

*Tervetuloa kuuntelemaan CSC:n For a Better Future -podcastia!*

### **Teemu Roos**

Tänään keskustellaan datasta – nykypäivän kullasta tai öljystä, ajan arvokkaimman raaka-aineen hallinnasta ja hallinnan merkityksestä tulevaisuudelle ja tulevaisuudessa. Mun nimi on Teemu Roos ja mä olen hostina tässä jaksossa. Ja meillä on vieraana Aalto-yliopiston Metsähovin radio-observatorion johtaja, tähtitieteilijä Joni Tammi ja Jessica Parland-von Essen, joka on tutkimushallinnan kehityspäällikkö CSC:llä. Tervetuloa Jessica ja Joni!

### **Joni Tammi**

Kiitos!

### **Teemu Roos**

No jos ihan aloitetaan siitä, että mitä te teette työksenne? Haluatko Joni aloittaa? Metsähovin radio-observatorio, mitä siellä tapahtuu ja mitä sä teet siellä?

### **Joni Tammi**

Joo. Metsähovin radio-observatorio on nykyään Suomen ainoa tähtitieteellinen ammattitason radio-observatorio tai tähtitieteellinen observatoriahavaintolaitos ylipäänsä, ja me tutkitaan siellä avaruutta. Lyhyesti, no oikeastaan kolme tutkimuskohdetta, mihin me keskitytään ovat auringon muuttuvuus ja pitkäaikainen toiminta, toisena erilaiset mustat aukot, aktiiviset galaksit ja sen sellaiset, ja sitten kolmantena ovat erilaiset radioteleskooppeiverkostot, kun me pistetään radioteleskooppeja ja observatorioita ympäri maailmaa yhteen, niin me saadaan siitä sellainen maapallon kokoinen mittalaitte, mikä sitten mahdollistaa ihan uusia juttuja. Mutta tosiaan radioastronomia on se meidän ala eli radioaaltoilla tehtävä tähtitiede ja. Ja sitten myös ylipäänsä avaruustutkimukseen liittyvät asiat, ja Aalto-yliopistossa radioastronomian ja tähtitieteen opetus.

### **Teemu Roos**

Eli paljon on muuttunut. Me ei olla hirveän kaukana tuosta Tähtitorninmäestä, ja siellä on jotain observatoriota ollut. Niissäkö nyt sitten ei enää kaukoputkilla tähyillä sinne taivaaseen?

### **Joni Tammi**

Sinne pääsee katsomaan observatoriomuseoon ja tutustumaan vanhaan mittalaitteistoon ja muuhun, mutta tosiaan Suomessa ei enää ole muita tähtitieteellisiä mittalaitteita. Se johtuu toisaalta meidän tästä ilmastosta, sillä ilmakehässä on yleensä aika paljon vettä sumuna, pilvinä tai muuten tuollaisena räntänä. Radioaaltoja esimerkiksi ne pilvet ja muut ei haittaa, eli me pystytään vielä toimimaan ja tekemään hyvälaatuista tutkimusta, mutta kaikki muu tähtitiede tai mittalaitteet oikeastaan kannattaa nykyään viedä jonnekin aavikolle tai vuoren huipulle.

### **Teemu Roos**

No jos me nyt puhutaan sitten datasta tänään... mitä se data sitten on? Onko se jotain kohinaa ja suhinaa, mitä radiosta voi nuppia kääntämällä saada korviinsa?

### **Joni Tammi**

Joo, tavallaan on. Se meidän suhina on sellaista... mikähän voisi olla hyvä vertaus sille meidän radion herkkyydelle? Nämä meidän kohteet ovat niin kaukaisia ja niin hiljaisia mitä me yleensä tutkitaan, että jos me vietäisiin tällainen normaali kännykkä kuuhun, niin se olisi maasta katsottuna maailmankaikkeuden toiseksi kirkkain säteilijä. Eli ainoastaan aurinko olisi sitä kirkkaampi. Kaikki muut on niin äärimmäisen kaukana, puhutaan miljardeista valovuosista. Mutta loppujen lopuksi meidän mittalaite on tavallaan suunnattava radio.

### **Teemu Roos**

Okei. Tää on aika pysäyttävä ajatus, että kuussa oleva kännykkä blokkaisi muita lähteitä avaruudessa.

### **Joni Tammi**

No just tän takia muun muassa tää meidän alue on radiohiljainen, ja meillä ei saa olla mitkään kännykät päällä. Meillä ei saa käyttää langatonta verkkoa koko talossa, mulla on 029-alkuinen lankapuhelinnumero kuten kaikilla meillä, ja ei saa olla langattomia hiiriä... Ei saa olla mitään tuollaista. Meillä ei saa käyttää mikroaaltouunia tiettyinä viikkoina vuodesta, ja kun me ostetaan led-lamppuja tai tai -valaisimia, niin meidän pitää tutkia niitä ensin viikko, että ne ei aiheuta liikaa radiohäiriötä.

### **Teemu Roos**

Ei tule sitten tutkittua, että mikäs avaruusolento siellä piipittää ja osoittautuu, että se olikin omassa taskussa oleva joku...

### **Joni Tammi**

Joo näin kävi kollegoille tuolla Australiassa muutama vuosi sitten, niin sen jälkeen ollaan oltu vähän tarkempia mikroaaltouunien kanssa.

### **Teemu Roos**

Kuulostaa muutenkin mielenkiintoiselta paikalta, että jos siellä voi sitten pitää tämmöistä retriittiä, että ei ainakaan puhelin piipittele.

### **Joni Tammi**

Joo, monet tiimit haluaa tulla nimenomaan sinne pitämään virkistyspäivää tai kehityspäiväänsä, koska kukaan ei ole puhelimen tavoitettavissa sinä päivänä.

### **Teemu Roos**

Noniin, tässä kaikille protip, eli sinne vaan retriittiin. Ja teillä on kuulemma siellä joku sellainen hälytys, joka tulee sitten, kun näitä vierailijoita tulee kännykät taskussa, joka lähtee soimaan.

## **Joni Tammi**

Joo, aina silloin tällöin jollekin unohtuu puhelin, tabletti, älykello tai muu vastaava päälle. Unohtuu mennä lentotilaan alueen portilla, niin sitten meillä lähtee semmoinen 80 desibelin varoitusääni sieltä muistuttamaan tästä, että nyt kännykkä pois.

*Varoitusääni: "Interfering radio signal detected! Turn off all mobile devices! Interfering radio signal detected! Turn off all mobile devices!"*

## **Teemu Roos**

Jessica, sun toimenkuvaa CSC:llä liittyy niin sanottu FAIR data. Kuulostaa reilulta, mutta voitko avata vähän tarkemmin, että mitä tällainen FAIR data tarkoittaa ja mitkä ne periaatteet on siinä?

## **Jessica Parland-von Essen**

Joo, FAIR on agronomi, joka tulee sanoista findable, accessible, interoperable ja reusable, eli löydettävä, saavutettava, yhteentoimiva ja uudelleenkäytettävä. Ja ne on sellaiset periaatteet, joista on sovittu, että näitä kohti pitäisi pyrkiä kun tuotetaan dataa, eli se käytännössä esimerkiksi findable FAIR:ssa tarkoittaa sitä, että sillä datasetillä on tunniste, jotta sen pystyy identifioimaan, joka on pysyvä, ja sitten sillä on kuvailuja - metatietoja, eli jotain tietoja, mikä kertoo siitä, että mitä se on.

## **Jessica Parland-von Essen**

Ja sitten saavutettavuus liittyy siihen avoimuuteen tässä yhteydessä, että se on avointa sekä ihmisille että koneille. Itse asiassa koneluettavaa, mutta että myös ihminen pystyy ymmärtämään mistä on kyse. Ja interoperable - se yhteentoimivuus on ehkä kaikkein vaikein, ja mä mietin tuossa kun kerroit siitä, että että tällaista samanlaista dataa tai samantapaista dataa kerätään muualla, että niitä pitäisi sitten pystyä oikeasti yhdistämään niitä datasettejä oikealla tavalla. Siitä ois kiva kuulla kanssa.

## **Teemu Roos**

Tässä tulee sellaisia termejä mieleen kuten MyData ja avoin data, nyt on FAIR data. Liittyykö nämä jotenkin toisiinsa?

Jessica Parland-von Essen

Liittyy, liittyy. FAIR data on avointa yleensä joo, mutta sitten toisaaltahan meillä on paljon tutkimusta, jota ei voi avata eri syistä. Se saattaa liittyä johonkin uhanalaisiin lajeihin tai henkilötietoihin tai johonkin kaupalliseen yhteistyöhön. Joten aina se ei voi olla avointa silleen, että se olisi niinku avoimesti verkossa, mutta silti se avoimuus siinä yhteydessä voi tarkoittaa sitä, että se on esimerkiksi just koneluettavaa.

## **Teemu Roos**

Luettelit englanninkielisiä termejä siinä, niin tämä kuulostaa kansainväliseltä aloitteelta. Eli se ei oo ihan oma keksintö.

### **Jessica Parland-von Essen**

Joo, tää on kansainvälinen yhteistyö ja tämä on hyvin vahvasti keskiössä myös Euroopassa, kun me rakennetaan tutkimusinfrastruktuureja Euroopan tasolla.

### **Teemu Roos**

Ja sä puhut tutkimusdatasta erityisesti eli ilmeisesti on jotain etuja, minkä takia tai miksi kaikkien tutkijoiden tai kaikki kaikkien, jotka käsittelee tutkimusdataa, kannattaisi tästä innostua?

### **Jessica Parland-von Essen**

Mä oon itse sitä mieltä, että on ehkä vähän vaikea ajatella sellaista tilannetta, että sä esittelet jonkun tutkimustuloksen, mutta et sitä evidenssiä tai sitä tausta-aineistoa mitä siinä on takana. Jotenkin tänä päivänä myös ainakin kuuluu mun mielestä hyvään tieteelliseen käytäntöön, tai ikään kuin vastuulliseen tieteen tekemiseen se, että sä hallitset sitä dataa ja sä pystyt esittelemään sitä ja jakamaan sitä, jos joku haluaa tarkistaa tai käyttää uudestaan sitä. Mutta kyllähän siihen liittyy myös muita asioita kuten toistettavuus, että sitä koko tutkimusta pystyy validoimaan sillä tavalla, ja sitten myös tutkijalle tietenkin semmoinen näkyvyys, että jos joku uudelleenkäyttää sun dataa, niin myös sä saat viittauksia siitä. Että siinä on monia, monia hyötyjä.

### **Teemu Roos**

No siis tota mä just itekin ajattelin, että tää varmaan on hyvä keino saada muutkin innostumaan siitä tutkimusaiheesta ja omasta tutkimuksesta, että varmasti jokaiselle tutkijalle ihan hyvä linkki.

### **Joni Tammi**

Mä voisin tohon sanoa vielä yhden puolen, eli se että joku muu katsoo ihan uudesta perspektiivistä sitä dataa. Jos itse katsoo ja oma tutkimusryhmä, ja esimerkiksi mun oma tutkimuslinja katsoo ja haluaa käyttää sitä dataa tavallaan omilla tutuilla menetelmillä ja omasta tutusta perspektiivistä, niin sitten kun se annetaankin täysin vaikka toisen alan ihmiselle... sanotaan vaikka, että meidän aurinkodata voi olla tosi mielenkiintoista taiteilijalle tai ilmakehäutkijalle. Ei meillä ollut mitään näkemystä siitä, miten sitä voisi käyttää, eikä meille ollut tullut mieleenkään sellainen. Mutta kun joku muu saa käsiinsä tai kuulee, että täällä on tällaista dataa käytettävissä, niin se voi heille olla tosi arvokasta sitten sellaisella tavalla mitä me ei oltaisi edes ikinä osattu tarjota.

### **Jessica Parland-von Essen**

Joo ja sitten vielä se, että hirveän usein tutkijalla menee ihan tajuttoman suuri työmäärä sen datan tuottamiseen. Siis paitsi että on kalliit laitteet, mutta myös joskus sen validointiin, siistimiseen ja organisoimiseen saattaa mennä helposti 80% työajasta tai resursseista. Ja onhan se sitten hassua, jos sen vaan heittää pois, varsinkin kun tutkimus tavallaan etenee koko ajan ja tavallaan ikään kuin vanhenee, että ne artikkelit ja tulokset saattaa muuttua, mutta se datahan on aina tavallaan ikuista, jos se on tarpeeksi hyvin dokumentoitu.

## **Teemu Roos**

Mä oon kyllä kuullut on 80 % monessa yhteydessä. Me ollaan kutsuttu aika usein datatieteilijöitä teollisuudesta puhumaan meidän opiskelijoille ja kertomaan, että millaista se datatieteilijän työ on. Ja sit yks kysymys, minkä mä kysyn aina ja laitan vastaukset sitten muistiin, eli kerään tällaista ikään kuin tilastoa siitä, niin kuinka paljon datatieteilijän työstä kuluu siihen datan putsaamiseen ennen kuin sen voi sitten tuupata jonnekin kivaan tilastotyökaluun tai koneoppimistyökalun sisään. Ja tämä 80 % on varmaan joku mediaani itse asiassa niistä luvuista mitä mä oon kerännyt. Ehkä pienin oli jotain tyyliin 60 %, mutta kyllä hyvin usein se on luokkaa 80, joskus se on ollut 95 %. Kyllähän se on älytön määrä työtä. Ja se on tosi hienoa, jos sitten sitä työtä tosiaan ei tarvitse heittää hukkaan ja jos siihen saa jopa sitten apua.

*For brilliant minds - For a Better Future!*

## **Teemu Roos**

Mut hei Joni, mä tiedän, että te olette saaneet viime vuonna tällaisen Avoimuuden edistäjät - palkinnon, kun me puhuttiin tässä avoimestakin datasta. Minkälainen palkinto se oli ja millä meriiteillä sellaisia palkintoja voi käydä pokkaamassa?

## **Joni Tammi**

Meille se annettiin... perustelu oli muistaakseni ainutlaatuisen kasvavan data setin tuomisesta kansainvälisen ja kansallisen tutkijayhteisön käyttöön. Eli juuri tämä, että meillä on paljon pitkäaikaista dataa. Ja sitten tämä mihin mä viittasin, että sen sijaan, että me itse käytetään sitä omaan tutkimuksemme, niin todetaan, että katsotaan mitä muut saa tästä irti omilla resursseilla ja omista lähtökohdistaan ja omista työkaluistaan. Se data, mitä me avattiin, oli oikeastaan koko Metsähovin radio-observatorion historian aikana kerätty, ja vieläkin kerättävä joka päivä, kuten aurinkodata. Eli me ollaan tehty -78 vuodesta lähtien oikeastaan parhaimmillaan päivittäin aurinkohavaintoja eli havaitaan radioteleskoopeilla, ensin vaan yhdellä - meidän isoimmalla, 14 metrisellä – tavallaan karttoja auringosta ja kuvia auringon radiokirkkaudesta ja rakenteesta. Ja sitten siellä on tullut vuosien aikana kaikkea muuta uutta mittalaitetta mukaan.

## **Joni Tammi**

Ja sitten tossa viimeiset 10-15 vuotta se on ollut oikein aktiivista, että on kerätty valtava määrä dataa. Sen tuominen sellaiseen muotoon, että se saadaan tuota... oikeastaan juuri tähän, mitä Jessica sanoi näistä FAIR-periaatteista, että se on löydettävissä ja ymmärrettävissä.

## **Joni Tammi**

Meillä oli muun muassa ongelma oli se, että datasta iso osa oli vanhalla tällaisella kynäpiirturilla tehtyä. Eli ei printterillä, vaan se, missä ikään kuin robottikäsi siirtää kynää paperilla edestakaisin ja piirtää sellaista karttaa. Sitten niiden päälle oli ihmiset laskeneet kahvikuppeja ja kirjoittaneet lyjiykynällä omia muistiinpanoja ja muita. Ja näistä oli alkuperäiset digitaaliset datat kokonaan kadonnut.

## **Joni Tammi**

Meillä oli nää kyllä myös skannattuna PDF:nä, mutta se ei vielä paljon auta, että meidän piti kehittää siihen itse asiassa juuri tällainen koneoppimiseen ja konenäköön perustuva kokonainen uusi softa, joka pystyy lukemaan ja tulkitsemaan ne ja muuttamaan moderneja karttoja vastaavaksi dataksi. Se oli aika mielenkiintoinen, useamman opinnäytetyön projekti. Ja se ei ole mitään sellaista ikään kuin helposti pakattavaa tai kompressoitavaa dataa, vaan se on käytännössä kohinaa niin kauan, että se ollaan saatu käsiteltyä.

## **Teemu Roos**

Kuvittelisin, että on nimenomaan haasteellista tavallaan tiivistää tai heittää mitään pois, jos ei ihan tarkkaan tiedetä, että mitä siitä aiotaan myöhemmin pyrkiä selvittämään. Että ei oo vahingossa menty heittämään pois just sitten sitä tärkeitä tiedon palasta.

## **Joni Tammi**

Joo, ja näissä yhteismittauksissa käytännössä se data on kohinaa niin kauan, että se saadaan käsiteltyä. Sillä ei ole itsessään edes mitään merkitystä, se data vasta muodostuu siinä niiden eri teleskoopeilta tulleiden datojen korreloinnissa tai ikään kuin yhteismitoittamisessa. Mutta sitten just tämä, mitä mä puhuin vaikka noista meidän radiohäiriöistä tai muusta, niin esimerkiksi tällainen aktiivinen tutkimusongelma on se, että jos me vaikka löydettäisiin ei pelkästään ne mielenkiintoiset kohdat, missä tulee vaikka häiriöitä tai missä vaikka auringossa tapahtuu joku purkaus, vaan ihan nekin päivät tai tunnit, jolloin mitään ei tapahdu, että me voidaan huoletta deletoida ne, niin tämä on tavallaan monisyinen ongelma, että tavallaan halutaan löytää ne mielenkiintoista kohdat, mutta halutaan löytää myös ne, missä varmasti ei ole mitään mielenkiintoista, koska muuten niitä kovalevyjä tosiaan kertyy se muutama giga per sekunti.

## **Jessica Parland-von Essen**

Tämä kuulostaa hyvin tyyppilliseltä mun mielestä. Tai mä kuulen jatkuvasti siitä just, että miten nää datamäärät kasvavat todella paljon, kun hankitaan lisää sensoreita ja ne on yhä tarkempia kaikilla tieteenaloilla joka paikassa. Se tosiaan vaatii sitä aika tarkkaa suunnitelmaa ja miettimistä, että mitä kannattaa säästää ja mitä ei kannata, koska ainakaan näillä näkymin me tuskin pystytään tallentamaan kaikkea ja kaikkia versioita. Mä joskus leikilläni puhun siitä, että siirrymme datan hallinnasta datan hillintään, koska se on haaste.

## **Teemu Roos**

Toi kuulostaa itse asiassa aika kinkkiseltä ongelmalta, jos ei ihan tarkkaan tiedetä, miten sitä dataa haluttaisiin pyöritellä ja mitä sieltä haluttaisiin löytää. Auttaako tää FAIR data -ajattelu nyt jotenkin suunnittelemaan, että mitä dataa kerätään ja mitä heittää pois?

## **Jessica Parland-von Essen**

Mä olen itse taustaltani historian tutkija ja mä tiedän mun omalta alalta sen, että me yleensä säästetään vääriä asioita. Että usein yhtäkkiä just ne kahviläiskät olisikin juuri se, mikä olisi itse asiassa se relevantti asia ja sitä ei sit voi mennä muuta kuin parhaimman arvauksen pohjalta ja miettiä, että mikä on laadukasta. Tai sitten pitää valita joku strategia, että millä tavalla säästetään raaka data tai joku otos tai

joku muu... Mutta juuri tämä, että se vaatii sitä kunnon suunnittelua ja sitten dokumentointia, että ”me tehtiin näin, koska näin ja näin”, jotta sitten viidenkymmenen vuoden päästä joku pystyy ymmärtämään, että mitä tää on ja miksi tämä on tällaista.

### **Teemu Roos**

Mielenkiintoista, että historioitsijan tausta. Toi aikaperspektiivi, mikä meillä tavallaan on, on kyllä mielenkiintoista, että mitä kaikkea on mokailltu historian saatossa siinä, että mitä lähteitä on ehkä säilytetty tai mitä ei, ja mitä on ehkä säilytetty vahingossa. Eikö se ole niin, että meillä on aika suuri osa esimerkiksi Suomessa kirjoitetuista tekstidokumenteista säilynyt jossain tilikirjan tai verokirjan kansissa, kun niitä kierrätettiin ja ne säilyi siksi vahingossa.

### **Jessica Parland-von Essen**

Joo.

### **Teemu Roos**

Et mielenkiintoista, että miten sitten tähtitieteilijä tavallaan voi hyötyä siitä, että historioitsija on huomannut, että no näin tässä on vähän toillailtu ihan eri alan datalla.

### **Joni Tammi**

Tosi hyvä kysymys on se, että mitä dataa kannattaa säilöä. Me ollaan tosiaan mahdollisuuksien mukaan ylipäänsä pyritty säilömään kaikki, mikä on vielä jotenkin ollut ikään kuin hoidettavissa.

### **Teemu Roos**

Mutta tosiaan kuten puhuin tuosta puolitoista gigaa dataa sekunnissa ja jatkossa vielä enemmän, niin se tosiaan meinaa petatavua viikossa, niin se ei ole budjetillisesti ihan helppo hoitaa. Loppujen lopuksi sitä datan käsittelyä täytyy kehittää esimerkiksi niin, että se automatisoidaan etsimään sieltä se varsinainen, mielekäs informaation sisältö, mitä me käytetään. Näin esimerkiksi meillä voi gigojen havainto tiivistyä loppujen lopuksi yhteen datapisteeseen, mikä me voidaan esittää yhtenä lukuna ja yhtenä virhearviona ja aikaleimana. Eli sinänsä joissain määrin tämä tiivistäminen on mahdollista, mutta sen automatisoituminen ja sen tekeminen reaaliaikaisesti on etenkin meidän alalla se ongelma.

### **Jessica Parland-von Essen**

Ja just tuonkin takia mun mielestä on niin äärettömän fiksumaa ja hyvä, että avataan sitä dataa nyt jo. Ja annetaan sitä myös muiden tutkijoiden katsottavaksi, että jos nyt jo ilmenee jotain ”Hei meille olisikin ollut tää XYZ tärkeä”, niin sitten pystyy ottaa sen huomioon, kun lähtee tekemään niitä säilytysuunnitelmia. Että se on kyllä kaikin tavoin fiksumaa avata dataa jos vaan voi.

### **Teemu Roos**

Ainakin oppia siitä, jos on tehty joku virhe, että nyt tuota tietoa vaan ei tullut tallennettua, niin ehkä sen sitten voi jatkossa ottaa talteen.

## Teemu Roos

Näitä harmillisia tarinoita on, että jotkut keräsi 20 vuotta jotain dataa jonkun ilmiön ymmärtämiseksi, ja sitten kun ne rupeesi analysoimaan sitä, niin sitten ne huomasi, että "Ah... tämä avainpalanen tietoa olisikin pitänyt olla. Ei mitään... aloitetaan nyt sitten koko homma alusta". Että näiden välttämiseksi kuulostaa hyvältä, että se huomattaisiin aikaisemmin kuin silloin vasta kahdenkymmenen vuoden päästä, kun se hanke päättyy sinne lopputulemaan.

*For brilliant minds - For a Better Future!*

## Teemu Roos

Onko ajatus, että ne säilyy ikuisesti sitten ne datat ja ja tuleeko siitä jotain ongelmia? Meillähän on kokemuksia myös siitä, että pitkäaikaissäilytys ei ole ihan triviaali juttu, että se data, formaatit ja säilytysmuodot vaihtuu. Ja ne mitkä on ajateltu olevan tosi hyviä tämmöisiä, esimerkiksi jotkut magneettinauhat, että onko enää edes semmoista laitetta, millä niitä pystyy lukemaan? Että miten tää, onko tämä selvä juttu? Mitä kannattaa hoitaa?

## Jessica Parland-von Essen

Ei oo selvä, mutta meillä on siis Suomessa tosi hyvä tilanne, koska meillä on opetus- ja kulttuuriministeriö satsannut isosti siihen, että CSC:lle on rakennettu pitkäaikaissäilytyspalvelu, joka tehtiin ensin arkistoille, museoille ja kirjastoille kulttuurihistorian tai perinnön puolelle, mutta nyt sitä on laajennettu osaksi tätä FAIR data -palvelua, että siihen on tullut Fairdata-PAS. PAS on pitkäaikaissäilytys tai -saatavuus, miten sen haluaakaan sanoa. Elikkä meillä on Suomessa tosi hyvä tilanne. Meillä on hyvää osaamista ja just sitä, että mietitty tosi paljon ja just tämmöistenkin tahojen kanssa kuin arkistot, jotka on tehnyt tätä muutaman 100 vuotta. Metatiedot, identifiointi ja kaikki tämmöinen. Ja sitten koska digitaalisessahan on se haaste, että ne vanhenee. Kaikki mediat vanhenee, ne happanee. Formaattit, teknologiat, kaikki. Joten tää tarkoittaa, että meidän täytyy myös sitä suunnitella, että ne on riittävän teknisesti dokumentoitu, että me pystytään **mikroimaan**, siirtämään tai uudelleen tuottamaan. Ykkösiä ja nollia vielä jotenkuten pystyy ehkä säilyttämään ja kopioimaan näin ikään kuin mekaanisesti, mut sit se, että sä oikeasti ymmärrät sen formaatin ja mitä se sisältö on, niin se vaatii hirvittävästi dokumentaatiota. Ja tosiaan meillä on siitä aika hyvää osaamista CSC:llä ja Suomessa.

## Teemu Roos

Tässä kun puhuttiin näistä, mitkä on hyviä vertauksia datalle, että onko se nyt sähköä tai öljyä tai kultaa, niin tuosta pitkäaikaissäilytyksestä tulee mieleen, että sä taisit joskus sanoa, että uraani on sellainen vertaus mitä sä olet käyttänyt.

Jessica Parland-von Essen

Joo, no mun mielestä se on vähän niinku ydinjätettä. Että sitä ei voi just jättää vaan jonnekin kaapin nurkkaan, esimerkiksi jotain kovalevyä.



## **Teemu Roos**

Vähän kyllä kiehtoisi kuulla näistä gravitaatiolinseistä ja mustista aukoista ja näistä. Jotenkin hohdokkaita...

## **Jessica Parland-von Essen**

Siis mulla on maailman paras duuni. Siis tää on niin ihanaa, kun saa olla eri alojen huippututkijoiden kanssa tekemisissä ja kun kaikilla on sitä dataa ja pystytään siinä toivottavasti vähän auttamaan.

## **Teemu Roos**

Joo, mulla on usein sama olo, kun on tekoälyn ja koneoppimisen kanssa tekemisissä. Mä olen itsekin tehnyt historioitsijoiden kanssa, että me ollaan täällä jotain Punahilkka-sadun historiaa ihmetelty tai jotain keskiaikaisia käsikirjoituksia ja sitten välillä meilläkin jossain kokouksissa puhutaan jostain... Mikäköhän se oli, joku intergalaktinen väliaine? Ja sen ominaisuuksista. Mä en juuri paljon siitä ymmärtänyt, mutta joka tapauksessa se datahan on just näitä kuuluisia nollia ja ykkösiä. Ja niitä voidaan sitten analysoida tilasto- ja koneoppimismenetelmillä. Mut jos sä Joni voit vähän kuitenkin meille kiinnostuneille avata, että minkä tyyppisillä menetelmillä vaikka niitä mustia aukkoja voidaan sitten niin sanotusti kuvata tai valokuvata sitten niistä nolista ja ykkösistä?

## **Joni Tammi**

Joo, no meillä on tavallaan pari tapaa. Yksi on tosiaan se, mitä mä mainitsin, eli tää että kytketään radioteleskooppeja yhteen ja niillä saadaan muodostettua sellainen kuva. Oikeastaan se tekniikka on sellainen, että tavallaan kytketään ympäri maailmaa laitteet yhteen, niin me saadaan tavallaan verkosto, joka toimii samalla resoluutiolla kuin että jos meillä olisi sen koko verkoston kokoinen teleskooppi, eli just niinku vaikka maapallon kokoinen teleskooppi. Tai yhdessä vaiheessa meillä oli 10 metrin radioteleskooppi maata kiertävällä radalla. Kun me kytkettiin se vaikka Metsähovin muiden teleskooppien kanssa maan päällä, niin me saatiin tavallaan resoluutio, mikä vastaa sitä, että meillä olisi maan kuusysteemin kokoinen teleskooppi eli se resoluutio on siellä sellainen, mitä millään muulla alalla ei voi saavuttaa todennäköisesti hyvin pitkään aikaan edes tulevaisuudessa. Mutta tää on vaan yks tapa. Ja esimerkiksi tää mustan aukon kuva, mikä saatiin pari vuotta sitten, se oli ensimmäinen, mutta me ollaan tutkittu 40 vuotta Metsähovissa mustia aukkoja. Eli me tutkittiin paljon ennen kuin me edes saatiin ensimmäisiä kuvia. Se mitä me tutkitaan on se, että me seurataan näiden kirkkauksia.

## **Teemu Roos**

Hetkinen, nyt on pakko keskeyttää. Siis nehän on mustia, miten ne voi olla kirkkaita?

## **Joni Tammi**

Siis mustat aukot ei säteile, mutta mustiin aukkoihin tippuva aine säteilee. Ja sitten siellä on myös valtavat magneettikentät ja muut erikoisuudet, mitkä sitten aiheuttaa sen, että ne mustat aukot heittelee ikään kuin ympäriinsä ainetta tavallaan tosi tarkkaan suunnatuissa esimerkiksi suihkussa. Ja esimerkiksi käytännössä jokaisen galaksin, eli myös meidän linnunradan, keskustassa on.. puhutaan supermassiivisista mustista aukoista, jotka painaa enemmän kuin miljoonat tähdet. Mutta meillä on semmoinen tylsä, mikä ei tee mitään kauhean erikoista, mutta sitten joissain galakseissa, etenkin tosi kaukana, se onkin kaikkea muuta kuin tylsä. Puhutaan aktiivisista galakseista. Ja siellä se musta aukko

syö hirveästi ainetta, sinne voi tippua tähtiä ja muuta, ja osa siitä aineesta tosiaan ikään kuin sinkoutuu ulos sieltä semmoisissa suihkuissa, mitkä on tosi kirkkaita. Ne suihkut säteilee ihan valtavasti. Ne liikkuu käytännössä valon nopeudella. Ja kun mä sanon ”käytännössä”, mä ihan tarkoitan konkreettisesti eli ei kuvainnollisesti. 99.995% valon nopeudesta tai jotain muuta vastaavaa. Ja tätä säteilyä me sitten pystytään tutkimaan käytännössä melkein maailmankaikkeuden halki. Se on äärimmäisen himmeää ja heikkoa, joten me tarvitaan siihen valtavat radioteleskoopit ja tällaiset, että me pystytään ikään kuin nauhoittamaan sitä tai keräämään sitä. Mutta loppujen lopuksi esimerkiksi se havaintotyyli, mitä me tehdään Metsähovissa, mihin me erikoistutaan eli tällainen monitorointi on se, että me muutamaa sataa meidän radiotaajuuksilla kirkkainta mustaa aukkoa. Maailmankaikkeus on täynnä mustia aukkoja ja ne on meille ihan yhtä arkisia kuin joku tähti tai planeetta on tuolla. Niillä on tietyt nimet ja ominaisuudet, ja me tiedetään missä ne on ja suurin piirtein mitä ne tekee. Ja niitä on tosiaan siis miljoonia ja miljardeja, ja me seurataan näistä niin kuin muutamaa sataa mielenkiintoisinta, ja tarkoitus on seurata mitä ne tekee. Jotkut on tosi aktiivisia, eli niissä on koko ajan jotain meneillään, eli esimerkiksi ne syö tähtiä tai planeettoja tai muuta vastaavaa. Sen takia ne kirkastuu ja himmenee ja muuta vastaavaa. Ja sitten kun me seurataan niitä kirkastumisia ja himmenemisiä, niiden aikaskaaloja ja sitä, millä tavalla se kirkastuminen tapahtuu, niin me pystytään saamaan aika paljon informaatiota irti siitä, mitä siellä fyysikaalisesti tapahtuu, mikä on mustan aukon koko, mitä se tekee ja muuta vastaavaa. Ja sitten siinä on myös se, että nää jotkut tosiaan muuttuu ihan valtavissa aikaskaaloissa. Vuoden tai kuukauden aikaskaala on tosi kiva, se on tosi lyhyt. Mut sit puhutaan kymmenistä tai sadoista vuosista, ja jotkut näistä kasvaa tosiaan miljoonia vuosia, niin me ei voida ikinä ihmisajan tai sanotaan vaikka meidän 40-vuotisen datan perusteella tutkia niitä kaikkia tai edes yhtä kokonaan. Se on vähän sama, kuin että me haluttaisiin tutkia ihmistä, mutta me ei ikinä olla nähty ihmistä aikaisemmin tai jotain muuta eläintä, ja meillä on 10 minuuttia aikaa tutkia sitä ja siitä päätellä sen koko elämä ja elinkaari. No se ei nyt onnistu. Mutta me voidaan mennä tonne rautatieasemalle ja ottaa siellä 10 minuutin aikana kuva tuhannesta erilaisesta ihmisestä, ja sitten me voidaan niistä päätellä esimerkiksi että onko nää eri populaatiota. Mitä ne tekee? Onko näissä jotain yhteistä? Onko esimerkiksi nää pienet, jotka on tällaisissa ihme toppahaalareissa, niin onko ne ihan oma lajinsa, vai tuleeko niistä myöhemmin tällaisia isompia? Ja muuta vastaavaa. Ja sitten on tää kalju populaatio, mikä täysin erilainen kuin tämä millä on tukkaa. Ja niillä on todennäköisesti täysin eri ominaisuudet, ei mitään mitään fyysistä yhteyttä ja muuta. Tää on tavallaan se meidän tutkimuksen taso tällä hetkellä. Tai siis se datan taso, koska ne kohteet on niin kaukaisia, että me ei pystytä menemään lähelle. Me ei pystytä saamaan niistä oikeasti aitoa kuvaa. Eikä oikeastaan halutakaan mennä ihan liian lähelle mustaa aukkoa. Mutta sitten sen sijaan me tutkitaan valtavaa määrää ja yritetään jotain suodattaa siitä käyttäytymisestä tänä lyhyenä aikana, mitä me vaikka parinkymmenen vuoden aikana ollaan saatu selville.

### **Teemu Roos**

Mielenkiintoisia vertauksia - rautatieaseman kaljut toppahaalarityypit, vai mitä nää oli? Mikä on sun suosikki musta aukko? Onko sillä joku hassu nimi?

### **Joni Tammi**

Nyt mun pitää olla varovainen, että mä en suututa ketään...

### **Teemu Roos**

Ketään mustaa aukkoa, sitä kakkossuosikkia?

## **Joni Tammi**

Eikun tutkijoita, jokaisella tutkijalla on aina suosikki. Mutta sanotaan vaikka 3C279.

## **Jessica Parland-von Essen**

Ai se?

## **Teemu Roos**

Joo siis sehän on ihan älyttömän symppis kyllä, eikä yhtään tylsä. Ei ainakaan nimen perusteella kuulosta siltä. Voitaisiin keskustella varmaan pidempäänkin näistä, mutta itse asiassa mulla olisi sellainen yks kysymys... Nyt kun me ollaan puhuttu paljon pitkistä aikaskaaloista, jotka on alkanut 70-luvulla. Monet suuret tapahtumat ovat saaneet alkunsa 70-luvulla. Ja jopa henkilöt. Tästä pitkästä aikaperspektiivistä tulee mieleen tulevaisuuteen katsominen, ja tulevaisuudessa tulee tapahtumaan muun muassa se, että kun CSC:llä on tänä vuonna 50-vuotisjuhlat, niin viidenkymmenen vuoden kuluttua tästä – voidaan ehkä päätellä, ei tullut suurena yllätyksenä – on satavuotisjuhlat. Ja CSC:llä on tällainen suunnitelma, että tallennetaan aikakapseliin erilaisia muistoja tästä nykyajasta. Mitä dataa tai mitä terveisiä te haluaisitte lähettää sinne aikakapseliin? Jos ja kun tämä podcast myös suljetaan sinne aikakapseliin, ja ne tulevaisuuden ihmiset vuonna 2071 tulee kuuntelemaan meidän terveisiä. Jessica, mitä terveisiä lähettäisit?

## **Jessica Parland-von Essen**

Sanoisin, että olkaa armollisia. Yritimme parhaamme.

## **Teemu Roos**

Se on hyvä terveinen.

## **Jessica Parland-von Essen**

Niin ja jos dataakin ajattelee, niin kyllähän mä luulen, että meillä on aika hyvä yritys täällä päällä kyllä säilyttää heille tietoa ja auttaa tutkimusta kaikin tavoin.

## **Teemu Roos**

Toivotaan, että silloin on kaikki datat saavutettavia ja interoperatiivisia ja niin edespäin. No Joni, mitä terveisiä lähettäisit sinne tulevaisuuteen?

## **Joni Tammi**

No, toivottavasti meiltä löytyy 90 vuotta tai vähän päälle 90 vuotta vuotta sitten jatkuvaa aurinkodataa ja näiden aktiivisten galaksien monitorointia, ja silloin todennäköisesti tiedetään aika paljon enemmän auringosta ja mustista aukoista. Mut tosiaan ehkä mäkin voisin vielä, jos joku katsoo sitä meidän aikaista... Sanotaan vaikka meidän 2020 tai 2021 vuoden julkaisutöitä, niin toi on tosiaan ihan hyvä pohjustus, että se vaikutti hyvältä idealta silloin tehdä se näin.

## **Jessica Parland-von Essen**

Joo. Ja siis just parempi yrittää kun ei yrittää. Siis että me nyt edes jotain saadaan pelastettua ja varmasti on hyödyllistä jo. Olen ihan varma siitä.

## **Joni Tammi**

Ja se, että jos me nyt oltaisiin tässä 2020-luvun tietämissä keksitty se paras tapa tehdä se, ja viidenkymmenen vuoden päästä ei olisi tullut parempaa tapaa ja silloin ihmiset eivät katselisi meitä ja miettisi, että miten ne tuollaiseen päätyi. Että nyt tämä vaikuttaa ihan naiivilta. Niin silloinhan se olisi tavallaan surullista. Mutta se, että me nyt sanotaan, että tämä on parasta mitä nyt tällä hetkellä on, ja sitten toivotaan, että ollaan näkemässä vielä sitten viidenkymmenen vuoden päästä, millä tavalla se homma silloin on.

## **Teemu Roos**

No niin, eli odotukset on nyt sitten tulevaisuuden ihmisille, että heidän on parempi tietää vielä paremmin kuinka se olisi kannattanut tehdä kuin meidän. Toivotaan näin! Eli terkkuja sinne tulevaisuuteen ja terkkuja myös kaikille nykyisyyteen. Mukava, että kuuntelitte ja kuulemisiin!

*Tämä podcast tallennetaan CSC:n 50-vuotisjuhlan kunniaksi aikakapseliin, joka avataan seuraavan kerran 100-vuotisjuhlan yhteydessä vuonna 2071. Kiitos kun kuuntelit For a Better Future -podcastia!*