



PAS-yhteistyöryhmä

26.9.2023



Kokouksen avaus

- Todetaan läsnäolijat
- Esityslistan hyväksyminen
- Edellisen kokouksen muistio



Määrittelyiden päivittäminen



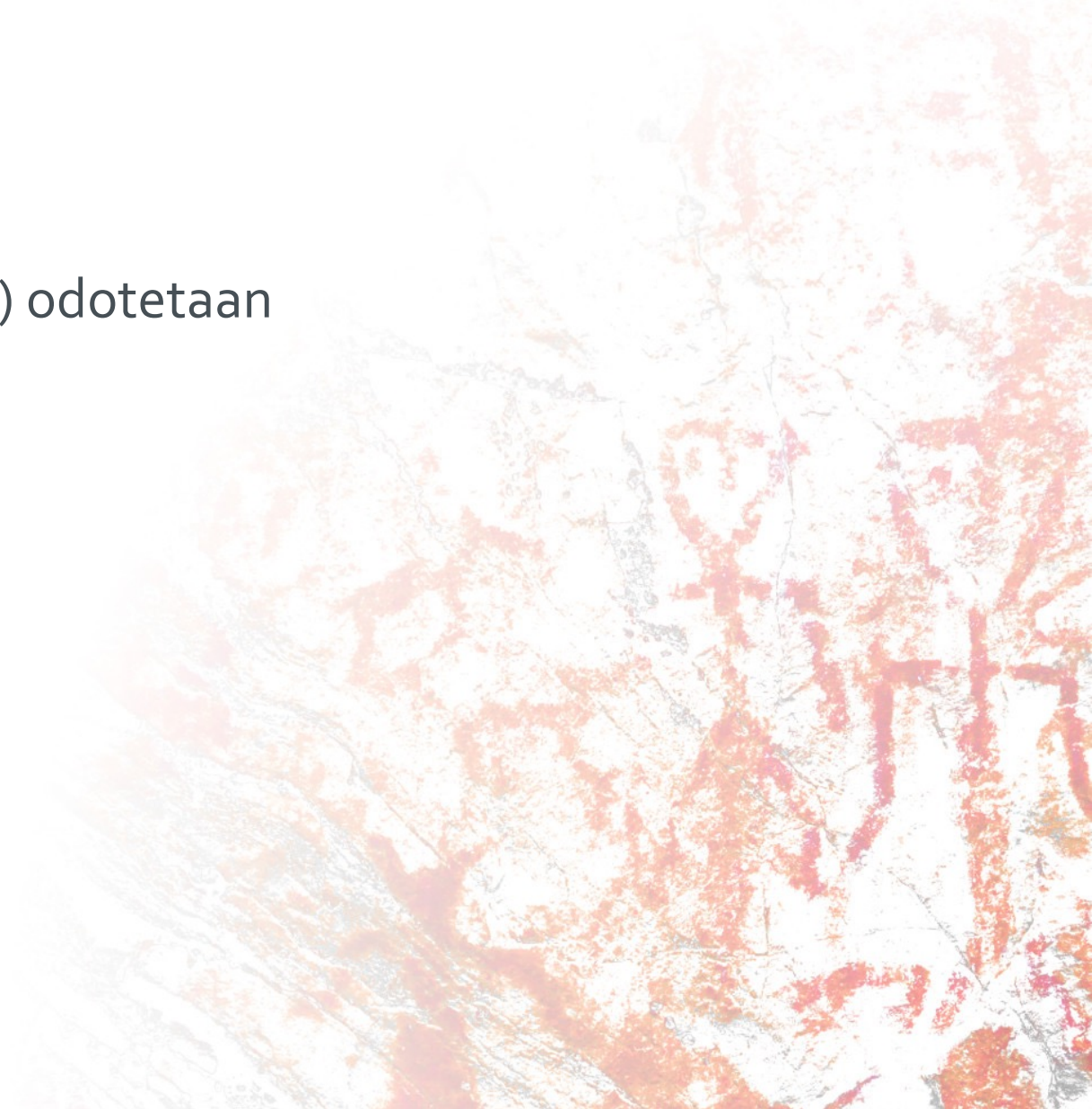
PAS-määrittelyiden päivittäminen

- Kysely käynnissä 5.6. – 30.8.2023
 - Kuusi vastausta
 - Tuttuun tapaan, palautetta saatu myös päivittäisessä PAS-työssä
- Dokumenttien päivitys
 - CSC päivittää
 - Päivitetyt versiot yhteistyöryhmälle kommentoitavaksi 28.11.2023 mennessä
- Hyväksyminen
 - Yhteistyöryhmä (12.12.2023)
- Julkaisu
 - Heti kun muutokset on toteutettu skeemoihin ja viety tuotantoon
 - Tavoite: Helmikuun aikana (muutokset tuotantoon maaliskuun huoltokatossa)

PAS-määrittelyiden päivittäminen

Metatiedot/paketointi

- Ei kommentteja
- Uutta paketointityökalua (dpres-siptools-ng) odotetaan



PAS-määrittelyiden päivittäminen

Tiedostomuodot



- Uusia tiedostomuotoja
 - WEBP, WEBM: Evaluoidaan
 - HEVC: Evaluoidaan, säilytyskelpoisen AVC:n seuraaja
 - AC-3: CSC avustaa normalisoinnissa
- Uusia versioita
 - IFC 4.3: Odotetaan lopullista versiota, versio 4.0.2.1 on jo säilytyskelpoisen.
- Muutoksia
 - DNG:n mukana ei tarvita TIFF- tai JPEG2000-tiedostoa
- Muuta
 - 3D-muodot: Suunnitellaan erikseen
 - Esimerkit ei-PAS-kelpoisen tiedostomuotojen paketointiin

PAS-määrittelyiden päivittäminen

Tiedostomuodot - jatkuu



- Lisäkysymyksiä:
 - **Siirtokelpoisten tiedostomuotojen arviointi:** lista on varsin vanha, tiedostomuodot on sovittu teknisessä jaoksessa hyödyntävien organisaatioiden kanssa heidän tarpeidensa perusteella
 - Keskeistä on, että niitä on paljon usealla organisaatiolla, jolloin keskitetty migraatio PAS-palvelussa on ajateltu järkeväksi ratkaisuksi
 - **Säilytyssuunnitelmat:** tarkoitus edistää näitä KV-yhteisön tuella tulevina vuosina
 - Loogisen tason säilyttämisen vaatimusmäärittely asemoi säilytyssuunnitelmat PAS-palveluissa
 - Sopivia malleja pohditaan kansainvälisesti
 - **Tiedostomuotojen tuki PAS-palveluiden vastaanotossa:** tukea toteutetaan tarvelähtöisesti
 - Kun on tiedossa, että tiettyä tiedostomuotoa on tulossa, kannattaa olla yhteydessä PAS-tukeen
 - Julkisen listan ongelmana on esimerkiksi tiedostomuotojen aliprofiilien kirjo
 - Lista voidaan toteuttaa esimerkiksi osana tiedostojen analysointityökalua

PAS-määrittelyiden päivittäminen

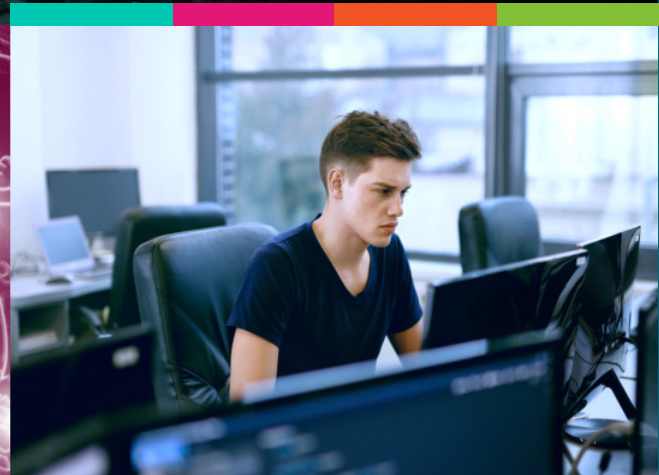
Rajapinnat

- Toivottu aineiston poistotoimintoa
- Overall statistics -ominaisuus ei ole tällä hetkellä toiminnassa

Ehdotukset työpajoiksi, koulutuksiksi, ohjeiksi

- Yleisesti PAS-asioiden esilläpito seminaarein
- Työpajat/Koulutukset
 - Optimoidut työprosessit aineistojen saattamisesta PAS-määrittelyiden mukaisiksi
 - Säilytyksessä olevan aineiston päivittäminen
 - Erilaisten tiedostomuotojen validointi ja PAS-kelpoiseksi muuttaminen
 - Kevään 2023 kaltaiset koulutukset
 - Esim. lyhytwebinaari, kun uusia työkaluja ilmestyy tai ne muuttuvat merkittävästi
 - REST-rajapinnan käyttö
- Ohjeita
 - AlmaLinux-asennus
 - Tapoja PDF-tiedostojen validointiin ja virheiden korjaukseen
- (Fairdata PAS: Tarkempi ohjeistus aineiston siirtoon ja noutoon sekä Metax API:n käyttöön)

Aineiston poistaminen



Aineiston poistaminen PAS-palvelusta

- Aineiston poistaminen PAS-palvelusta -toiminnallisuutta on toivottu yhteistyöryhmässä
- Aineiston poistamisessa joitain teknisiä haasteita
 - Säilytyspakettien kopioita ei kaikki voida poistaa lennosta tallennusinfrastruktuurista johtuen (nauhoja ei esim. voi/kannata ylikirjoittaa)
 - Pimeään arkistoon ei voi tehdä muutoksia
 - Kaikkien säilytyspakettien kopioiden sijaintitieto on pidettävä ajan tasalla
- Säilytyksessä oleva aineisto ei siis poistu PAS-palvelusta hetkessä, mutta poistettavan aineiston jakelu voidaan estää

Aineiston poistaminen PAS-palvelusta

- Suurempi haaste ovat kuitenkin hallinnolliset prosessit
 - Kuka (organisaation sisällä) saa tehdä aineiston poistopyynnön
 - Vaaditaanko kahdennettua hyväksyntää (toinen henkilö, joka hyväksyy pyynnön)
 - Vaaditaanko karenssiaikaa ennen kuin aineistoa varsinaisesti poistetaan
 - Vaaditaanko tiettyjä syitä tai perusteluita poistamiseen (jotka dokumentoidaan)
 - Poistetaanko vain tiettyä säilytyspaketin versiota (jos aineistoa on päivitetty)
 - Kuinka säilytysvastuun purkaminen poistettavan aineiston kohdalta pitää raportoida
- Onko poistamisessa aineisto- tai organisaatiokohtaisia erityisvaatimuksia?
- Mitä PAS-palvelun pitää (tai saa) dokumentoida poistetusta aineistosta?
- Onko tarvetta automaatiikalle?

Esimerkki: Aineiston poistaminen SAPA-palvelusta

- Aineiston poistaminen SAPAssa poistaa aineiston digitaalisten ilmentymien metatiedot, aineiston käyttökopiot, ja estää tallekappaleiden jakelun PAS-palvelusta
- Aineiston poistopyynnön voi luoda vain SAPAssa ylläpitäjän roolissa toimiva Kansallisarkiston työntekijä
- Aineiston poistopyynnön pitää hyväksyä toinen ylläpitäjä Kansallisarkistosta (omia poistopyyntöjä ei voi hyväksyä)
- Poistopyynnot ja niiden hyväksyminen/hylkääminen tehdään SAPAn käyttöliittymässä REST-rajapintoja hyödyntäen

Esimerkki: Aineiston poistaminen SAPA-palvelusta (jatkuu)

- Aineiston poistopyynnössä valitaan joku alla-olevista syistä:
 - Digitaalinen tietoaineisto on liitetty väärään kuvailevaan metatietoon
 - Digitaalinen tietoaineisto sisältää puutteita
 - Digitaalinen tietoaineisto sisältää tietoja, joita ei voida säilyttää tietoturvasyistä
 - Digitaalinen tietoaineisto on virheellisesti siirretty kaksoiskappale jo vastaanotetusta/säilytettävästä aineistosta
 - Digitaalinen tietoaineisto sisältää tietoja, joita ei voida säilyttää sopimussyistä
 - Digitaalinen tietoaineisto sisälsi tietoa, joka korjataan tai poistetaan tietosuoja-asetuksen perusteella
 - Muu syy, mikä:

Esimerkki: Aineiston poistaminen SAPA-palvelusta (jatkuu)

- SAPAan tallentuu seuraavat tiedot poistosta:
 - Poistopyynnön tekijä
 - Poistopyynnön hyväksyjä
 - Poiston syy
 - Aikaleimat
 - Aineiston tunnisteet
 - PAS-palvelussa olevat säilytyspakettien tunnisteet
 - Aineiston tila PAS-palvelussa (poistettu/ei poistettu)
 - Poiston status
- Tämä mahdollistaa tulevaisuudessa aineiston poiston PAS-palvelusta sitten, kun se on tuettu

Aineiston poistaminen

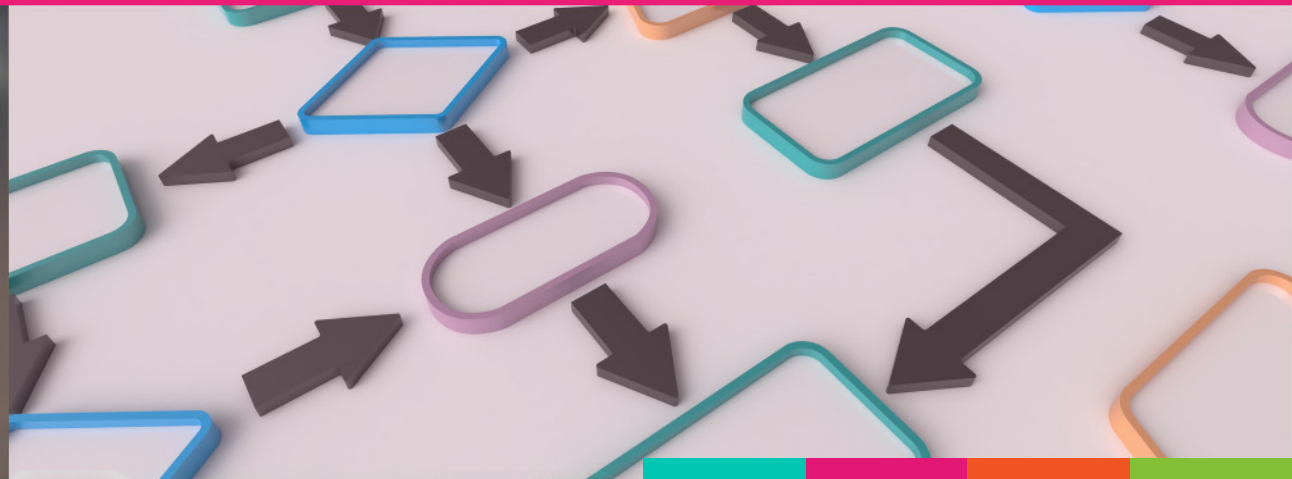
- Keskustelua aineiston poistamisen vaatimuksista ja käyttötapauksista hallinnollisen prosessin hahmottamista varten



Kysymyksiä keskustelun tueksi

- Hallinnolliset prosessit
 - Kuka (organisaation sisällä) saa tehdä aineiston poistopyynnön
 - Vaaditaanko kahdennettua hyväksyntää (toinen henkilö, joka hyväksyy pyynnön)
 - Vaaditaanko karenssiaikaa ennen kuin aineistoa varsinaisesti poistetaan
 - Vaaditaanko tiettyjä syitä tai perusteluita poistamiseen (jotka dokumentoidaan)
 - Poistetaanko vain tiettyä säilytyspaketin versiota (jos aineistoa on päivitetty)
 - Kuinka säilytysvastuun purkaminen poistettavan aineiston kohdalta pitää raportoida
- Onko poistamisessa aineisto- tai organisaatiokohtaisia erityisvaatimuksia?
- Mitä PAS-palvelun pitää (tai saa) dokumentoida poistetusta aineistosta?
- Onko tarvetta automaatiikalle?

Konferenssi kuulumisia



iPres 2023 -konferenssi 19.9–23.9

- Konferenssin teemat:
 - Digital Accessibility, Inclusion, and Diversity
 - Sustainability: Real and Imagined
 - We're All in this Together
 - From Theory to Practice
 - Immersive Information
- CSC:llä esitys PAS-palveluiden hiilijalanjäljestä *Calculating the Carbon Footprint of Digital Preservation - A Case Study*
- Myös METS 2 tutoriaali

Tiedostojen validointipalvelu

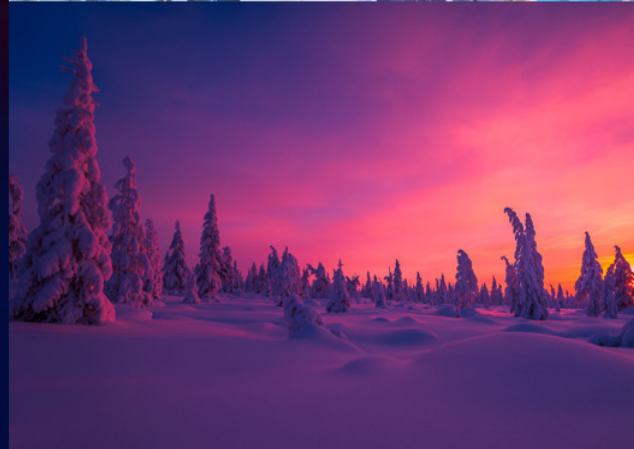


Tiedostojen validointipalvelu

- Demotaan uusi graafinen tiedostojen validointipalvelu
- Se julkaistaan osana PAS-palveluja piakkoin



PAS-palveluiden ohjelmistokerros



RHEL9 / AlmaLinux 9 -siirtymä

- RHEL:in ja AlmaLinuxin yhteensopivuuteen on tullut maailmalla mutkia
 - Jatketaan nyt kuitenkin entisellä suunnitelmalla
- PAS-palveluiden työkaluista on tarkoitus perustaa RPM-jakelu saataville
 - Tällöin asennus ja työkalujen päivittäminen on helppoa
 - Asennuksessa käytetään Linuxin omaa pakettinhallintaohjelmaa
 - Riippuvuudet asentuvat automaattisesti mukana
- RPM-jakelun toteutuksessa tulee huomioida monenlaisia asioita
 - Esim. 3. osapuolen työkalujen lisenssit
- Lisätietoa syksyn aikana

PAS-palveluiden uudet ominaisuudet Q3/2023

- ARC/WARC-migraation kehitystyö
 - Koemigraation suoritus
 - Työvuoro migroitujen koepakettien validointiin
 - Pieniä korjauksia
- Paketointi ja validointi
 - Metatietojen käsittely usean kuvan kuvatiedostoille
 - Bittitason säilytys ei-tuetulle muodolle, jos mukana on tuettu muoto
- Nauha- ja levytallennuksen migraatioiden edistyminen
- Tiedostomuotojen analysointityökalu
 - DPX-muodon validoinnin korjaus
 - Validointituloksen varmempi käsittely
 - Konfiguraatioiden muuttamismahdollisuus
- RHEL9-migraatiotyön edistyminen
- Työkalujen RPM-jakelun suunnittelu ja toteutuksen aloitus
- Bugikorjauksia ja muita kehitystöitä
- SAPA-kehitystyö omassa projektiryhmässään
- Fairdata PAS:
 - Hallintaliittymän käyttöliittymä uudistus
 - Aineiston siirtämisen prosessin muokkaus
 - Autorisointiuudistus
 - RHEL9-migraatiotyön edistyminen
 - Bugikorjauksia ja muita kehitystöitä

Sopimukset – KP-PAS (1/2)

(12.9.2023)



Organisaatio	Kapasiteetti (Tt)	Aineistoa (Tt)	Täyttöaste	Säilytyspaketteja
Celia	110	88,33	80,30	32 607
Kavi	2 400	1 602,53	66,77	2 022
Kansallisarkisto Kansallisarkiston vastaanottamat alkujaan digitaaliset valtiorhallinnon asiakirjalliset aineistot	41	1,56	3,80	1 971
Kansallisarkisto VAPA-järjestelmään siirretyt tietoaineistot	1	0,14	14,00	389
Kansallisarkisto Kansallisarkiston massadigitointi-hankkeen aineistot	114	92,64	81,26	77 360
Kansallisarkisto Kansallisarkiston digitaaliarkistosta siirrettävät aineistot ja takautuvan digitoinnin aineistot	805	302,60	37,59	586 540
Kansallisarkisto Kansallisarkiston yksinomaan digitaalisessa muodossa olevat yksityisarkistoaineistot	27	0,00	0,00	0

Sopimukset – KP-PAS (2/2)

(12.9.2023)



Organisaatio	Kapasiteetti (Tt)	Aineistoa (Tt)	Täyttöaste	Säilytyspaketteja
Kansallisgalleria Kiasman mediataiteen teosten pitkäaikaissäilytys	20	6,32	31,60	478
Kansalliskirjasto Kansalliskirjaston digitoimat kulttuuriperintöaineistot	355	224,46	63,23	668 571
Kansalliskirjasto Kulttuuriaineistolain nojalla kerätyt aineistot	175	201,32	115,04	1 982 369
KOTUS	60	9,83	16,38	407
Museovirasto	1	0,53	53,00	40 186
Musiikkiarkisto	70	7,60	10,86	123
Svenska litteratursällskapet i Finland	50	1,14	2,28	166
Yhteiskuntatieteellinen tietoaarkisto	1	0,06	6,00	9 217
Yhteensä	4 230	2 539,06	60,03	3 402 406

Säilytyspäätökset – FD-PAS (1/2)

(12.9.2023)



Organisaatio	Kapasiteetti (Tt)	Aineistoa (Tt)	Täyttöaste	Säilytyspaketteja
Geologian Tutkimuskeskus GTK:n tomografialaitteen tuottamat tietoaineistot	12	6,72	56,00	142
Helsingin yliopisto FIRE (The Finnish Reflection Experiment)	1	0,00	0,00	0
Helsingin yliopisto Helsingin yliopiston SMEAR-aineistojen valikoima meteorologisia - ja ilmanlaatumittauksia	2	0,01	0,50	13
Helsingin yliopisto M. cinxia and C. melitaeorum in the Åland metapopulation system	2	0,00	0,00	1
Helsingin yliopisto Luomuksen aineistot	150	0,00	0,00	0
Itä-Suomen yliopisto SENSOTRA	1	0,00	0,00	0

Säilytyspäätökset – FD-PAS (2/2)

(12.9.2023)



Organisaatio	Kapasiteetti (Tt)	Aineistoa (Tt)	Täyttöaste	Säilytyspaketteja
Jyväskylän yliopiston kiihdytinlaboratorio 250-Nobeliumin hajoamisspektroskopia	1	0,00	0,00	0
Oulun yliopisto, Sodankylän geofysikaalinen observatorio <u>Havaintoaineistot</u>	30	0,00	0,00	0
Tampereen yliopisto Yhteiskuntatieteiden tiedekunnan Kansanperinteen arkiston A-K-kokoelma	2	1,08	54,00	5 361
Turun yliopisto Historian, kulttuurin ja taiteiden tutkimuksen arkiston aineistot (HKT-arkisto)	20	0,25	1,25	1
Åbo Akademi Samlingar vid Åbo Akademis bibliotek	10	0,00	0,00	0
Yhteensä	231	8,06	3,49	5 518

Ilmoitusasiat

- PAS-kokonaisarkkitehtuuri julkaistu
 - <https://www.digitalpreservation.fi/documents/enterprise-architecture>
- 21.8.-23.8. Research Service Days 2023, Aalto-yliopisto
 - Esitys: On Preserving Research Data - What to Consider Before Ingesting Data?
- 4.10.-5.10. Symposium: The Born-Digital Challenge in Private Archives: Research Perspectives and Archival Practices, SLS

Kokouksen päättäminen

- Muut asiat
- Seuraava kokous
 - Joulukuu: Tiistai 12.12. klo 12:30-15:00
- Kokouksen päättäminen

