

Fairdata-palvelut

Säilytä, kuvaile, hae, pitkäaikaissäilytä



Tutkimusdatan hyödyntämisen kansallisten kehityshankkeiden lyhyt historia

- Kansallinen digitaalinen kirjasto (KDK) 2008 - 2017
- Tutkimuksen tietoaaineistot (TTA) 2011 - 2013
- Avoin tiede ja tutkimus (ATT) 2014 - 2017
- Fairdata.fi 2018 -



Tee tutkimusaineistostasi

- Löydettävä
- Saavutettava
- Yhteentoimiva
- Uudelleenkäytettävä



FAIR

force11.org/group/fairgroup/fairprinciples

Miten tutkimusaineistosta FAIR?

- Löydettävä

Aineisto kuvaillaan ja aineistolle annetaan yksilöivä ja ikuisesti kestävä tunniste.

- Saavutettava

Tunnistautuminen ja valtuutus ovat mahdollisia protokollan avulla. Hyödynnä olemassaolevia tunnistautumis- ja valtuutuspalveluja, kuten REMS.

Huolehdi kaikkien aineistojesi ja julkaisujesi mahdollisimman laajasta kuvailusta ja metatietojen avoimuudesta.

- Yhteentoimiva

Metatietojen merkityksen tulee säilyä, vaikka tietoja siirrettäisiin tietojärjestelmästä toiseen. Suosi valmiita sanastoja, ontologioita ja yleisessä käytössä olevia termejä löytyvyyden takaamiseksi.

Aineistossa käytetään hyväksytyjä viittauksia muihin aineistoihin.

- Uudelleenkäytettävä

Lisenssi/käyttöehdot ovat tärkeä metatieto, koneluettava formaatti kuten CC-lisenssi on suositeltava. Käytä mahdollisimman sallivaa lisenssiä.

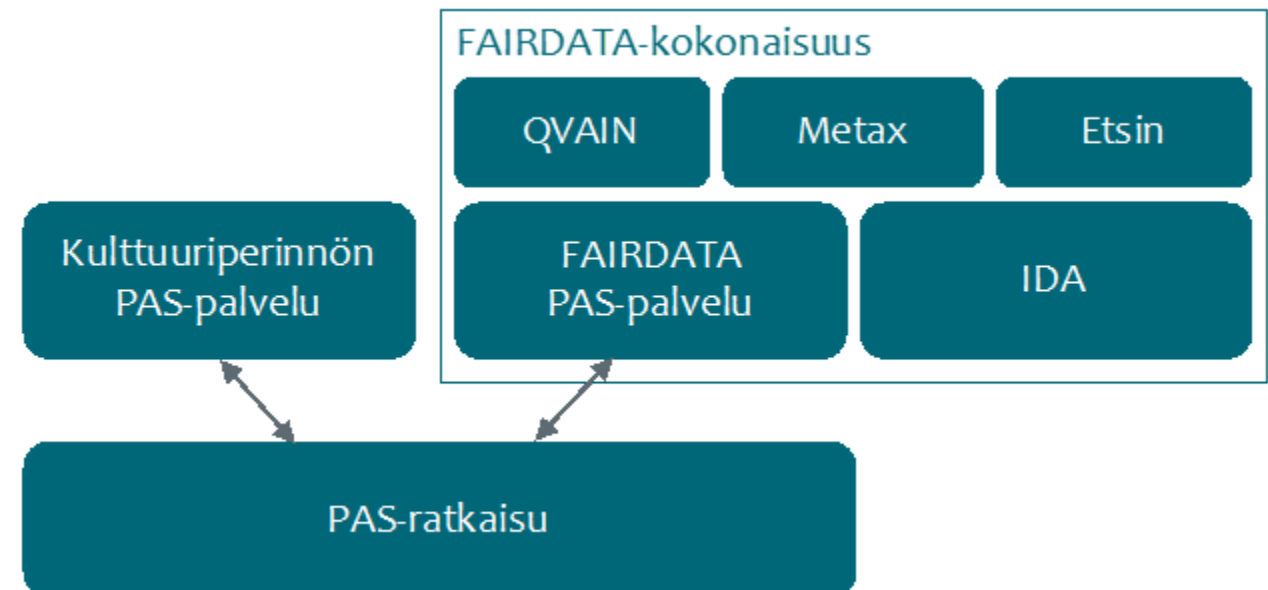
Aineiston riittävän kattava ja täsmällinen kuvailu sekä tehdyn kuvailun ymmärrettävyys, yksiselitteisyys ja -käsitteisyys, ovat merkittäviä asioita metatiedon laadun ja sitä kautta aineiston uudelleenkäytön kannalta.

Datan avaaminen: eri tarpeita eri aineistotyypeille

- Pysyvä vs. kumuloituva ja muuttuva data
- Itsenäinen data vs. metodologiset vaatimukset, lähdekoodi, vuorovaikutteisuus
- Rakenteisen datan satunnaishaku: tietokannat, API:t
- Avoin, luvitettu ja sensitiivinen data
 - Arkaluontoinen henkilötieto: etninen alkuperä, vakaumus, rikoshistoria, terveystieto, seksuaalinen suuntautuminen, sosiaalihuoltoon liittyvä
- Tieteenalakohtaiset vaatimukset: datatyypit, palvelut, tutkimusinfra

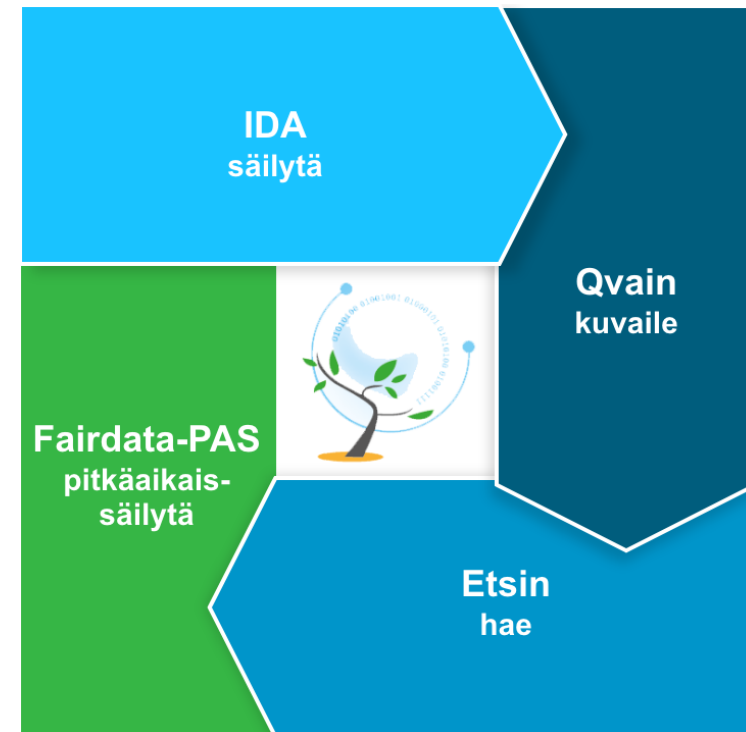
Fairdata.fi –palveluiden kehitys

- Vuonna 2018 valmistuu uudistunut, integroitu palvelukokonaisuus *Fairdata.fi*
- Sisältää palvelut kaikkiin tutkimusprosessin vaiheisiin: säilytys, aineistojen haku, kuvailu, julkaisu sekä pitkäaikaissäilytys
- Palvelut tukevat järkevää aineistonhallintaa

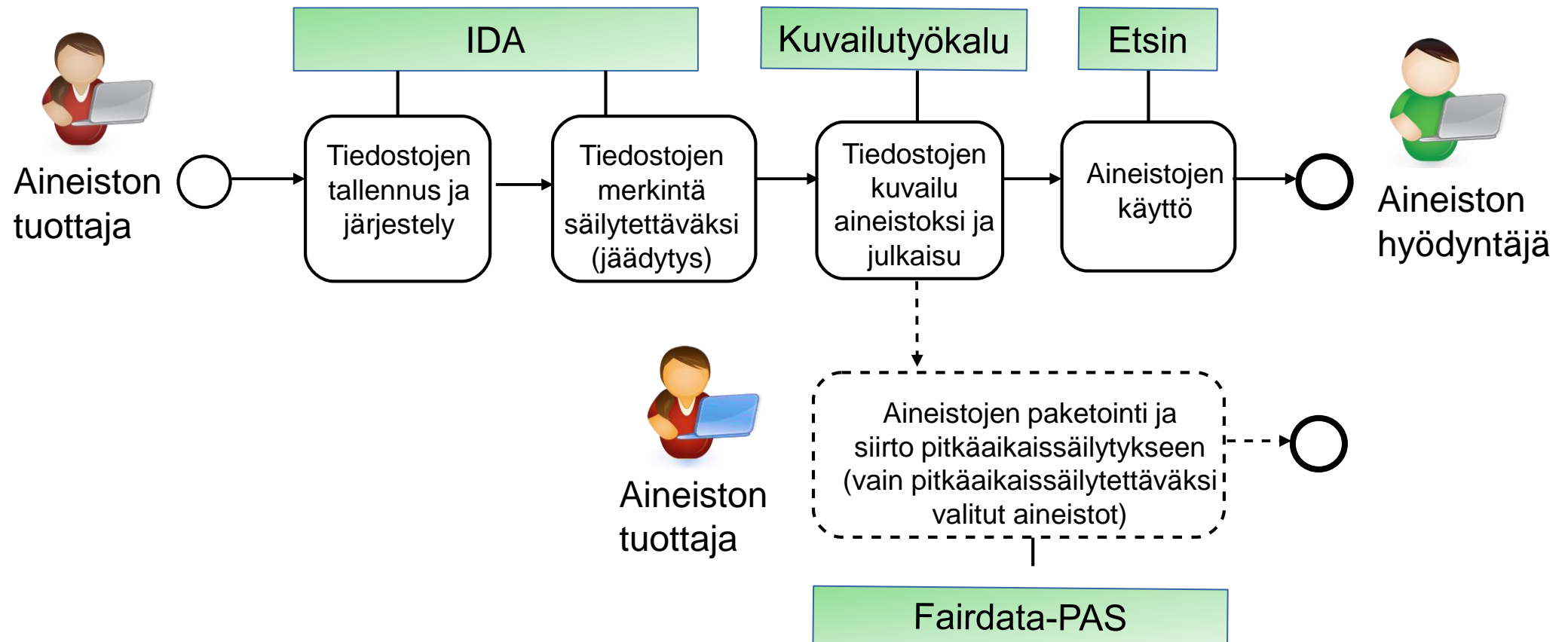


Fairdata.fi –palvelut tutkimusprosessin tukena

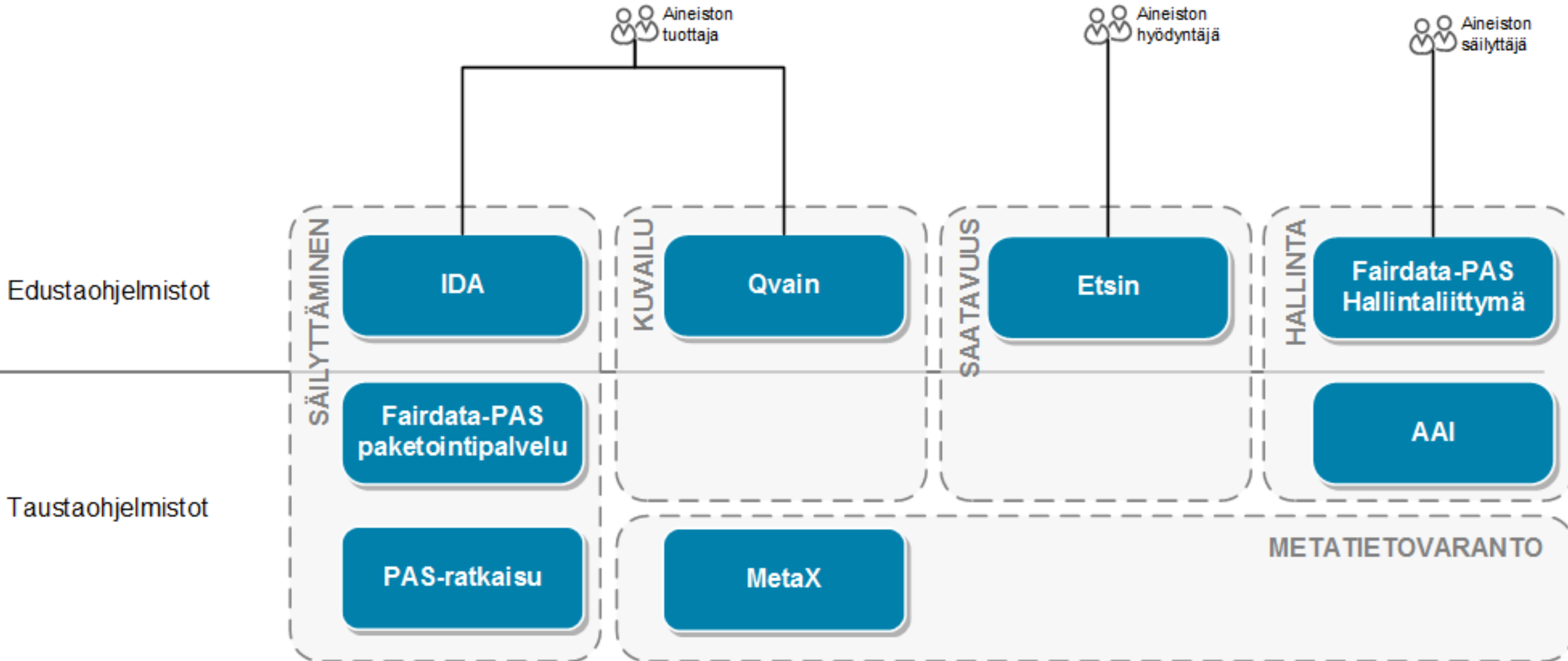
- IDA – säilytä
- Qvain – kuvaile
- Etsin – hae
- PAS – pitkäaikaissäilytä



Uusi Fairdata.fi -palveluiden toimintamalli

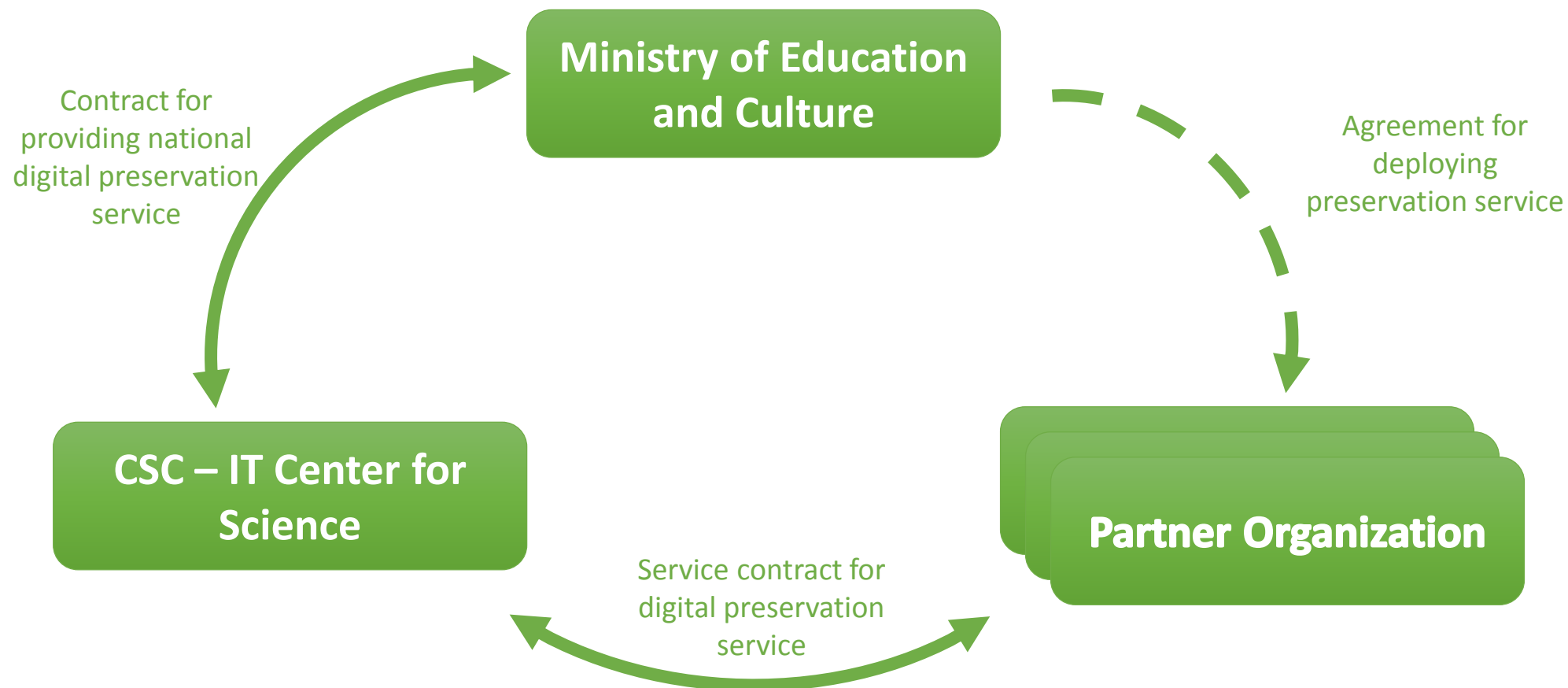


Palvelun komponentit ja toimijat



Yhteistyöstä

Fairdata/PAS –sopimusten periaate



Fairdata yhteystiedot palveluiden käyttäjille

- Fairdata-palveluiden käytössä sinua tukevat kotiorganisaatiosi Fairdata-palveluyhteyshenkilöt
 - Yhteistyö palveluiden ja palveluyhteyshenkilöiden kanssa työstetään arviolta toukokuussa
- Käyttöohjeet palveluihin löytyvät verkosta
 - <https://www.fairdata.fi> (tekeillä)
- Fairdatan käyttäjätukeen saa yhteyden sähköpostitse
 - asiakaspalvelu@csc.fi tai servicedesk@csc.fi
 - Ota yhteyttä, jos havaitset palvelussa teknisiä ongelmia tai sinulla on muuta kysyttävää palvelun toiminnasta, jossa kotiorganisaatiosi Fairdata-palveluyhteyshenkilö ei osaa auttaa

Palaute ja muutoksenhallinta

- Ohjelmistokehitys
 - Viankorjaukset ja ylläpito
 - Ominaisuuksien parantelu
 - Uudet toiminnallisuudet ja integraatiot
- Syötettä eri lähteistä
 - Käyttäjäpalaute ja käytettävyytestaus
 - Organisaatioiden ja palvelun omistajan ohjaus
 - Sisäiset kehityskohteet
 - Strategiset kehityskohteet
- Muutoksenhallinta
 - Palveluohjelmistojen julkaisut (release) ja kehityksen syklit
 - Prosessi suunnitteilla – ideoita?

Fairdata-palvelut

IDA- tutkimusaineistojen säilytyspalvelu

- Tutkimusdatan säilytyspalvelu

Opetus- ja kulttuuriministeriön tarjoama Suomen korkeakoulujen ja Suomen Akatemian rahoittamille tutkijoille
Käyttöosuuksia myönnetään projekteille organisaatioiden hallinnoimasta säilytyskapasiteetista
IDA-projektin kesto ei tutkimusprojektin elinkaareen sidottu

- Käyttäjä voi siirtää tiedostoja IDAan säilytettäväksi

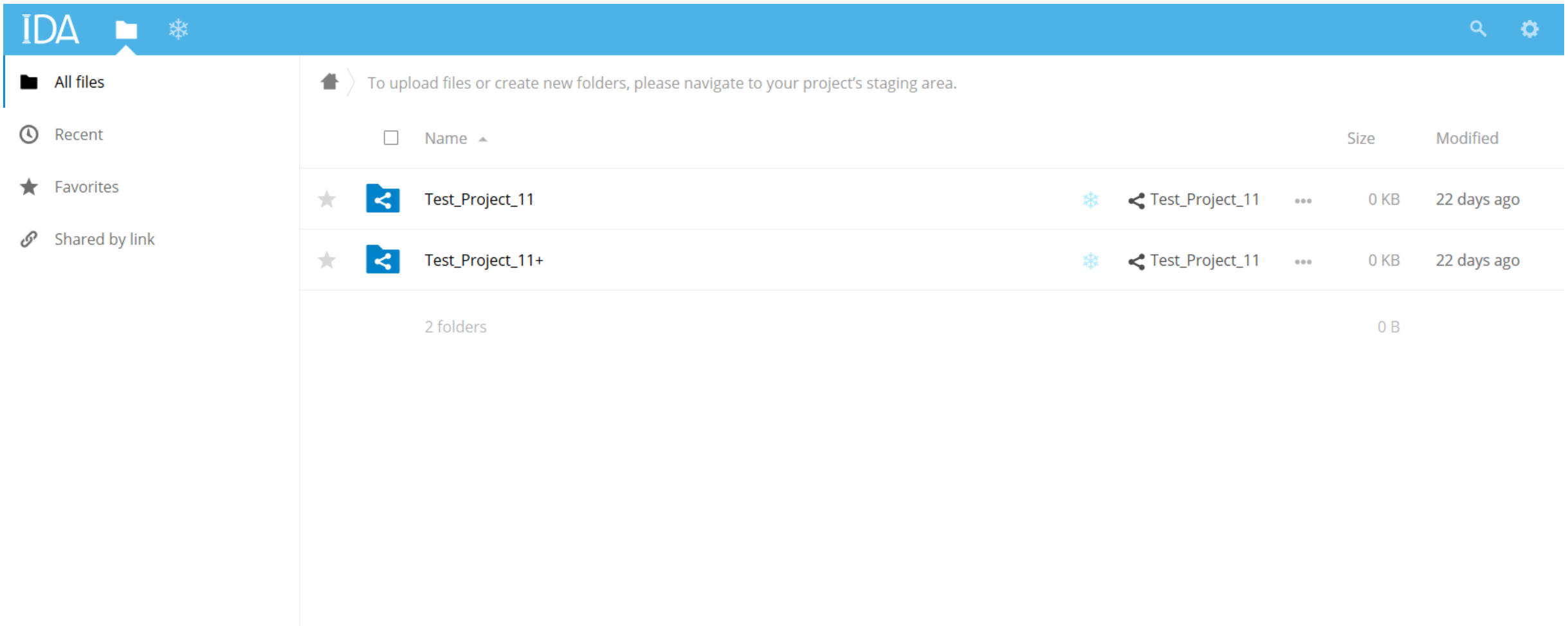
Ydintoiminnot ja tarkoitus pysyvät samoina
Tiedostojen käsittelyyn ja käyttöliittymiin tulee muutoksia

- Uusi toimintamalli tiedostojen tallennukseen ja kuvailuun



Tiedostot siirretään ensin **valmistelutilaan**, jossa niitä voi muokata ja järjestellä
Käyttäjä voi merkitä haluamansa tiedostot säilytettäväksi (**jäädytys**)
Kuvailutyökalulla näistä tiedostoista voidaan koota ja kuvailla aineisto, joka julkaistaan Etsin-palvelun kautta

- IDA integroidaan PAS-palveluun, jonne voidaan paketoita ja siirtää pitkäaikaisäilytettäväksi valitut aineistot
- Käytetyt teknologiat: NextCloud, RabbitMQ, GlusterFS
- Kehittäjä ja ylläpitäjä; CSC

IDA – tutkimusaineistojen säilytyspalvelu



The screenshot shows the IDA web interface. The top navigation bar is blue and contains the IDA logo, a folder icon, a snowflake icon, a search icon, and a settings icon. The left sidebar has four menu items: 'All files', 'Recent', 'Favorites', and 'Shared by link'. The main content area shows a message: 'To upload files or create new folders, please navigate to your project's staging area.' Below this is a table with columns for 'Name', 'Size', and 'Modified'. The table lists two folders: 'Test_Project_11' and 'Test_Project_11+'. Both folders are marked with a star, a share icon, a snowflake icon, and a share icon. The size for both is '0 KB' and the modified date is '22 days ago'. At the bottom of the table, it says '2 folders' and '0 B'.

	Name	Size	Modified
★	 Test_Project_11	0 KB	22 days ago
★	 Test_Project_11+	0 KB	22 days ago

2 folders

0 B

Qvain – tutkimusaineistojen kuvailutyökalu

- Työkalu tutkimusaineistojen kuvailuun

Metatietomalli e.g. <https://tietomallit.suomi.fi/model/mrd/CatalogRecord/>

- Kuvailut julkaistaan Etsimessä (Metaxin kautta)
- Muokkausoikeudet voidaan jakaa ja siirtää
- Käytetään määrittämään tutkimusaineiston käyttöoikeuksia

Integraatio IDA:aan, REMS valtuuksien hallinta

- Käytetyt teknologiat: Go, javascript
- Kehittäjä ja ylläpitäjä: Kansalliskirjasto

Metax – tutkimusaineistojen metatietovaranto

- Fairdata.fi -palvelujen välinen ”liima”
- Säilyttää metadataa, antaa pysyvän tunnisteiden
- Kaavat validointia varten esim.
<https://tietomallit.suomi.fi/model/mrd/CatalogRecord/>
- Haravointi ulkoisista tutkimusaineistojen arkistoista
- Tarjoaa metadatan rajapinnan kautta
- Käytetyt teknologiat: Nginx, Gunicorn, Django, RabbitMQ, Elasticsearch
- Kehittäjä ja ylläpitäjä: CSC

Metax Research Datasets

Näytä tietomallin tiedot

Sisältö englanniksi ▾

LUOKKA ATTRIBUUTTI ASSOSIAATIO

- Activity
- Agent
- Catalog Record**
- Checksum
- Concept
- Concept Scheme
- Directory in IDA
- Documented link
- Editor
- Entity relation
- File characteristics
- File in IDA
- Identifier
- License Document
- Linguistic System
- Location

Catalog Record

Luonnos

Näytä historia

Lataa ▾

Tarkennetun luokan tiedot

Tarkennetun luokan nimi ⓘ

Catalog Record

Tarkennetun luokan tunnistus ⓘ

mrd:CatalogRecord 📄

Tarkennettu luokasta ⓘ

dcat:CatalogRecord

Käsitettä ei ole määritelty

Tarkennetun luokan ominaisuudet ⚙️

Research data catalog ➔

Kuvaus ⓘ

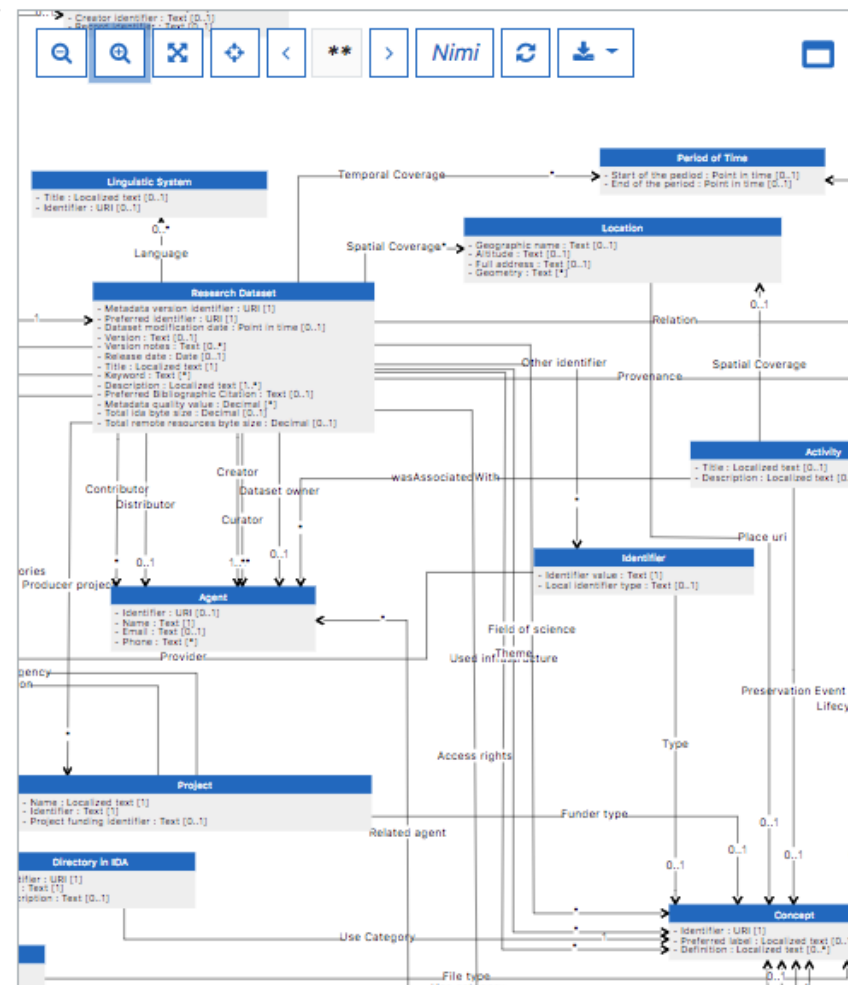
A record in a data catalog, describing a single dataset.

Tila

Luonnos

Määritely soveltamisprofiilissa ⓘ

Metax Research Datasets



Etsin – tutkimusaineistojen hakupalvelu

- Tutkimusaineistojen hakupalvelu suomalaisille tutkimusaineistoille (haku Metaxiin tallennetun metadatan kautta)
- Lisää näkyvyyttä integroinnin (haravointi) ja indeksoinnin (esim. google)
- Tarjoaa pääsyn tutkimusaineistoihin IDA:n ja REMS:in kautta
- Käytetty teknologia: React.js, Nginx, Gunicorn, RabbitMQ, Elasticsearch, CKAN (harvesting)
- Kehittäjä ja ylläpitäjä: CSC

Etsin – tutkimusaineistojen hakupalvelu

HOME DATASETS ORGANIZATIONS HELP & ABOUT Add dataset fi

Search datasets

Search term

634 results Best Match

Organization ▼

Creator ▼

Field of Science ▼

Keyword ▼

Opus ECB Corpus Restricted Access

The corpus, containing the European Central Bank sub-corpus of the Opus open parallel corpus (<http://opus.lingfil.uu.se/>), will be made available for download at <https://korp.csc.fi/download/>

Corine Land Cover 2000 Restricted Access

CORINE Land Cover 2000 dataset provides information on Finnish land cover and land use on 2000. The data was produced as a part of the European CLC 2000 project. Dataset includes several spatial layers: • CLC raster (resolution of 25x25 m) • CLC vector (minimum mapping unit 25 hectares and

Fairdata-PAS – tutkimusaineistojen pitkäaikaissäilytyspalvelu

- Hallintaliittymä ja paketointipalvelu rakennettu yhteisen kansallisen pitkäaikaissäilytys-ratkaisun (PAS) päälle
- Organisaatiot hallitsevat tutkimusaineistoja hallintaliittymän kautta
- Paketointipalvelu on työkalu, jolla tutkimusaineistoista tehdään PAS-kelpoisia paketteja
 - IDA ja Metax yksi lähde, muutkin mahdollisia
- Käytetty teknologia: Angular.js, Flask, Luigi
- Kehittäjä ja ylläpitäjä: CSC



Fairdata-PAS – tutkimusaineistojen pitkäaikais säilytyspalvelu

Search datasets

Dataset Status

Waiting for Approval

In Preservation

In Ingestion

Rejected

All

Status	Dataset	Owner	Contract	Submitted	Size	Action
Waiting for approval (mid-term)	Wonderful Title		Title of Contract 1	2014-01-17T08:19:58Z	12 KiB	<input type="button" value="Preserve"/> <input type="button" value="Reject"/>
Waiting for approval (long-term)	Wonderful Title		Title of Contract 1	2014-01-17T08:19:58Z	12 KiB	<input type="button" value="Preserve"/> <input type="button" value="Reject"/>
In mid-term preservation	Wonderful Title		Title of Contract 1	2014-01-17T08:19:58Z	12 KiB	
In ingestion	Wonderful Title		Title of Contract 1	2014-01-17T08:19:58Z	12 KiB	

Mitä PAS tarjoaa?

- Aineistojen säilymisen ymmärrettävänä kymmeniä tai jopa satoja vuosia
 - Laitteistojen, ohjelmistot ja käyttötavat muuttuvat, mutta informaatio säilyy
- Aineistojen ja niiden metatietojen laadun tarkastuksen
- Paketointipalvelun ja hallintaliittymän
- Rajapinnat integrointiin organisaation omien järjestelmien kanssa
- Tukea aineistojen valmisteluun PAS-kelpoiseksi

Mitä PAS edellyttää?

- Aineistokokonaisuus on muille tutkijoille käyttökelpoinen
 - Kokonaisuuteen tulee sisältyä kaikki ymmärrettävyyden kannalta olennaiset tiedot, mukaan lukien tutkimusmenetelmien ja tiedostojen dokumentaatio
 - Tavoite: käytettävyys myös tieteenalarajojen yli
- Kokonaisuuteen kuuluvat tiedostot ja niiden suhteet on kuvattu PAS-ratkaisun määrittysten mukaisesti
- Voidaan tehdä kuvailutyökalun ja paketointipalvelun avulla tai itsenäisesti
- Tiedostot ovat hyväksytyissä tiedostomuodoissa
- Käyttöoikeustiedot ja lisenssit on ilmoitettu

Käyttöönoton vaiheet

1) Pitkäaikaissäilytettävän aineiston valinta

- Valitaan merkittävät aineistot (kansalliset arviointikriteerit) ja sovitaan mitkä niistä tulevat PASiin
- Valitaan säilytettävät osat aineistosta (käyttökelpoisuus)

2) Aineiston valmistelu PAS-kelpoiseksi

- Kokonaisuuteen kuuluvien tietojen koostaminen
- Tiedostomuotojen valinta yhteistyössä PAS-palvelun kanssa
- Kuvailu ja paketointi

3) Siirtopolun valinta ja järjestelmien integrointi

- Siirto IDA:n kautta vai suoraan PASiin, aineistojen käyttö

4) Testauksen kautta tuotantoon

Fairdata tunnistus ja luvitus (AAI)

- **Identiteetin hallinta ja linkitykset**
 - Useita tunnistusmenetelmiä, esim. Haka-federaatio
 - Kertakirjautuminen
 - Käytetty teknologia: Shibboleth IdP & SP, MPASS-ID, Open Id Connect
- **Oikeudet, roolit ja ryhmät**
 - Käytetty teknologia: Perun identiteetin ja pääsynhallintajärjestelmä
- **Aineistojen saatavuuden hallinta**
 - Pääsy avoimiin tai luvitettuihin aineistoihin hakemusten ja työvoiden kautta
 - Käytetty teknologia: REMS (kehitetty CSC:llä)
- **Latauskomponentti**

Jatko

- Fairdata.fi tuotantoon kesällä 2018
- Yhteistyön suunnittelu organisaatioiden kanssa
- Palvelun jatkokehitys
 - Parannukset ja uudet ominaisuudet
 - Uusia integrointimahdollisuuksia?
 - Tunnistusmenetelmät, haravoinnit, sensitiivinen data, yhteydet muihin palveluihin (laskentaympäristöt, EUDAT, EOSC, Elixir, tieteenalaspesifit palvelut ja infrat)
 - → Osallistu työpajaan!

Kiitos!

- Kysymyksiä?