Digitaalse säilitamise teenused ja töövahendid: DC-Net uuring

Raivo Ruusalepp
Eesti Äriarhiiv
12.01.2012

Eesti ÄriARHIIV
ESTONIAN BUSINESS ARCHIVES

Teemad

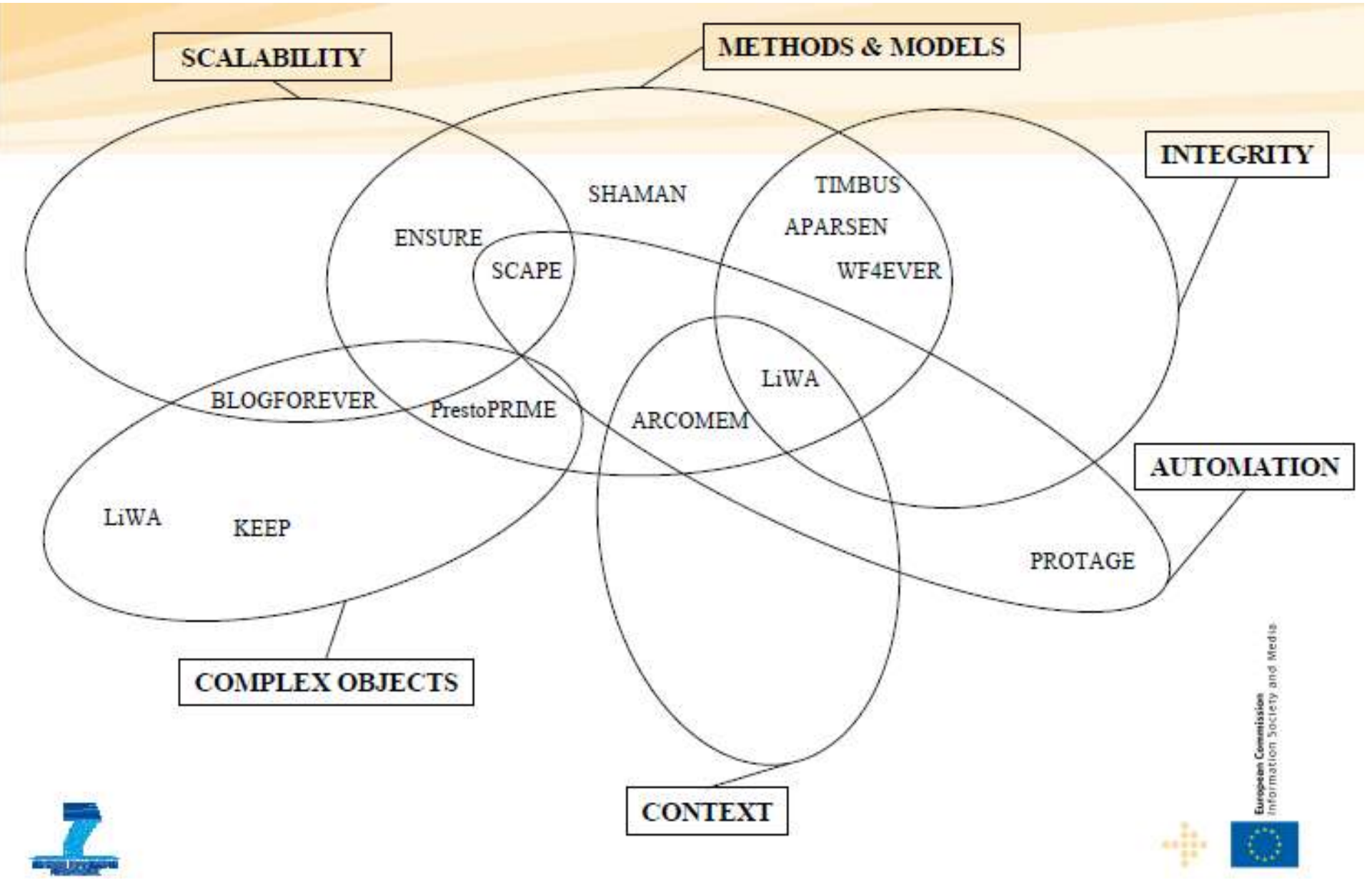
- Digitaalse säilitamise mudelid
- Uued suunad e-teenuste osutamisel
- Ülevaade digitaalse säilitamise teenustest
- Järeldused

I. Sissejuhatuseks

Digitaalne säilitamine on probleem

- Tehnoloogia uuenemine / iganemine põhjustab raskusi vanema digitaalse info kasutamisel ja mõistmisel
- See põhjustab usaldamatust digitaalsel kujul info vastu, mis ei ole kuigi ratsionaalne
- Euroopa Komisjon on alates 2006 a. rahastanud 15 uurimisprojekti €86 miljoniga, et uurida digitaalse säilitamise probleemi olemust ja leida lahendusi selle erinevatele aspektidele
- Kevadises ICT Call 9 voorus lisandub veel €30M

Peamised uurimussuunad



Digitaalne säilitamine

Digital preservation (DPE):

- a set of activities required to make sure digital objects can be located, rendered, used and understood in the future

Digital curation (DCC):

- maintaining, preserving and adding value to digital research data throughout its lifecycle

Digitaalne säilitamine jaguneb:

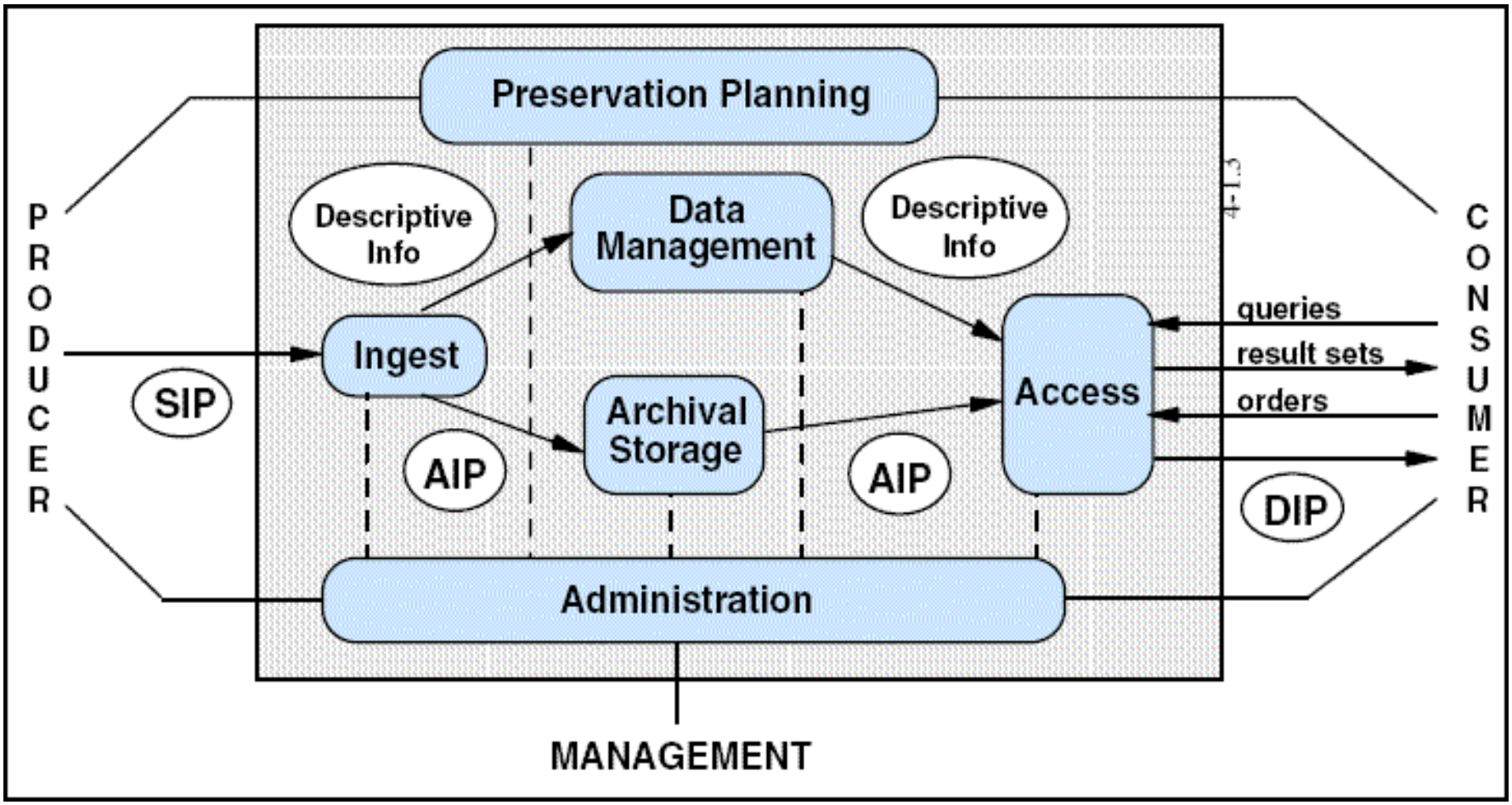
- Passiivne digitaalne säilitamine – sarnane analoog-materjali säilitamisega, ehk bittide alles püsimise tagamine
- Aktiivne digitaalne säilitamine – sisu loetavuse ja mõistetavuse tagamine läbi digitaalse säilitamise tehnikate (migreerimine, emuleerimine) rakendamise

II. Digitaalse säilitamise senine mudel

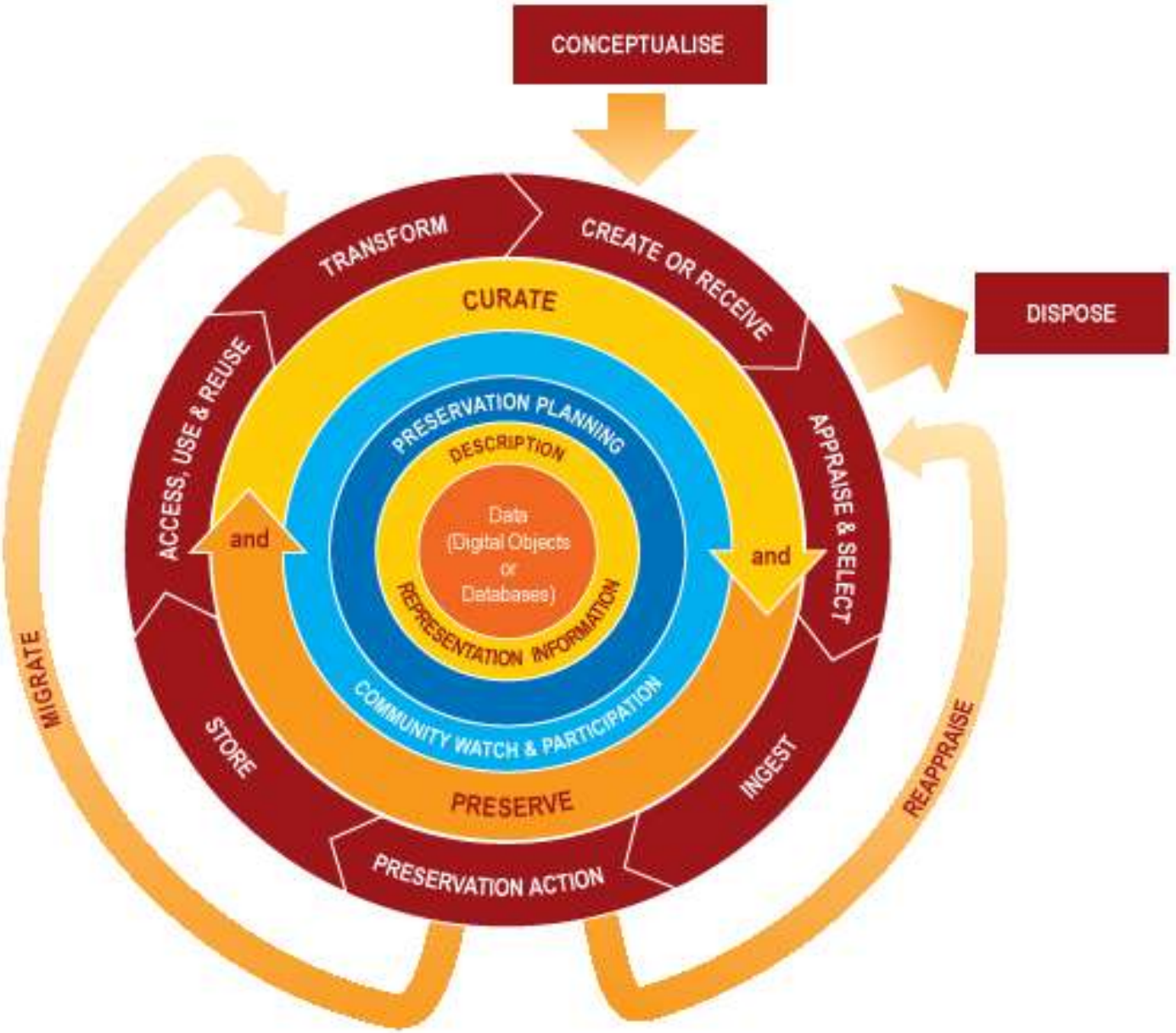
Hoidlakeskne säilitamine

- Senini on püsinud arusaam, et digitaalne säilitamine eeldab digitaalhoidlat
- Digitaalhoidlaid on mitmesuguseid, aga aktiivse säilitamise komponent on kõigis sarnane või samane
- Mõtteviisi juurdumist on toetanud digitaalarhiivi arhitektuuri mudel OAIS (ISO 14721:2003) ja arvukad digitaalhoidla tarkvaratooted, milles tegelik digitaalse säilitamise tugi on minimaalne või olematu

OAIS funktsionaalne mudel



DCC Digital Curation Lifecycle Model



„Igaühel oma“ lahenduse nõrkused

Pikaajalise digitaalse säilitamise mudel, kus iga (mälu)asutus arendab välja omaenda digitaalse säilitamise tarkvaralahenduse, ei ole:

- Majanduslikult mõistlik
- Koosvõimeline
- Jätkusuutlik (pikemas perspektiivis)

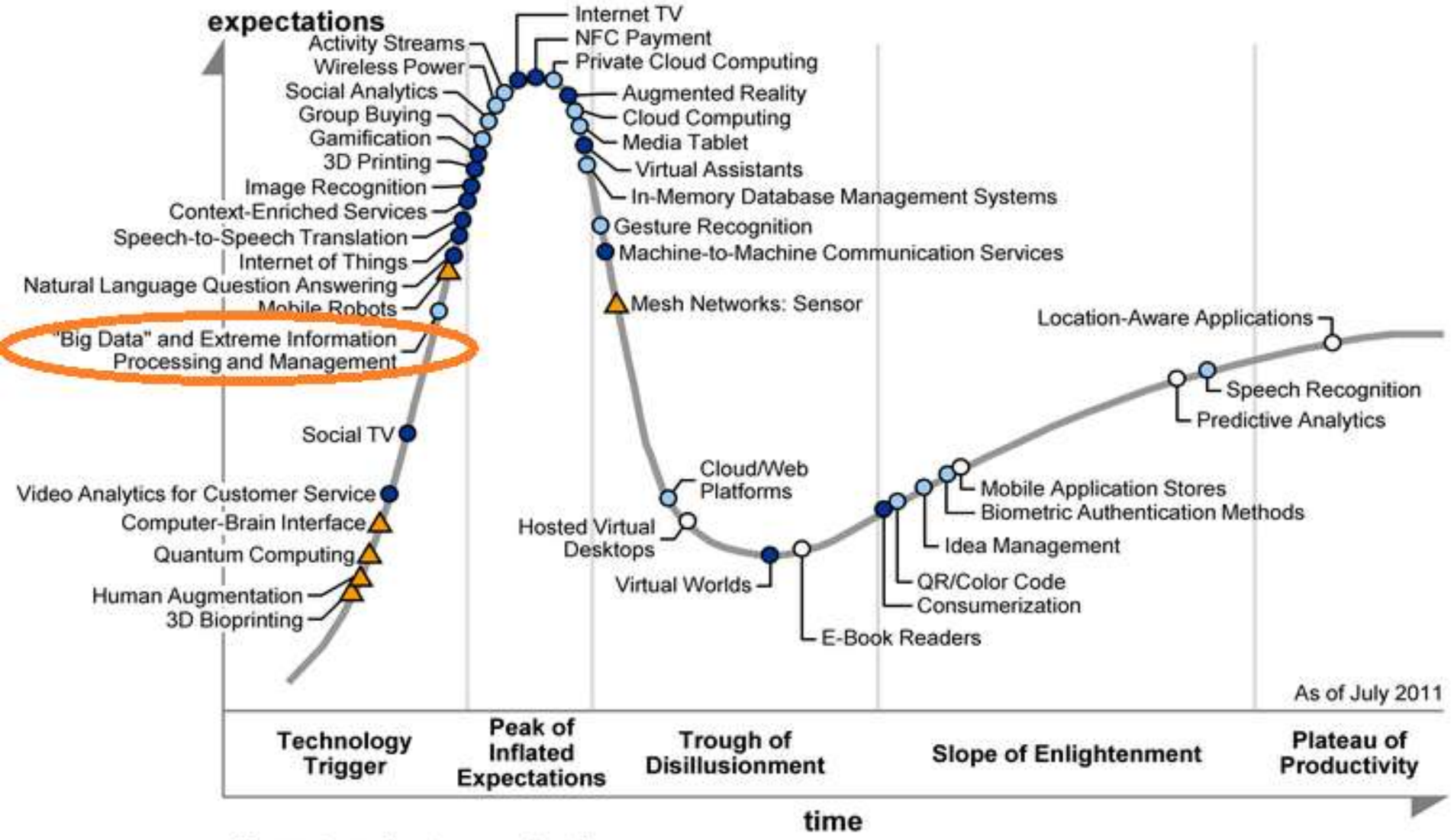
- Liigne fragmenteeritus ei taga kokkuhoidu...

III. Uued võimalused digitaalse säilitamise teostamiseks

Trendid IT maailmast

- Teenusepõhine arhitektuur (SoA)
  - API (Application Programming Interfaces)
  - Veebiteenused (web services)
- Teenusplatvormid
  - GRID
  - Cloud
  - SaaS (software as a service)
  - PaaS (platform as a service)
  - IaaS (infrastructure as a service)
- Mikroteenused (microservices)

Gartner Hype Cycle of Emerging Technologies (July 2011)

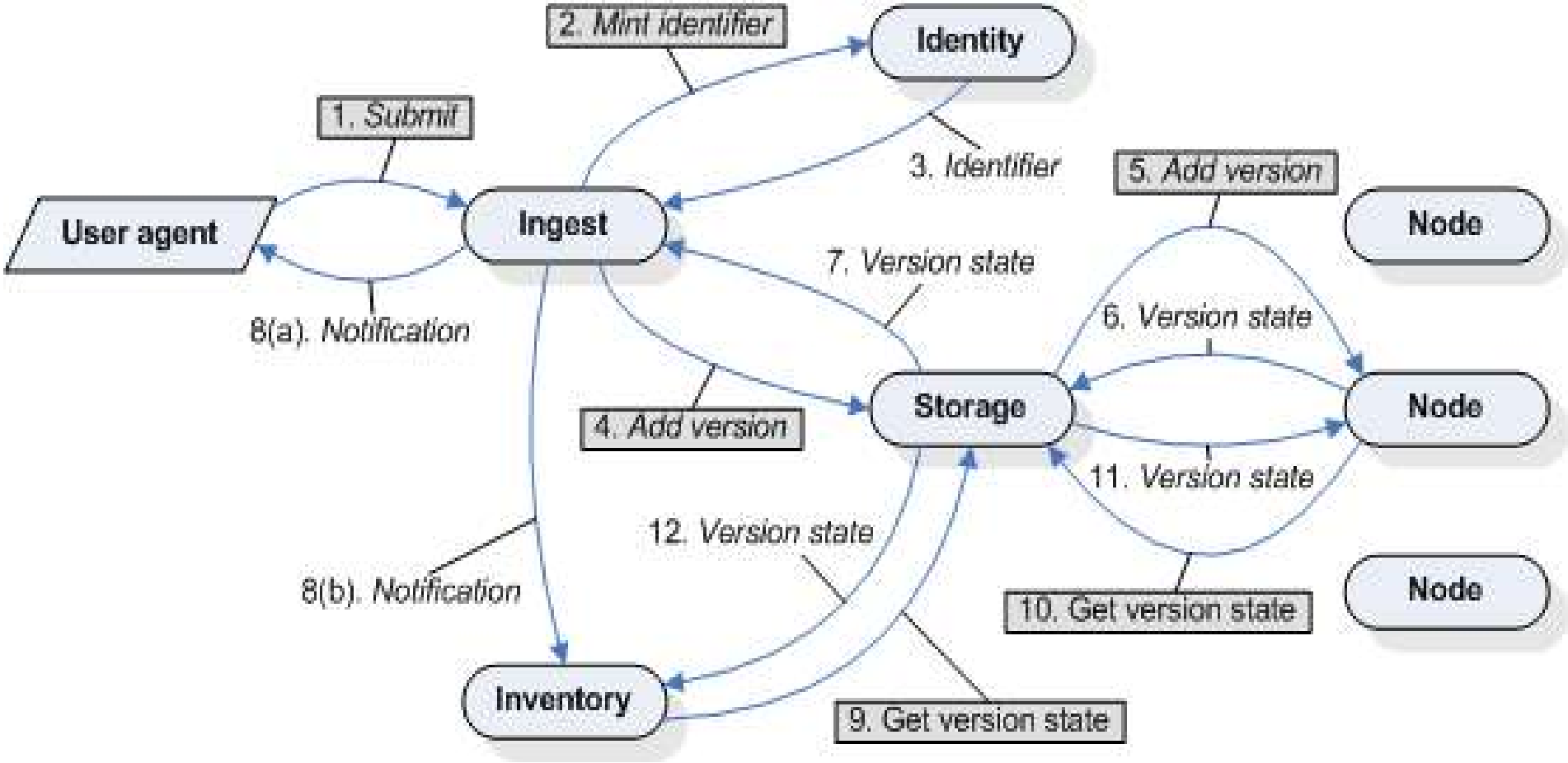


Mikroteenused digitaalses säilitamises

- Micro-services are an approach to digital curation based on devolving curation function into a set of independent, but interoperable, services that embody curation values and strategies. Since each of the services is small and self-contained, they are collectively easier to develop, deploy, maintain, and enhance. Equally as important, they are more easily replaced when they have outlived their usefulness. Although the individual services are narrowly scoped, the complex function needed for effective curation emerges from the strategic combination of individual services.

[Curation Micro Services website]

Mikroteenuste näide



UC Curation Center (UC3)

Trendid digitaalses säilitamises

- Säilitamisvalmidusega süsteemid – preservation-ready systems (Borbinha 2010)
- Säilitamise funktsionaalsuse integreerimine erinevatesse infosüsteemidesse
- Ise-säiluvad objektid – self-preserving objects; durable objects (Billenness 2011)
- Ühised süsteemid ja hoidlad digitaalse arhiveerimise ja säilitamise korraldamiseks (LOCKSS, MetaArchive, DISTARNET)
- Pilvetehnoloogia kasutamine (Askhoj, Nagamori & Sugimoto 2011)

Digitaalse arhiivi toimimismudelid

- *Tarkvara mudel* – säilitamine on integreeritud digitaalhoidla tarkvarasse
- *Institutsionaalne mudel* – organisatsioonil on mitmeid erinevaid digihoidlaid, millest üks on kujundatud säilitamise jaoks
- *Hajamudel* – erinevad organisatsioonid haldavad ühte hajutatud digitaalhoidlat, millel on keskne säilitamise moodul
- *Võrgumudel* – põhineb algselt Storage Resource Broker (SRB) ja seejärel iRODS lahenduse ning GRID ja pilvetehnoloogiatel
- *Teenusepakkuja mudel* – kasutatakse teenusepakkuja teenuseid; teenusepakkujaks võib olla ka GRID teenuse pakkuja

- Ilmselt kõik need mudelid kasutavad aktiivse digitaalse säilitamise teostamiseks mikroteenuseid

IV. Digitaalse säilitamise töövahendid

Mõisted

Services

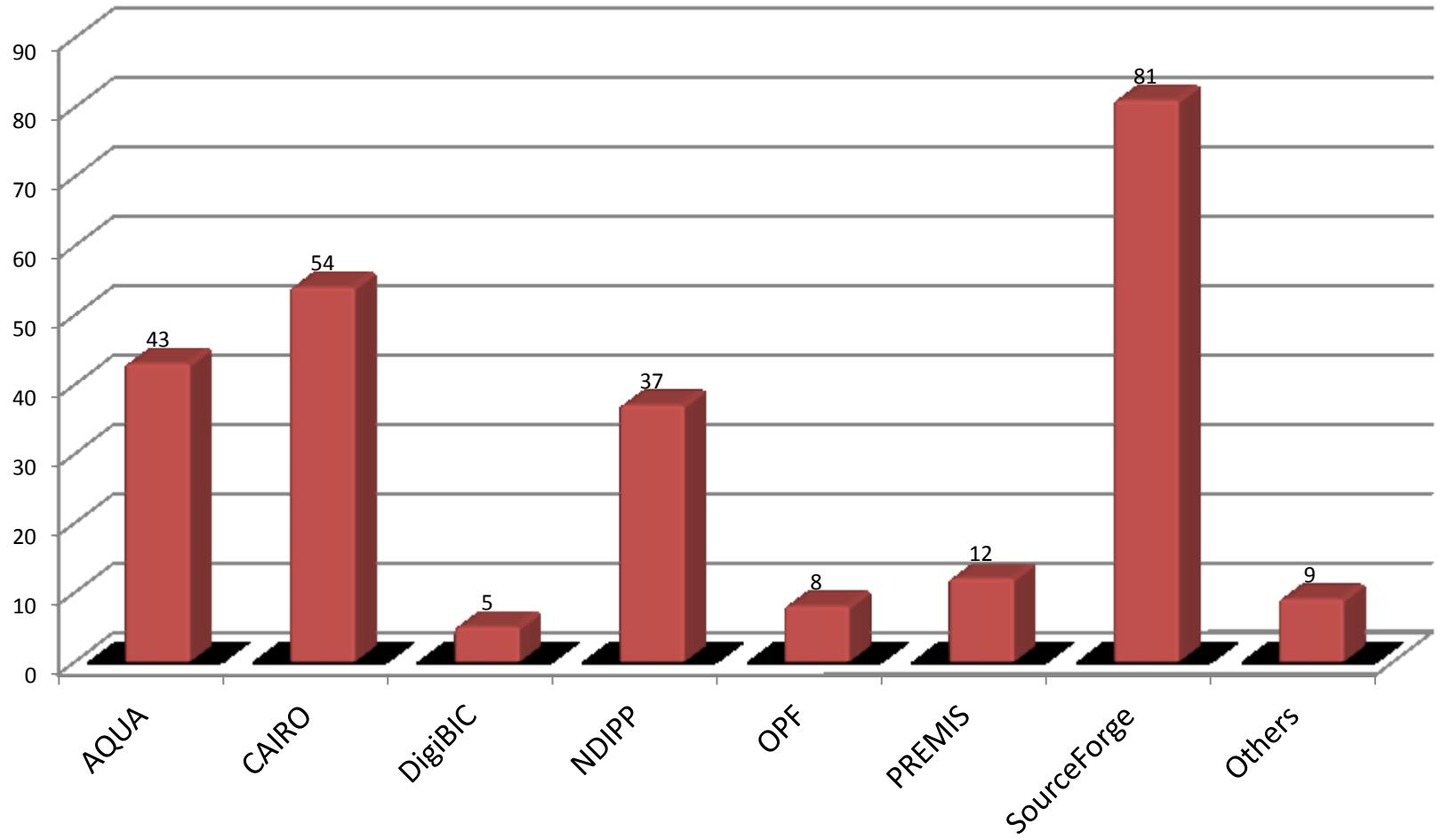
- unitary software components for which there is a body which offers customer support, there are clearly defined access conditions, and user documentation is available.

Tools

- Tarkvaravahendid, mis on välja töötatud digitaalse säilitamise jaoks ja on kättesaadavad

Allikad

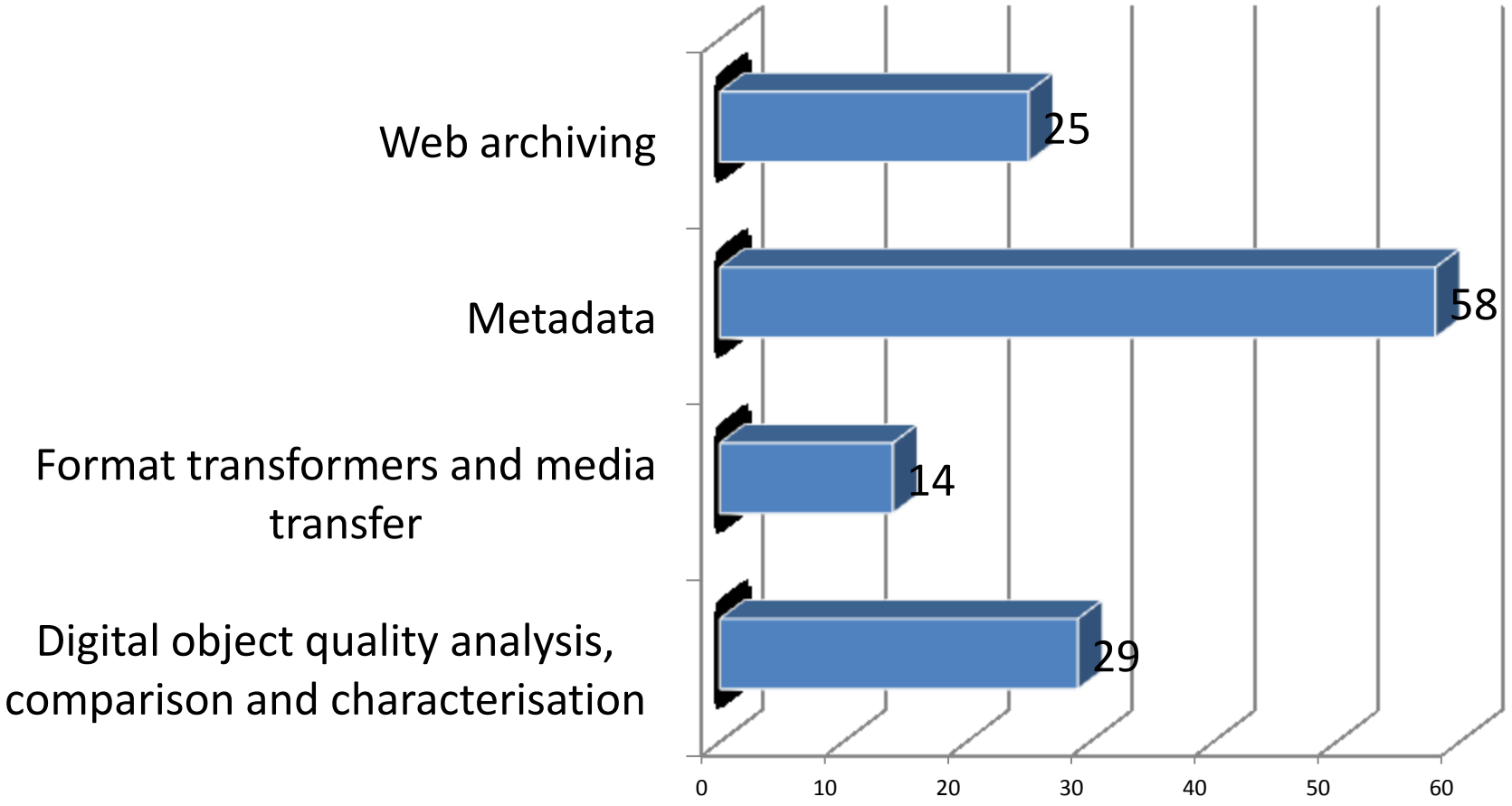
- Kokku 191 töövahendit ja teenust



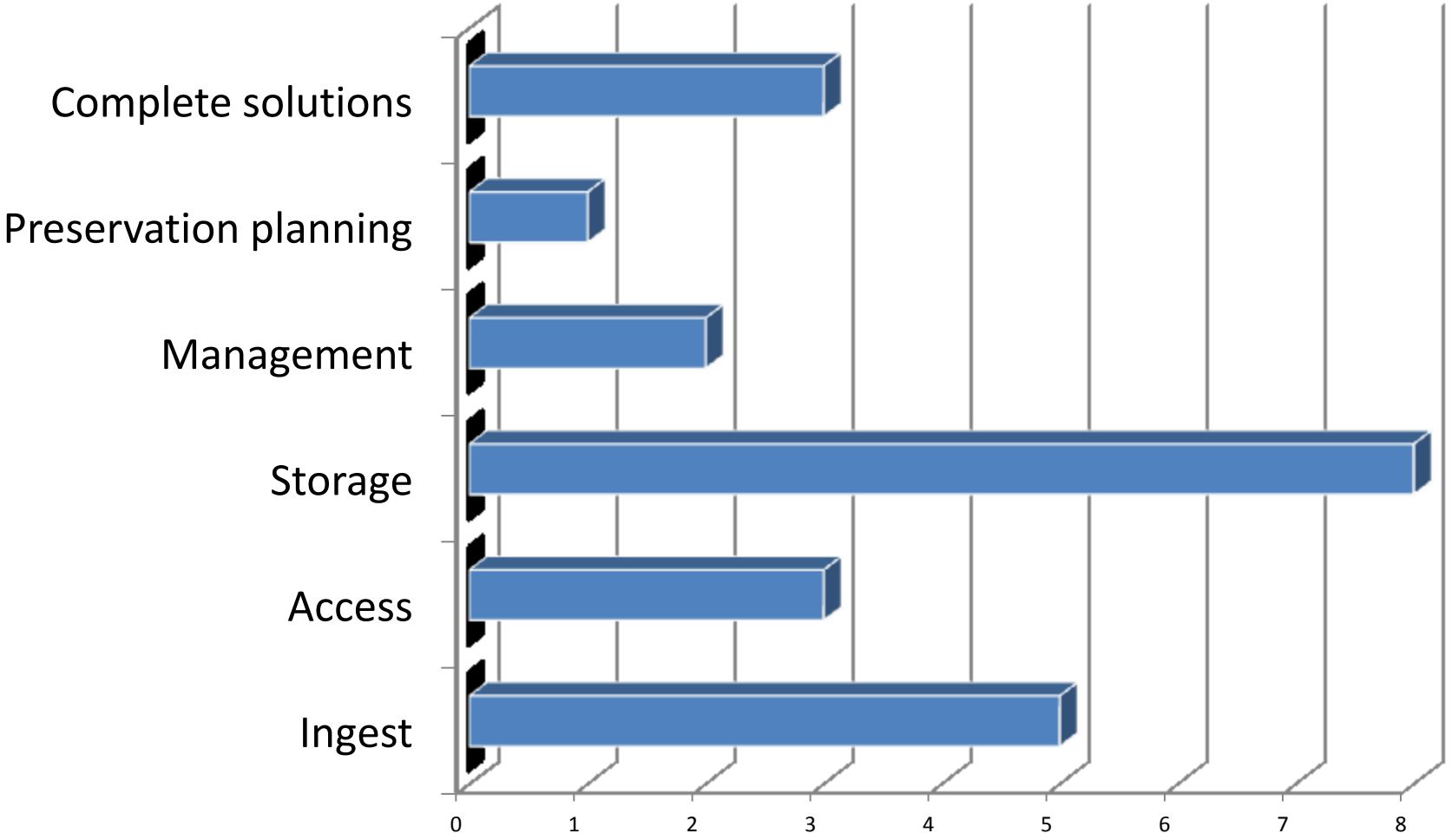
Teenuste jaotumine elukäigule

- Digitaalse säilitamise protsessid (nt. OAIS funktsionaalsed olemid) ja saadaolevad tarkvaravahendid ei kattu päris üks-ühele
- Töövahendid võivad olla mõeldud konkreetse organisatsiooni digihoidla jaoks (nt. TIAMAT ingest) või globaalsed (PERPS monitoring tool)

Populaarsemad tegevused



OAIS funktsionaalsete olemite kaetus



Ingest teenuste näited

Name	Description	Creator
Characterize and extract metadata	Microservice in Archivematica. Identifies and validates formats and extracts object metadata using the File Information Tool Set (FITS). Adds output to the PREMIS files.	Project consortium (Artefactual systems, Inc.: UNESCO, et al.)
Merritt Characterization service	Provides a mechanism for the automated examination of digital objects to determine their significant properties. Addresses four aspects: Identification, Validation, Feature extraction, and Assessment. Based on JHOVE.	University of California, CDL
Merritt Fixity Service	Service checking for file integrity and corruption and related to authenticity. The Fixity service verifies the bit-level integrity by testing two values: filesize and message digest (such as an MD5 checksum).	University of California, CDL
msiGetDataObjAIP	iRODS microservice that gets the AIP of a data object in XML format	iRODS
msiGuessDataType	iRODS microservice which guesses the data type of an object based on its file extension.	iRODS

Ingest

- Erinevaid töövahendeid, mida kasutatakse digitaalse säilitamise elukäigu erinevates etappides:
  - Failivormingu tuvastamine
  - Metaandmete eraldamine
  - Digitaalse objekti iseloomustamine (object characterisation)
- Terviklikud *ingest*-vahendid on tüüpiliselt saadaval mõne konkreetse digitaalhoidla tarkvara jaoks (nt. DSpace, Fedora)

Archival storage

- Arhitektuurselt on lahendused tihti üsna keerukad, kuna lisaks sisuobjektidele on vaja hoida ka kirjeldust nende kohta
- Põhiprobleemideks on suur objektide arv ja hoitava ainese kettamaht
- Virtualiseerimine ja pilve tüüpi teenused on ilmselt paratamatud, kuna andmemahud kasvavad väga kiiresti
- Selgelt defineeritavaid töövahendeid pole palju (8)

Säilitamise planeerimine

- PLATO on domineeriv töövahend ja enamus arendustööd keskendub selle ümber
- Digitaalsete objektide keerukuse kasv, eriti kokku-lingitud objektid, muudavad digitaalse säilitamise planeerimise keerukamaks
- Säilitamise planeerimine on inkrementaalne tegevus – migreerimised järgnevad üksteisele ja teadmised varasemate sammude kohta on vajalikud järgmiste sammude tegemisel

V. Uuringu järeldused

Järeldused

- Enamus seniseid töövahendeid on välja töötatud arhiivide ja raamatukogude poolt, samas kui muuseumid ei ole pea üldse selles töös osalenud
- Väljatöötatud töövahendite sisu ja eesmärgid on tihti kattuvad – erinevad organisatsioonid on teinud dubleerivat tööd
- Töövahendite ja mikroteenuste tasand (*granularity*) vajab selgemat piiritlemist, et tagada kõigi vajalike funktsionaalsuste olemasolu ja vältida infokadu objektide töötlemise käigus

Järeldused

- GRID ja pilvetehnoloogiat kasutatakse juba humanitaaria teadusandmete hoidmiseks ja säilitamiseks, seega sobivad need kindlasti ka mäluasutuste jaoks
- Praegune seis digitaalse säilitamise „tarkvaraturul“ on üsna killustunud, domineerivad vabavaralised „vidinad“, millel puudub arvestatav kasutajatugi ja dokumentatsioon
- Teenuste turg on väga algeline; turu nõudluse hindamine on raske, kuna säilitamist teostatakse erinevatel eesmärkidel
- Töövahendite võrdlemiseks on väga vähe vahendeid – kuidas otsustada, millist vahendit eelistada või usaldada?

Küsimused?

raivo@eba.ee

Eesti ÄriARHIIV
ESTONIAN BUSINESS ARCHIVES