



Tutkimuksen tietoineistot



TTA

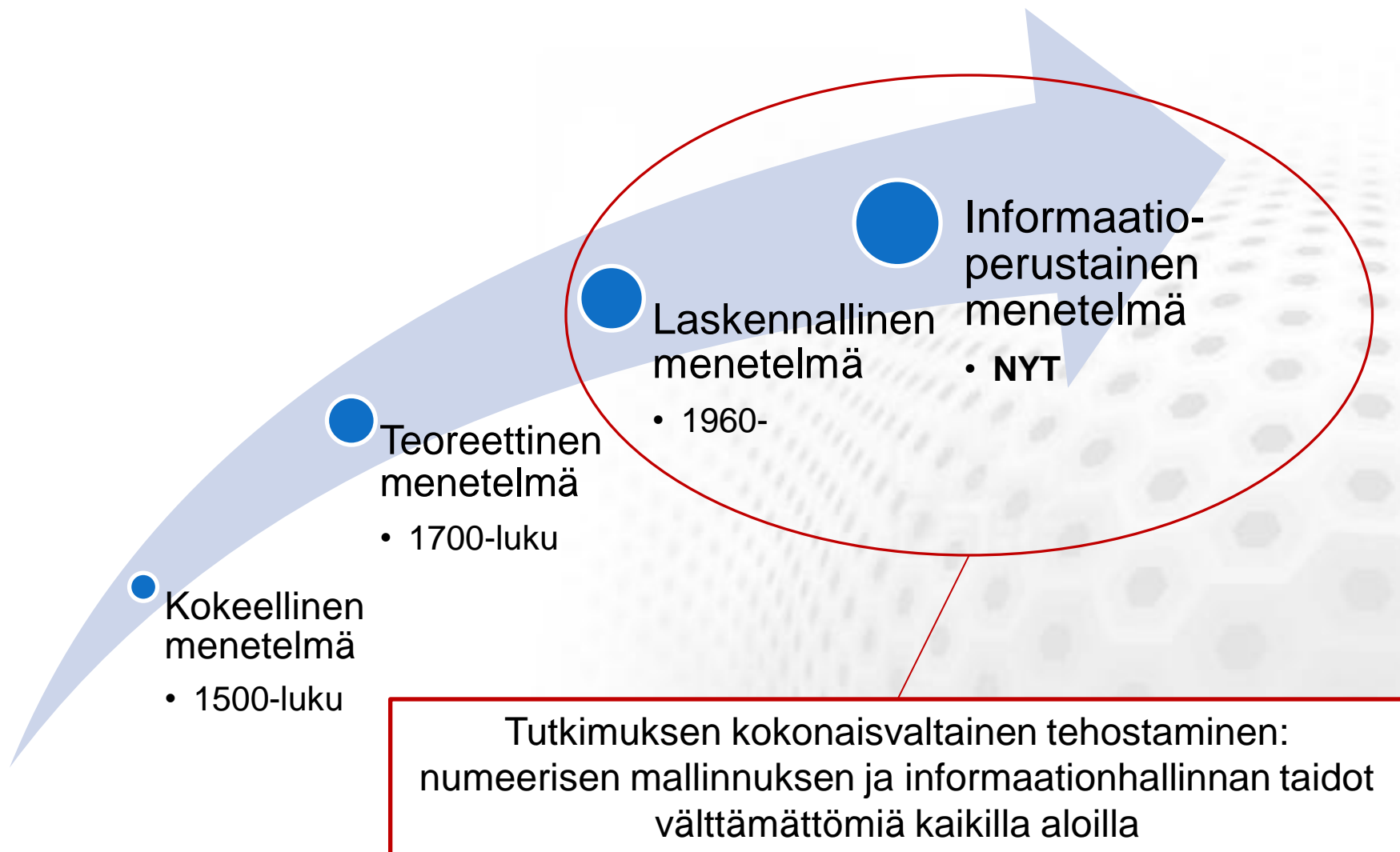
TUTKIMUKSEN TIETOAINIESTOT

Tutkimuksen trendejä

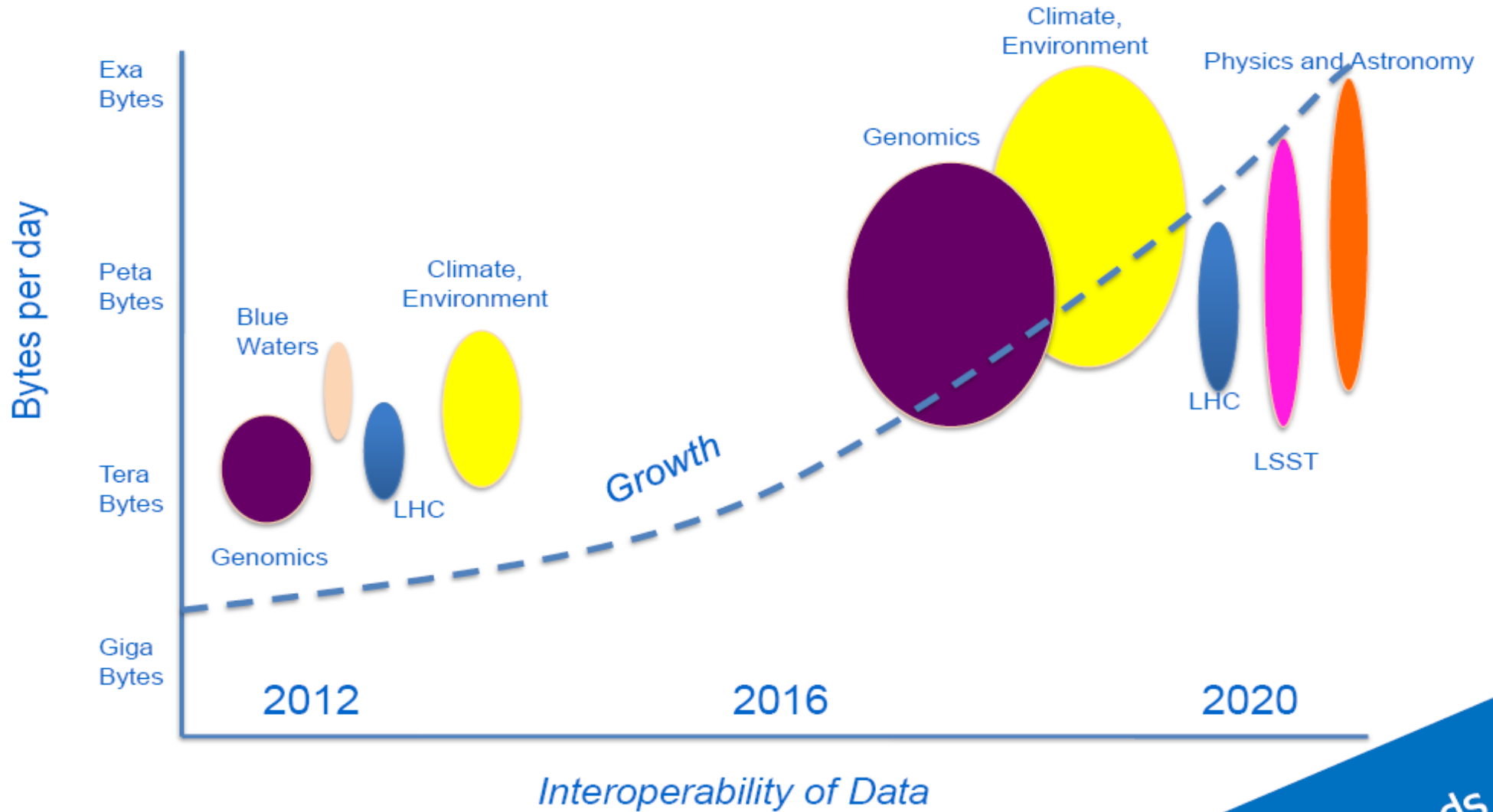


- Yksilövetoisuudesta tutkimusryhmä/organisaatiotasolle
- Moniosaajien yhteistyö, joukkoistaminen
- Avoin tiede
- Yhteiskunnallinen hyöty
- Kustannustehokkuus
- Tietomassojen kasvu

Neljäs paradigma



Data Challenges



- **Miten selviämme valtavan tietovirran kanssa?**
 - tuottaminen, kerääminen, siirtäminen, ...
- **Miten löydämme tarvitsemamme tiedon?**
 - yhteentoimivuus, sisällön ymmärrys, yhteistyörakenteet, ...
- **Miten osaamme päättää mitä säilyttää, mitä tuhota?**
 - rakenteet päätöksenteolle, rahoitusratkaisut
- **Miten turvaamme oleellisen tiedon säilymisen?**
 - resursointi, osaaminen, tekniset ratkaisut, laadun varmistus,
...



- Tutkimuksen tuottamat tietoaineistot käytettävissä
- Avoin pääsy tutkimuksen tuottamiin julkaisuihin
- Mahdollisuus käyttää tutkimusmenetelmiä joilla aineisto tuotettu
- Avoin infrastruktuuri jatkokäyttöä varten

Avoin data ei ole arvo yksinään - Avoin tiede on enemmän.

Tehokkaseen tutkimukselliseen tiedonvaihtoon tarvitaan älykästä avoimuutta.

Datan tulee olla:

- Saatavilla
- Ymmärrettävää
- Arvioitavissa
- Uudelleen käytettävissä



METATIETO

Vasta kun nämä neljä ehtoa täyttyvät, on data kunnolla avointa.
Metatiedon täytyy olla käyttäjäherkkää.

- Systemaattinen hallinta oleellista jotta tieto ja sen evidenssiketju voidaan turvata
 - Tukee tutkimuksen ydintoimintoja
 - Vaikuttaa moniin tutkimusprosessin vaiheisiin
 - mahdollistaa luotettavuuden arvioinnin, ja siten tuo uskottavuutta tutkijalle, tutkimusryhmälle, organisaatiolle, rahoittajille



Antaa kaikkien kukkien kukkia

Liian paljon liian nopeasti

Haasteet

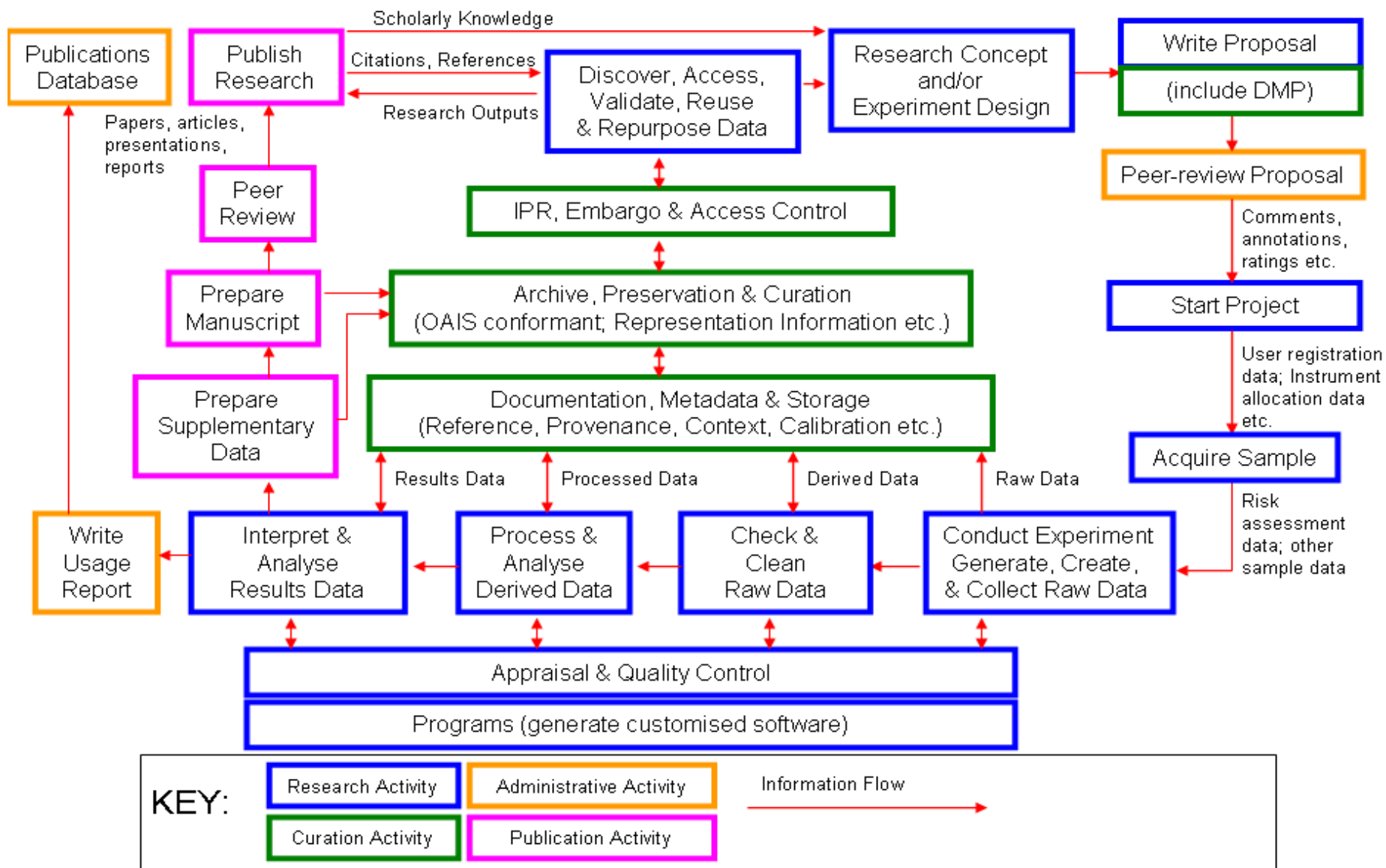
- Tieteen itsensä korjaamisen tukeminen (Ylitetään käsite – data kuilu)
- Vastaaminen kansalaisten vaatimukseen ymmärrettävästä tutkimuksesta.

Mahdollisuudet

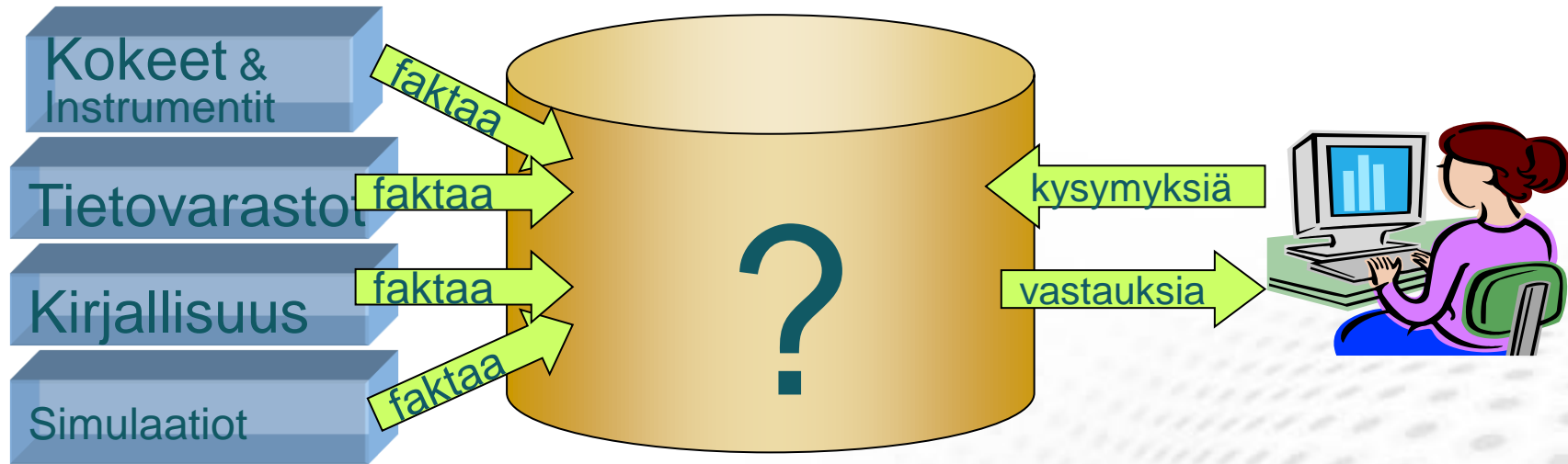
- Tieteen neljännen paradigman realisointi
- Linkitetyn datan potentiaali
- “Data is the new raw material for business”
- Tieteellisten petosten ja väärennösten löytäminen
- Kansalaistieteen stimuloitni
- Toinen “avoimen tieteen” vallankumous?

- Koulutuksella uutta osaamista
 - **Tutkimusprosessi muuttuu** dramaattisesti digitalisoitumisen myötä (niin sanottu e-Tiede)
 - Avoimet toimintamallit, avoin data, avoin julkaiseminen ym. kannustavat **raja-aitojen ylittämiseen**
- Tutkimus tuottaa arvokasta uutta tietoa
 - **Tutkittu tieto** on yhteiskunnan toimivuuden kannalta kriittistä
 - Tieto käyttöön! - tietovarantojen potentiaalia pitää pystyä **hyödyntämään**
- Digitaalisen infrastruktuurin kehittyminen
 - Digitalisoituvan maailman perustana toimiva **tietoinfrastruktuuri** mahdollistaa uusien asioiden tapahtumisen
 - Verkostojen voimasta saadaan enemmän hyötyä kun **pääsy tietoon** voidaan taata
 - **Tutkimuksen tietoaaineistot** tulee saada talteen ja hyötykäyttöön

Research Process



eTutkijan ongelma



Datan koostaminen
Petataavumäärien hallinta
Yhteiset skeemat
Miten organisoida?
Miten uudelleenorganisoida?

Rinnakkaiselo ja yhteistyö muiden tutkijoiden kanssa?

Kyselyjen ja visualisoinnin työkalut?

Tuki ja koulutus

Suorituskyky

Vastaukset kyselyihin riittävän nopeasti
Erittäin suurten massojen analysointi ja hallinta

Ongelman ulottuvuuksia

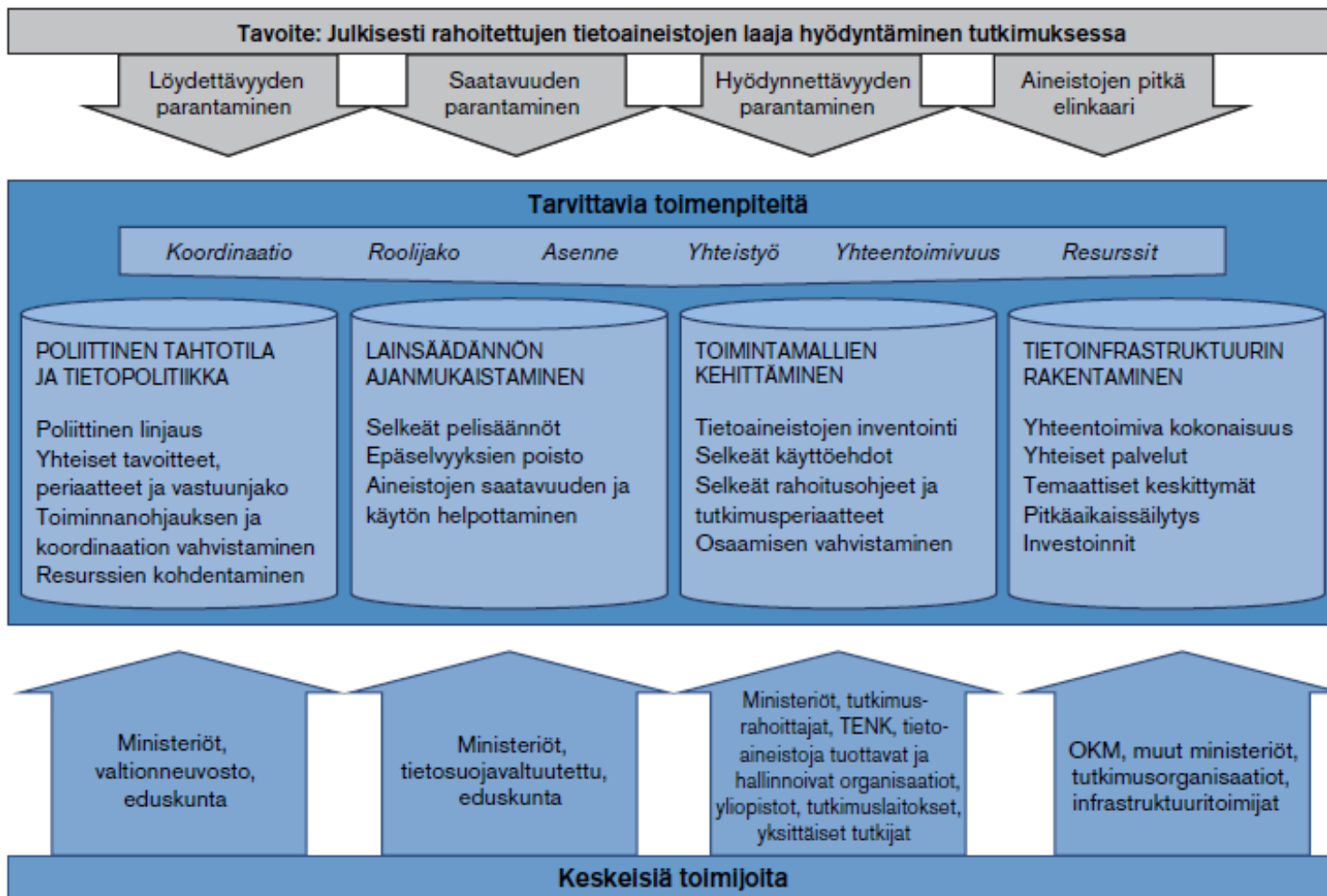


Kyvykkyydet datahallintaan

14 kategoriaa

Dataset and workflow discovery	Data access services	Tools to Probe, Validate, Verify, and Visualize Data	Data security and trust
Workflow execution management	Data management within workflows	Metadata for workflow and data sets	
Numerical methods and software engineering	Modeling standards and frameworks	Modeling capabilities within Cloud, Grid, HPC, and Science Portals	
Policy enforcement processes	Best practices & Governance Models for definitions & standards	Broad Participation: NGO, industry, domain partners, international	Continuity, Sustainability, & Evolution

Toimenpiteitä tietoaineistojen hyödyntämisen parantamiseksi



Mistä puhumme kun puhumme digitaalisesta tiedosta?



Miten tieto tuotettu?
Mitä varten tieto tuotettu?

- Koko
- Rakenne
- Elinikä

Tiedon pyramidi



Tietoaineistojen käyttötapoja



TTA-hanke edistää tietoineistojen **kuvausten yhtenäistämistä, säilytystä ja käyttöä.**

TTA-hankkeessa tuotetaan tätä tukemaan:

- *tutkimuksen tietoineistojen prosessikartta*
- *tietoinfrastruktuurin palvelukokonaisuuden suunnittelu*
- *metatietomalli tutkimustiedon hallinnalle*
- *yhteinen metatietokatalogi/hakupalvelu tutkimusaineistoille*
- *tallennuspalvelu tutkimuksen tietoineistoille*
- *yhteisen pitkäaikaissäilytysratkaisun valmistelu*

TTA-palvelukokonaisuus edistää tutkimuksen tietoineistojen kansallista, eurooppalaista ja kansainvälistä **yhteentoimivuutta.**



Koordinaatioryhmä

Ohjausryhmä

Metatieto-
työryhmä

Tietoinfrastrukt
uuri-työryhmä

TURE-
pienryhmä

PAS

Tiedon
saatavuus-
työryhmä

TUTKIMUS-IDA

DATAKATALOGI

DATANHALLINNAN
PILOTTI

- Keskitettyä tietoa tutkimuksen tietoaaineistoista: helpompi löytää, helpompi käyttää
- Yhtenäisempiä käytäntöjä aineistojen hallintaan
- Yhteentoimivuuden lisääntyminen: metatietomalli, rajapinnat
- Monipuolisen palvelukokonaisuuden kehitys
- Tietoaaineistojen säilymisen turvaaminen



Tier 1 – International data services

Tier 2 – National data services

Tier 3 – Institutions (Universities & Institutes)

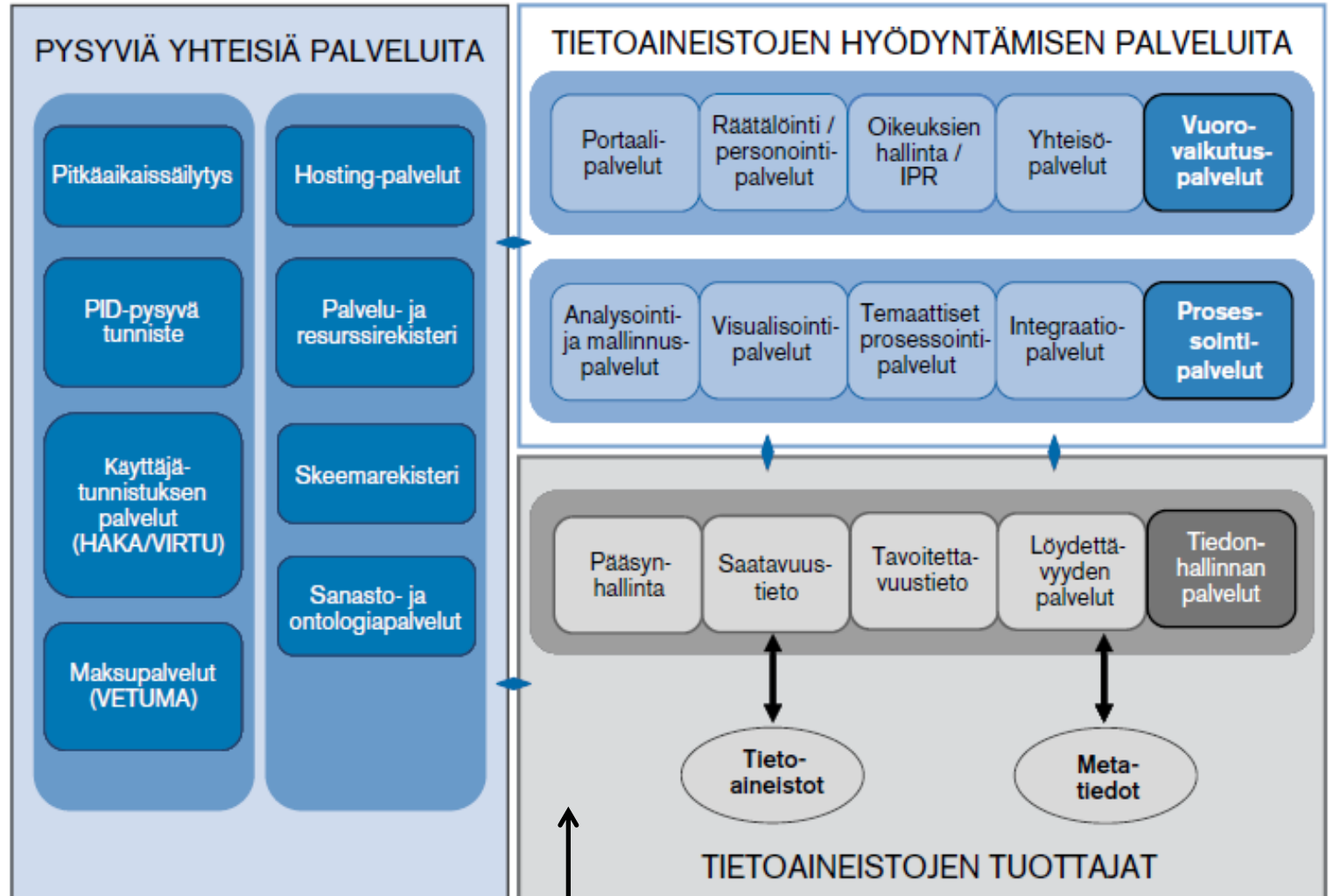
Tier 4 – “Small science” researchers & research groups

Tutkimuksen tietoaineistopyramidi



Laadullinen hyöty	Merkittävyys	Aikajänne	Todennäköisyys	Hyötyindeksi
Hyödyntävät organisaatiot pystyvät suoriutumaan ydintehtävistään ja velvoitteistaan	3	2	3	81
Poistaa päällekkäisiä toimintoja	2	2	3	63
Mahdollistaa saumattoman yhteistyön ja yhteiskäytön organisaatorajojen yli	3	1	2	38
Parantaa mahdollisuuksia osaamisen kehittämiseen ja henkilöriippumattomuuteen	2	2	3	63
Hyödyntävät organisaatiot pystyvät keskittymään ydinosaamiseensa	3	2	3	81
Digitoitujen aineistojen säilyvyys voidaan varmistaa	3	1	3	63
Prosessien ja palveluiden laatu paranee	2	2	3	63
Mahdollistaa tuleville sukupolville aineistojen hyödyntämisen	3	1	3	63

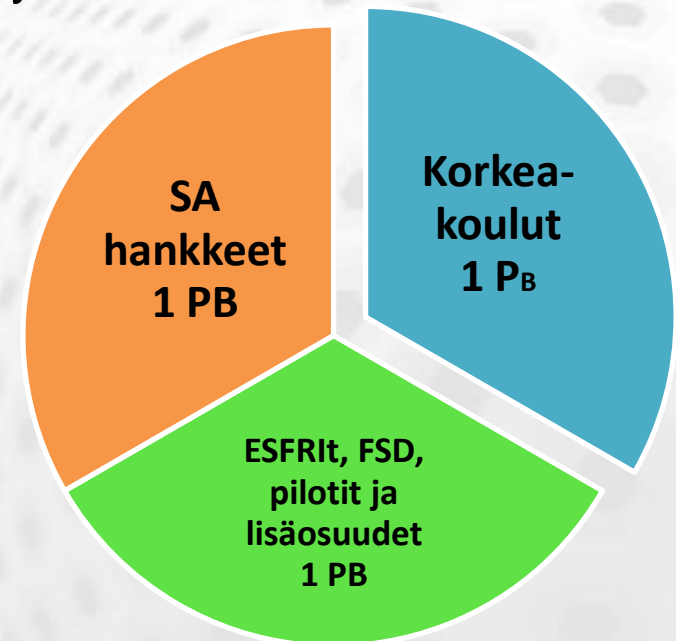
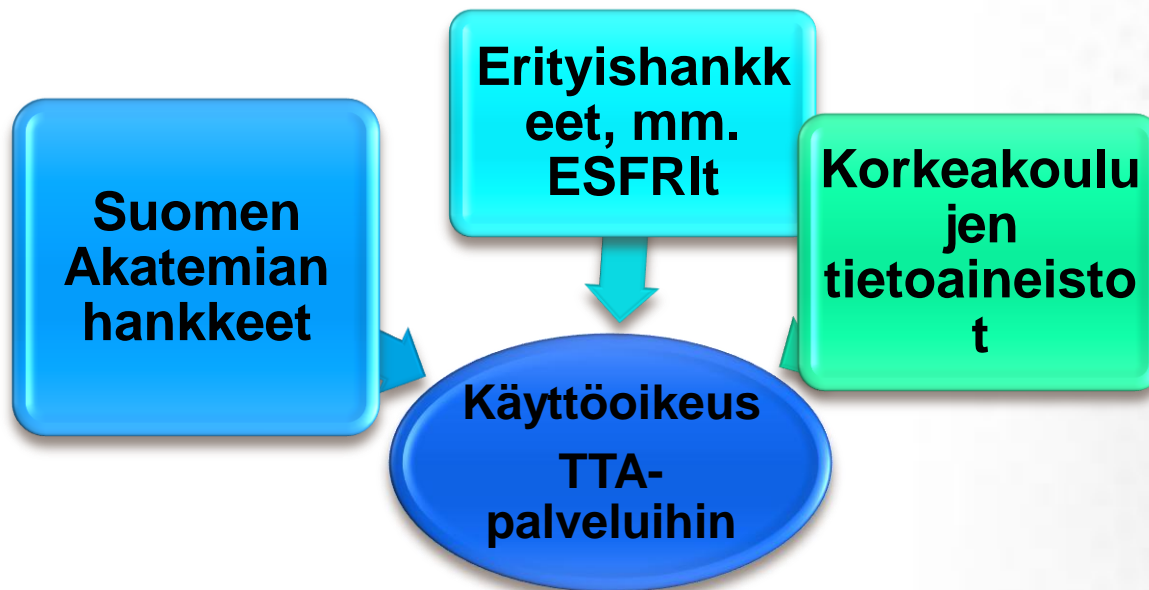
TUTKIMUKSEN TIETOINFRASTRUKTUURI



TTA-palvelut:
Tutkimus-IDA, Datakatalogi,
Datanhallinnan pilotti

IDA :yhteinen tallennuspalvelu tutkimuksen digitaalisten tietoaaineistojen (jatkossa data) säilyttämiseksi ja uudelleenkäytön lisäämiseksi. Ida mahdollistaa tutkimuksen tuottamien tietoaaineistojen ja niihin liittyvän metatiedon turvallisen säilytyksen.

- Datan säilyminen eheänä ja muuttomattomana sekä asiakkaan helposti saatavilla on palvelun ensisijainen tavoite. Säilytettävän datan muuttumattomuus turvataan kopioiden ja niiden eheyden hallinnalla.



Yliopisto	Osuus	Teratavuina
Aalto-yliopisto	0,12	120
Helsingin yliopisto	0,27	270
Itä-Suomen yliopisto	0,08	80
Jyväskylän yliopisto	0,07	70
Kuvataideakatemia	0	
Lapin yliopisto	0,01	10
Lappeenrannan teknillinen yliopisto	0,03	30
Oulun yliopisto	0,09	90
Sibelius-Akatemia	0	
Svenska handelshögskolan	0,01	10
Tampereen teknillinen yliopisto	0,06	60
Tampereen yliopisto	0,06	60
Teatterikorkeakoulu	0	
Turun yliopisto	0,14	140
Vaasan yliopisto	0,01	10
Åbo Akademi	0,04	40

Ammattikorkeakouluille varataan 10 TT kiintiö ammattikorkeakoulujen yhteisiin tarpeisiin.

Käyttöoikeudet: Suomen Akatemian kautta



- IDA-palvelussa tarjotaan Suomen Akatemian hankkeiden tarpeisiin vuoden 2013 loppuun asti vähintään noin 1 petatavun (PT) tallennuskapasiteetti.
- Käyttöoikeus on maksuton Suomen Akatemian rahoituksen hankkeille (akatemiahankeet, huippuyksiköt, tutkimusohjelmat, tutkimusinfrastruktuurit).
- K.o. hankkeille tarjotaan 2 teratavun (TT) tallennusoikeutta hanketta kohden .

Käyttöoikeudet: Lisäosuus

- ESFRI-hankkeiden sekä Suomen Akatemian rahoituksen hankkeiden keskuudesta ehdotetaan valikoitavaksi tarpeen ja kyvykkyyden mukaan hankkeet, joille voidaan myöntää tasakiintiötä suurempi kiintiö.
- Valikointiin nimitetään vastuutaho.
- Tähän varataan n. 1 PT kapasiteetista

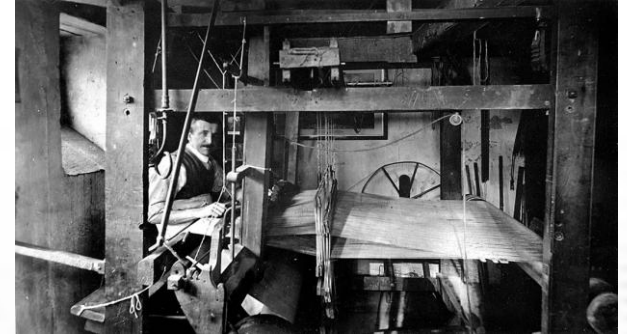
Aineistokatalogi KATA

TAVOITTEET

- Aineistokatalogin keskeisin tehtävä tutkimuksen tietoinfrastruktuurissa on helpottaa tutkimuskäyttöön saatavilla olevien aineistojen löytämistä.
- Aineistokatalogin toinen tärkeä tehtävä on OKM:ään kohdistuvien tietohallintolain mukaisten inventointivaatimusten toteuttaminen: tuottaa tietoa aineistojen olemassaolosta rahoittajille
- Aineistokatalogin kolmas tehtävä on mahdollistaa yhtenäisen käyttöehto- ja käyttöoikeuskulttuurin luominen. Tämä tarkoittaa selkeiden omistajuus- ja hyödyntämisoikeustietojen kirjaamista aineistokatalogiin.
- Neljäs tehtävä on mahdollistaa meriitin syntyminen tutkijalle.
- Viides tehtävä on auttaa tunnistamaan ja löytämään pitkäaikaissäilytykseen ”valmiita” tietoaineistoja.

täytyy tukea:

- koko tutkimusprosessia
- tietoaineistojen siirtoa tieteenalojen välillä
- avoimen datan tuottoa
- työvuopohjaista toimintaa
- julkaisujen ja tietoaineistojen linkitystä
- tiedepolitiikan kehikkoa



Tutkimusaineistojen pitkäaikaissäilytys

- Tutkimustiedon **kumuloituminen** on tutkimuksen ja tutkimusorganisaatioiden toiminnan ydin
 - Nykyisin ei ole käytössä hallittua ja toimivaa tapaa huolehtia digitaalisesta tiedosta pitkällä aikavälillä
- Digitaalisten aineistojen pitkäaikaissäilytys tarkoittaa digitaalisen informaation **luotettavaa säilyttämistä kymmenien tai jopa satojen vuosien ajan**
 - Laitteet, ohjelmistot ja tiedostomuodot vanhenevat, mutta tästä huolimatta informaation täytyy säilyä ymmärrettävänä.
- Tutkimusaineistojen pitkäaikaissäilytyksessä tehdään tiivistä **yhteistyötä** Kansallisen Digitaalisen Kirjaston (KDK) kanssa

PITKÄAIKAISSÄILYTYS

“Leave a trail of evidence for scientists of the future”



Digitaalinen pitkäaikaissäilyttäminen

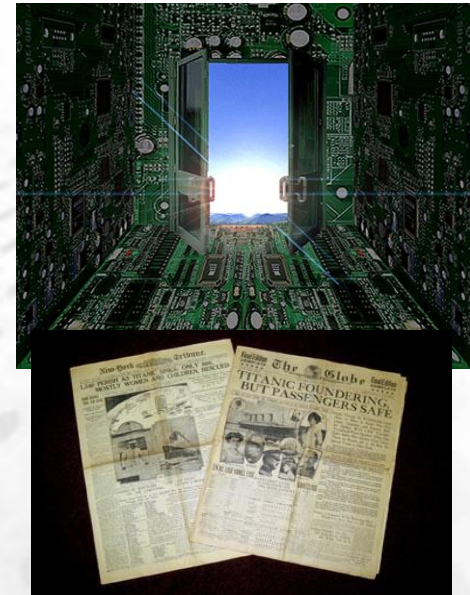


- prosessi, jossa digitaalinen kokonaisuus irroitetaan luontiympäristöstään, ja sen olemassaolo varmistetaan säilytysympäristössä autenttisuus ja eheys säilyttäen

- Kriittiset asiat:

- Autenttisuuden ja eheyden varmistaminen
- Teknologisten riskien hallitseminen
- Kustannusten hallitseminen

- Päämääränä infrastruktuuririippumattomuus, jotta voidaan käyttää mitä tahansa tallennusratkaisua



HAASTEET:

Mitä pitää säilyttää?

- Materiaali joka pitää "pelastaa"
- Materiaali jota arvioimme tarvitsevamme pitkällä aikavälillä

Kuinka se tulisi säilyttää?

- Formaatit
- Tallennusmediat
- Kuka on vastuussa, kuka tekee?

Kuka maksaa?

- Sisällön tuottajat?
- Laitokset ja instituutiot?
- Käyttäjät?

Kolme aineistoluokkaa

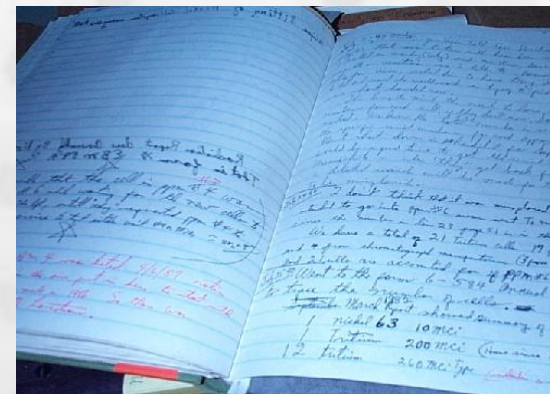
1. Tärkeät säilytettävät aineistot

- Välttämättömiä organisaation toiminnan jatkuvuudelle
- Liittyy korkea uudelleen käytön potentiaali
- Ei liity merkittäviä käyttörajoituksia

2. Mahdollisesti säilytettävät aineistot

- organisaation tehtävään keskeisesti liittyvät aineistot
- Liittyy uudelleen käytön potentiaalia
- Sallitaan jonkin verran käyttörajoituksia

3. Muut organisaation aineistot





Kansallinen digitaalinen kirjasto

Pitkäaikaissäilytys

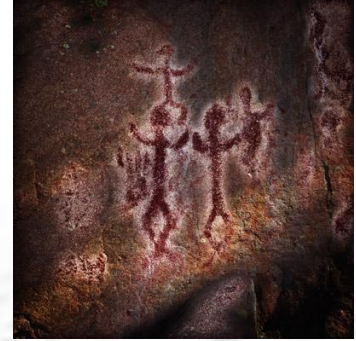
Yhdessä enemmän!



TTA

TUTKIMUKSEN
TIETOAINIESTOT

- KDK ja TTA: pitkäaikaissäilytyksen osalta mahdollisimman laaja yhteistyö
- OKM: pitkäaikainen sitoutuminen ja riskien hallinta, CSC:n kyvykkyyden varmistaminen
- CSC: toteuttamisen vastuut ja tehtävät, riskien minimointi, läpinäkyvyys kustannuksissa



“But I am a scientist, NOT a records manager!”



Kiitos!

”Päämäärä ja keinot ovat
synonyymeja”

Mahatma Gandhi

