



CSC

ICT Solutions for
Brilliant Minds



CSC:n palvelut ammattikorkeakouluille

TuAMK 12.2.2020



Päivän ohjelma

Klo 10-12 Datanhallinnasta ja CSC:n palveluista

Klo 10-13 Lounas

Klo 13-15 Työpaja

- Hyödynnetään mallia TKI-projektin elinkaaresta:
<https://verkkolehti.seamk.fi/index.php/arkisto/maaliskuu-2017-seamk/avoimen-tki-toiminnan-prosessimalli/>
- Osallistujamäärästä riippuen jakaudutaan pienryhmiin, joissa jokaisessa CSC:n edustaja fasilitoimassa
- TKI-projektin elinkaarimallin pohjalta tunnistetaan sekä olemassa olevia että puuttuvia palveluja, rooleja, prosesseja, yms.

Tässä esityksessä

1. Mikä on CSC?
2. Datanhallinnan merkitys ja hyvät käytännöt
3. CSC:n palvelut TKI-projektin elinkaaren vaiheisiin
4. Tietoa koulutuksista

CSC on korkeakoulujen ja valtion omistama yhtiö



Omistajuuden ansiosta

- ▶ CSC:n osaaminen on joustavasti ja tehokkaasti korkeakoulujen, tutkijoiden ja opettajien käytettävissä
- ▶ In house -toimijana tarjoamme tukea ja toteutamme ratkaisuja omistajiemme yhteisiin ja räätälöityihin tarpeisiin



Maksuttomia palveluja opettajille ja tutkijoille

- ▶ CSC:n laskenta-, pilvi- ja tallennuspalveluiden käyttö on maksutonta esimerkiksi suuren datamäärän jakamiseen yksittäisen kurssin aikana, opinnäytetöiden tekemiseen ja ohjaamiseen tai laskentatehoa vaativaan kurssityöhön

Kuvat: Thinkstock

Tavoitteet

1

Mahdollistaa maailmanluokan datanhallinta ja laskenta

2

Maksimoida datan arvo

3

Hyödyntää tekoälyn vipuvoima

4

Tasoiittaa tie jatkuvalla oppimiselle

CSC katalysoi asiakkaiden menestystä

- Luomalla lisäarvoa yhdessä asiakkaiden ja kumppanien kanssa
- Toimimalla läpinäkyvästi, ketterästi ja kokeillen
- Kehittämällä yhteentoimivuutta ja yhteistyötä
- Osoittamalla edelläkävijyyttä ja vastuunkantoa
- Vaikuttamalla paikallisesti ja maailmanlaajuisesti

Missio

CSC osana kansallista tutkimusjärjestelmää kehittää, integroi ja tarjoaa korkeatasoisia tietotekniikkapalveluja tutkimukselle, koulutukselle, kulttuurille, julkishallinnolle ja yrityksille

Visio 2020

CSC on asiakkaiden arvostama, kansainvälisesti korkeatasoinen digitaalisten palvelujen tarjoaja toimialallaan

Arvot

Yhdessä

Osaamme

Välitämme

Vastuullisesti

Lupaus

Valjastamme osaamisemme, verkostomme ja tietotekniikan edistämään asiakkaan menestystä ja koko yhteiskunnan etua

Korkeakoulujen ja CSC:n yhteinen ideapankki.csc.fi: löytää ideoille edistäjiä



Idean kirjaaminen
ideapankki.csc.fi



CSC:n palvelukehityslinjat

Idean esiselvitys ketterää kokeilukulttuuria hyödyntäen

Muut palvelukehityslinjat

Ideapankki kerää tutkimuksen ja korkeakoulutuksen palveluiden kehitysideoita, jakaa tietoa sidosryhmille ja välittää ideoita oikeille toteuttajille.

Ideoita voivat jättää, kommentoida ja seurata kaikki.





AVOIN TIEDE
JA TUTKIMUS



AMK
AVOIN TKI



Avoin tiede



Tieteellisten seurain
valtuuskunta



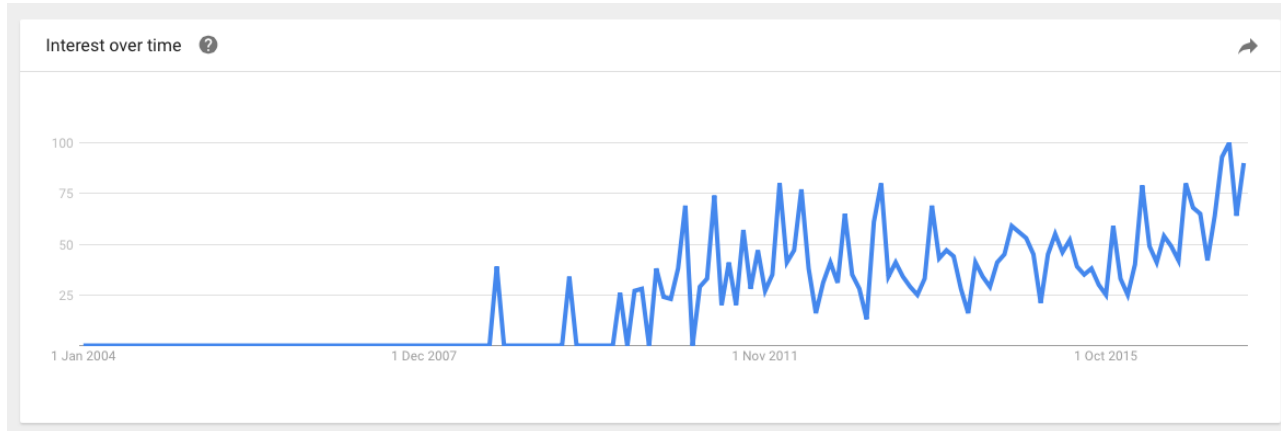
UNIFI

AVOIN TIEDE JA DATA

TOIMENPIDEOHJELMA
SUOMALAISELLE TIEDEYHTEISÖLLE

Mikä datassa kiinnostaa?

"Data on uusi öljy. Se on arvokasta, mutta jalostamattomana sitä ei voi käyttää. Se on muutettava bensiiniksi, muoviksi, kemikaaleiksi, ja niin edelleen, jotta syntyy jotain arvokasta joka synnyttää tuottavaa toimintaa; samalla tavalla dataa on käsiteltävä, analysoitava, jotta sillä olisi arvo." – **Clive Humby**, matemaatikko ja arkkitehti



Lainauksen ja kuvaajan lähde: <https://medium.com/project-2030/data-is-the-new-oil-a-ludicrous-proposition-1d91bba4f294>

Datan hyödyntämisen kaksi pullonkaulaa

What data scientists spend the most time doing

- Building training sets: 3%
- Cleaning and organizing data: 60%
- Collecting data sets: 19%
- Mining data for patterns: 9%
- Refining algorithms: 4%
- Other: 5%

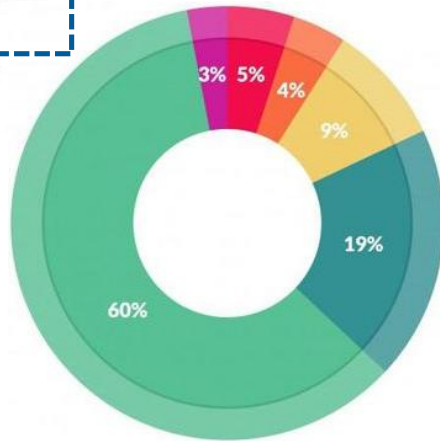
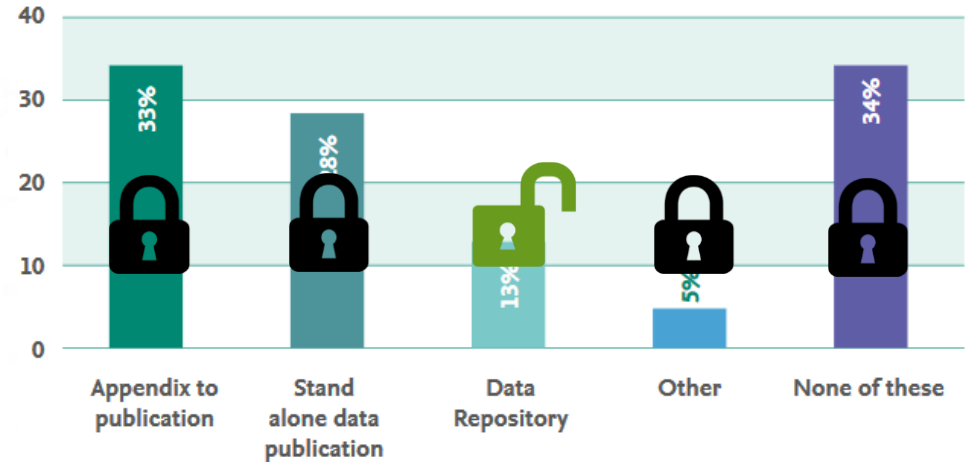


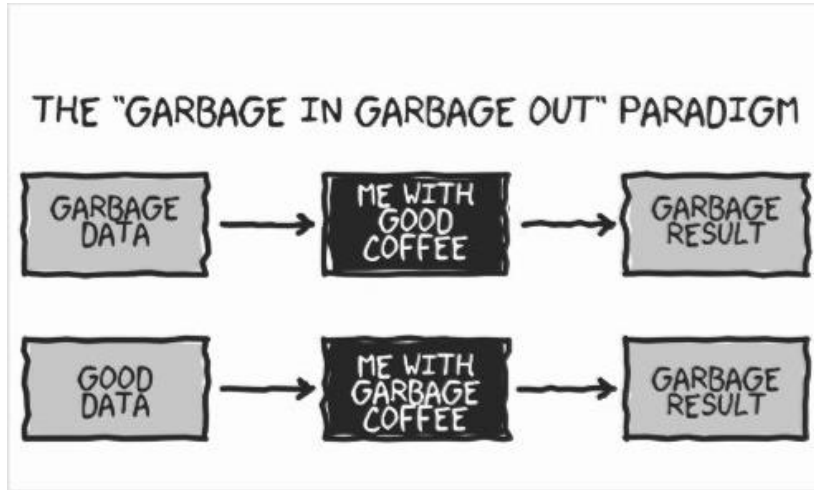
Figure 1. Dissemination of research data (% , n=1162)



Liide: <https://www.fakes.com/insights/2016/03/23/data-preparation-most-time-consuming-but-enjoyable-data-science-task-survey.aspx?h3473663>

Liide: <https://www.electro.com/about/open-science/research-data/open-data-report>

Data on avain digitalisaatioon



Data on digitalisaation ytimessä. Tekoälyn ja esineiden internetin kaltaiset uudet mahdollisuudet perustuvat *laadukkaaseen* dataan.



Dataan liittyvien työvöiden ja tietovirtojen hahmottaminen on avain CSC:n palveluiden täysimittaiseen hyödyntämiseen.

Data TKI-projektin elinkaaren vaiheissa

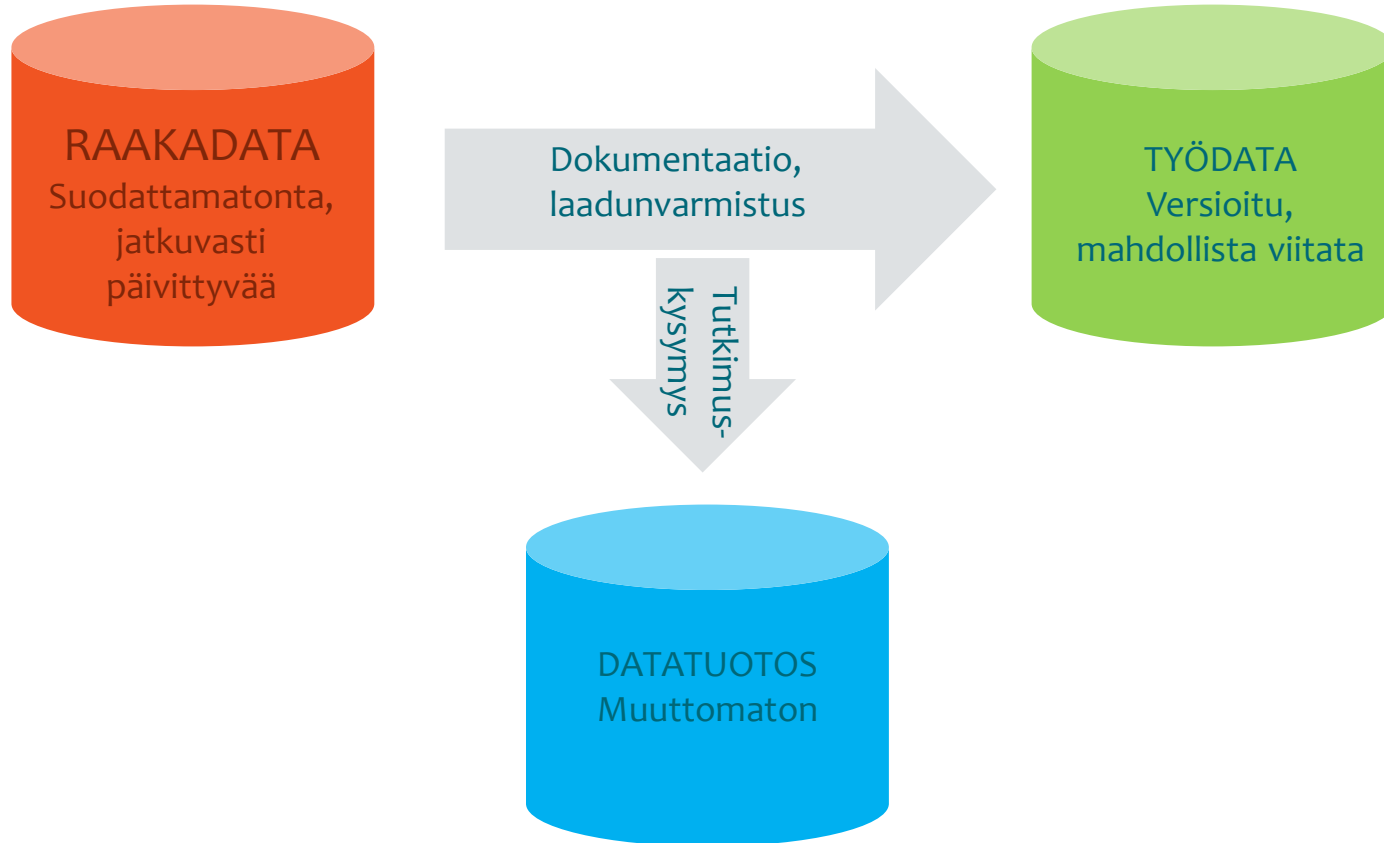


Kuvan lähde: <https://verkkolehti.seamk.fi/index.php/arkisto/maaliskuu-2017-seamk/avoimen-tki-toiminnan-prosessimalli/>

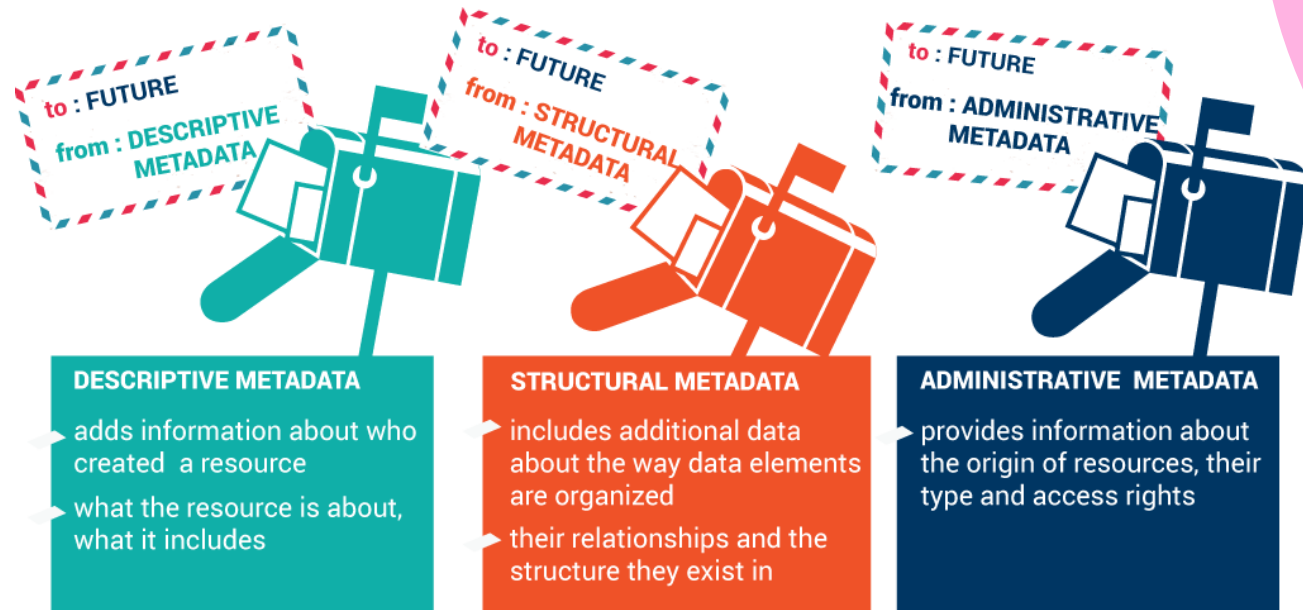
Mitä on tutkimusdata



Tutkimusdatatyyppiä



Mitä on metadata



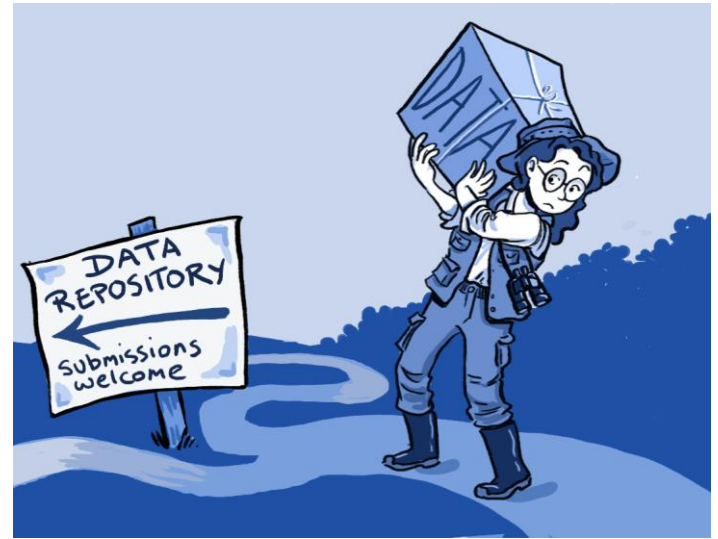
”Metadata on rakkauskirje tulevaisuuteen.”
– Jason Scott

Lähde:
<https://www.ontotext.com/knowledgehub/fundamentals/metadata-fundamental/>

Mitä on datanhallinta

Datanhallinta tarkoittaa kerätyn ja tuotetun information käsittelyä. Datanhallinta lisää datan arvoa tekemällä siitä yhteentoimivaa, helpommin löydettävää ja vaikeammin hukattavaa, ymmärrettävää ja todennäköisemmin ja helpommin uudelleen käytettävää. Datanhallinnan vakiintuneisiin käytäntöihin kuuluu mm. datanhallinnan suunnittelu, datan ja sen elinkaaren dokumentoiminen, järjestäminen, tallettaminen, jakaminen, julkistaminen, säilyttäminen, tarpeen tullen jopa tuhoaminen

More info: <https://wiki.eduuni.fi/x/4pioBg>



Kuvan lähde:
https://www.openaire.eu/images/easyblog_articles/1120/Registry-of-Research-Data-Repositories.png

Mistä tunnistaa arvokkaan, hyvin hallitun datan

Noudattaa
tutkimusalan
standardeja

Runsaasti
metadataa

Tunnisteellinen
ja viitattava

Tietoturvallisesti
ja kestävästi
säilytetty

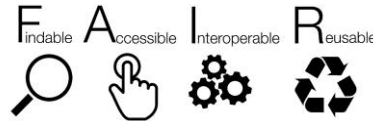


Versioitu (jos
muuttuvaa tai
kertyvää)

Asuu
yhteentoimivien
palveluiden
ekosysteemissä

Data

Lisensoitu





Linkki videoon: <https://youtu.be/K-kEvfaUJdA>

Niin avointa kuin mahdollista, niin suljettua kuin välttämätöntä



Perusteltuja syitä rajoittaa avoimuutta

- Sopimukset – esim. yritysten liikesalaisuudet ja immateriaalioikeuksiin liittyvät syyt
- Henkilötietosuojat – yksityisyyden kunnioittaminen
- Henkilökohtaiset syyt – informanttien suostumuksen asettamat rajat
- Kansalliset intressit – kansallinen turvallisuus, ulkosuhteet, jms.
- Oikeudenkäyttö – käynnissä olevat oikeusprosessit ja niihin liittyvät intressit
- Ympäristönsuojelu – rajoitettu pääsy luonnonympäristöjen suojelemiseksi
- Tekijänoikeus – rajoitettu pääsy tekijänoikeuden haltijan oikeuksien nojalla
- Kulttuuriset syyt – rajoitettu pääsy kulttuuristen arvojen ja/tai alkuperäiskansaoikeuksien turvaamiseksi
- Tutkimus – käyttö rajoitettu tutkimukseen sopimuksen nojalla
- Koulutus – käyttö rajoitettu koulutukseen sopimuksen nojalla
- Muut syyt – Lista ei ole tyhjentävä!

Sensitiivinen data

- Sensitiivistä dataa on mikä tahansa informaatio, jota halutaan suojella asiattomalta pääsylvä
- Datan suojaamisen perusteena voi olla esim. lakiin perustuva tai eettinen syy, yksityisyyden suoja tai aineettoman omaisuuden suoja



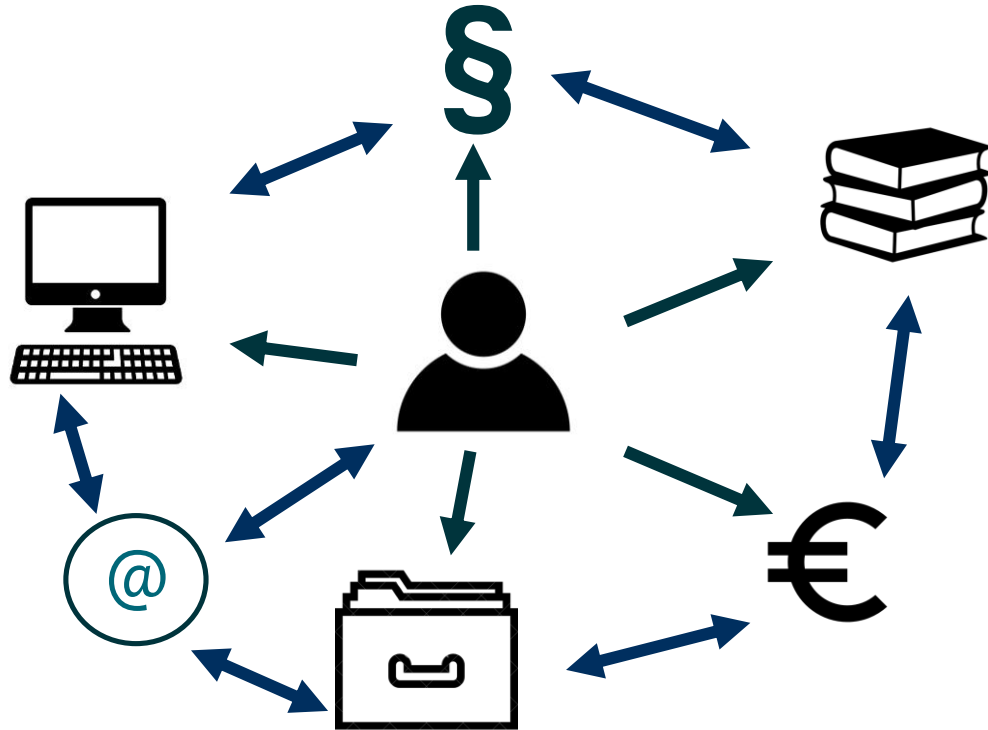
GDPR ja erityiset henkilötietoryhmät

Tällaisista tiedoista ilmenee henkilön

- rotu tai etninen alkuperä
- poliittisia mielipiteitä
- uskonnollinen tai filosofinen vakaumus
- ammattiliiton jäsenyys
- terveyttä koskevia tietoja
- seksuaalinen suuntautuminen tai käyttäytyminen
- geneettisiä ja biometrisia tietoja henkilön tunnistamista varten.

<https://tietosuoja.fi/erityisten-henkilotietoryhmien-kasittely>

Datanhallinnan tukeminen organisaatiossa



CSC:n datapolitiikka

Periaatteet

1. Datatn omistajuus on selvä ja datanhallinnan vastuista on sovittu.
2. Data on dokumentoitu ja datan elinkaari on määritelty
3. Datatn avoimuus ja suojaus hoidetaan asianmukaisesti.
4. Data on mahdollisimman avointa ja yhteentoimivaa.
5. Datanhallinnan ratkaisuisssa suosimme avoimia standardeja, protokollia ja ohjelmistoja.
6. Aineistonhallinnan osaaminen on tunnistettu ja sitä kehitetään näiden periaatteiden ja niiden toimeenpanon edistämiseksi.

Politiikka kokonaisuudessaan: <https://www.csc.fi/fi/data-policy>

CSC:n
palvelut ja
TKI-projektin
elinkaari

Valmistelu

Yhteistyö

Tutkimus

Datan
hallinta

Tulosten
avaaminen

Tiedolla
johtaminen

Liiketoimintaa



Valmistelu



DMPTuuli on työkalu aineistonhallintasuunnitelmien laatimiseen. OKM rahoittaa ja Helsingin yliopiston kirjasto ylläpitää DMPTuulia sekä Tuuli-verkostoa. CSC tukee DMPTuulin kehittämisessä.

Luodessaan käyttäjätunnuksen käyttäjä valitsee listasta oman organisaationsa ja saa itselleen räätälöityjä ohjeita. Tuulissa ohjeita myös eri tutkimusaloille ja sensitiiviselle datalle sekä pohja 10 tutkimusrahoittajan vaatimusten mukaiselle suunnitelmalle (esim. Business Finland).

Yhteistyö



Eduuni on palvelukokonaisuus joka koostuu workspace-alueesta (Microsoft SharePoint), wikistä ja Jira-tehtävienhallintatyökalusta. Ns. Software as a Service -pilvipalvelu. Auditoitu suojaustason IV aineistolle soveltuvaksi. Tallennustilaa oletuksena 100 Mt / käyttäjä, mutta voi kasvattaa.

Eduuni-tunnuksen voi luoda mm. Hakan ja Virtun lisäksi esim. Google- tai Facebook-tunnuksella, eli mahdollistaa organisaatio- ja maarajat ylittävän yhteistyön. Palvelu on organisaatioille maksullinen.

Funet FileSender on suurten tiedostojen lyhytaikaiseen ja kohdennettuun jakamiseen tarkoitettu selainpohjainen verkkopalvelu. Edellyttää organisaatiolta Funet-asiakkuutta. Kirjautuminen Haka-tunnuksella, vastaanottaja voi olla kuka vain. Pisin säilytysaika on 14 vuorokautta ja suurin tiedostokoko 300 gigatavua.

Eduuni-wikin sovellusesimerkki: <https://wiki.eduuni.fi/x/Q4Hy>

Tutkimus



Puhti on CSC:n uusin supertietokone ja osa uutta kansallista datanhallinnan ja laskennan ympäristöä yhdessä Altaan ja toukokuussa käynnistyvän Mahdin kanssa. Puhti-AI on tekoälysovelluksille ja tekoälytutkimukseen laskentaan suunnattu osio

cPouta ja **ePouta** ovat ns. Infrastructure as a Service -palveluita. Ne mm. mahdollistavat virtuaalikoneiden käytön etätyöpöydän kautta, eli eräänlaisen tietokoneen tietokoneessa. Virtuaalikoneeseen voidaan räätälöidä tarvittavat ohjelmistot. ePoudan pilottikäytössä oleva etätyöpöytä soveltuu sensitiiviselle datalle.

[Notebooks](#) on selaimen kautta käytettävä ohjelmointiympäristö jonne kirjaututaan Haka-tunnuksilla. Se on kehitetty ensisijaisesti opetuksen tarpeisiin, mutta sovellettavissa tutkimustarpeisiin. Opettaja voi rakentaa palveluun reseptin, jonka avulla kaikki oppilaat saavat käyttöönsä samanlaisen ympäristön.

Datan hallinta

B2DROP on eurooppalainen mutta CSC:n kehittämä henkilökohtainen pilvipalvelu, johon voi tallentaa ja jossa voi jakaa dataa tutkimusdatan elinkaaren varhaisessa vaiheessa.

Allas on tallennuspaikka työdatalle. Sen kapasiteetti on 12 Pb. Toistaiseksi se soveltuu parhaiten laskentaprojektin väliaikaiseksi datasäilöksi. Dataa ei versioda tai varmuuskopioida. Dataan voi jakaa pääsyn verkkolinkin kautta.

Fairdata-IDA on selaimen kautta käytettävä tallennuspaikka datatuotoksille. Pääsyä dataan voi jakaa 90-päivän väliaikaisilla linkeillä. Datalle voi määritellä salasanan. Organisaatioiden IDA-yhteyshenkilöt hyväksyvät IDA-hakemukset ja säilytyskiintiön koon.

Tulosten avaaminen



Linkki videoon: <https://youtu.be/g6pYZ4MkxRo>

Liiketoimintaa

Avoimesti ja helposti löydettävissä olevat laadukkaat aineistot ja projektitiedot johtavat todennäköisemmin liiketoimintaan, kuin vaikeasti löydettävät ja huonolaatuiset. Esimerkiksi Eduunin wikin kautta voi jakaa tietoa projektista ja Fairdatan kautta edistää aineistojen löydettävyyttä. Etsin-hakupalvelusta löytyvät aineiston hallinnoijan yhteystiedot ja tieto siitä onko aineisto käytettävissä kaupalliseen toimintaan.

Tiedolla johtaminen

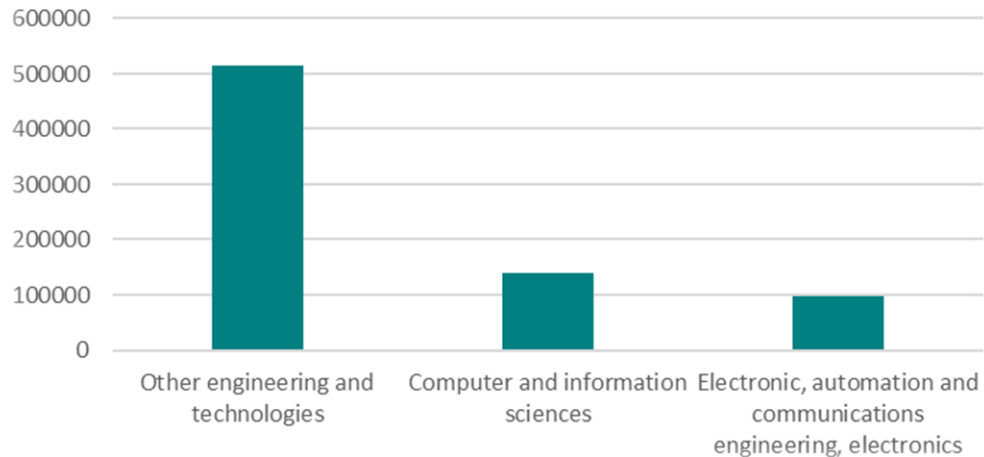


Tutkimustietovaranto kokoaa ja välittää tietoja esimerkiksi julkaisuista, tutkimusaineistoista, tutkimusinfrastruktuureista, tutkijoista ja tutkimushankkeista. Tutkimusta kuvaavia tietoja käytetään tutkimustyössä ja tutkimukseen liittyvässä hallinnollisissa prosesseissa, esim. tiedonhauissa, rahoitushauissa ja -raportoinnissa, julkaisutoiminnassa, tilastoinnissa ja muussa tutkimustiedon levittämisessä. Tutkimustietovarannon research.fi-portaali julkaistaan kesäkuussa 2020.

Justus on ammattikorkeakouluille kehitetty palvelu julkaisutietojen keräämiseen. Yhdistettynä Tutkimustietovarantoon Justusta voi hyödyntää tutkimustietojärjestelmänä (CRIS), joka helpottaa ja automatisoi tutkimustuotteiden seuranta ja raportointia organisaatiossa.

Turun ammattikorkeakoulun käyttötiedot

Turun ammattikorkeakoulun käyttötiedot (BU)
tieteenaloittain 2018-



TurunAMK:n projekteja, jotka hyödyntävät/ovat hyödyntäneet CSC:n palveluita:

AIRBORNE

SaVOC

Embedded Linux Workshop

Tietokannat2017

e3Power

Battery cell cooling UQ

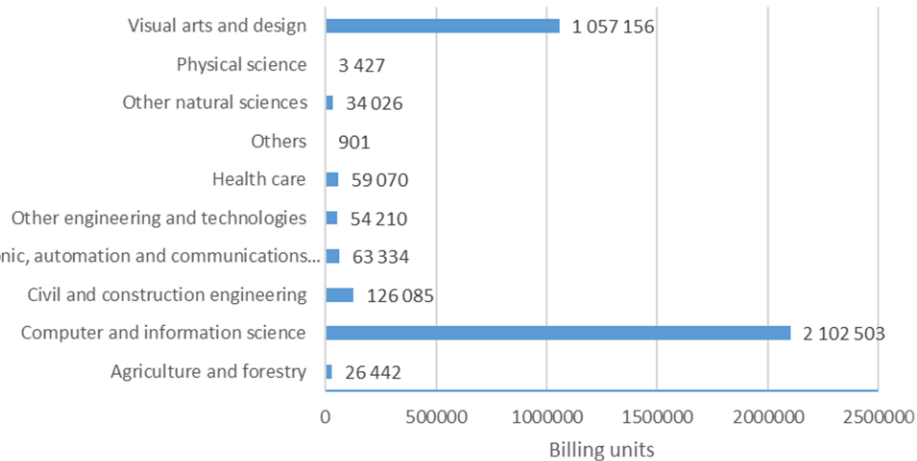
H₂-deflagration

FAN RPM CFD

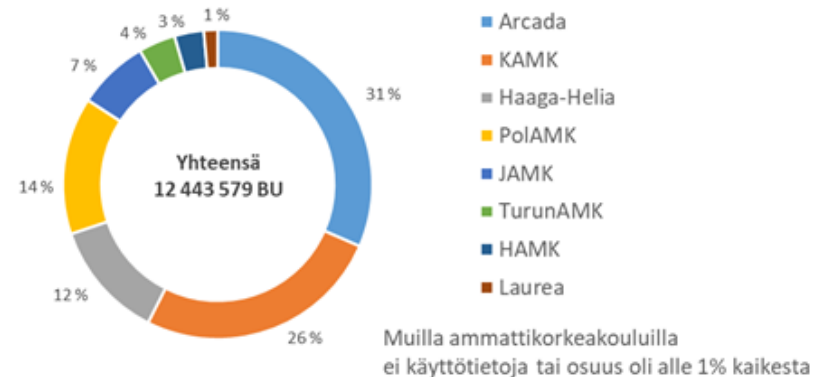
CNN training for signal classification

Ammattikorkeakoulujen käyttötiedot vuodelta 2019

Ammattikorkeakoulujen käyttötilastot tieteenaloittain 2019



Ammattikorkeakoulujen käyttötiedot 2019



Sovellusesimerkkejä

- Kerätään ja kehitetään osoitteessa <https://wiki.eduuni.fi/x/TySeBw>
- Esimerkki jossa hyödynnetty CSC:n datanhallinnan palveluja:
Kaleria-AMKin Ravinnerenki-projekti
<https://wiki.eduuni.fi/x/fCSeBw>

CSC:n palvelut opetuskäyttöön

CSC:n palvelut ovat saatavilla korkeakoulujen opetus- ja tutkimuskäyttöön samojen käyttöpolitiikkojen perusteella: se mikä on maksutonta tutkimukselle on maksutonta myös opetukselle. Lisäksi opetuksen tarpeisiin kehitetään omia palveluja, niistä lisätietoja:

<https://www.csc.fi/web/education/ratkaisut-oppijalle-ja-opettajalle>



Aki Reinimäki

Asiakasratkaisupäällikkö
Opetuksen ja koulutuksen palvelut
puh. +358 50 509 8719

Sopivan palvelun löytäminen ja käyttö



Csc.fi

Tietoa CSC:stä ja palveluistamme: millaisia palveluita on, mihin käyttöön ne soveltuvat, mitkä ovat käyttöehdot.



Fairdata.fi

Ohjeita datanhallintaan ja Fairdata.fi-palveluiden käyttöön.



My.csc.fi

CSC:n laskennan ja datanhallinnan palvelujen käyttäjäportaali. Täällä voit hakea laskenta-aikaa ja tallennustilaa.



Docs.csc.fi

CSC:n palveluiden yksityiskohtaiset käyttöohjeet.

Mitä koulutusta CSC järjestää?

- Fairdata-palveluiden käyttökoulutuksia organisaatioiden datatukihenkilöille ja IDA- Fairdata –yhteyshenkilöille
- Datanhallinnan perusteiden ja FAIR-periaatteiden koulutusta
- Organisaatiovierailuja, työpajoja
- Esityksiä seminaareissa
- Koulutukset loppukäyttäjille
- Verkkokoulutukset



ETUSIVU MIKSI FAIRDATA? FAIRDATA-PALVELUT PALVELUN HYÖDYNTÄMINEN KOULUTUS

Koulutukset

<https://www.fairdata.fi/koulutus/koulutukset/>

LISÄTIETOA

Koulutukset

Koulutuksen tallenteet

Ennakkotietoa Fairdata-koulutuksista:

Vuonna 2019 tullaan järjestämään useita koulutuksia / webinaareja koskien uudistuvia Fairdata-valmistunutta uudistunutta, integroitu palvelukokonaisuus Fairdata.fi, joka sisältää palvelut kaikkiin tu aineistojen haku, kuvailu, julkaisu sekä pitkäaikaissäilytys. Palvelut tukevat järkevää aineistonhalli



<https://www.csc.fi/web/training>

Koulutus ja tapahtumat

Tilaisuudet

Oppimateriaalit

Verkko-opiskelu

Webinaarit

Info

Aihepiiri



Tyyppi



Näytä avainsanat



Arkisto

Marraskuu 2019

19.11.

Webinar: Data migration from Taito to Allas (2/2)

The second webinar in the 2-part

19.11. - 20.11.

Practical Machine Learning

This course gives a practical introduction to machine learning,

27.11.

Webinar: From Taito to Puhti - Changes to users

Taito.csc.fi will be closed by the

27.11. - 28.11.

Big Data Analysis with Apache Spark

Spark is a high-performance

Tallenteet järjestetyistä koulutuksista

Koulutuksen tallenteet

LISÄTIETOA

Koulutukset

Koulutuksen tallenteet

Fairdata-palvelukoulutusten tallenteet kerätään tälle sivulle.

<https://www.fairdata.fi/koulutus/koulutuksen-tallenteet/>

Webinaarit ja tutoriaali-videot

Webinaari: [Webinaari: Fairdata PAS-palvelu](#), (41:41 min), [esitysmateriaali](#)

Webinaari: [Qvain \(Light\) - tutkimusaineistojen kuvailutyökalu](#) (42:03 min), [esitysmateriaali](#)

Webinar: [Qvain \(Light\) Research Dataset Metadata Tool](#) in English (36:53 min)

Webinar: [Metax - Metadata Warehouse](#) (22:04 min), [esitysmateriaali](#) (download m

Webinar: [Webinaari: FAIR-datan yhteentoimivuus osa 2](#), (43:30 min) [esitysmateriaali](#)

Webinar: [Vinkkejä datanhallinnan kouluttajille](#) (14:37 min), [esitysmateriaali](#) (down

Webinar: [FAIR-datan yhteentoimivuuden tukeminen](#) (50:38 min) [Esitysmateriaali](#) (

<https://www.csc.fi/web/training/arkisto>

CSC youtube Playlist Webinars



Ota yhteyttä!

HEIDI LAINE
Asiakasratkaisupäällikkö
Datanhallinnan palvelut

CSC - IT CENTER FOR SCIENCE LTD.
Keilaranta 14, P. O. Box 405,
FI-02101 Espoo, Finland
Tel. +358 40 513 95 93
heidi.laine@csc.fi
www.csc.fi



facebook.com/CSCfi



twitter.com/CSCfi



youtube.com/CSCfi



linkedin.com/company/csc--it-center-for-science



github.com/CSCfi