurgakivi kui ka tulevikus tekkiva pärandi alus.”
– Seamus Ross, 2000
- “Peab rõhutama, et digitaalsele informatsioonile pika-ajalise ligipääsu pakkumine nõuab pika-ajalist pühendumist.”
– Bernard Smith, 2003



- “Tuginedes moodsale arusaamale on (staatilised) lingitud andmed täiesti uued seosed – kasutajad jälgivad ja otsivad informatsiooni ning peavad olema kaasatud. “
– Tim Berners-Lee, 2009



Pikaajaline digitaalne säilitamine mäluasutuste perspektiivis.



Andres Uueni
12.01.2012



Ennistuskoda Kanut (EKK)

- EKK on asutatud 1986. a. riikliku konserveerimiskeskusena
 - 28 riigi muuseumit, sh 9 keskmuseumit
 - Muinsuskaitseamet
 - Projektipõhised kliendid
 - Eratellijad
- maali ja polükroomse materjali konserveerimise osakond
- mööbli ja metalli konserveerimise osakond
- esemete (keraamika ja klaas, paber ja fotod, nahk ja köide ning tekstiil) konserveerimise osakond
- dokumenteerimise ja digiteerimise osakond



Pika-ajaline digitaalne säilitamine

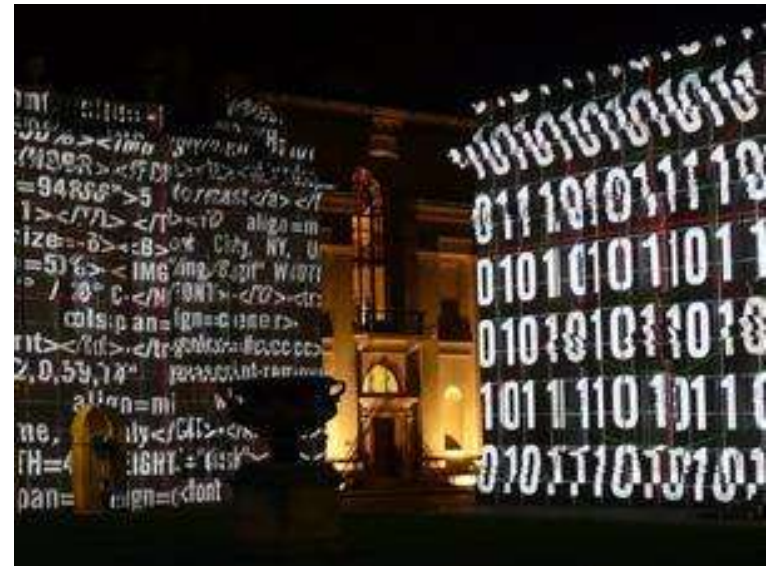
Kanutis

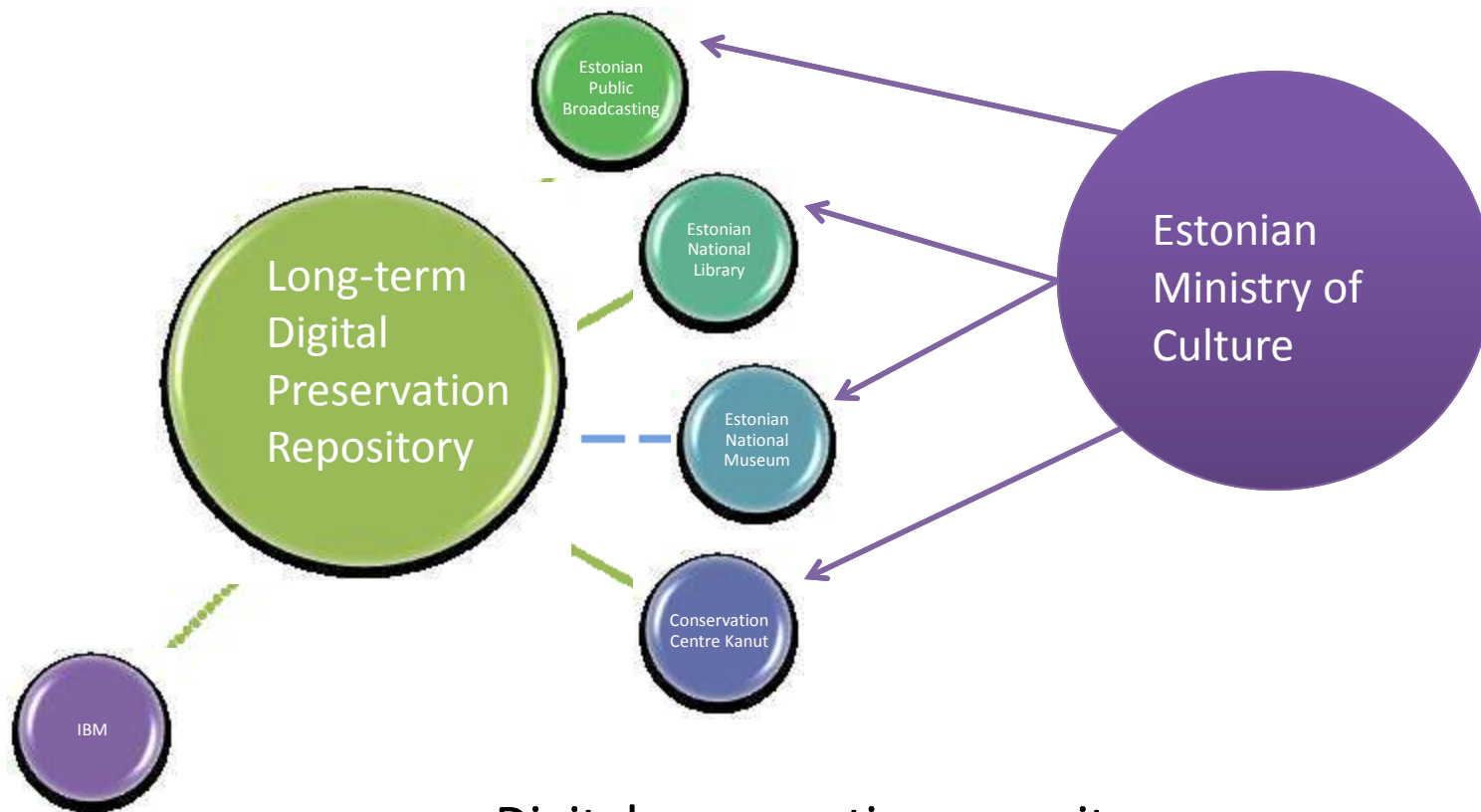
- Ülesanne: pikaajaline säilitamine, mis oleks turvaline ja töökindel
 - Kuni 2004 *online* ja *offline* meediate kombinatsioonid
- 2005 – serveri-põhine lahendus
 - (900 Gb)
- 2008 – uued digiteerimislahendused
- 2008 – serveri ümber ehitamine
 - (RAID5 3 Tb)
- 2009 – digiteerimistehnika täiendamine
- 2010 – serveri ümber ehitamine
 - (RAID5 6,5Tb)
- 2011 – digiteerimislahenduse täiendamine



Digihoodla

- Digihoodlat haldab Eesti Rahvusringhääling
- Tehniline lahendus:
 - Tivoli Storage Manage (IBM) *client-server* tarkvara
 - Lindirobot
 - LTO-4 Ultrium WORM tehnoloogia
 - Kettamassiiv
 - Teenusserverid
 - Teised seadmed
- Maht: 560Tb





Digital preservation repository



Digitaalne säilitamine

- Digitaalne aines on habras
- Pidev kontroll ja haldamine
- Kasutatavus on oluline pika-ajalise säilitamise osa
- Läbipaistvus
- Milline on tulevik?

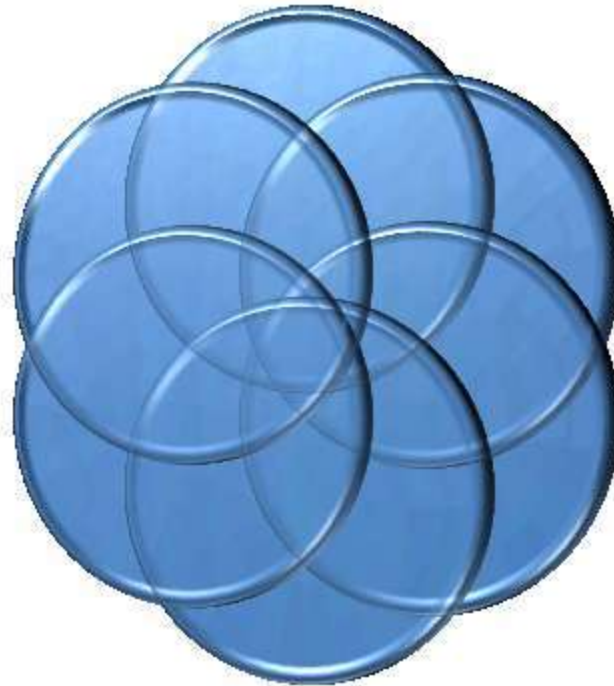




SISU

**PIKA-AJALINE
SÄILITAMINE**

**DIGITEERITUD
MATERJALID**



KASUTAJA

LIGIPÄÄS

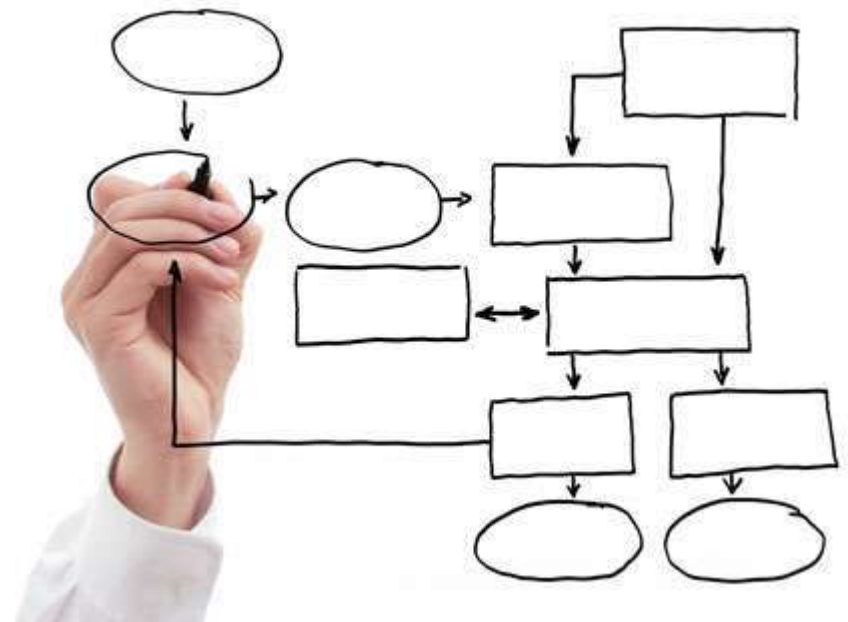
PÜSI ID



- Haldamine on (uus?) väljakutse
- Säilitamist peetakse lisaväärtuseks
- Vaja on:
 - oskusteavet
 - resursse
- Pika-ajalise säilitamisele peab lähenema:
 - süsteemselt
 - struktuurselt



- Kavandamine
- Toimiv töövoog
 - säilitamine – jagamine – jätkuvus





Checklist

- Millised ja kui palju on digitaalseid andmeid
- Teadmine, mida andmetega tehakse
- Eelistatud formaadi kriteeriumid
- Metaandmeid
- Kriitiline lähenemine
- Digitaalse säilitamise dokumenteerimine



veebi versioon

algne fail (Master copy)



800x800px@72 dpi

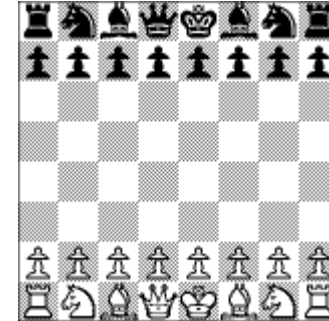


2896x3872px@300 dpi

KONRAD MÄGI (1878-1925) Daami portree, 1923-24, õli, lõuendil, 63,5x53,5 cm, Tartu Kunstimuuseum



- Emuleerimisstrateegia
- Migreerimisstrateegia
- Kombinatsioonid



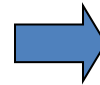
Käesolev hetk

Tarkvara
Metaandmed
Kujutised
Riistvara



Kavandamine

'wish list'



Testimine

Riistvara ja tarkvara
katsetamine
Andmete testimine
uues keskkonnas



Rakendamine

Säilitamise strateegiad



Emuleerimine

- KEEP – EL projekt



- Emulation Framework (version 1.1.0)
- open source (Apache v2 litsents)
- Vaba kasutus

- <http://emuframework.sf.net>
- <http://www.keep-project.eu>



Digitaalse ainese säilitamine peab olema organisatsioonis teadvustatud sh:

Kindlad põhimõtted

Kirjeldatud erandjuhud

Süsteemne kvaliteedi kontroll

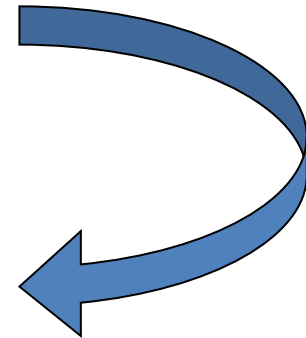
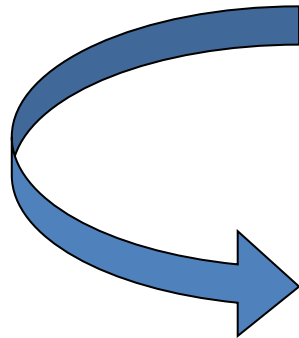
Ainult objekti digiteerimine pole pika-ajaline säilitamine

**Olenemata (andme-)mahtudest algab arhiveerimine
ning pikaajalise säilitamise tagamine
tööde kavandamise hetkel**



• formaat

○ maksumus



loomine
haldamine
töötlemine
kontroll
riskide hindamine
ligipääs



Metaandmete loomine

- Metaandmed (metadata) on mingeid andmeid kirjeldavad andmed ehk nii-öelda andmed andmete kohta.
- Failide puhul võivad metaandmeteks olla selleks nt: autor, loomise kuupäev, faili loomiseks kasutatud tarkvara jne. Selline info võib paikneda eraldi andmebaasis või ka failis sees

Metaandmed

- *Päritolu*: kes on digitaalse objekti eest vastutav/ omanik?
- *Ehtsus*: kas digitaalne objekt on see, mis tähendus sellel olema peab?
- *Säilitustegevused*: mida on tehtud, et antud digitaalset objekti säilitada?
- *Tehniline keskkond*: mida on vaja, et antud digitaalset objekti esitada, kasutada ja muuta?
- *Õiguste haldus*: milliseid intellektuaalse omandi õigusi peab antud digitaalse objekti puhul tähele panema?



Haldamine

- Kujutiste haldamise süsteemid on erinevad
- Rakenduse paindlikus
- Võimalused: muutmise, kopeerimine, liigutamine , (automaatne) ümbernimetamine, metaandmete lisamine, virtuaalsed kaustad + etc.



Vastutus

- Määrata ära vastutus:
 - Kellele kuulub digitaalne kogu?
 - Kes on vastutav süsteemi toimimise eest?
 - Milline on emuleerimise/migreerimise puhul otsustusprotsess, kes teeb otsuse?
 - Milline on rahastus?
 - Pika-ajalise säilitamisega seotud kulud?
 - + etc.



Turvalisus

- Eri tasemel ligipääsud
 - administraator, toimetaja/moderaator, kasutaja
 - individuaalsed õigused
- Digitaalsete õiguste haldus
- Garantiid pole





Kasutamine

- Koopiad
- Kui vanad on kasutatavad meediad?
- Võimalus salvestada uuemasse versiooni?
- Formaatide registrid, nt [PRONOM](#)



Ligipääs

- Edasine andmete töötlus
- Uue sisu loomine
- Läbipaistvus
- Ligipääs
- Mitmekeelsus
- Koostöö
 - Kasutajad
 - Mäluasutused
 - Teadusasutused
- Privaatsuse probleem
- Lingitud andmemasiivid



- 8” vöi 5 ¼” floppy disk – mis see on?
– IBM 23FD, 1971





Seotud andmed

- METS, TEI, EAD, OAI-PMH, MARC, z39.50, DC, ESE, CIDOC-CRM, LIDO, +etc.
- Kasutades standardeid ja/või parimaid praktikaid on võimalik siduda eri sisu kandvad andmemassiivid
- Institutsioonid, riigid, EL, UNESCO



LIDO

- CIDOC Conceptual Reference Model (CIDOC-CRM / ISO 21127) rakendusmudel
- LIDO (Lightweight Information Describing Objects)
 - SPECTRUM
 - CDWA Lite/museumdat
- Kogu kirjeldav informatsioon on struktureeritud ja standardiseeritud



Metaandmete ühtlustamine

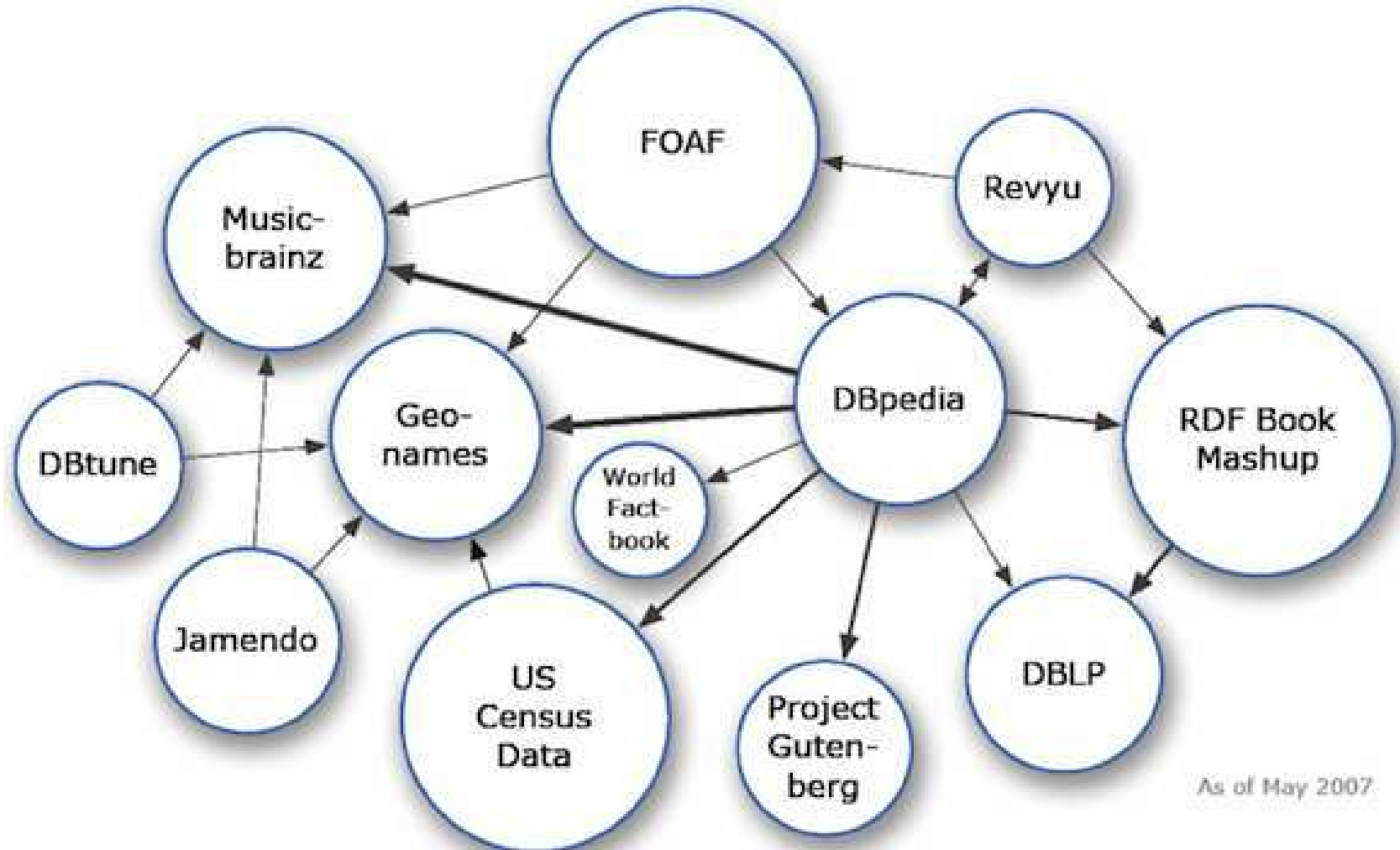
- Google-refine – *messy data*
 - <http://code.google.com/p/google-refine/>

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:foaf="http://xmlns.com/foaf/0.1/" xmlns:dc="http://purl.org/dc/elements/1.1/">
  <rdf:Description rdf:about="http://www.tmm.ee/?s=70155">
    <dc:title>Teatrifoto kunsti ja ajaloo jäädvustajana digisajandil
  </dc:title>
  <dc:publisher>Eesti Teatri- ja Muusikamuuseum
  </dc:publisher>
  <foaf:primaryTopic>
    <foaf:Event>
      <dc:title>Teatrifoto kunsti ja ajaloo jäädvustajana
      digisajandil</dc:title>
    </foaf:Event>
  </foaf:primaryTopic>
  </rdf:Description>
</rdf:RDF>
```

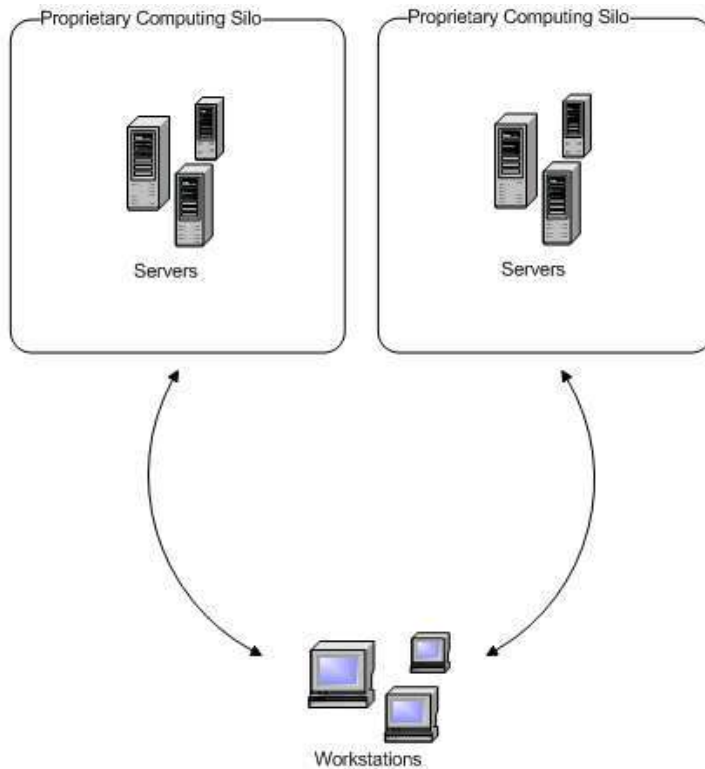




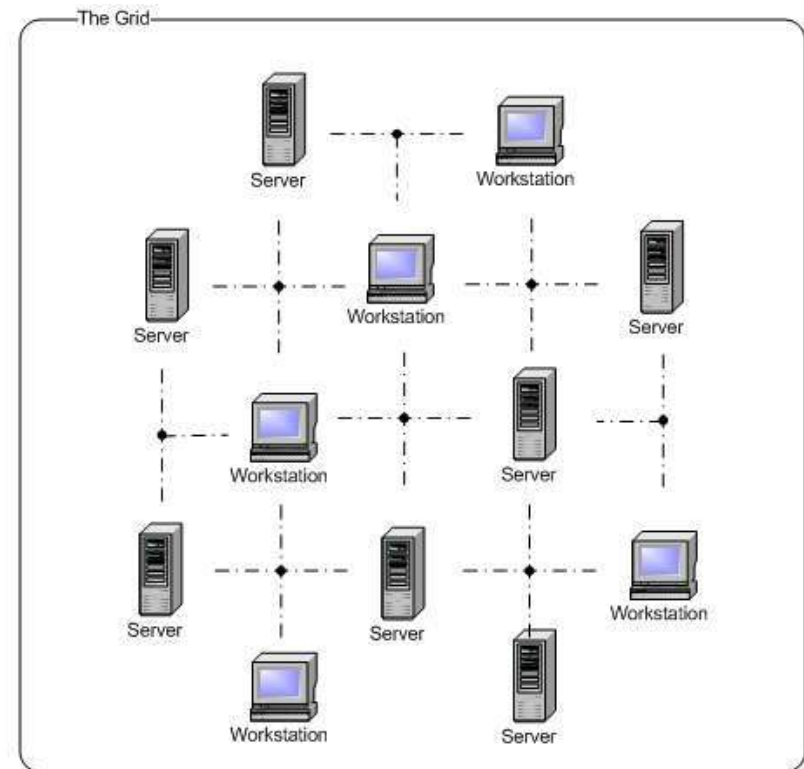
Linked Open Data (LOD), 2007



As of May 2007



Traditional n-tier architecture. Solution application servers operate within individual computing silos and balance tasks across the cluster. To improve the computing capacity of a solution domain, additional servers must be added to the computing silo.

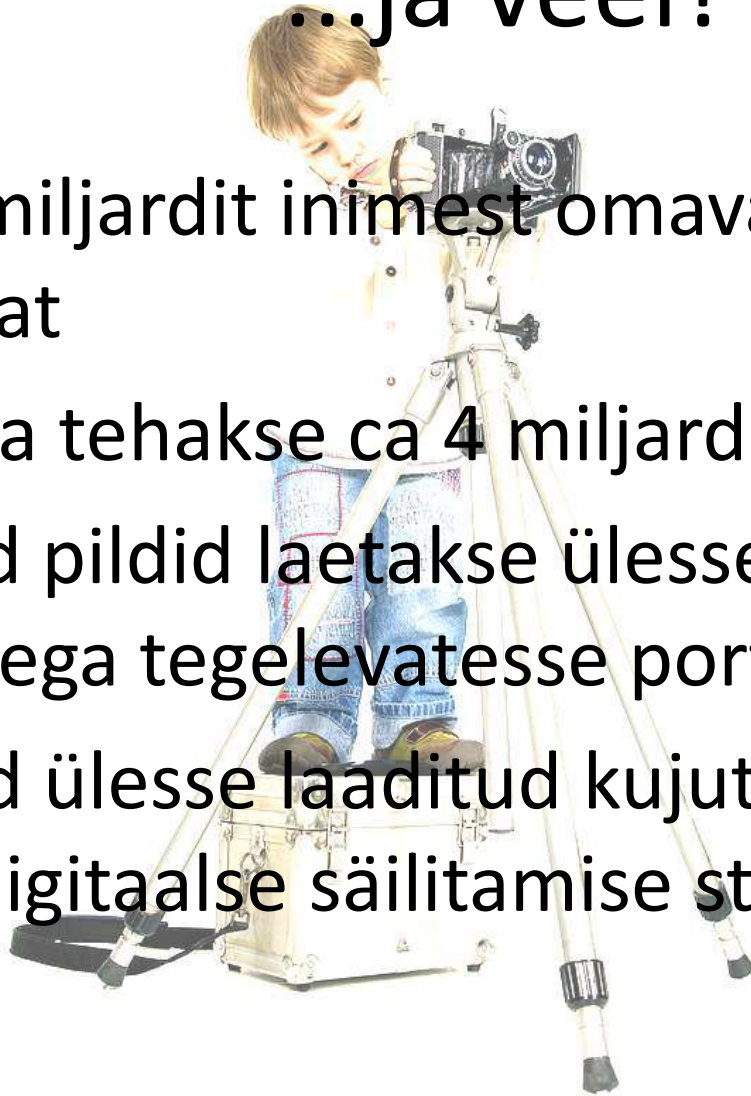


Grid Computing architecture. Computation is balanced across all nodes within the grid which span multiple solution domains and eliminate computing silos. To improve the computing capacity for a solution domain, nodes are configured to "listen" for computing tasks, or additional nodes can be added.



...ja veel?

- Ca 2,5 miljardit inimest omavad digitaalset kaamerat
- Iga aasta tehakse ca 4 miljardit pilti
- Miljonid pildid laetakse ülesse piltide jagamisega tegelevatesse portaalidesse
- Kas neid ülesse laaditud kujutisi saab võtta osana digitaalse säilitamise strateegiast?





- +
 - Laetakse ülesse parimad/lemmik pildid
 - Võimalik hilisem töötlus
 - Vahendid iga faili nimetamiseks, kirjeldamiseks ja tag'imiseks
- -
 - Online äri stabiilsus
 - Pole vastutust kadunud kujutiste puhul
 - Pole informatsiooni/ühtset poliitikat kuidas pilte säilitatakse, varundatakse
- Jeffrey Friedl's Online Metadata Viewer:
<http://regex.info/exif.cgi>



Säilitamine

- Vähemalt kaks koopiat
- Arvuti – offline meedia – välised kettad – pilv
- Säilitamine erinevates kohtades
- Turvalisuse kontrollimine
- Piltide valikuline kontroll vähemalt 1x aastas
- CD/DVD puhul regulaarne kontroll ja iga 5 aasta tagant uute koopiade loomine.



- Offline varukoopiad on ajutised
- Online hoidlad vajavad pidevat tööd



Kokkuvõte

- Tehnoloogia vananemine ja digitaalsete andmete haprus
 - Afganistani sõda (2001-...) vs. Krimmi sõda (1853-56)
- Digitaalse informatsiooni omadus – kerge andmete kopeerimine ja muutmine on paralleelne sellega, et andmeid on ka kerge rikkuda või hävitada



- Oluline on pöörata digiteerimise planeerimise faasis tähelepanu ka pika-ajalise säilitamise küsimustele
- Digitaalse ainese online hoidmine keskkondades nagu Google Docs, Slideshare, Flickr, vms eeldab, et need materjalid on säilitatud nendest keskkondadest ka väljaspool



Vastavalt olemasolevatele ressurssidele tuleb pika-ajalise säilitamisega tegeleda pidevalt ja süsteemselt.



- http://en.wikipedia.org/wiki/Digital_preservation
- <http://www.digitalpreservation.gov/>
- <http://www.digitizationguidelines.gov/>

if Küsimused?

then

vastan

proovin vastata

else

TÄNAN TEID!

fi

Andres Uueni

andres.uueni@kanut.ee

Digiteerimise ja dokumenteerimise osakonna juhataja

Ennistuskoda Kanut, 2011