



Kansallinen digitaalinen kirjasto

Digitaalinen pitkäaikaissäilytys

30.5.2012

Kimmo Koivunen

CSC – Tieteen tietotekniikan keskus Oy



Esityksen sisältö

- Mitä tarkoittaa digitaalinen pitkäaikaissäilyttäminen?
 - › Säilyttämisen vaatimustasot ja menetelmät
- KDK:n PAS-ratkaisu
 - › Kokonaisarkkitehtuuri
 - › Toteuttamissuunnitelma
 - › Vaiheistus
 - › Määritykset
- Yhteenveto



Ymmärrettävyyden säilyttäminen

Erään huomattavan kirjailijan perikunta lahjoittaa hänen jälkeenjääneet paperinsa kotipitäjänsä kirjastolle. Aineistoon kuuluu myös pahvilaatikollinen lerppuja eli 5¼ -tuumaisia taipuisia muistilevykkeitä.

Kirjasto löytää varastosta vanhan 5¼ -tuumaisen levykeaseman sekä tietokoneen, johon asema saadaan kiinni. Koneessa on onneksi myös riittävän vanha käyttöjärjestelmä, joka osaa hyödyntää asemaa. Osa lerpuista on vahingoittunut, mutta osa saadaan avattua. Tiedot kopioidaan nykyaikaisemmille tallennusvälineille.

Ongelmaksi muodostuu kuitenkin levykkeillä olevien tiedostojen avaaminen. Selvittelyissä käy ilmi, että kyseessä on valtionhallinnossa käytössä ollut TEKO-ohjelmisto. Pitkällisen jäljityksen tuloksena löydetään TEKOn asennuslevyke. Valitettavasti käyttöohjeita ei löydy.

Ohjelmiston asennus ei onnistu, vaikka sitä yritetään monta kertaa.

Lopulta asiaan ei enää voida käyttää enempää aikaa. Lerppulaatikko viedään varastoon odottamaan aikaa parempaa. Bitit on jotenkuten säilytetty – ymmärrettävyyttä ei.



Uhat ja niihin varautuminen

- Säilytysmedian viat
 - › Datasta useampi kopio erilaisilla medioilla
- Toimittajakohtaiset systemaattiset viat
 - › Datasta kopiot eri toimittajien ratkaisussa
- Toiminnallinen tai hallinnollinen virhe
 - › Data hallinoidaan vähintään kahdessa järjestelmässä
- Luonnonkatastrofit
 - › Säilytys toteutetaan maantieteellisesti hajautettuna
- Pahantahtoinen käyttäjä
 - › Säilytystä varmennetaan ns. pimeällä arkistolla
- Osaamisen ja henkilöstön riittämättömyys
 - › Toimintaa keskittämällä ja suunnitelmallisuudella varmistetaan osaamisen kehittyminen ja riittävä henkilöstö



Yleiset vaatimukset pitkäaikaissäilytyksen tekniikalle

- Järjestelmässä ei saa olla yksittäistä kohtaa, jonka pettäessä koko järjestelmä pettää (single point of failure)
 - › Koskee erityisesti elintärkeitä osioita
- Järjestelmän täytyy toimia, vaikka mediat, ohjelmistot ja laitteistot vaihtuvat tasaisena virtana
 - › Järjestelmässä tulee hyödyntää erilaisia tekniikoita ja välttää lukkiutumista tiettyjen laite- tai ohjelmistotoimittajien ratkaisuihin
- Järjestelmän täytyy jatkuvasti seurata tietopakettien eheyttä
 - › Järjestelmässä on aineistoa, jota käytetään harvoin, mikä aiheuttaa suuren riskin piilevien virheiden kerääntymiseen.
- Järjestelmän ylläpitotoimet on oltava mahdollisia ilman käyttökatoja



Alkuperäisen käyttökokemuksen säilyttäminen

- Vaativin, mutta käyttäjän kannalta usein mielenkiintoisin taso
- Esim. Tietokonepelin pelaaminen samoin kuin 30 vuotta sitten

Sisällön ymmärrettävyyden säilyttäminen

- Teksti, kuvat ja muu olennainen sisältö voidaan tulkita
- Ymmärrettävyys riippuu myös kohdeyleisöstä
- Käytön ja uuden tiedon tuottamisen kannalta yleensä tärkein taso

Bittien säilyttäminen

- Varmistetaan alkuperäisten ykkösten ja nollien säilyminen ja luettavuus, sisällön tulkinnasta ei takeita
- Säilyttämisen perusta ja minimitaso säilytyksen aloittamiseen:
Suunnitelmallisesti tehtynä vaatimustasoa voidaan nostaa myöhemmin



Alkuperäisen käyttökokemuksen säilyttäminen

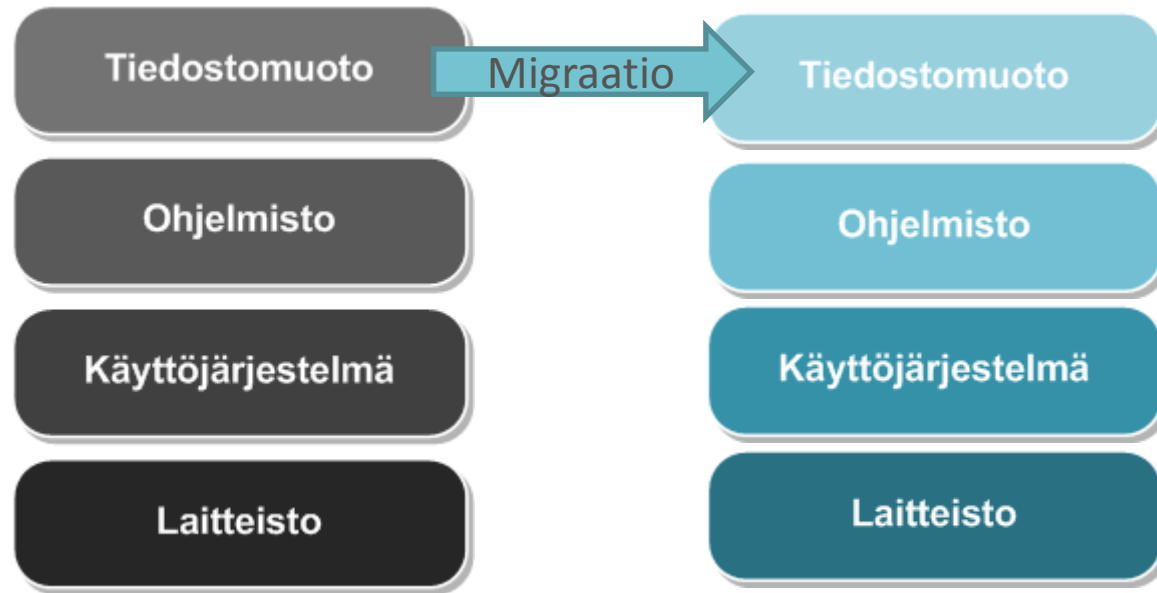
- **Emulaatio**
 - Emulaattoreiden kehittäminen uusille alustoille, formaattien seuranta jne.
 - Aktiivinen testaaminen ja seuranta

Sisällön ymmärrettävyyden säilyttäminen

- **Migraatio**
 - Tiedostomuotojen kehityksen seuranta (formaattikirjasto)
 - Muunnosten kehittäminen, testaaminen, toteuttaminen ja valvonta
 - Varautuminen palautuksiin

Bittien säilyttäminen

- **Eheyden varmistaminen**
 - Tiedostojen validointi ja valvonta
 - Kopioiden hallinta
 - Koskee sekä objekteja että metadataa



- Migraatiossa mahdollistetaan digitaalisten aineistojen hyödyntäminen uudistuneessa tietoteknisessä ympäristössä
- Informaatioisällön säilymiseksi migraatioon suunnitteluun, testaamiseen ja arviointiin erityistä huomiota

Emulaatio



Kansallinen digitaalinen kirjasto

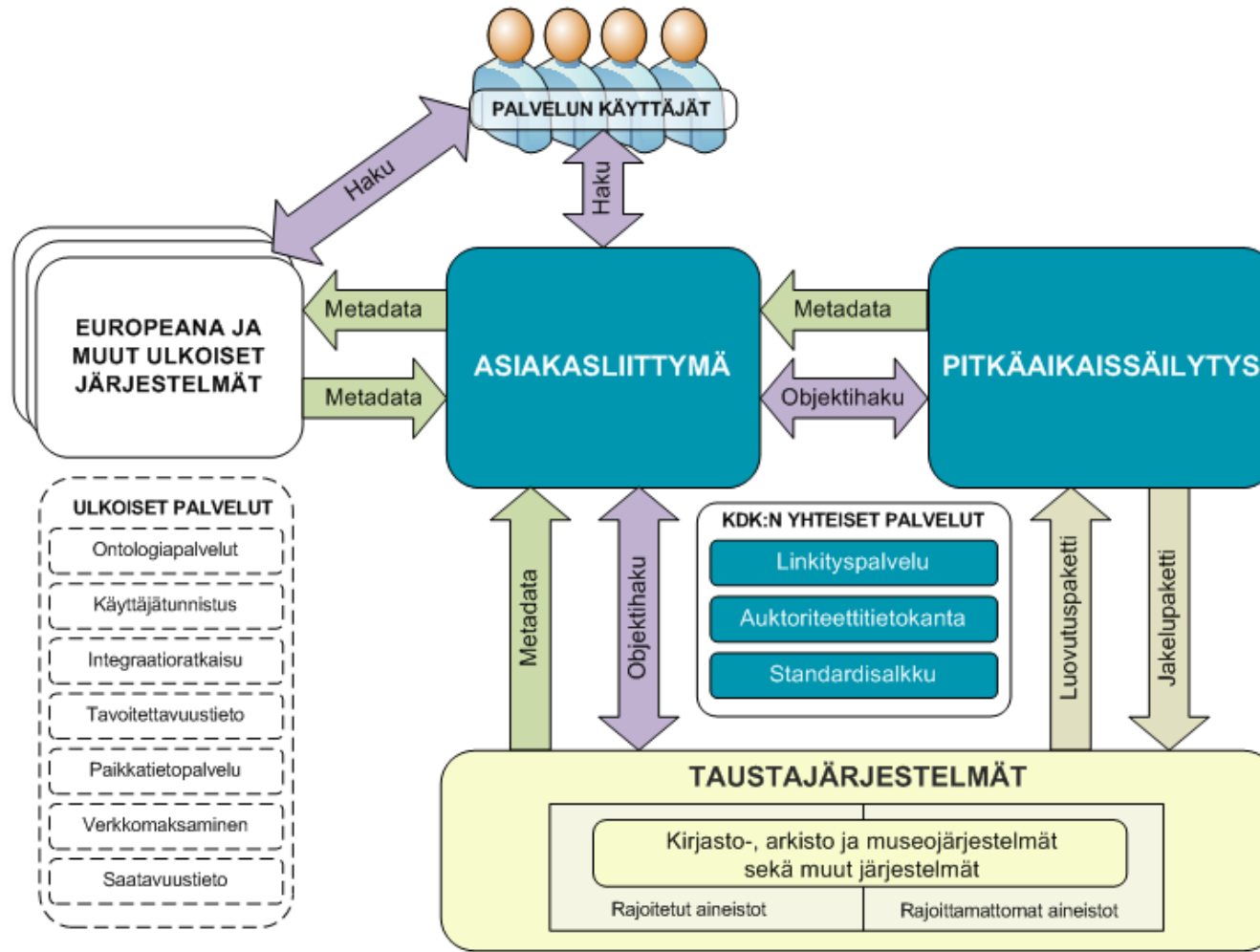


- Emulaatiossa mahdollistetaan (vanhentuneen) ympäristön käyttö uudistuneilla laitteistoilla
- Emulaatiossa tulee myös ratkaista kuinka informaatiota voidaan hyödyntää uuden tiedon tuottamisessa (vrt. copy-paste)



KDK:n kokonaisarkkitehtuuri

Kansallinen digitaalinen kirjasto



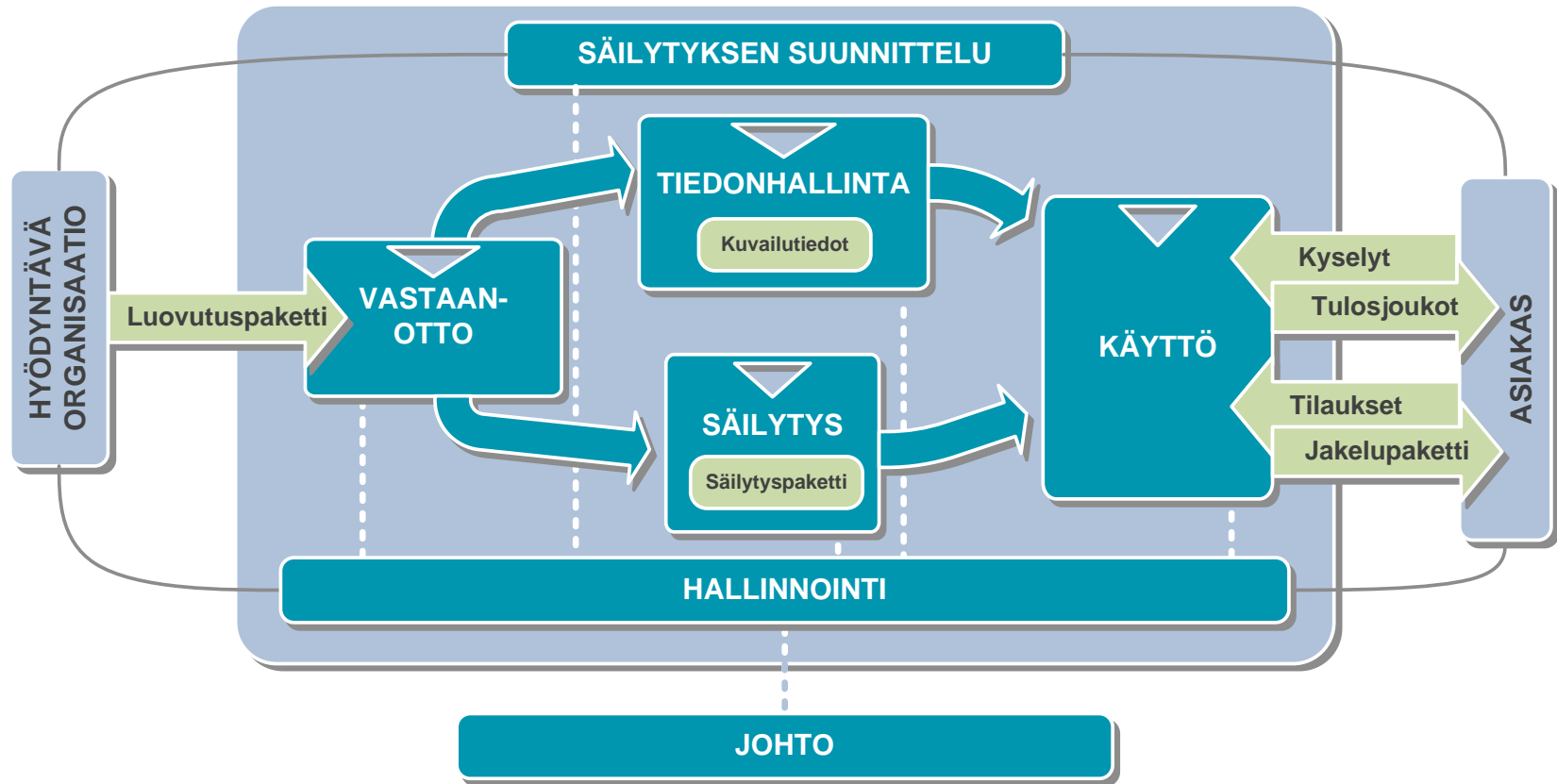


Pitkäaikaissäilytysjärjestelmä on palvelujärjestelmä, johon opetus- ja kulttuuriministeriön hallinnonalalla toimivat, kulttuuriperintöä säilyttävät organisaatiot siirtävät pitkäaikaisesti tai pysyvästi säilytettäviä aineistoja. Aineistot säilyvät myös pitkäaikaissäilytysjärjestelmässä tallentavien organisaatioiden aineistoina. Tavoitteena on, että pitkäaikaissäilytysjärjestelmä otettaisiin käyttöön vuonna 2016. Kokonaisuuden suunnittelussa tulee huomioida, että järjestelmää on mahdollista hyödyntää tulevaisuudessa myös tutkimuksen tietoaineistojen säilyttämisessä.

KDK-PAS noudattaa OAIS-mallia

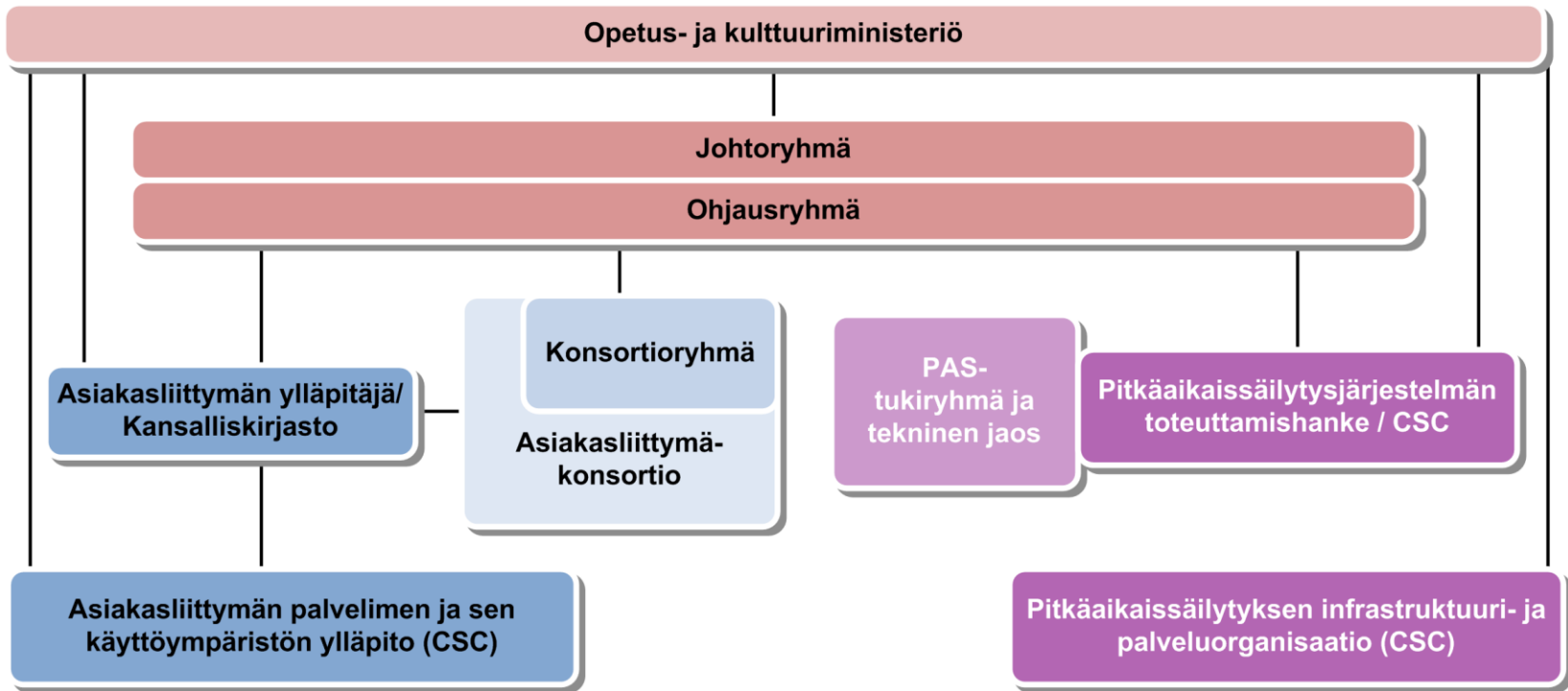


Kansallinen digitaalinen kirjasto





KDK:n hallinnointi 1.4.2011–31.12.2013



Tukiryhmä:

- Tuomas Alaterä, Yhteiskuntatieteellinen tietoaarkisto
- Juha Hakala, Kansalliskirjasto
- Vesa Hongisto, Museovirasto
- Minna Karvonen, opetus- ja kulttuuriministeriö
- Istvan Kecskemeti, Kansallisarkisto
- Esa-Pekka Keskitalo, Kansalliskirjasto
- Markus Merenmies, Kansallisarkisto
- Johan Kylander, Svenska Litteratursällskapet i Finland SLS
- Kari Peiponen, Valtion taidemuseo
- Pekka Tähtinen, Kansallinen audiovisuaalinen arkisto
- Kimmo Koivunen, puheenjohtaja, CSC
- Kuisma Lehtonen, sihteeri, CSC

Tekninen jaos

- Vesa Hongisto, Museovirasto

PAS-toteuttamishanke CSC:llä

- Pirjo-Leena Forsström, Heikki Helin, Kimmo Koivunen, Juha Lehtonen, Kuisma Lehtonen

Aineiston laajuus ja säilytysmenetelmät- kartoitus 2011

Taulukko 1: Arviot aineistojen laajuudesta vuoden 2011 kartoituksessa

	2010		2011		2015		2020	
	Objektien määrä (milj. kpl)	Koko (Tt)	Objektien määrä (milj.kpl)	Koko (Tt)	Objektien määrä (milj.kpl)	Koko (Tt)	Objektien määrä (milj.kpl)	Koko (Tt)
Asiakirjat ja dokumentit	11,6	328	15,4	394	25,6	646	48,7	1301
Valokuvat	1,7	18	2,1	30	3,9	68	6,1	120
Elokuvat	0,1	495	0,2	1143	0,8	3055	1,2	8020
Äänitallenteet	1,2	606	1,5	771	2,4	1418	3,7	2176
Viitetiedot	19,5	1,2	21	1,5	27	2,4	34	3,4
Verkkoarkisto	496	20	646	27	1396	59	2300	97
Radio- ja TV-arkisto	0,8	95	1,2	142	2,9	327	5,0	558
YHTEENSÄ	530	1 563	687	2 509	1458	5 575	2400	12 275

Pitkäaikaissäilytysratkaisun toteuttamissuunnitelma - sisällys



Kansallinen digitaalinen kirjasto

Suunnitelma saatavilla KDK:n www-sivuilla

TIIVISTELMÄ

1 JOHDANTO

2 PITKÄAIKAISSÄILYTYKSEN PERUSTEET

3 PITKÄAIKAISSÄILYTYSRATKAISUN TAVOITTEET

4 PITKÄAIKAISSÄILYTYKSEN PALVELUT

- 4.1 Aineiston valmistelun ja vastaanoton palvelut
- 4.2 Säilytyspalvelut
- 4.3 Aineiston käyttöpalvelut
- 4.4 Aineiston hallintapalvelut
- 4.5 Neuvonta- ja tukipalvelut

5 AINEISTON VALMISTELU JA VASTAANOTTO

6 AINEISTON SÄILYTTÄMINEN

7 AINEISTON KÄYTTÖ JA HYÖDYNTÄMINEN

8 HALLINNOINTI JA SÄILYTYKSEN SUUNNITTELU

9 PAS-RATKAISUN VAIHEISTAMINEN

- 9.1 Vaihe I: PAS-ratkaisun valmistelu (2011-2013)
- 9.2 Vaihe II: PAS-ratkaisun toteuttaminen (2014-2016)

LIITTEET



KDK:n pitkäaikaissäilytyksen tavoitteena on (s.18-19):

1. Varmistaa digitaalisten kulttuuriaineistojen säilyvyys
2. Varmistaa aineistojen hyödynnettävyys ja yhteiskäyttö
3. Laatia kokonaisuus kustannustehokkaasti
4. Edistää organisaatioiden yhteistyötä
5. Rakentaa avoimella yhteistyöllä parempia palveluita
6. Laajentua erilaisille aineistoille



4.1 Aineiston valmistelun ja vastaanoton palvelut

- › Metatietojen määrittely
- › Säilytys suunnitelmien laadinta
- › Paketointipalvelu
- › Luovutuspaketin siirto ja tarkastaminen
- › Siirtokelpoisten tiedostomuotojen käsittely

4.2 Säilytyspalvelut

- › Säilytysmenetelmien kehittäminen ja seuranta
- › Säilytystoimenpiteiden suorittaminen
- › Toimintaympäristön seuranta

4.3 Aineiston käyttöpalvelut

- › Hakutoiminnot
- › Aineiston jakelu

4.4 Aineiston hallintapalvelut

- › Metatietojen päivittäminen
- › Aineiston päivittäminen
- › Aineiston poistaminen
- › Säilytys suunnitelman päivittäminen

4.5 Neuvonta- ja tukipalvelut

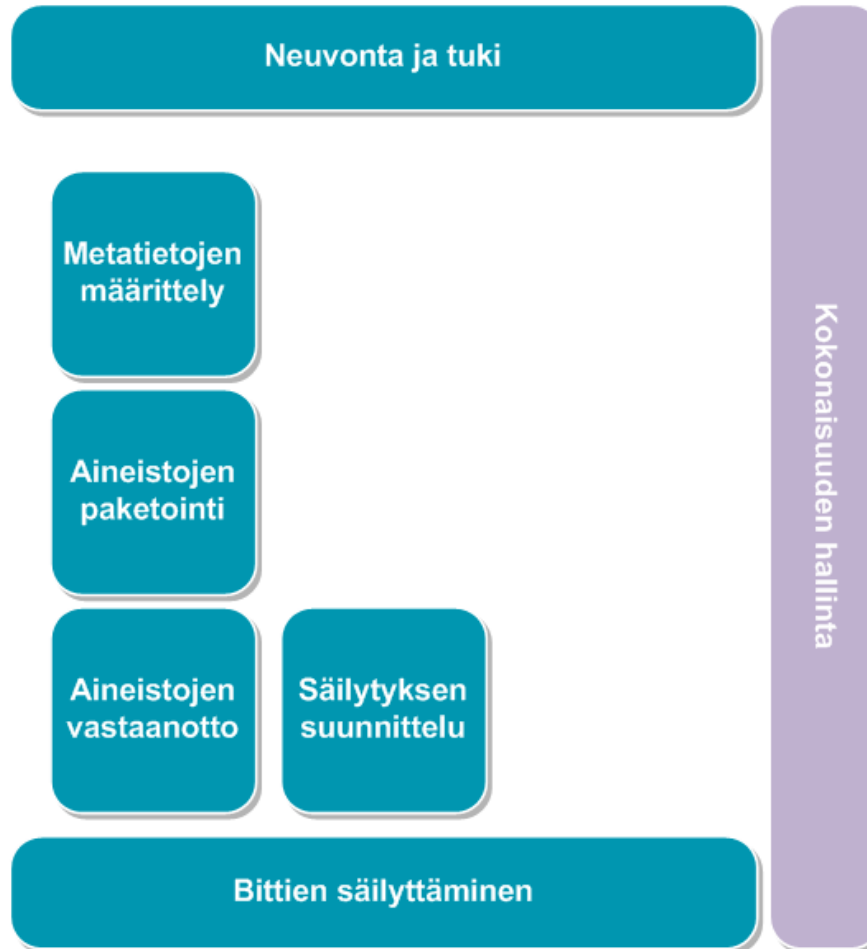
- › Palveluiden ja PAS-järjestelmän käytön tuki
- › Hallinnollinen tuki
- › Koulutus ja tiedotus

- PAS-ratkaisun toteuttamissuunnitelman mukaisesti edetään kahdessa vaiheessa
- **Vaihe I : PAS-ratkaisun valmistelu (2011-2013)**
 - › Varmistetaan digitaalisen datan (bittijono) muuttumattomuus käynnistämällä bittien säilyttäminen. Rakennetaan aineistojen vastaanotto tarvittavine (metatieto)määrityksineen huomioiden ymmärrettävyyden säilyttäminen.
- **Vaihe II : PAS-ratkaisun toteuttaminen (2014-2016)**
 - › Varmistetaan aineiston ymmärrettävyyden säilyminen (informaatioisältö) tuleville sukupolville. Rakennetaan hyödyntävien organisaatioiden yhteiseen käyttöön tähän tarvittavat välineet, palvelut ja prosessit.

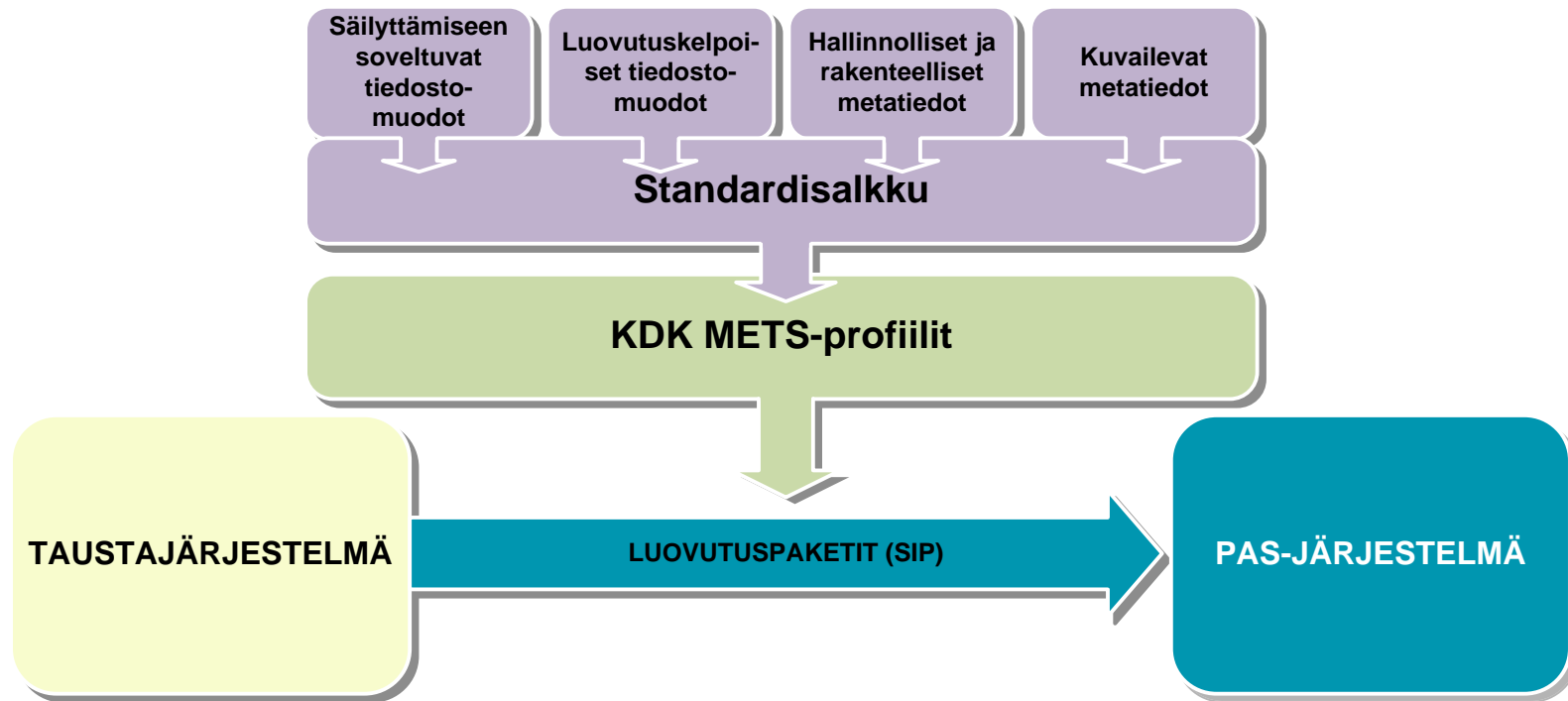
Vaihe I: PAS-ratkaisun valmistelu 2012 - 2013



Kansallinen digitaalinen kirjasto



 Vaihe I:
PAS-ratkaisun
valmistelu



- › KDK:ssa käytettävät standardit, tiedostomuodot sekä pakolliset ja suositeltavat metatiedot määritellään standardisalkussa
- › Standardisalkun mukaisesti aineisto siirretään taustajärjestelmästä PAS-järjestelmään käyttäen METS vaihtomuotoformaattia

Keskitetyn pitkäaikaissäilytysjärjestelmän kustannukset ja kustannussäästöt



Nationaalinen digitaalinen kirjasto

Vuonna 2010 laadittu analyysi:

- Yhteisen PAS-järjestelmän kokonaiskustannukset 12 ensimmäisen toimintavuoden aikana ovat noin 42 miljoonaa euroa
- Suunnittelu- ja käyttöönottovaiheen (17-18 milj. euroa) jälkeen kustannukset ovat noin 4 milj. euroa / vuosi
- Jos säilyttämistä varmennetaan ns. pimeällä arkistolla 6. toimintavuodesta alkaen, lisäkustannukset olisivat noin 3,4 miljoonaa euroa (6.-12. toimintavuosina)
- Yhteinen PAS-järjestelmä saavuttaa muihin toteutustapoihin verrattaessa (SADe-ohjelman hyötyanalyysi):
 - › tuotantovaiheessa yli 10 miljoonan vuosittaiset säästöt ja
 - › käyttöönottovaiheessa n. 30 miljoonan kertaluontoiset säästöt
 - › Kumulatiivinen hyöty 12 ensimmäisen vuoden aikana 100 miljoonaa euroa



Julkaistuja suunnitelmia

- KDK:n pitkäaikaissäilytyksen toteuttamissuunnitelma
- KDK:n hallinnolliset ja rakenteelliset metatiedot ja aineiston paketointi
- Säilytys- ja siirtokelpoiset tiedostomuodot
- Digitaalisten aineistojen laajuus ja säilytysmenetelmät
- Pitkäaikaissäilytyksen kustannus- ja hyötyanalyysi

<http://www.kdk.fi/fi/pitkaaikaissailytys/>

Yhteenveto: yksittäisen digitaalisen objektin säilyttäminen





Kansallinen digitaalinen kirjasto

Kiitos!

www.kdk.fi/pitkaaikaissailytys

Kimmo.Koivunen@csc.fi