



# Tiedustelu avoimista lähteistä metatiedon avulla

---

Glad, Timo

2013 Leppävaara

Laurea-ammattikorkeakoulu  
Leppävaara

## Tiedustelu avoimista lähteistä metatiedon avulla

Timo Glad  
Turvallisuusalan koulutusohjelma  
Opinnäytetyö  
Maaliskuu, 2013

Glad, Timo

**Tiedustelu avoimista lähteistä metatiedon avulla**

Vuosi	2013	Sivumäärä	48
-------	------	-----------	----

---

Tämän opinnäytetyön tarkoituksena on selvittää avoimiin lähteisiin perustuvaa tiedustelua ja sen hyödyntämistä organisaatioiden julkaisemien dokumenttien metatiedon analysoinnissa. Tutkimuksen menetelmänä on dokumenttianalyysi. Avoimiin lähteisiin pohjautuvan tiedusteluprosessin lähteenä käytetään NATO:n sekä YK:n julkaisemia ohjeistuksia.

Työ toteutettiin valitsemalla viisi yritystä joiden julkaisemat, ja Google-hakupalvelun indeksoimat, dokumentit kerättiin ja analysoitiin metatiedon osalta. Kerättyjä ja analysoituja dokumentteja oli yhteensä 1567.

Työn lopputuloksena todetaan etteivät organisaatiot vaikuta kiinnittävän riittävää huomiota dokumentteihin tallennetun metatiedon poistamiseen. Valtaosa kerätyistä ja analysoiduista dokumenteista sisälsi metatietoa joka liittyi dokumentin tekijään, tekijän käyttäjätunukseen, dokumentin luomiseen käytettyyn ohjelmaan tai ohjelmaversioon. Työssä esitetään, miten tallennettua metatietoa voidaan hyödyntää ”spear-phishing”-tyyppistä hyökkäystä tehtäessä.

Glad, Timo

**Open Source Intelligence Gathering using document metadata**

Year	2013	Pages	48
------	------	-------	----

The objective of this thesis is to study how Open Source Intelligence Gathering can be used in the discovery of documents released by organizations and in the analysis of metadata saved in those documents. The method used in this thesis is document analysis. For the process of Open Source Intelligence Gathering techniques guidebooks released by NATO and UN were used.

Five organizations were selected, and the metadata of every document file located on each of the selected organizations official website was analyzed. 1567 document files in total were located and collected using Google search engine.

From the results of this thesis it would appear that organizations do not pay enough attention to the removal of metadata from publicly released document files. Nearly every collected and analyzed document had metadata about the creator, his or her username and program or program version used to create or edit the document. The thesis will show how a malicious attacker can use excess metadata when he or she is scoping potential targets for a “spear-phishing”-attack.

Keywords: Open Source Intelligence Gathering, Metadata

## Sisällys

1	Johdanto.....	6
1.1	Työn lähtökohdat ja aihealueen rajaukset.....	7
1.2	Työhön liittyvät keskeiset käsitteet.....	8
1.3	Käytetyt menetelmät.....	9
2	Avoimiin lähteisiin perustuva tiedustelu .....	9
2.1	NATO: Avoimiin lähteisiin perustuva tiedustelu.....	10
2.2	UNODC: Tiedusteluprosessi .....	12
3	Tiedon luotettavuuden arviointi ja analysointitekniikat .....	13
3.1	Kerätyn tiedon 4x4- ja 6x6-arviointimenetelmät .....	14
3.2	Kerätyn tiedon analysointiprosessi .....	17
3.3	Verkostoanalyysi .....	18
3.4	Tapahtumakaavioanalyysi .....	19
3.5	Vuoanalyysi .....	21
3.6	Puheluanalyysi.....	22
3.7	Analysoidusta tiedosta johdetut päätelmät.....	23
3.8	Tiedon keräysvaiheeseen liittyvät ongelmat.....	24
3.9	Tiedon analysointivaiheeseen liittyvät ongelmat .....	25
3.9.1	Syy- ja seuraussuhteiden virhetulkinnat .....	26
3.9.2	Todennäköisyyksien arviointiin liittyvät virhetulkinnat.....	27
3.9.3	Tiedon jälkikäteisen tulkinnan aiheuttamat virhetulkinnat.....	28
3.10	Tiedoston metatiedon tutkiminen .....	29
4	Avoimiin lähteisiin perustuvan tiedustelun hyödyntäminen käytännössä .....	31
4.1	Käytetyt tietolähteet ja tiedon kerääminen.....	32
4.2	Dokumenttitiedostojen etsimiseen käytetyt hakutermit .....	34
4.3	Metatiedon keräämiseen käytetyt valinnat .....	35
4.4	Kerätty tulokset .....	35
5	Johtopäätökset .....	37
6	Oman työn arviointi.....	42
	Lähteet .....	44
	Kuvat .....	46
	Kuviot.....	47
	Taulukot .....	48

Ajatus opinnäytetyölle syntyi New York Times-lehdessä olleen artikkelin ”How Companies Learn Your Secrets” pohjalta (Duhigg 2012). Artikkelin teemana oli se kuinka yritykset seuraavat ja analysoivat asiakkaidensa ostokäyttäytymistä. Ostokäyttäytymisen pohjalta yritykset pystyivät selvittämään asiakkaidensa sen hetkiseen elämäntilanteeseen liittyviä yksityiskoh-  
tia. Artikkelissa kuvattiin, kuinka Target-yritys pystyi selvittämään ostokäyttäytymisen perusteella sen oliko asiakas raskaana.

Sosiaalisten yhteisöpalvelujen sekä niitä käyttävien ihmisten määrä on kasvanut valtavasti muutaman viime vuoden aikana. Yhtenä tunnetuimmista sosiaalisista yhteisöpalveluista on Facebook, jota The Economist kutsui 2010 julkaistussa artikkelissaan maailman kolmanneksi suurimmaksi valtioksi 500 miljoonalla käyttäjällä. Listalta löytyy myös MySpace 300 miljoonalla käyttäjällä sekä Twitter 124 miljoonalla käyttäjällä. (Facebook population 2010.) Palveluiden käyttäjämäärät ovat vain kasvaneet, ja Facebookin odotetaan rikkovan yhden miljardin käyttäjän rajan 2012 vuoden kuluessa (MacMillan 2012). Palveluissa jaetaan huolettomasti käyttäjiin liittyviä henkilötietoja itse palvelun sekä muiden käyttäjien kanssa. Tällaisia tietoja voivat olla valokuvat, videot, paikkatiedot sekä elämäntilanteeseen, poliittiseen näkökulmaan ja mielipiteisiin ja vakaumukseen liittyvät asiat. Työnantajat ovat myös aktivoituneet sosiaalisissa verkkopalveluissa, ja kärjistyneimmissä tapauksissa työnhakijoilta on vaadittu sosiaalisten verkostoitumispalveluiden tunnuksia, ennen näiden palkkaamista. Toinen mielenkiintoinen aihe koskee metatietoa, jota tallennetaan eri ohjelmien toimesta automaattisesti niiden kautta luotuihin tai niillä muutettuihin tiedostoihin. Metatieto sisältää runsaasti tietoa tiedoston luoneesta käyttäjätilistä sekä järjestelmästä, mutta se ei kuitenkaan tyypillisesti näy tiedostoa käsittelevälle tai sen avaavalle henkilölle.

Työn aiheen valinta tapahtui pitkälle omista kiinnostuksen kohteista sekä avoimiin lähteisiin perustuvaa tiedustelua että tarkemmin perinteistä tiedusteluprosessia kohtaan. Toisaalta ottamalla mukaan tietoturvallisuutteen liittyviä seikkoja, työhön on pyritty tuomaan myös ajan-  
kohtaisuutta. Ihmiset tallentavat itsestään valtavan määrän tietoa Internetiin, pohtimatta tarkemmin mihin ja miten tätä tietoa voidaan käyttää väärin. Toisaalta valtava määrä tietoa voidaan saattaa julkisesti saataville tahattomasti. Erilaiset tiedoston luomiseen ja muokkaamiseen tarkoitetut ohjelmistot lisäävät runsaasti ohjelman käyttäjään liittyvää tietoa, eikä käyttäjä välttämättä tiedä tai tule ajatelleeksi sitä. Yritykset julkaisevat runsaasti eri dokumentteja, jotka käsittelevät muun muassa yrityksen toimintaa, sen tarjoamia tuotteita tai taloudellista tilannetta. Potentiaalisesti kaikki julkaistavat dokumentit voivat sisältää, ja usein sisältävät, tiedoston luoneen ja sitä käsitelleisiin tahoihin liittyvää metatietoa.

Työn tavoitteena on käydä läpi perinteistä tiedusteluprosessia sekä avoimiin lähteisiin perustuvaa tiedustelua. Tiedusteluprosesseja on tarkoitus hyödyntää yrityksen julkaisemien dokumenttiedostojen keräämiseen sekä näissä olevan metatiedon analysointiin. Keräämiseen käytetyt menetelmät käydään läpi omassa osiossa, jonka jälkeen tutustutaan tarkemmin kerätyistä tiedostoista löydetyn metatiedon sisältöön.

### 1.1 Työn lähtökohdat ja aihealueen rajaukset

Tutkielmatyyppisen opinnäytetyöni aihe on ”Tiedustelu avoimista lähteistä metatiedon avulla”. Tällä tarkoitetaan tiedon keräämistä erilaisista julkisista ja kaikille avoimista, joko ilmaisista tai maksullisista, lähteistä. Kerättyä tietoa yhdistetään, analysoidaan ja tätä kautta siitä pyritään johtamaan uudenlaista tietoa sekä havaintoja. Avoimiin lähteisiin perustuva tiedustelu on käsitteenä vanha, ja sitä on hyödynnetty pääasiassa valtiollisten toimijoiden keskuudessa. Näissä tapauksissa painotuksena on ollut sotilaallinen toiminta ja siihen liittyvä tiedustelu. Viime vuosina tapahtunut Web 2.0-palveluiden nousu on johtanut siihen, että yksittäisistä ihmisistä on saatavilla yhä enemmän tietoa verkon välityksellä (Madig 2012).

Opinnäytetyö on rajattu koskemaan avoimista lähteistä kerättävää metatietoa ja tämän aiheuttamia uhkia organisaatioille ja yksilöille. Organisaatioon voi kohdistua uhkia yksilön toiminnan tai laiminlyöntien takia, ja toisaalta yksilöön voi kohdistua riskejä myös organisaation jäsenyyden takia. Avoimet lähteet ovat käsitteenä hyvin laaja, käsittäen lukuisia eri medioita ja julkaisuja. Tästä syystä työssä on päädytty siihen, että tässä työssä avoimilla lähteillä tarkoitetaan kaikkea sellaista tietoa jota voidaan kerätä Internetistä. Esimerkkejä tällaisista lähteistä ovat tietoa indeksoivat hakupalvelut, sosiaalinen media sekä erilaiset viranomaisten, yritysten tai organisaatioiden ylläpitämät avoimet tietokannat.

Työssä esitetty avoimiin lähteisiin perustuvan tiedusteluprosessin malli nojaa vahvasti NATO:n sekä YK:n julkaisemiin ohjeistuksiin. Lähdeaineiston monipuolisuuden puute on tiedostettu ongelma, joskin sen pääasiallisena syynä on se, että tiedusteluprosessit on yleisesti kehitetty tukemaan valtiollisten tasojen tiedustelu- ja sotilasorganisaatioiden tarpeita eikä niitä ole varsinaisesti suunniteltu yksityisten organisaatioiden käyttöön. Tämä tarkoittaa käytännössä sitä että prosessikuvauksia ja lähdemateriaalia on saatavilla rajoitetusti.

Olen valinnut työtä varten viisi tunnettua Suomessa toimivaa yksityisen- tai julkisen sektorin organisaatiota, joiden kohdalla tarkastelen organisaatiosta saatavilla olevaa tietoa. Valitut organisaatiot edustavat viestintäteknologian, tietoturvallisuuden, kaivosteollisuuden sekä julkisen sektorin toimijoita. Tarkastelukulma keskittyy pääasiassa tietolähteisiin, joiden kautta yrityksistä, niiden toiminnasta ja niiden jäsenistä voidaan kerätä tietoa edellä mainittujen tahojen sitä havaitsematta. Näistä työn kannalta merkittävimpinä tietolähteinä toimivat or-

ganisaatioiden julkaisemassa materiaalissa oleva metatieto, sekä erilaisten sosiaalisten verkopalveluiden ja hakupalveluiden hyödyntäminen.

Yhtenä opinnäytetyöni tarkoituksena on osoittaa, että vaikka metatietoon liittyvät mahdolliset ongelmat on tunnettu jo vuosikymmenten ajan, organisaatiot eivät edelleenkään pyyhi metatietoja julkaisemistaan dokumenteista.

## 1.2 Työhön liittyvät keskeiset käsitteet

**Avoimiin lähteisiin perustuvalla tiedustelutiedolla** tarkoitetaan erilaisista julkisesti saatavilla olevista lähteistä kerättävää tietoa. Tyypillisiä esimerkkejä avoimista lähteistä ovat sanomalehdet, radio, televisio, julkaistut tutkimukset sekä julkisyhteisöjen toiminnan yhteydessä tuotettu materiaali. (Intelligence Collection Disciplines 2012.) Tässä työssä avoimiin lähteisiin perustuvalla tiedustelutiedolla tarkoitetaan vain Internetin eri palveluiden kautta kerättävissä olevaa tietoa. Tarkastelu painottuu pääasiassa yrityksen julkaisemiin dokumenttitiedostoihin, sekä näihin, joko tarkoituksella tai tahattomasti, tallennettuun metatietoon.

**Digitaalisella jalanjäljellä** tarkoitetaan tietoverkkoon tallennettua tietoa, joka kuvaa ja yksilöi henkilön muista tietoverkkoa käyttävistä henkilöistä (Intelligence Exploitation of the Internet 2002, 54). Työn kontekstissa digitaalisella jalanjäljellä tarkoitetaan kaikkea sitä yksilöitävää, organisaatiota tai henkilöä kuvaavaa tietoa, joka on julkisesti saatavilla.

**Metatiedolla** tarkoitetaan tietoa kuvaavaa tietoa. Varsinainen metatiedon sisältö vaihtelee tiedostokohtaisesti, mutta yleisesti ottaen metatieto sisältää muun muassa tiedostotyyppin ja tiedostonimen, tiedoston sisällön tai sen ominaisuuksien muuttamiseen liittyviä aikaleimoja sekä tiedoston käyttöoikeuksiin liittyviä määrytyksiä. (Altheide, Carvey & Davidson 2011, 41.)

**Tiedon analyysillä** tarkoitetaan kerätyn tiedon yksityiskohtaista tutkimista, jonka kautta voidaan selvittää tietoon liittyvät ominaispiirteet sekä tiedon merkitys ja tarkoitus (Criminal Intelligence 2011, 13). Analyysi toimii myös toimintaa ohjaavana tekijänä, paljastaen kerätyn tiedon laatuun tai määrään liittyvät ongelmat, ja toisaalta sen kautta on mahdollisuus havaita lopputuloksen kannalta keskeinen puuttuva tieto.

**Tiedustelutiedolla** tarkoitetaan kerättyä ja prosessoitua tietoa, jonka merkitystä on pyritty tulkitsemaan tiedon analysoinnin avulla. Koska tiedosta johdetaan tuloksia sen sisällön ja merkityksen tulkinnan avulla, se sisältää lähtökohtaisesti aina tietoa analysoivan henkilön spekulatiota sekä tiedon tulkintatavasta aiheutuvia epätarkkuuksia. Tiedon luotettavuuteen ja siinä oleviin epätarkkuuksiin vaikuttavat analysointia varten kerätyn tiedon määrä, laatu ja luotettavuus. (Criminal Intelligence 2011, 9.)



### 1.3 Käytetyt menetelmät

Työssä käytetyt tutkimusmenetelmät pohjautuvat Hirsjärven, Remeksen ja Sajavaaran Tutki ja kirjoita- (2009) sekä Denscomben The Good Research Guide (2010)-kirjoihin. Opinnäytetyö on luonteeltaan tutkielmatyyppinen. Tiedonkeräyksen menetelmäksi on valittu laadullinen eli kvalitatiivinen tutkimus, jonka aineisto on kerätty dokumenttianalyysin avulla.

Kvalitatiivisella eli laadullisella tutkimuksella tarkoitetaan kerättävän tiedon laatuun keskittävää suuntausta, jonka tarkoituksena on tutkia tutkimuksen kohdetta kokonaisvaltaisesti. Laadullisessa tutkimuksessa tarkoituksena onkin, olemassa olevien väittämien todentamisen sijaan, pyrkiä paljastamaan tutkimuksen kohteeseen liittyviä tosiasioita. (Hirsjärvi ym. 2009, 160 - 161.)

Dokumenttianalyysi on tiedonkeräysmenetelmä, jossa tiedonlähteenä käytetään eri tyyppisiä dokumentteja. Nämä voivat tyyppiltään kirjallisia dokumentteja, kuvia, valokuvia tai musiikkia. Dokumentit nähdään itse yhtenä tiedon muotona tallennustavasta riippumatta, eikä eroa ole oli kyseessä sitten sähköinen tai paperille tulostettu tallenne. Analysoitavan dokumentin luotettavuuden arviointiin käytetään neljää kriteeriä, jotka ovat autenttisuus, edustuksellisuus, tarkoitus sekä uskottavuus. (Denscombe 2010, 216 - 222.) Työssä tehtävä dokumenttianalyysi kohdistuu sähköisiin dokumenttitiedostoihin tallennettuihin metatietoihin. Dokumenttitiedoston varsinainen sanallinen tai muu asiasisällöllinen tieto ei ole analyysin kohteena.

## 2 Avoimiin lähteisiin perustuva tiedustelu

Työn teoriaosuuden runkona toimii kaksi tiedusteluprosessia kuvaavaa julkaisua. Ensimmäisenä kuvattu NATO:n 2001 julkaisema ”Open Source Intelligence Handbook”-ohjeistus toimii avoimiin lähteisiin perustuvan tiedusteluprosessin esittelijänä. Ohjeistus kuvaa perinteisen tiedustelusyklin vaiheineen pintapuolisesti, keskittyen niihin eroihin joita avoimiin lähteisiin perustuvassa tiedusteluprosessin sekä viranomaislähteisiin perustuvan tiedusteluprosessin välillä esiintyy. Toisena yleistä tiedustelusykliä kuvaavana dokumenttina on käytetty YK:n huum- ja rikollisuutta valvovan viraston julkaisemaa, ”Criminal Intelligence: Manual for Analysts”-ohjeistusta. Siinä missä NATO:n ohjeistus kuvaa avoimiin lähteisiin perustuvaa tiedustelua ja sen erityispiirteitä, YK:n ohjeistus keskittyy perinteisen tiedustelusyklin lisäksi tarkemmin varsinaisen kerätyn tiedon luotettavuuden arviointiin ja sen merkityksen analysointiin. Dokumentissa käydään läpi useita kerätyn tiedon luotettavuuden arviointimenettelyjä sekä analysointitekniikoita.

## 2.1 NATO: Avoimiin lähteisiin perustuva tiedustelu

Avoimiin lähteisiin perustuva tiedustelusykli vastaa eri vaiheiltaan perinteistä tiedustelusykliä. Keskeisin ero on tiedustelutiedon ja analyysien lähteenä käytetyn materiaalin julkaisu- ja keräystavassa, jotka perustuvat avoimiin lähteisiin perustuvassa tiedustelussa julkisiin ja vapaasti kerättäviin tietoihin. Avoimiin lähteisiin perustuva tiedustelu pyrkii tarjoamaan samanaikaisesti lopputuotteen käyttäjälle vapaan pääsyn analyysissä käytettyyn lähdemateriaaliin sekä analyytikon lähdemateriaalista tuottamaan analyyttiseen lopputuotokseen. (Open Source Intelligence Handbook 2001, 30.) Prosessin yhteydessä puhutaan neljästä D:stä, jotka ovat Discovery, Discrimination, Distillation ja Dissemination. Tarkoituksena on selvittää tutkimuksen kannalta keskeiset lähteet, erotella prosessin tavoitteiden kannalta keskeiset teemat ja löytää näiden teemojen kannalta keskeiset asiat sekä keskeiset toimijat. (Open Source Intelligence Handbook 2001, 15.)

Ensimmäisenä avoimiin lähteisiin perustuvan tiedusteluprosessin vaiheena on toimeksiantajan antama tutkittavan ongelman ja tähän liittyvien tiedustelutarpeiden määrittely. Toimeksiantokuvauksen tarkoituksena on ohjata analyytikon työtä muissa prosessin vaiheissa. Koska toimeksiantokuvaus toimii keskeisenä tietotarpeen määrittelijänä ja tätä kautta analyytikon työtä ohjaavana tekijänä, tulee toimeksiantokuvauksen määrittely tehdä yksityiskohtaisesti ja tarkasti. Analyytikon tulee tietää lopputuotteeseen liittyvät odotukset ja tietotarpeet, jotta lopputuotteen on mahdollista vastata toimeksiantajan tietotarpeeseen. Toimeksiantajan olisi suositeltavaa myös ohjata, osallistua ja antaa palautetta tiedusteluprosessista ja sen tuloksista prosessin eri vaiheissa. (Open Source Intelligence Handbook 2001, 16 - 17.)

Toisena vaiheena avoimiin lähteisiin perustuvassa tiedusteluprosessissa on tiedustelutiedon kerääminen. Tämän vaiheen tarkoituksena on löytää keskeisiä tietotarpeeseen liittyviä lähteitä, joiden avulla voidaan luoda toimeksiantajan tietotarpeeseen vastaava lopputuote. Tiedon onnistuneen keräämisen kannalta on keskeistä, että tietotarve on ensin muunnettu tiedusteluvaatimuksiksi, jotka toimivat tietotarpeeseen vastaavana toimenpidelistana. Tätä listaa hyödynnetään tiedonkeräyksen strategian luomisessa, tavoitteisiin soveltuvien lähteiden valitsemisessa ja tiedon varsinaisessa keräämisessä. (Open Source Intelligence Handbook 2001, 17.)

Tiedustelutarpeet voivat olla analyytikko- tai tapahtuma-vetoisia tai ne voivat perustua ennalta määritettyyn aikatauluun. Siinä missä analyytikkovetoinen tietotarve perustuu asiakkaalla oleviin yksilöityihin tarpeisiin ja vaatimuksiin, tapahtumavetoinen tietotarve puolestaan syntyy yleensä jonkun tietyn keskeisen tapahtuman seurauksena. Kolmantena vaihtoehtona on ennalta määritettyyn aikatauluun perustuva tiedustelutarve, jossa voidaan seurata jonkin tietyn tapahtuman kehitystä säännöllisin väliajoin. Koska tiedon keräämiseen käytetty

aika on pois tiedon analysointiin käytettävästä ajasta, on tiedustelutiedolta ja lopputuotteelta edellytettävä tarkkuus ja yksityiskohtaisuus määriteltävä ennen tehtävän aloittamista. (Open Source Intelligence Handbook 2001, 17 - 19.)

Tiedon prosessoinnin ja hyödyntämisen vaihe on prosessin keskeisin vaihe, jossa analyytikko käyttää arvostelukykyyään ja analyysitekniikoita kerätyn tiedon merkityksen arvioimiseen. Selkeät analyysimallit eri tyyppisen ja -tasaisen tiedon analysoimisessa helpottavat toisaalta analyytikon työtä ja toisaalta tuovat esiin kerätyssä tiedossa olevia puutteita tai sen yhteyksiä muuhun tietoon. Avoimista lähteistä kerätyssä tiedossa ja sen analysoimisessa on useita ongelmia, jotka liittyvät muun muassa analyytikon ennakoasenteisiin sekä julkaisijan ja julkaisun tiedon luotettavuuteen. Näistä syistä on keskeistä että analyytikko selvittää mahdollisimman tarkasti alkuperäisen tiedon lähteen sekä arvion tiedon luotettavuudesta kerätessään tietoa. Ilman riittävää lähteen ja kerätyn tiedon luotettavuuteen liittyvää arviota analyytikko voi käyttää ennakkoluuloista tai vääristynyttä tietoa osana lopputuotetta, joka johtaa joko osittain tai kokonaan virheelliseen lopputulokseen. Tyypillisinä tiedon arviointikokonaisuuksina käytetään tarkkuutta, luotettavuutta, ja auktoriteettia, ajankohtaisuutta, objektiivisuutta sekä relevanttiutta. (Open Source Intelligence Handbook 2001, 23 - 25.)

Kerätyn ja analysoidun tiedustelutiedon lopputuotteet jaetaan neljään eri ryhmään. Ensimmäisenä ryhmänä ovat raportit, jotka ovat tyypiltään yleisluotoisia tilannekuvauksia tai määritellyn tiedustelutarpeen täyttäviä kuvauksia. Raportin varsinaiselle sisällölle ei ole asetettu vaatimuksia ja se voidaan rakentaa kunkin tiedustelutarpeen mukaisesti. Raportin tulisi kuitenkin sisältää analyttinen yhteenveto raportin lähdeaineiston tiedoista, lyhyet yhteenvedot raportin eri teemoista sekä suoria lainauksia lähteenä käytetystä materiaalista. Samoin raportista tulisi käydä ilmi sen kattama ajanjakso sekä milloin tiedustelutiedon kerääminen raporttia varten on päättynyt. (Open Source Intelligence Handbook 2001, 29 - 30.)

Toisena keskeisenä tyyppinä on linkkitaulukko, jonka pääasiallinen tarkoitus on toimia tukiaineistona tiedustelutietoa keräävälle analyytikolle. Taulukko voidaan rakentaa tarpeiden mukaan mutta sen tulisi sisältää kuvaukset lähteen arvosta, lähteen verkko-osoite sekä lyhyt yhteenveto lähteen sisällöstä. (Open Source Intelligence Handbook 2001, 31.) Kolmantena tyyppinä on etäopiskeluportaali, joka sisältää organisaation toiminnan kannalta merkittäväksi luokiteltua tietoa. Portaalin tarkoituksena on toimia lähtökohtana muun tiedustelutiedon keräämiselle. (Open Source Intelligence Handbook 2001, 31 - 32.) Neljäntenä ja viimeisenä tyyppinä ovat joko suljetut tai avoimet keskusteluryhmät. Ryhmiin kerätään eri aloja ja alueita tuntevia asiantuntijoita, jotka osallistuvat tiedon analysoimiseen ja vastavuoroisesti saavat hyödyntää ryhmän tuottamaa materiaalia. (Open Source Intelligence Handbook 2001, 32.)

Avoimiin lähteisiin perustuvan tiedustelun viimeisenä vaiheena on lopputuotteen jakelu. Koska lopputuotteeseen kerätty tieto perustuu avoimiin lähteisiin, sitä voidaan jaella vapaasti eri tahoille lopputuotteen tilaajan harkinnan sekä tehtävänannon tarkoitusten mukaisesti. Tyypillisiä tahoja ovat yksityisen sektorin kumppaniyritykset ja eri kansalaisjärjestöt. Varsinainen jakelu voidaan suorittaa eri organisaatioille muun muassa hyödyntämällä Internetin palveluita. (Open Source Intelligence Handbook 2001, 33 - 35.)

## 2.2 UNODC: Tiedusteluprosessi

Perinteisen tiedusteluprosessin sykli jaetaan tyypillisesti seitsemään eri vaiheeseen. Näitä vaiheita ovat ohjeistaminen, tiedon kerääminen, kerätyn tiedon arviointi, tiedon tallentaminen, analysointi, johtopäätösten tekeminen sekä tulosten esittäminen tiedustelutyön toimeksiantajalle. Ohjeistamisvaiheessa tarkoituksena on kartoittaa ja löytää asiakkaalla olevat tarpeet ja vaatimukset tiedusteluprosessin lopputuotteelle. Tästä johtuen tiedustelutiedon analysoimiseen ja sen ohjaamiseen vaikuttavat lopputuotteen toimeksiantajan antamat tehtävät. Näillä tehtävillä on kaksi päätyyppiä, joista ensimmäisessä asiakas määrittelee tietyn aiheen tai aihealueen, joihin hän haluaa vastauksen lopputuotteessa. Toisena tyyppinä on asiakkaan yleisluontoinen ennako-odotus liittyen johonkin tälle haitalliseen tai edulliseen tapahtumaan, jonka mahdolliseen toteutumiseen asiakas on etsimässä vastausta. (Criminal Intelligence 2011, 10 - 11.)

Tiedon keräysvaiheessa tarkoituksena on kerätä materiaalia jonka avulla toimeksiantajan määrittelemä tehtävän tai ongelman ratkaisu voidaan selvittää. Hyödynnettävää materiaalia voidaan kerätä lukuisista julkisista ja ei-julkisista lähteistä. Lähteiden ja lähdeaineiston runsaus voi kuitenkin johtaa helposti tilanteeseen, jossa tietoa kerätään liian paljon tai toimeksiantannon kannalta keskeistä tietoa jää havaitsematta. Tämän torjumiseksi analyytikon tulee ennen tiedonkeräyksen aloittamista, laatia suunnitelma kerättävästä tiedosta ja siihen liittyvistä tietotarpeista sekä kerättävän tiedon mahdollisista lähteistä. (Criminal Intelligence 2011, 11 - 12.)

Tiedon arviointivaihe on keskeinen osa tiedusteluprosessin sykliä, sillä tehtyjen päätelmien tarkkuus riippuu olennaisesti päätelmien tukena olevan tiedon laadusta ja tarkkuudesta. Tärkeää on että tiedon laatua ja tarkkuutta arvioidaan välittömästi tietoa kerättäessä, jotta keräämisessä voidaan huomioida se minkälaisessa yhteydessä kyseinen tieto on kerätty. UNODC esittää raportissaan kaksi erilaista kerätyn tiedon arviointimenettelyä jotka ovat 4x4- ja 6x6-menetelmät. (Criminal Intelligence 2011, 13.) Näitä menetelmiä käydään tarkemmin läpi jäljempänä.

Tiedon tallentamisvaiheessa kerätty tieto käydään läpi ja siitä poistetaan lopputuotoksen kannalta tarpeeton tieto. Jäljelle jäänyt materiaali luetteloidaan ja tallennetaan esimerkiksi tietotokantaan, jotta sen tietosisältöä voidaan käyttää analyysissa. (Criminal Intelligence 2011, 13.)

Analysointivaihe on koko prosessin keskeisin vaihe, koska siinä tutkitaan lopputuotosta varten kerättyä tietoa ja sen merkitystä. Tarkoituksena on arvioida kerätyn tiedon laatua ja siinä olevia vahvuuksia sekä sen mahdollisia puutteita. Arvioinnin perusteella analyttikko voi ohjata ja suunnitella työn muita vaiheita ja tarvittaessa palata prosessin aikaisempaan vaiheeseen. Tiedon analysointivaihe jaetaan kahteen eri ala-vaiheeseen, jotka ovat kerätyn tiedon yhdistäminen sekä tiedon tulkinta. Tiedon yhdistämisessä tarkoituksena on koota ja lajitella eri lähteistä kerättyä tietoa erilaisiksi kokonaisuuksiksi, jotta tietoa on helpompi hyödyntää ja ymmärtää. Seuraavan vaiheen, eli tulkitsemisvaiheen onnistumisen kannalta on keskeistä, että analyttikolla on käytössään mahdollisimman paljon tietoa jotta tämä voi arvioida tiedon merkitystä ja yhteyttä liittyen muuhun aineistoon sekä tehtävänantoon. Ensimmäinen varsinainen analyttikon tekemä analyttinen tuotos on johtopäätös, joka perustuu koostetusta tiedosta kerättyihin perusteisiin. Näitä johtopäätöksiä on neljää eri tyyppiä, jotka ovat hypoteesi, ennustus, arvio ja päätelmä. (Criminal Intelligence 2011, 13 -15.)

Tulosten esittely lopputuotteen asiakkaalle on tiedusteluprosessin viimeinen vaihe. Sen varsinainen muoto riippuu asiakkaan tarpeista tai tehtävänannon tyypistä. Yleisimpiä tapoja analyysin esittämiseen on joko kirjallinen raportointi tai työn tulosten suullinen esittely. Nämä esittelytavat korostavat kuitenkin työn kertaluonteisuutta. Mikäli työ on jatkuva, voidaan muina tulosten esittelytapoina käyttää säännöllisiä suullisia tiedoksiantoja tai kirjallista raportointia. (Criminal Intelligence 2011, 15.)

Yhtenä ylimääräisenä vaiheena voidaan mainita jatkuva tiedusteluprosessin uudelleenarviointia, jonka tarkoituksena on prosessin eri vaiheiden yhteydessä pohtia tapoja kehittää toimintaa tehokkaammaksi ja luotettavammaksi. Kehityskohteet voivat liittyä niin resursointiin, tiedon keräämiseen, raportointiin kuin myös analysointiin. (Criminal Intelligence 2011, 15 - 16.)

### 3 Tiedon luotettavuuden arviointi ja analysointitekniikat

Toisin kuin NATO:n ”Open Source Intelligence Handbook”-ohjeistuksessa, UNODC:n ”Criminal Intelligence: Manual for Analysts”-ohjeistuksessa käydään läpi useita eri tapoja joiden avulla kerätyn tiedon luotettavuutta voidaan arvioida ja analysoida. Ohjeistuksessa käydään läpi kaksi erilaista tiedon luotettavuuden arviointimenetelmää ja neljä eri tiedon analyysitekniikkaa. Läpikäytyt tiedon arviointimenetelmät ovat 4x4- ja 6x6-menetelmät, joilla arvioidaan

tiedon luotettavuuden lisäksi myös tietolähteen luotettavuutta (Criminal Intelligence 2011, 25.)

Ensimmäinen ohjeistuksessa esitelty analyysitekniikka on verkostanalyysi, jonka tarkoituksena on selvittää analyysin eri kohteiden välisiä suhteita ja tämän jälkeen havainnollistaa niitä graafisessa muodossa analyysityön helpottamiseksi (Criminal Intelligence 2011, 35.) Tapahtumakaaviot ovat toinen esitelty analyysitekniikka. Niiden tarkoituksena on tunnistaa analyysin kannalta keskeiset tapahtumat, henkilöt ja organisaatiot, sekä esittää ne analyysin helpottamiseksi aikajanalla. (Criminal Intelligence 2011, 49.)

Kolmantena analyysitekniikkana esitellään vuorokauden analyysi, jonka tarkoituksena on havainnollistaa analyysin kohteena olevan tapahtumaketjun etenemistä. Vuorokauden analyysin vahvuutena on mahdollisuus tunnistaa analysoitavan kohteen ja siihen kuuluvien jäsenten toimintaa sekä näiden merkitystä toiminnalle. (Criminal Intelligence 2011, 53.) Viimeisenä esiteltynä analyysitekniikkana on puheluanalyysi jota voidaan käyttää joko viestinnässä olevien kaavojen tunnistamiseen tai puhelun mahdollisen merkityksen ja sen sisällön päättelyyn (Criminal Intelligence 2011, 59.) Puheluanalyysi ei kuitenkaan ole avoimiin lähteisiin perustuvan tiedustelun kannalta merkityksellinen, sillä sen hyödyntäminen edellyttää tyypillisesti viranomaisvaltuuksia, eikä kyseessä ole kaikille avoin lähde.

### 3.1 Kerätyn tiedon 4x4- ja 6x6-arviointimenetelmät

UNODC esittää kaksi eri tiedon luotettavuuden arviointijärjestelmää joita kutsutaan 4x4- ja 6x6-järjestelmiksi. 4x4- ja 6x6-järjestelmien tarkoituksena on selvittää tietolähteen luotettavuutta tähän liittyvien ominaisuuksien ja tunnuspiirteiden kautta. Tämän jälkeen arvioidaan varsinaisen kerätyn tiedon luotettavuutta hyödyntämällä tiedon lähteen ja varsinaisen tiedon välistä suhdetta. Lähteenä toimivan tahon sekä lähdeaineiston luotettavuus tulee arvioida molemmat erillisissä vaiheissa, ja arvioinnin tulee perustua ammattitaitoiseen ja perusteltuun näkemykseen eikä tiedon arvioijan henkilökohtaisten tunteiden tai mielipiteiden saa antaa vaikuttaa arvioinnin tulokseen. Arviointi tulee tehdä myös mahdollisimman lähellä alkuperäistä lähdeaineistoa. (Criminal Intelligence 2011, 25.)

A	Lähde on osoittautunut aikaisemmin aina luotettavaksi tai lähteen aitoudesta, eheydestä, luotettavuudesta tai pätevyydestä ei ole epäilystä.
B	Lähde, jolta tieto on kerätty, on osoittautunut suurimmassa osassa tapauksista luotettavaksi.
C	Lähde, jolta tieto on kerätty, on osoittautunut suurimmassa osassa tapauksista epäluotettavaksi.
X	Lähteen luotettavuutta ei voida arvioida.

Taulukko 1. 4x4-järjestelmän lähteen luotettavuuden arviointikriteerit (Criminal Intelligence 2011, 26).

Taulukko 1 on ensimmäinen neliosainen 4x4-järjestelmässä käytetty kriteeristö tietolähteen luotettavuuden arvioimiseksi. Asteikko on neliportainen, jossa A kuvaa luotettavaksi arvioitavaa tietoa ja X tietoa jonka luotettavuutta ei voida arvioida. Väliasteet B ja C kuvaavat joko sitä, että tieto on aikaisemmin osoittautunut tyypillisesti joko luotettavaksi tai epäluotettavaksi. 4x4-järjestelmässä korostetaan tietolähteestä aikaisemmin kerättyjen tietojen luotettavuutta.

1	Tiedon tarkkuudesta ei ole epäilystä.
2	Lähde tuntee tiedon, mutta tiedon välittänyt taho ei. Tieto on loogista ja yhdenmukaista muun aiheeseen liittyvän tiedon kanssa.
3	Lähde ei tunne tietoa henkilökohtaisesti, mutta muu kerätty tieto tukee kyseisen tiedon todenmukaisuutta.
4	Lähde ei tunne tietoa henkilökohtaisesti eikä tiedon oikeellisuutta pystytäkään varmistamaan muista lähteistä.

Taulukko 2. 4x4-järjestelmän tiedon luotettavuuden arviointikriteerit (Criminal Intelligence 2011, 26).

Taulukko 2 arvio kerätyn tiedon luotettavuutta ja tarkkuutta. Vaikka kriteeristöllä arvioidaan tiedon luotettavuutta, painotetaan siinä samalla myös tietolähteen roolia tiedon välittämisessä. Asteikko on neliportainen, jossa 1 määrittää tiedon luotettavaksi ja 4 määrittää tiedon sellaiseksi ettei tietolähteen luotettavuutta voida varmistaa. UNODC:n esittämä toinen tiedon arviointijärjestelmä, 6x6-järjestelmä, jakautuu myös kahteen eri osa-alueeseen, jossa kummassakin on kuusi arviointikriteeriä.

A Täysin luotettavaa	Lähde osoittautunut täysin luotettavaksi. Ei epävarmuutta liittyen lähteen aitoutta, eheyttä, luotettavuutta tai pätevyyttä kohtaan.
B Yleisesti luotettavaa	Lähde osoittautunut yleisesti luotettavaksi. Hieman epävarmuutta lähteen aitoutta, eheyttä, luotettavuutta tai pätevyyttä kohtaan (yksi kohta).
C Kohtalaisen luotettavaa	Lähde osoittautunut kohtalaisen luotettavaksi. Epävarmuutta lähteen aitoutta, eheyttä, luotettavuutta tai pätevyyttä kohtaan (kaksi tai useampi kohta).
D Usein epäluotettavaa	Lähde osoittautunut menneisyydessä usein epäluotettavaksi. Perusteellista epäluuloa lähteen aitoutta, eheyttä, luotettavuutta tai pätevyyttä kohtaan.

E Epäluotettavaa	Lähde osoittautunut epäluotettavaksi menneisyydessä. Varmuus lähteen aitoudessa, eheydessä, luotettavuudessa tai pätevyydessä olevista puutteista.
F	Lähteen luotettavuutta ei voida arvioida.

Taulukko 3. 6x6-järjestelmän lähteen luotettavuuden arviointikriteerit (Criminal Intelligence 2011, 26).

Taulukossa 3 nähdään 6x6-järjestelmän lähteen luotettavuuden arviointiin käytettävä kriteeristö. Arviointikriteeristö noudattaa useammasta kohdasta huolimatta pääpiirteittäin 4x4-järjestelmän tietolähteen arvostelukriteerejä. Asteikko on kuusiportainen, ja siinä A-luokitellusta lähteestä kerättyä tietoa pidetään täysin luotettavana ja F-luokitellun tietolähteen luotettavuutta ei ole pystytty arvioimaan.

1 Vahvistettu todenmukaiseksi	Tieto on vahvistettu luotettavaksi muiden itsenäisten lähteiden kautta. Tieto on loogista ja yhteneväistä muun aiheesta kerätyn tiedon kanssa.
2 Todennäköisesti todenmukaista	Tieto on loogista ja yhteneväistä muun aiheeseen liittyvän tiedon kanssa, mutta sen luotettavuutta ei ole vahvistettu muista riippumattomista lähteistä.
3 Mahdollisesti todenmukaista	Tiedon todenmukaisuutta ei ole vahvistettu, mutta se on kuitenkin loogista ja ainakin osittain yhtenevää muun aiheeseen liittyvän tiedon kanssa.
4 Todenmukaisuus epävarmaa	Tieto ei ole epäloogista. Tiedon todenmukaisuutta ei ole vahvistettu eikä siihen ole uskottu tietoa vastaanotettaessa, mutta tieto voi kuitenkin pitää paikkansa.
5 Todenmukaisuus epätodennäköistä	Tieto on epäloogista tai se on ristiriidassa muun aiheeseen liittyvän vahvistetun tiedon kanssa.
6	Tiedon luotettavuutta ei voida arvioida.

Taulukko 4. 6x6-järjestelmän tiedon luotettavuuden arviointikriteerit (Criminal Intelligence 2011, 27).

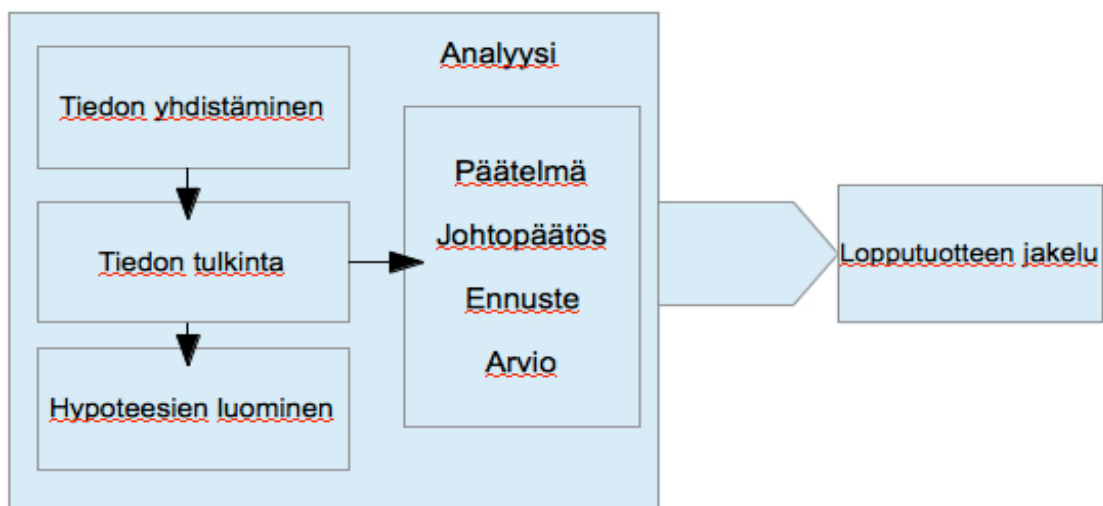
Taulukossa 4 on kuvaus 6x6-järjestelmän tiedon luotettavuuden arviointikriteeristöstä. Tässä kohtaa voidaan havaita erot 4x4- ja 6x6-järjestelmien välillä. 4x4-järjestelmän tiedon luotettavuuden arviointikriteeristö korostaa 6x6-järjestelmää enemmän tietolähteen itsensä tuntemaan tiedon merkitystä. Kun taulukon 2 ja 4 kriteeristöjä verrataan, voidaan havaita että 6x6-järjestelmän tiedon arviointikriteeristö käsittelee tiedon luotettavuutta erillisenä asiana lähteen luotettavuudesta. 4x4-järjestelmä puolestaan korostaa tiedon luotettavuuden arviointivaiheessa tietolähteen omakohtaisesti tuntemaan tietoa, ja arvioi toisen käden tietoja kriittisemmin. (Criminal Intelligence 2011, 27.)



### 3.2 Kerätyn tiedon analysointiprosessi

Tiedon analysointivaihe nähdään lopputuotoksen onnistumisen kannalta keskeisenä vaiheena, koska sen avulla on mahdollista tunnistaa analysoitavassa tiedossa olevia heikkouksia sekä vahvuuksia ja tulokset ohjaavat tarvittaessa koko tiedusteluprosessin suuntaa (Criminal Intelligence 2011, 29). Prosessin ensimmäisenä vaiheena toimii kerätyn tiedon yhdisteleminen eri kokonaisuuksiin ja teemoihin sekä näiden perusteella alustavien hypoteesien rakentaminen. Koska kyseessä on syklinen prosessi, ensimmäinen vaihe toimii myös tiedon keräystoimintaa ohjaavana tekijänä, paljastaen kerätyssä tiedossa olevia puutteita ja ohjaten tiedon keräystä eteenpäin tietotarpeiden mukaisesti. (Criminal Intelligence 2011, 29 - 30.)

Seuraavana analyysin vaiheena on kerätyn tiedon tulkitseminen. Siinä kerätylle tiedolle annetaan merkitys ja konteksti analyysin kohteessa. Tässä vaiheessa yksittäinen päivämäärä, henkilö tai tapahtuma yhdistetään osaksi isompaa analysoitavaa tapahtumaa, jonka kautta voidaan havaita tällaisen yksittäisen tiedon merkitys suuremmassa yhteydessä. Apuna voidaan käyttää esimerkiksi seuraavassa osiossa esiteltyjä tiedon analysointi- ja visualisointitekniikoita. Kerätyn tiedon ja tiedon merkityksen perusteella voidaan luoda hypoteeseja analysoitavasta kohteesta. Tyypillisesti hypoteesit sisältävät runsaasti spekulatiivista tietoa, joten hypoteesin luotettavuuden vahvistamiseksi näitä varten tulee kerätä sekä hypoteesia tukevaa että sen kanssa ristiriidassa olevaa aineistoa. Hypoteesia tukevan tai sen kanssa ristiriidassa olevan materiaalin perusteella voidaan erottaa toisistaan hypoteesit joita voidaan pitää todennäköisenä ja hypoteesit jotka ovat epätodennäköisiä. (Criminal Intelligence 2011, 29 - 31.)



Kuvio 1. Analyysiprosessin vaiheet (Criminal Intelligence 2011, 30).

Kuten kuviosta 1 voidaan havaita, varsinainen analyysivaihe koostuu kerätyn tiedon käsittelystä ja valmistelusta tiedon analysointia varten. Kerätyn ja käsitellyn tiedon pohjalta luodaan varsinainen analyttinen tuote, joka koostuu analyttikon käsittelemän tiedon ja hypoteesien pohjalta tekemistä päätelmistä, johtopäätöksistä, ennusteista ja arvioista.

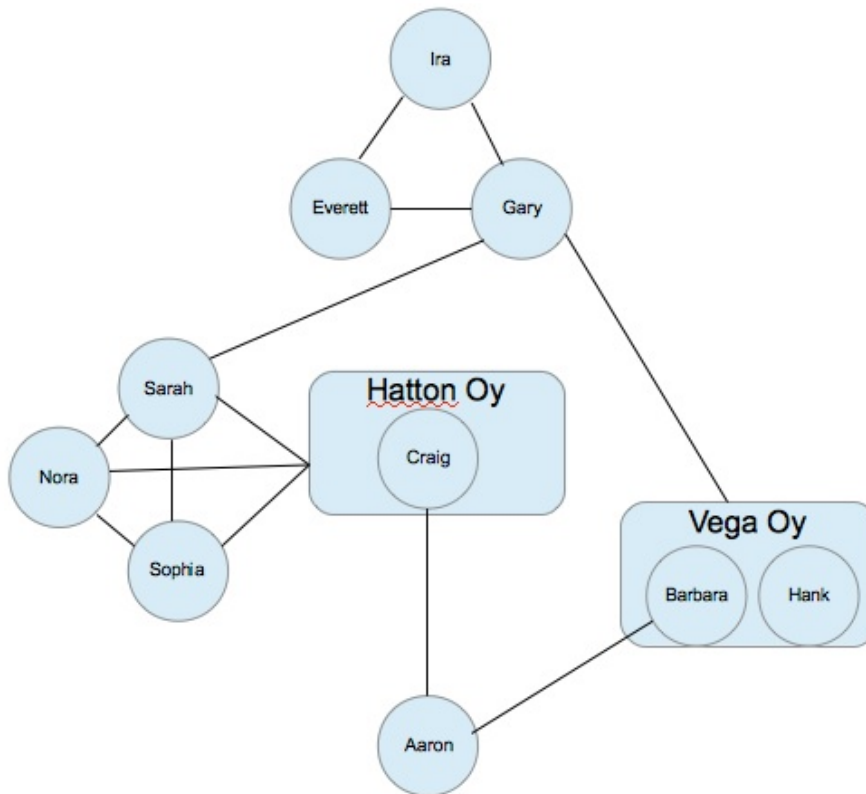
### 3.3 Verkostoanalyysi

Verkostoanalyysin tarkoituksena on esittää analyysin kohteeseen liittyvien eri organisaatioiden ja henkilöiden välisiä suhteita visuaalisessa muodossa (Criminal Intelligence 2011, 35). Analyysin avulla saadaan tietoa organisaatioiden ja henkilöiden välisistä suhteista sekä näiden suhteiden merkityksestä ja mahdollisesta sisällöstä. Sillä voidaan havaita myös verkostossa tapahtuvia muutoksia, jotka viestivät verkoston toiminnan, jäsenten tai tavoitteiden siirtymisestä. Verkostoanalyysi on riippuvainen vähintään yhdestä luotettavasta lähteestä. Ongelmana on myös analysoitavan verkoston rajaaminen, joka liian laveana voi johtaa analyysin kohteen paisumiseen sekä verkoston päivittämisen hidastumiseen tai liian suppeana verkostoanalyysistä saatetaan karsia tahattomasti analyysin kannalta keskeisiä osia pois. (Heuer & Pherson 2010, 68 - 69.)

Verkostoanalyysi jaetaan tyypillisesti useampaan vaiheeseen. Heuer & Pherson (2010) esittelevät kolmiportaisen jaon, jossa analyysi voidaan päättää siinä vaiheessa kun se antaa vastauksen analyttikon ongelmaan. Ensimmäisenä vaiheena esitellään verkostokaavion rakentaminen, jossa analyysin kannalta keskeiset toimijat ja paikat tunnistetaan ja näiden välille merkitään toimijoiden ja paikkojen suhteita sekä suhteiden laatua kuvaavat yhteydet. Toisena vaiheena on ensimmäisessä vaiheessa luodun kaavion analysointi, jossa toimijoiden ja paikkojen väliset suhteet lajitellaan niiden tyyppin mukaan. Tämän jälkeen niistä etsitään analyysin kannalta keskeisiä säännönmukaisuuksia. Viimeisenä vaiheena on sosiaalisen verkoston analyysi, jossa toimijoiden ja paikkojen yhteyksien välimatkoja toisistaan sekä suhdetyyppejä mitataan matemaattisesti. Tarkoituksena on kerätä tarkempaa tietoa verkoston eri osapuolten välisistä suhteista sekä näihin liittyvästä osapuolten toisiinsa käyttämästä vaikutusvallasta. (Heuer & Pherson 2010, 68 - 69.)

UNODC:n verkostoanalyysimallissa on seitsemän vaihetta. Ensimmäinen vaihe on raakatiedon kokoaminen analyysiä varten, jossa analyttikko kerää kaiken analyysin kannalta tarpeellisen tiedon yhteen. Toisessa vaiheessa eli verkostokaavion tarkoituksen määrittelyssä analyttikko tunnistaa raakatiedossa olevia ja analyysin kannalta keskeisiä henkilöitä, paikkoja tai muita vastaavia tunnistetietoja, joihin verkostoanalyysissä keskitytään. Kolmas ja neljäs vaihe sisältävät toimijoiden suhteiden ja niiden laadun rakentamista matriisiin, jonka tietoja käytetään myöhemmin hyväksi varsinaista verkostokaaviota piirrettäessä. Viidennessä vaiheessa analyttikko laskee toimijoihin liittyvien yhteyksien määrät. Kuudes ja seitsemäs vaihe käsittävät

varsinaisen verkostoaalyysikaavion piirtämisen kerättyjen tietojen pohjalta. (Criminal Intelligence 2011, 35 - 40.)



Kuvio 2. Esimerkki verkostoaalyysista.

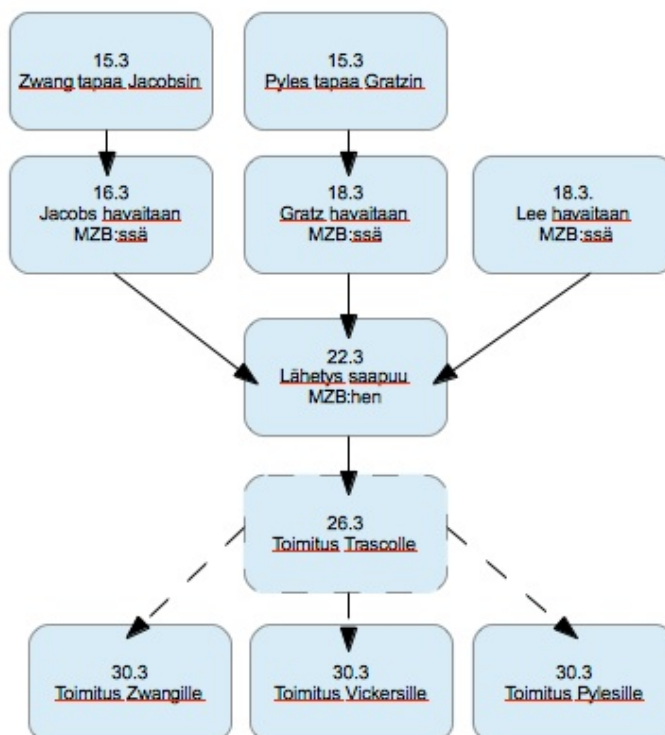
Kuviossa 2 on yksinkertainen esimerkki verkostoaalyysikaaviosta. Kaavioon on merkitty kaksi yritystä joita kuvataan laatikoilla, sekä lukuisia henkilöitä joita kuvataan ympyröillä. Henkilöiden ja yritysten väliset suhteet on merkitty viivoilla. Kuvasta nähdään, että Hatton Oy:n työntekijän Craig lisäksi yritykseen on yhteydessä henkilöiden Sarah, Nora ja Sophia muodostama ryhmä. Hatton Oy:llä on myös yhteys Vega Oy:lle Aaron nimisen henkilön kautta. Vega Oy:llä on yhteys henkilöön Gary, joka muodostaa toisen ryhmän yhdessä Iran ja Everettin kanssa. Garylla on myös yhteys ensimmäiseen ryhmään Sarahin kautta.

### 3.4 Tapahtumakaavioanalyysi

Tapahtumakaavioanalyysin vahvuus perustuu erilaisten ajallisesti eroavien tapahtumaketjujen ja niiden eri osien välisten yhteyksien visualisointiin, sekä yksittäisten tapahtumien merkityksien ymmärtämiseen osana isompaa kokonaisuutta (Criminal Intelligence 2011, 49). Tällaiset aikajanat auttavat myös trendien, merkittävien muutosten ja kehityksen sekä tapahtuman perusmalliin liittyvien poikkeamien havaitsemista. Tekniikka soveltuu käytettäväksi niin tässä hetkessä etenevien tapahtumien kuin myös jo tapahtuneiden tapahtumien analysoinnissa. Jäl-

kikäteen tapahtumasta tehtävässä tapahtumakaavioanalyysissä painopiste on tätä varten kerättyyn tiedustelutietoon ja virheelliseen analyysiin johtaneiden syiden havaitseminen ja analysointi. (Heuer & Pherson 2010, 52.) Morgan D. Jones (2009) toteaa ihmisillä olevan pa-konomainen tarve luokitella ja nähdä tapahtumat aikajärjestyksessä, tarkemmin katsottuna syy-seuraus-suhteessa muihin tapahtumiin nähden. Tapahtumien kronologinen järjestely on analysoitavan tiedon kontekstin ja merkityksen ymmärtämisen kannalta korvaamaton (Jones 2009). Toisaalta samalla tulee huomata, etteivät myöhemmät tapahtumat ole välttämättä syy-seuraussuhteessa aikaisempiin tapahtumiin (Heuer & Pherson 2010, 52 - 53).

Tapahtumakaavio voidaan rakentaa useammalla eri tavalla, mutta yksinkertaisimmillaan se koostuu analysoitavan tapahtuman tai tapahtumaketjun aikajanasta ja tähän aikajärjestykseen liitettyjen merkityksellisten tapahtumien kuvauksista ja tapahtumien välisistä suhteista. Kaavio rakentuu tyypillisesti kolmesta eri osasta joita ovat ajankohta tai tapahtuman järjestysluku, tapahtumakuvaus sekä tapahtumien välisiä suhteita kuvaavat janat ja nuolet. Lopputuloksen kannalta tapahtumakaavion tulisi antaa mahdollisimman selkeä kuva tapahtumista, olla mahdollisimman yksinkertainen ja sisältää vain analysoitavan tapahtuman kannalta keskeiset tiedot. (Criminal Intelligence 2011, 49.)



Kuvio 3. Tapahtumakaavioanalyysin esimerkki (Criminal Intelligence 2011, 50).

Kuviossa 3 on yksinkertainen esimerkki tapahtumakaavioanalyysistä. Kuvassa analyysin kannalta merkityksellisiä tapahtumia kuvataan laatikoilla joihin on merkitty tapahtuma-aika sekä

lyhyt kuvaus tapahtuman sisällöstä. Tapahtumat on järjestetty aikajärjestykseen jossa ensimmäiset tapahtumat on merkitty kuvan yläosaan ja viimeiset kuvan alaosaan. Hypoteettisia tapahtumia, joiden uskotaan tapahtuneen mutta joiden tueksi ei ole toistaiseksi kerätty tai löytynyt riittävästi todisteita, kuvataan katkonaisin viivoin.

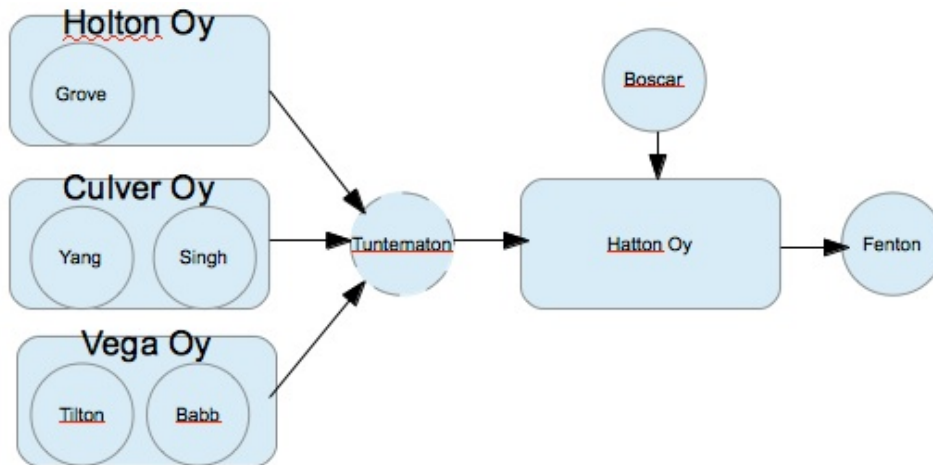
### 3.5 Vuoanalyysi

Vuoanalyysiä voidaan käyttää hyväksi tapahtumaketjun eri osien merkityksen ja toiminnan tunnistamisessa. Analyysitekniikalla selvitetään analyysin kohteena olevan tapahtuman osatapahtumia jotka ovat syy-seuraus-suhteessa varsinaiseen päätapahtumaan. (Jones 2009). Tiedustelutoiminnassa vuoanalyysiä voidaan käyttää organisaatioiden toimintatapojen tunnistamiseen ja tulkitsemiseen (Heuer & Pherson 2010, 82). Rikostorjunnan puolella analyysitekniikkaa käytetään rikollisorganisaation avainhenkilöiden ja -roolien sekä erilaisten hyödykevirtojen tunnistamiseen. Tällaisia hyödykevirtoja voivat olla muun muassa raha, huumausaineet tai aineettomat hyödykkeet. Kolme tyypillistä vuoanalyysitekniikkaa ovat hyödykkeitä, toimintaketjuja tai tapahtumia kuvaavat kaaviot. (Criminal Intelligence 2011, 53 - 54.)

Hyödykkeitä kuvaavassa vuoanalyysissä seurataan materiaalistien hyödykkeiden tai palveluiden liikkumista yritysten, henkilöiden tai tapahtumapaikkojen välillä. Analyysillä pyritään selvittämään hyödykkeiden siirtämisen merkitystä. Keskeisiä tietoja analyysin kannalta ovat siirrettävien hyödykkeiden määrät, tyypit, päivämäärät, toimintaan osallistuvat henkilöt ja organisaatiot sekä hyödykevirran suunnat. Hyödykevuokaavio voidaan rakentaa joko hyödyntäen verkostanalyysiä johon on merkitty hyödykevirrat, tai matriisia johon on kerätty analysoitavat tiedot. Näiden tietojen pohjalta rakennetaan varsinainen vuokaavio. (Criminal Intelligence 2011, 54 - 56.)

Toimintaketjuja kuvaavia vuokaavioita käytetään luomaan yleiskuva tapahtumaketjuun kuuluvista välttämättömistä vaiheista. Näiden eri vaiheiden pohjana käytetään aikaisempia vastaavia tapahtumaketjuja ja niissä olevia osavaiheita. Tämä tarkoittaa sitä, että analyysin näkökulma on yleisluontoinen. Toimintaketjuja kuvaavia vuokaavioita voidaan hyödyntää eri toimintaketjujen vertailuun ja samankaltaisuuksien havaitsemiseen tai monimutkaisten toimintaketjujen esittämiseen. (Criminal Intelligence 2011, 56 - 57.)

Tapahtumavuoanalyysin tarkoituksena on analysoida tiettyä tapahtumaa edeltäneitä tapahtumia sekä kyseiseen tapahtumaan liittyviä ja sitä seuraavia tapahtumia. Varsinainen analysoitava tapahtuma luo kontekstin muille siihen liittyville tapahtumille, joka mahdollistaa tällaisten osatapahtumien havaitsemisen ja merkityksen tulkitsemisen osana kokonaisuutta. (Criminal Intelligence 2011, 57 - 58.)



Kuvio 4. Esimerkki hyödykevirran vuoanalyysistä (Criminal Intelligence 2011, 53).

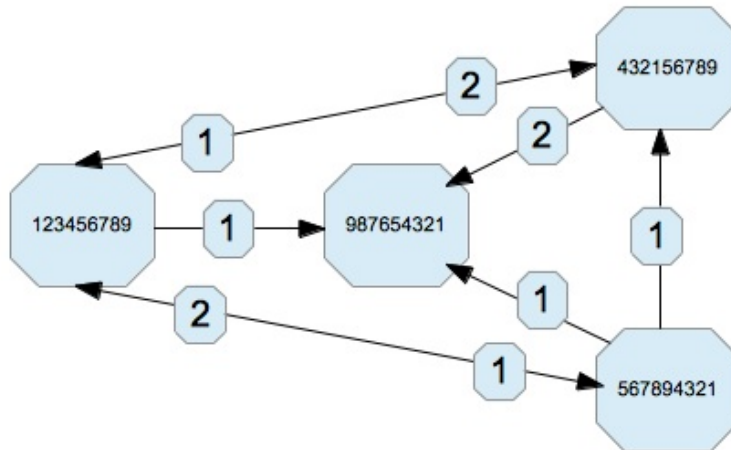
Kuviossa 4 nähdään henkilöitä ja yrityksiä sisältävästä vuoanalyysistä, jossa yritykset on merkitty laatikoilla ja henkilöt ympyröillä. Kuvassa oleva katkoviivoilla merkitty ympyrä kuvaa vuokaavioon kuuluvaa mutta toistaiseksi tuntematonta tekijää. Kuvasta nähdään että hyödykevirta on lähtöisin kolmesta eri yrityksestä ja näissä toimivista viidestä eri henkilöstä. Hyödykevirta kulkee tuntemattomaksi jäävän henkilön sekä Hatton Oy:n kautta päätyen Fenton nimiselle henkilölle. Kuvion 4 pohjalta voidaan havaita, että hyödykevirran toiminnan kannalta tuntemattomaksi jäävä tekijä on keskeinen, ja mikäli tämä tunnistetaan ja saadaan poistettua hyödykevirta henkilölle Fenton keskeytyy joko hetkellisesti tai pysyvästi.

### 3.6 Puheluanalyysi

Puheluanalyysillä tarkoitetaan puhelua kuvaavien tietojen keräämistä ja analysointia. Tällaista tietoa ovat puhelun päivämäärään ja kellonaikaan liittyvä tieto, puhelujen lukumäärä, soitetut ja vastaanotetut puhelut, puhelujen kestot sekä puhelinnumerot. Käytettäviä analyysitapoja ovat tilastollinen, laadullinen sekä assosiaatioanalyysi. Tilastollinen analyysi keskittyy kerätyssä materiaalissa olevien viestintäkaavojen etsimiseen, kun taas assosiaatioanalyysin avulla on mahdollista muodostaa hypoteeseja viestinnän osapuolten välisestä suhteesta, puhelun tarkoituksesta sekä sisällöstä, käyttäen hyväksi tilastollisia puhelutietoja. Keskeistä puhelutietojen tulkitsemisen kannalta on ymmärtää sen olevan suunnattua tietoa puhelun soittajan ja vastaanottajan välillä. Laajemmin ajateltuna puheluanalyysin ei tarvitse rajoittua vain puhelimella tehtyyn viestintään, vaan se voi periaatteessa kattaa kaikenlaisen elektronisen viestinnän joka koostuu viestin lähettäjästä ja vastaanottajasta. (Criminal Intelligence 2011, 59 - 60.)

Puheluanalyysin kannalta keskeistä tietoa ovat puhelun soittajan ja vastaanottajan puhelinnumerot sekä puhelujen tiheys. Analysoinnin apuna voidaan käyttää matriisia, jossa soittava puhelinnumero merkitään vasemmalle reunalle ja puhelun vastaanottava numero yläreunalle.

Oikealle reunalle merkitään soittajan puheluiden kokonaismäärä ja alareunalle vastaanottavan puhelinnumeron kokonaismäärät. Kun matriisi on saatu täytettyä, luodaan tämän tietojen pohjalta verkostanalyysikaavio johon merkitään soittavan ja vastaanottavan puhelinnumeron lisäksi puheluiden suunnat sekä määrät. (Criminal Intelligence 2011, 60 - 63.)



Kuvio 5. Puheluanalyysin vuokaavion esimerkki (Criminal Intelligence 2011, 63).

Kuviossa 5 nähdään esimerkki neljän eri puhelinnumeron välisten yhteydenottojen suunta sekä lukumäärä. Puhelinnumeroita on kuvattu isoilla kahdeksankulmioilla ja tehtyjen puheluiden määriä pienillä kahdeksankulmioilla. Yhteydenottojen suuntaa kuvataan nuolilla. Kuten kuvasta voidaan havaita, puheluanalyysin vuokaavio muuttuu analyysiin kuuluvien puhelinnumeroitten kasvaessa sekavaksi. Tästä johtuen laajemmissa puheluanalyysissä suositellaan käytettäväksi erityistä analyysiin soveltuvaa ohjelmistoa, analyytikon itsensä piirtämien kaavioiden sijaan (Criminal Intelligence 2011, 64).

### 3.7 Analysoidusta tiedosta johdetut päätelmät

Yllä kuvatut tiedon analysointitekniikat auttavat analyttikkoa kerätyn tiedon luokittelussa, järjestämisessä, siinä olevien yhteyksien ja puutteiden havaitsemisessa sekä tiedon visualisoinnissa. Analyysitekniikoilla pyritään muokkaamaan tietoa yksinkertaisempaan ja helpommin tulkittavampaan muotoon, jotta analyttikko voi muodostaa ja testata erilaisia johtopäätöksiä ja teorioita käsiteltävänä olevaa tietoa vastaan. Teorioiden ja hypoteesien kokeilulla voidaan poistaa analysoitavasta joukosta epätodennäköiset vaihtoehdot ja keskittyä todennäköisempiin, kerätyn aineiston tukemiin, vaihtoehtoihin. (Criminal Intelligence 2011, 65.) Päätelmätyyppejä on kaksi, jotka ovat oletamus ja johtopäätös (Criminal Intelligence 2011, 66 - 67).

Olettamuksilla tarkoitetaan argumentin perustana toimivaa tietoa. Ne ovat tiedon analysoinnin kannalta keskeinen tekijä, sillä niiden avulla tuetaan analysoitavan tiedon perusteella

muodostettuja johtopäätöksiä. Olettamukset voidaan muodostaa joko yhdestä tai useammasta kerätystä tiedon yksiköstä joiden määrää ja laatua käytetään olettamuksen todennäköisyyden ja merkityksen arviointiin. Koska olettamukset muodostetaan käyttämällä lähteenä olevaa kerättyä tietoa, ne ovat tästä syystä usein raakatiedon läheisyydestä johtuen objektiivisin tulkinta tästä tiedosta. Toisaalta koska olettamukset toimivat johtopäätösten argumentteina, niiden avulla voidaan muodostaa useita erilaisia johtopäätöksiä. (Criminal Intelligence 2011, 66 - 67.)

Johtopäätöksellä tarkoitetaan raakatiedosta johdettujen olettamusten tukemaa tulkintaa analyysin kohteena olevasta asiasta. Kyseessä on siten tulkinta kerätyn tiedon merkityksestä. Johtopäätöksestä saatavan hyödyn kannalta on tärkeää että sen luotettavuudesta ja todennäköisyydestä esitetään arvio. Tätä todennäköisyydestä esitettyä arvioita voidaan käyttää hyväksi analyysin tai tiedonkeräyksen suuntaa ohjaavana tekijänä, riippuen sille asetetusta luotettavuuden arviosta. Johtopäätöstyyppejä on neljä: hypoteesit, ennusteet, arviot sekä tulokset. Hypoteeseilla tarkoitetaan selitystä jonka vahvistamisen tai kiistämisen tueksi tarvitaan lisää tietoa. Ennusteella puolestaan tarkoitetaan johtopäätöstä mahdollisesta tulevaisuudessa tapahtuvasta tapahtumasta. Kolmantena tyyppinä on kerätyn tiedon pohjalta tehtävä määrällinen arvio analyysin kohteena olevasta asiasta kuten resurssien tai ajan tarpeesta. Viimeisenä tyyppinä on tulos jonka tueksi on riittävästi todisteita ja jota voidaan pitää kattavana yhteenvetona muista arvioiduista hypoteeseista, ennusteista tai arvioista. Jotta johtopäätöksiä ja niiden luotettavuutta voidaan arvioida ja priorisoida, tulee arvioida näiden todennäköisyyttä. Todennäköisyys saadaan jakamalla analysoitavan tapahtuman lukumäärä sillä miten usein analysoitavan tapahtuman on mahdollista tapahtua. (Criminal Intelligence 2011, 67 - 68.)

### 3.8 Tiedon keräysvaiheeseen liittyvät ongelmat

Avoimista lähteistä kerättävään tietoon liittyy runsaasti mahdollisia ongelmia. NATO:n Open Source Intelligence Handbookissa (2001) tunnistettavat ongelmakohdat liittyvät operatiiviseen turvallisuuteen, tekijänoikeuksiin, vieraisiin kieliin sekä ulkopuolisiin verkostoihin. Toimintaan liittyvän turvallisuuden varmistamiseksi esitetään tiedon kerääjän henkilöllisyyden peittämistä tiedon keräysvaiheessa sekä Non Disclosure Agreement-tyyppisten sopimusten käyttämistä, mikäli yhteistyötä tehdään yksityisen sektorin toimijoiden kanssa. (Open Source Intelligence Handbook 2001, 20 - 21.) Operatiivinen turvallisuus on tärkeää, koska jokainen verkossa tehtävä haku jättää runsaasti digitaalisia jalanjälkiä jotka voidaan tietyin edellytyksin johtaa takaisin tiedon hakijaan (Open Source Intelligence Handbook 2001, 28).

Digitaalisella jalanjäljellä tarkoitetaan tietoa joka yksilöi käyttäjän muista palvelimella tarjottavalla verkkosivustolla vierailevista käyttäjistä. Internetin toimintaperiaatteesta johtuen jokainen verkkosivulla vieraileva käyttäjä jättää itsestään jälkeensä lukuisia erilaisia jälkiä.



Mikäli kiinnostusta tietoa kohtaan ja tästä johtuvaa tiedon keräystä halutaan salata, tulee näiden jälkien syntyminen ottaa tiedon keräämisessä ja keräyksen suunnittelussa huomioon. (Open Source Intelligence Handbook 2001, 27 - 28.) Yhtenä tällaisena esimerkkinä jätettävää jäljestä on tietoliikenteen reitittämisessä käytettävä Internet Protocol- eli IP-osoite, joka tietyin edellytyksin voidaan yksilöidä tiettyyn tietoliikennepalveluntarjoajan asiakkaaseen (TCP/IP Overview 2005). Toisena esimerkkinä on tehtyjen hakujen kellonaika, joita keräämällä on mahdollista päätellä hakujen suorittajan todennäköinen sijainti aikavyöhyketietojen avulla (Open Source Intelligence Handbook 2001, 28).

Tekijänoikeudet nähdään ongelmana pääasiassa lopputuotoksen levittämisen ja eettisen toiminnan kannalta. Koska avoimiin lähteisiin perustuvan tiedustelun yhtenä tarkoituksena on tuottaa yhteistyökumppaneille levitettäväksi tarkoitettua materiaalia, on tärkeää että keräysprosessissa noudatetaan tekijänoikeuksia. Tämä mahdollistaa sen että tiedustelun lopputuotetta voidaan jakaa laajalti eri yksityisen sektorin yhteistyökumppaneiden kanssa. Kolmantena ongelmana nähdään vieraat kielet, sillä avoimista lähteistä saatava ja lopputuotteen kannalta keskeinen tieto ei välttämättä ole saatavilla analyytikon osaamilla kielillä. Neljänneksi ongelmaksi on mainittu ulkoiseen verkostoitumiseen liittyvät ongelmat jotka koskevat yhteistoimintaa ja tiedonvaihtoa yksityisen sektorin toimijoiden ja asiantuntijoiden kanssa. (Open Source Intelligence Handbook 2001, 20 - 21.)

### 3.9 Tiedon analysointivaiheeseen liittyvät ongelmat

Tiedon analysointivaiheeseen liittyy runsaasti eri ongelmia, jotka saattavat johtua joko jo tiedon keräysvaiheessa tehdyistä ratkaisuksista tai ihmisten tavasta käsitellä tietoa. Yhtenä tiedon keräysvaiheesta johtuvana ongelmana on tiedonkeräyksen suppeus. Tiedon keräysvaiheessa löydetyn yksittäisen tiedon merkitystä kokonaisuudelle voi olla hankala arvioida. Tämä saattaa johtaa tilanteeseen jossa analyysin kannalta kriittistä tietoa on karsittu pois ennen analyysivaihetta, koska tiedon merkitystä ei ole ymmärretty vielä tiedon keräysvaiheessa. (Criminal Intelligence 2011, 14.) Analyytikon käyttämät lähteet saattavat myös sisältää puolueellista tai muuten vääristynyttä tietoa. Tästä syystä lähteen luotettavuus ja puolueettomuus tulee arvioida ennen siitä saatavilla olevan tiedon hyödyntämistä osana analyysia tai varsinaista lopputuotetta. (Open Source Intelligence Handbook 2001, 24.)

Kerätystä tiedosta johdettujen johtopäätösten tulisi olla perusteltavissa näiden tukena olevasta tiedosta. Tyypillisenä ongelmana on kuitenkin se että analyytikolla on ennen analyysin aloittamista ennako-oletus analyysin kohteena olevasta asiasta tai siihen vaikuttavista syistä, ja tiedon kerääminen ja analysointi keskittyy tätä ennako-oletusta tukevan tiedon etsimiseen ja käsittelyyn. (Criminal Intelligence 2011, 14.) Kyseessä on yksi kognitiivisten vääristymien tyypeistä, jotka liittyvät ihmisten virheelliseen tapaan käsitellä tietoa. Ongelmana on

myös ettei tietoisuus kognitiivisista vääristymistä tai niiden mahdollisuudesta johda yksin näistä vääristymistä vapaana oleviin tulkintoihin. Kognitiivisia vääristymiä voi esiintyä tiedon arvioinnissa, syy-seuraussuhteiden hahmottamisessa, todennäköisyyksien tulkitsemisessa sekä analyytikon itsensä esittämissä analyyseissä ja näiden luotettavuuden arvioinnissa. (Heuer 1999, 111 - 112.)

### 3.9.1 Syy- ja seuraussuhteiden virhetulkinnat

Tiedon kuvauksen eläväisyys on yksi tiedon arviointiin liittyvistä kognitiivisista vääristymistä. Siinä henkilökohtaisesti koettu tai värikkäästi kuvattu tieto muistetaan helpommin ja sille annetaan suurempi merkitys kuin tilastolliselle tai muuten pelkistetylle tiedolle. Tällainen tieto johtaa helposti virheellisiin tulkintoihin ja olettamuksiin, eikä sille tästä syystä tulisi antaa juurikaan painoarvoa varsinaisessa analyysivaiheessa. Puuttuvasta tiedosta johtuvat virheet on toinen tiedon arviointiin liittyvä vääristymä. Tiedusteluprosessin tyypillinen piirre on, ettei kaikkea siihen liittyvää keskeistä tietoa tunneta tai havaita. Analyytikon tulisi kuitenkin pystyä arvioimaan minkälaista analyysin kannalta keskeistä tietoa puuttuu, minkälainen merkitys tällaisella puuttuvalla tiedolla on analyysin lopputulokseen, ja muokata näiden perusteella omaan arviointiin liittyvää luottamusta. Herkkyys tiedon yhdenmukaisuudelle on kolmas tiedon arviointiin liittyvä vääristymätyyppi, jossa runsaalle päällekkäiselle tiedolle tai liian suppealle otannalle ja sen edustavuudelle annetaan analyysissa liiallista painoarvoa. Tällaisten virheiden välttämiseksi tulisi arvioida analysoitavan tiedon edustavuutta. Mikäli tiedon edustavuutta ei pystytä määrittelemään tai kerätty tieto perustuu yhdenmukaiseen ja suppeaan tietoaaineistoon, analyysin lopputulokseen liittyvä luottamus tulisi määritellä vähäiseksi. Neljäs vääristymätyyppi liittyy luotettavuudeltaan epävarman tiedon analysointiin. Tyypillisesti ihmiset joko hyväksyvät tai hylkäävät analysoitavan tiedon kokonaisuudessaan. Mikäli tieto hylätään, sitä ei enää huomioida myöhemmissä tiedon käsittelyvaiheissa. Toisaalta mikäli tieto hyväksytään, se hyväksytään kokonaisuudessaan välittämättä sen luotettavuuteen liittyvistä ongelmista. Vääristymän vaikutusta voidaan vähentää tekemällä arvio olettaen että analysoitava tieto on luotettavaa. Tämän jälkeen analyytikon tekemän arvion luotettavuutta vähennetään käytetyn tiedon arvioidun luotettavuuden mukaisesti. Viimeisenä tiedon arviointiin liittyvänä vääristymänä on vääräksi todetun tiedon säilyvyys, siitä huolimatta että sen tukena oleva todistusaineisto voidaan osoittaa virheelliseksi. Kun ihmisille esitetään uutta tietoa, he luovat lukuisia uusia oletuksia jotka joko tukevat tai torjuvat tiedon todenperäisyyttä. Nämä oletukset perustuvat syy-seuraus-suhteisiin, jotka selittävät esitettyä uutta tietoa. Mitä vahvempi syy- ja seuraussuhde voidaan luoda tiedon tueksi, sitä vahvempi vaikutelma syntyy tiedon todenperäisyydestä. (Heuer 1999, 116 - 126.)

Tapahtumien syy- ja seuraussuhteisiin liittyy useita eri vääristymätyyppejä. Ihmiset ovat totuneet etsimään ja löytämään tapahtumien välillä olevia säännönmukaisuuksia ja niihin johdaneita syitä. Mikäli tapahtumien välillä olevia yhdistäviä tekijöitä ei löydetä, syynä saatetaan pitää puuttuvaa tietoa tapahtumien satunnaisuuden sijaan. Säännönmukaisuuksia nähdään myös tiedossa, joka on satunnaista. (Heuer 1999, 127 - 130.) Tähän osittain liittyvänä toisena vääristymätyyppinä on suosia olettaa siitä, että tietty ryhmän tai organisaation toimintaa ohjataan keskitetysti. Tällöin tapahtumat nähdään seurauksena ryhmän tai organisaation keskitetysti ohjaamasta toiminnasta, eikä niitä tai niiden seurausten mahdollista tahattomuutta tai satunnaisuutta huomioida riittävästi (Heuer 1999, 131 - 132). Kolmantena syy- ja seuraussuhteisiin liittyvänä vääristymänä on olettaa, että oletetun tapahtuman syyn seurauksia verrataan tapahtuman oletetun seurauksen ominaisuuksiin. Tämä johtaa siihen että merkittävien ja laajojen seurausten oletetaan johtuvan merkittävistä tapahtumista, ja toisaalta siihen etteivät mitättömät tapahtumat voi aiheuttaa merkittäviä seurauksia. (Heuer 1999, 132 - 134.)

Kun henkilön käyttäytymisen syiden arvioinnissa painotetaan tähän sisäisesti vaikuttavia seikkoja kuten tämän henkilökohtaisia ominaisuuksia tai uskomuksia, on kyse kognitiivisesta vääristymästä. Ihmiset arvioivat tyypillisesti oman käyttäytymisensä ja siitä seuranneiden tekojen johtuvan kyseiseen tilanteeseen vaikuttavista ulkoisista seikoista. Tällaisia seikkoja ovat esimerkiksi vertaisryhmän tai sosiaalisen ympäristön aiheuttama paine sekä erilaiset tilanteisiin liittyvät roolit ja niiden aiheuttama käyttäytyminen. Kolmansien osapuolien tapauksessa kuitenkin voidaan tulkita näiden käyttäytymisen johtuvan ulkoisten ja tilanteeseen liittyvien vaikutteiden sijaan sisäisistä tekijöistä. Tämä johtaa siihen ettei analyysissä oteta riittävästi huomioon käyttäytymisen mahdollista tilannesidonnaisuutta. Heuer (1999) mainitseekin, että ihmisillä on tyypillisesti hyvin erilainen näkemys toisten toimijoiden käyttäytymiseen johtaneista syistä. (Heuer 1999, 134 - 138.)

Edelliseen kohtaan liittyvänä kognitiivisena vääristymänä voidaan pitää myös tapaa olettaa henkilön omien tekojen vaikutusta tai merkitystä kohteena olevan organisaation tai henkilön muuttuneissa päätöksissä, ja toisaalta vähätellä teoista riippumattomien ulkoisten tapahtumien vaikutusta muuttuneissa päätöksissä. Tämä aiheutuu siitä että henkilö tuntee omat vaikutuspyrkimyksensä kohteena olevaan tahoon nähden, mutta ei kuitenkaan näiden pyrkimysten ulkopuolelta kohdistuvia vaikutuksia ja niistä aiheutuvia käyttäytymiseen liittyviä muutoksia. (Heuer 1999, 138 - 140.)

### 3.9.2 Todennäköisyyksien arviointiin liittyvät virhetulkinnat

Tiedon saatavuuteen liittyvä virhetulkinta on yksi todennäköisyyksiin liittyvistä virhetulkintatyypeistä. Siinä tapahtuman yleisyyttä arvioidaan sen mukaan, kuinka helposti aikaisempia vastaavia tapahtumia pystytään palauttamaan mieleen. Tähän vaikuttavat muun muassa se,

onko analyysiä suoritettava taho kokenut tapahtuman henkilökohtaisesti, kuinka paljon aikaa aikaisemmasta tapahtumasta on kulunut ja kuinka tärkeänä tapahtumaa on pidetty tapahtumahetkellä. Ehkä osittain tästä johtuen analyytikon suorittama analyysi vahvistaa muistikuvaa aikaisemmista vastaavista tapahtumista. Tämä puolestaan voi johtaa saatavuuteen liittyvän virhetulkinnan mahdollisuuden kasvamiseen. (Heuer 1999, 147 - 150.)

Ankkurointiin liittyvä virhetulkinta aiheutuu siitä aikaisemman tiedon tai olettamuksen käyttämisestä lähtökohtana uudelle analyysin kohteena olevalle samankaltaiselle tiedolle. Tällainen tieto voi perustua aikaisemman analyysin tuloksiin, mutta se voi myös olla täysin irrationaalinen lähtökohta. Uutta tietoa analysoidessa lähtökohtaa muutetaan muiden tilanteeseen vaikuttavien seikkojen mukaisesti, joka saattaa johtaa virheelliseen vallitsevan tilanteen tulkintaan. (Heuer 1999, 150 - 152.)

Todennäköisyyksiin liittyvät virhetulkinnat voivat johtua myös kirjoitetun kielen epätarkkuudesta ja tämän aiheuttamasta virheellisestä tulkinnasta. Analyysin lopputuote sisältää tyypillisesti arvion siitä, kuinka todennäköisenä analyytikko pitää analysoimansa tapahtuman toteutumista. Tätä todennäköisyyttä kuvataan usein kielellisellä ilmaisulla lukuarvon sijaan, jonka todennäköisyyden lukija tulkitsee itse käyttämällä lähteenä asiayhteyttä sekä omaa näkemystään tapahtuman todennäköisyydestä. Näin analyytikon tarkoittama tapahtuman todennäköisyys voi poiketa analyysin lopputuotteen lukijan omasta todennäköisyyden tulkinnasta. Mikäli tapahtumaa kuvataan tapahtumaketjuna, kuvattujen yksityiskohtien määrä vaikuttaa tapahtumaketjun todennäköisyyden tulkintaan siitä huolimatta että varsinaiset yksityiskohdat eivät liittyisi mitenkään tapahtumaketjun todennäköisyyteen. (Heuer 1999, 152 - 157.)

Tapahtuman esiintyvyyteen liittyvä virhetulkinta aiheutuu, kun analysoitavaan tapahtumaan liittyy samalla kertaa sekä tapahtumaa yleisellä tasolla kuvaavaa tilastollista materiaalia yhdessä tiettyä yksittäistä tapahtumaa kuvaavaa tietoa. Tällaisissa tilanteissa virhetulkinta aiheuttaa sen ettei tilastollista materiaalia hyödynnetä, ellei se tue yksittäistä tapahtumaa ja siihen johtaneita syitä. (Heuer 1999, 157 - 160.)

### 3.9.3 Tiedon jälkikäteisen tulkinnan aiheuttamat virhetulkinnat

Tiedustelutieto ja siihen nojautuva analyysi on luonteeltaan joko nykyistä tilannetta selittävää tai mahdollisia tulevia tapahtumia arvioivaa. Analyysin laatuun ja luotettavuuteen liittyvä virhetulkinta syntyy, kun aikaisemmin tehtyä analyysiä verrataan jälkikäteen kerättyyn ja tilanteeseen liittyvään tietoon. Tällöin analyysin laatuun ja luotettavuuteen liittyvänä arviointikriteerinä käytetään sitä, mitä analyytikon olisi pitänyt tietää tilanteesta analyysihetkellä ja miten tämän olisi pitänyt tunnistaa nykyhetkeen johtaneet tapahtumat. Koska nykytilaa ja sitä edeltävää analyysiä katsotaan kuitenkin jälkikäteisesti, ei analyysiä ole kuitenkaan mah-

dollista arvioida kuten ennen nykyhetkeä kuvaavan uuden tiedon vastaanottamista. (Heuer 1999, 161 - 163.)

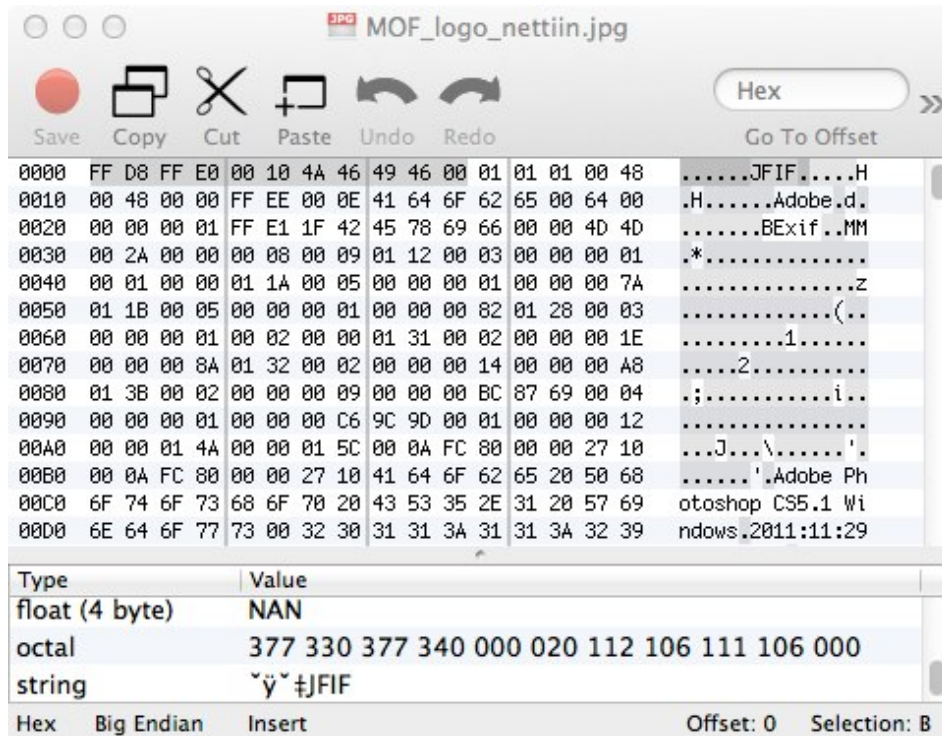
### 3.10 Tiedoston metatiedon tutkiminen

Dokumenttitiedostot sisältävät runsaasti metatietoa dokumentista, dokumentin luojaista tai organisaatiosta enemmän kuin sen julkaisija on tarkoittanut. Tällainen tieto voi olla dokumentin luomiseen tai editoimiseen käytetty ohjelmisto, ohjelmistoversio, dokumenttiin tehtyjen muokkauksien ajankohta, muokkauksien määrä tai dokumenttia käsitelleet käyttäjät. Ensimmäisessä vaiheessa käydään läpi metatietojen etsimistä dokumenttitiedostoista, jonka jälkeen käydään läpi käytettyjä hakutermejä.

Metatietoa tutkittaessa tulee huomioida se, että kaikki tiedostotyyppit eivät välttämättä tue metatiedon lisäämistä. Samoin tulee huomata se, ettei metatiedon lisääminen, poistaminen tai muokkaaminen vaikuta millään tavalla tiedoston tarkoitettuun toimintaan. (Altheide, Carvey & Davidson 2011, 173.) Metatiedon tutkimisen havainnollistamiseksi on valittu Laurea-ammattikorkeakoulun verkkosivuilta löydettyä kuvatiedostoa, joka on nimetty "MOF\_logo\_nettiin.jpg". Kuvatiedosto on tallennettu 7.12.2012 URL-osoitteesta "http://www.laurea.fi/SiteCollectionImages/Etusivu/MOF\_logo\_nettiin.jpg". Tarkoituksena on ensin käydä läpi tiedostoa ja sen rakennetta hex-editorin avulla. Tämän jälkeen havainnollistetaan saman tiedon keräämistä tätä varten tehdyn ohjelman avulla.

Tietokone käsittelee tietoa binääritasolla, jossa tiedon kuvaamiseen käytetään 1- sekä 0-numeroita. Kyseessä on base2-numerojärjestelmä jonka nimi tulee numerojärjestelmässä käytettävissä olevien numeroiden määrästä. Heksadesimaalinumerojärjestelmä on puolestaan base16-numerojärjestelmä. Siinä käytetään numeroita 0-9 ja kirjaimia A-F, joista desimaalilukuina A on 10 ja F on 15. Kahdella heksadesimaaliluvulla muodostetaan yksi tavu, joka koostuu kahdeksasta bitistä eli 1-tai 0-numerosta. (Carrier 2005, 17 - 21.) Heksadesimaalieditori on tietokoneohjelma joka näyttää tiedoston rakenteen sekä bitti että tavutasolla (Hex Editor Definition 2006).

Kuvassa 1 on näkymä "MOF\_logo\_nettiin.jpg"-tiedoston sisällöstä 0xED-nimisen heksadesimaalieditorin kautta katsottuna. Vasemmassa reunassa pystysuoralla linjalla nähdään heksadesimaaleina kuvattu offset-luku, jolla määritetään sijainti tiedostossa. Ensimmäisellä rivillä luku on "0000", koska ensimmäinen "0xFF"-tavu sijaitsee offsetissä 0. Seuraavalla rivillä luku on "0010", koska rivin ensimmäinen "0x00"-tavu sijaitsee offsetissä 16.



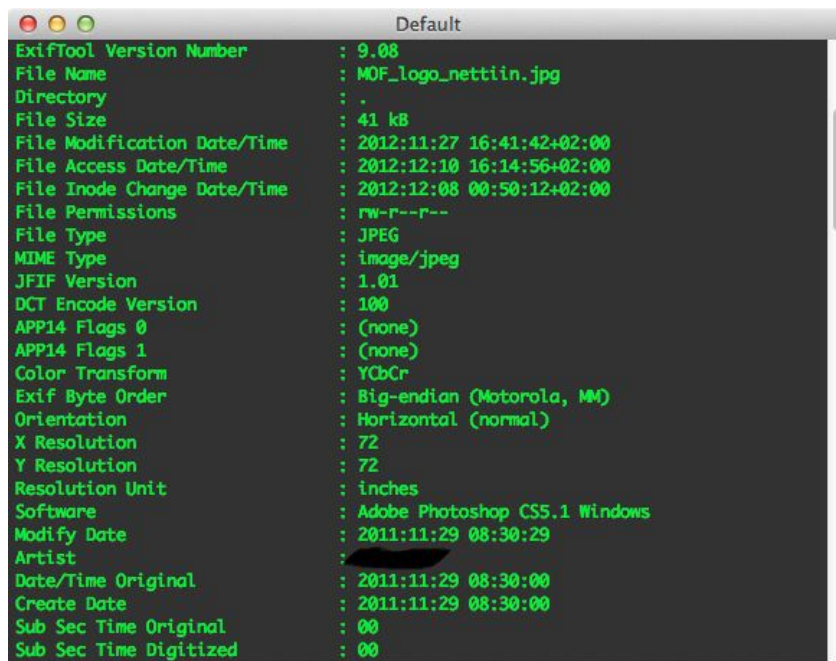
Kuva 1. JPEG-kuvatiedoston sisältöä heksadesimaalieditorin kautta katsottuna.

JPEG File Interchange Format-dokumentissa kuvataan JPEG-tiedoston yksinkertaista tiedostorakennetta (Hamilton 1992, 1). Määrittelyn mukaan kyseinen tiedostotyyppi alkaa Start of Image- eli SOI-merkinnällä, joka on heksadesimaalina 0xFFD8. Tätä seuraa APP0-merkintä 0xFFE0. Seuraavat kaksi tavua 0x0010 ovat APP0-merkinnän pituus tavuina ja tämän jälkeen tulee varsinainen JPEG-tiedoston tunnus eli JFIF-merkintä 0x4A46494600. Start of Image-, APP0- sekä JFIF-merkinnät muodostavat keskeisen osan JPEG-tiedoston ylätunnisteesta, jota tiedostoa käsittelevät ohjelmat käyttävät tiedostotyyppin tunnistamiseen. (Hamilton 1992, 5.) Mainittu ylätunniste näkyy kuvassa 1 tummennettuna. JPEG-tiedosto päättyy tiedoston lopusta löytyvään EOI:n eli End of Image-merkintään, jota kuvataan heksadesimaaleilla 0xFFD9 (JPEG File Interchange Format 2012). Kuten kuvan 1 oikeasta reunasta nähdään, ”MOF\_logo\_nettiin.jpg”-tiedosto sisältää muun muassa tiedostoa muokanneen ohjelman nimen ja versionumeron sekä muokkauksen päivämäärän.

Jokaisella tiedostoformaattilla on oma tapansa määritellä sisältönä olevan tiedon rakennetta. Yllä kuvattu JPEG-tiedoston rakenne esittää yleisen tiedostojen perusrakenteen, jossa tiedoston alkua kuvataan ylätunnisteella ja loppua alatunnisteella. Näiden tunnisteiden välissä on tiedoston varsinainen tietosisältö, jonka tiedoston avaava ohjelma näyttää käyttäjälle esimerkiksi kuvana tai tekstinä.

Seuraavaksi tutustutaan ”MOF\_logo\_nettiin.jpg”-tiedoston sisältämään metatietoon ExifTool-nimisen ohjelman avulla. Ohjelma toimii tutkimalla tiedoston rakenteisiin tallennettua meta-

tieto. Merkintöjen perusteella ohjelma tulostaa näytölle yksinkertaisen näkymän tiedoston sisältämästä metatiedosta. (ExifTool 2012).



Kuva 2. JPEG-tiedoston metatietoa ExifToolin kautta katsottuna.

Kuvasta 2 nähdään, että ”MOF\_logo\_nettiin.jpg”-tiedosto sisältää runsaasti siihen tallennettua metatietoa. Huomattavaa on muun muassa alkuperäisen tiedoston luontiajankohta sekä siihen käytetyn ohjelman nimi ja versionumero. Tiedostoon on tallennettu myös sitä muokanneen käyttäjän nimi, joka on tummennettu pois kuvasta.

#### 4 Avoimiin lähteisiin perustuvan tiedustelun hyödyntäminen käytännössä

Seuraavaksi tutustutaan avoimiin lähteisiin perustuvan tiedustelun toteuttamiseen käytännössä. Prosessissa hyödynnetään aikaisemmin työssä esiteltyjä NATO:n ja YK:n tiedusteluprosessin malleja, joiden vaiheita käytetään työvaiheiden jaksottamisessa ja toiminnan ohjaamisen suunnittelussa. Koska opinnäytetyö pyrkii painottamaan erityisesti tietoturvallisuuteen liittyviä uhkia, keskitytään kuvauksessa valittujen yritysten käytössä olevien tietojärjestelmiin liittyviin tietoihin. Tarkoituksena on esittää tilannekuvaus jossa simuloidaan organisaatiota vastaan hyökkäävän tahon tiedonkeräykseen ja sopivan kohteen löytämiseen tähtääviä toimintatapoja. Tavoitteena on löytää viidestä esimerkkiin valitusta yrityksestä mahdollisimman paljon sellaista tietoa, jota voidaan käyttää hyväksi mahdollisessa kohdistetussa hyökkäyksessä kyseistä organisaatiota vastaan. Kiinnostuksen kohteena ovat erityisesti yrityksen käyttämät käyttöjärjestelmät, ohjelmistot, ohjelmistoversiot sekä yrityksen työntekijöihin liittyvät tiedot.

#### 4.1 Käytetyt tietolähteet ja tiedon kerääminen

Työn kannalta keskeisimmäksi tietolähteeksi on valittu Yritys- ja yhteisötietojärjestelmän verkkosivusto. Kyseinen sivusto on Patentti ja rekisterihallituksen sekä Verohallinnon ylläpitämä palvelu, joka sisältää perustietoa edellä mainittujen viranomaisten ylläpitämiin rekistereihin merkityistä yksiköistä, kuten yrityksistä ja yhteisöistä. Palvelu sisältää tietoa kauppa- ja säätiörekisteristä, ennakkoperintärekisteristä, työnantajarekisteristä, arvonlisäverovelvollisten rekisteristä, vakuutusmaksuverovelvollisten rekisteristä sekä verohallinnon asiakasrekisteristä. (Mikä on YTJ 2012.) Työn kannalta merkittävin rekisteri on kaupparekisteri, joka julkisena rekisterinä sisältää ajankohtaista tietoa rekisteriin merkityistä yrityksistä. Yrityksillä on tietojen, sekä niissä tapahtuvien muutosten ilmoitusvelvollisuus kaupparekisteriin. (Kaupparekisterin esittely 2012.) Koska kyseessä on viranomaisten ylläpitämä rekisteri, johon yritykset ovat velvoitettuja ilmoittamaan tietonsa sekä niissä tapahtuvat muutokset, voidaan rekisteristä saatavia tietoja pitää lähtökohtaisesti luotettavina.

Toiseksi merkittäväksi lähteeksi on valittu Yritys- ja yhteisötietojärjestelmään ilmoitettu kunkin yrityksen virallinen verkkosivusto. Yritykset ja yhteisöt julkaisevat verkkosivustollaan tietoa toiminnastaan, sekä mahdollisesti myös erilaisia julkisiksi tarkoitettuja dokumentteja. Työn kannalta kiinnostavia ovat erityisesti julkaistut dokumentit, sillä nämä tyypillisesti sisältävät runsaasti metatietoa. Julkaistun tiedon oikeellisuutta valvotaan vain yrityksen itsensä toimesta joten sille ei voida antaa vastaavaa painoarvoa kuin viranomaisten ylläpitämille rekistereille. Toisaalta yrityksen intresseissä ei ole toimintaansa liittyvän virheellisen tiedon julkaiseminen joten verkkosivuilla oleva ja sieltä ladattava tieto on pääpiirteittäin luotettavaa.

Kolmantena keskeisenä lähteenä käytetään Google-hakukonetta. Pääasialliset syyt tälle ovat Googlen tehokas ja kattava verkkosivujen indeksointi sekä laaja tuki tarkennetuille hakuparametreille. Eri hakukoneita on valtava määrä eikä niiden kaikkien esittely tai läpikäynti ole työn kannalta tarkoituksenmukaista. Tavoitteena ei ole löytää kaikkea tietoa haun kohteena olevasta henkilöstä tai organisaatiosta vaan pikemminkin löytää kaikki tähän liittyvä keskeinen tieto, jota voidaan hyödyntää yritystä tai sen työntekijää vastaan kohdennetussa hyökkäyksessä yrityksen tietojärjestelmiä vastaan. Hakupalvelut indeksoivat verkkosivuja varmistamatta niissä olevan tiedon luotettavuutta joten löydetyn tiedon todenmukaisuus pyritään varmistamaan muista luotettavimmista lähteistä. Työssä hakupalvelut toimivatkin enemmän tiedonhaun ja työn suuntaa ohjaavina tekijöinä.

Viimeisenä yleistason lähteenä käytetään sosiaalista mediaa eli erilaisia henkilöiden verkostoitumiseen käytettäviä verkkopalveluita. Vaikka tunnetuimpia tällaisista palveluista on Facebook, työn luonteen takia keskitytään pääasiassa työelämään tiukemmin liittyviin palveluihin.



Tunnetuin näistä on LinkedIn-palvelu jossa on tällä hetkellä noin 200 miljoonaa jäsentä. Palvelun tarkoituksena on auttaa työntekijöitä verkostoitumaan toisten työntekijöiden sekä yritysten kanssa, ja palveluun ilmoitetaan myös avoimia työpaikkoja. (LinkedIn: About Us 2013.)

Ensimmäisessä vaiheessa valitut yritykset etsitään käyttämällä YTJ-tietopalvelua. Tarkoituksena on varmistaa että myöhempien vaiheiden hakutermeissä käytetään yrityksen virallista verkkosivustoa, jolle tallennettua tietoa voidaan pitää kyseisen yrityksen näkökulmasta luotettavana. Koska YTJ-tietopalvelu on viranomaisten ylläpitämä tietojärjestelmä, voidaan siellä olevaa tietoa pitää lähtökohtaisesti luotettavana. Tämän jälkeen vierailtiin yrityksen kotisivulla josta tallennettiin kaikki Google-hakupalvelun indeksoimat doc/docx-, pdf-, xls/xlsx- sekä ppt/pptx-dokumenttitiedostot. Kotisivuilta varmistettiin myös yrityksen työntekijöiden käyttämä sähköpostiosoitteen yleisrakenne, joka oli kaikissa tapauksissa [etunimi.sukunimi@yritys.fi](mailto:etunimi.sukunimi@yritys.fi) tai [etunimi.sukunimi@yritys.com](mailto:etunimi.sukunimi@yritys.com). Kun sähköpostiosoitteen muoto tunnetaan voidaan sähköpostia lähettää yrityksen jokaiselle työntekijälle, mikäli näiden nimi on tiedossa.

Seuraavassa vaiheessa Google-hakupalvelua hyödynnettiin seulomaan LinkedIn-verkkopalvelun sisältöä, ja etsimään sieltä valitun yrityksen palveluksessa olevat työntekijät. Tähän hakutapaan liittyy kuitenkin muutamia heikkouksia. Ensimmäisenä ongelmana on se, ettei Google välttämättä ole indeksoinut jokaista LinkedIn-palvelun profiilisivua, jonka syynä voi olla muun muassa se ettei profiilisivu ole julkinen. Toisena ongelmana on se, että hakutermeillä löydetään henkilöt, jotka työskentelevät tällä hetkellä kyseiselle organisaatiolle sekä henkilöt, jotka ovat jossain vaiheessa työskennelleet kyseiselle organisaatiolle, mutta jotka tällä hetkellä saattavat olla jonkun toisen organisaation palveluksessa. Viimeisensä ongelmana on se, että LinkedIn-palvelun sisältö on käyttäjän itsensä laatimaa eikä sitä tarkasteta, joten käyttäjä voi virheellisesti väittää työskentelevänsä yrityksessä. Toisaalta palvelun luonteen takia voidaan olettaa, että käyttäjien sinne syöttämät tiedot pitävät ainakin jossain määrin paikkansa. Koska palvelun luotettavuuteen liittyy useita ongelmia, siitä kerättäviä tietoja voidaan pitää pikemminkin vain suuntaa antavana tietona, joka pyritään varmistamaan muista luotettavimmista lähteistä.

**site:fi.linkedin.com "at yrityksen nimi" -inurl:jobs -inurl:dir -inurl:title**

Taulukko 5. Google-hakutermi LinkedIn-palvelun sisällölle.

Taulukossa 5 on hakutermi, jolla tiedot kerättiin LinkedIn-palvelusta. Käytetty hakutermi sisältää erityisiä hakuoperaattoreita, jolla voidaan vaikuttaa siihen miten ja mistä Google-hakupalvelu hakee tietoa. Ensimmäinen osa "site:fi.linkedin.com" rajaa haun koskemaan vain kyseistä domain-nimeä, joka koskee pääasiassa Suomesta rekisteröityjä LinkedIn-tilejä. Seuraava osuus "'at yrityksen nimi'" rajaa hakua etsien LinkedIn-palvelusta työntekijöitä jotka

ovat ilmoittaneet työskentelevänsä kyseiselle yritykselle. Tyypillinen esitystapa noudattaa kaavaa ”titteli at yrityksen nimi”, joten hakutermeillä saadaan rajattua käyttäjät jotka ovat rekisteröineet tilinsä Suomesta ja joiden työpaikaksi on merkitty kohteena oleva yritys. Kaksi viimeistä hakuoperaattoria poistaa hakutuloksista ne LinkedIn-palvelun URL-osoitteet, joissa esiintyy sanat ”jobs”, ”title” tai ”dir”. Edellä mainitut hakutermit poistetaan, koska ne eivät viittaa yksittäiseen käyttäjään.

#### 4.2 Dokumenttitiedostojen etsimiseen käytetyt hakutermit

Seuraavilla hakutermeillä etsitään Google-hakukoneen indeksoimat dokumenttitiedostot, jotka on saatavilla yrityksen verkkosivustolta, ja jotka Google-hakupalvelu on indeksoinut.

**site:yrityksendomain filetype:pdf**

Taulukko 6. Google-hakutermin pdf-tiedostojen löytämiseksi yrityksen verkkosivustolta.

Taulukossa 6 on hakutermin, joka rajaa hakutulokset domain-nimeä käyttävälle palvelimelle tallennettuihin pdf-tiedostoihin. PDF-tiedostotyyppi on yhtenäinen sekä avoin tiedostomuoto, ja sitä käytetään muun muassa kuva- teksti- tai taulukkolaskentatiedostojen tallentamiseen.

**site:yrityksendomain filetype:doc OR filetype:docx**

Taulukko 7. Google-hakutermin doc- ja docx-tiedostojen löytämiseksi yrityksen verkkosivustolta.

Taulukossa 7 on hakutermin, jolla hakutulokset rajataan koskemaan yrityksen domain-nimeä käyttävälle palvelimelle tallennettuihin Microsoft Office- ja Word-ohjelmiston eri tiedostomuotoja. Etsityt tiedostotyytit käyttävät joko doc- tai docx-tiedostopäätettä.

**site:yrityksendomain filetype:xls OR filetype:xlsx**

Taulukko 8. Google-hakutermin xls- ja xlsx-tiedostojen löytämiseksi yrityksen verkkosivustolta.

Taulukossa 8 on hakutermin, jolla haetaan Microsoft Office-ohjelmiston Excel- taulukkolaskentaohjelmalla luotavia tiedostoja. Nämä taulukkolaskentatiedosto käyttävät joko xls- tai xlsx-tiedostopäätettä.

**site:yrityksendomain filetype:ppt OR filetype:pptx**

Taulukko 9. Google-hakutermin ppt- tai pptx-tiedostojen löytämiseksi yrityksen verkkosivustolta.

Taulukossa 9 on hakutermi jolla haetaan Microsoft Office-ohjelmiston Powerpoint-ohjelmalla luotavia tiedostoja. Nämä tiedostot käyttävät ppt- tai pptx-tiedostopäätteitä.

#### 4.3 Metatiedon keräämiseen käytetyt valinnat

Taulukosta 10 nähdään exiftool-ohjelmassa käytetyt valinnat, joilla dokumenteista kerättiin tiedostoihin tallennettu metatieto. Jatkokäsittelyä varten kerätyt tiedot ovat tiedoston nimi ("-filename"-lippu), tiedoston luontipäivämäärä ("-createdate"-lippu), tiedoston sisällön viimeisin muokkausajankohta ("-modifydate"-lippu), tiedoston luoneen tai sitä käsitelleen käyttäjän tiedot ("-currentuser"-, "-author"-, "-creator"-, "-username"-, ja "-lastmodifiedby"-liput). Lippu "-title" kerää dokumentin otsikon ja hakemistopolun. Tiedoston luomiseen käytettyjen ohjelmien tiedot kerättiin "-producer"- ja "-creatortool"-lipuilla. Viimeisenä määritelty "-csv"-lippu määrittää kerätyn tiedon tallennusmuodon.

```
exiftool -filename -createdate -modifydate -title -currentuser -username -author -creator -producer -creatortool -csv
```

Taulukko 10. Exiftool-ohjelmassa käytetyt valinnat metatietojen tallentamiseen kerätyistä tiedostoista.

```
SourceFile,FileName,CreateDate,ModifyDate,Title,Author,Creator,Producer,CreatorTool
```

Taulukko 11. Exiftool-ohjelman luoman CSV-tiedoston rakenne.

Taulukossa 11 on exiftool-ohjelman valintojen perusteella luodon csv-tiedoston rakenne. Tiedoston metatieto kuvataan yksittäisellä rivillä, jossa tallennettu tieto on erotettu toisistaan pilkulla.

#### 4.4 Kerätyt tulokset

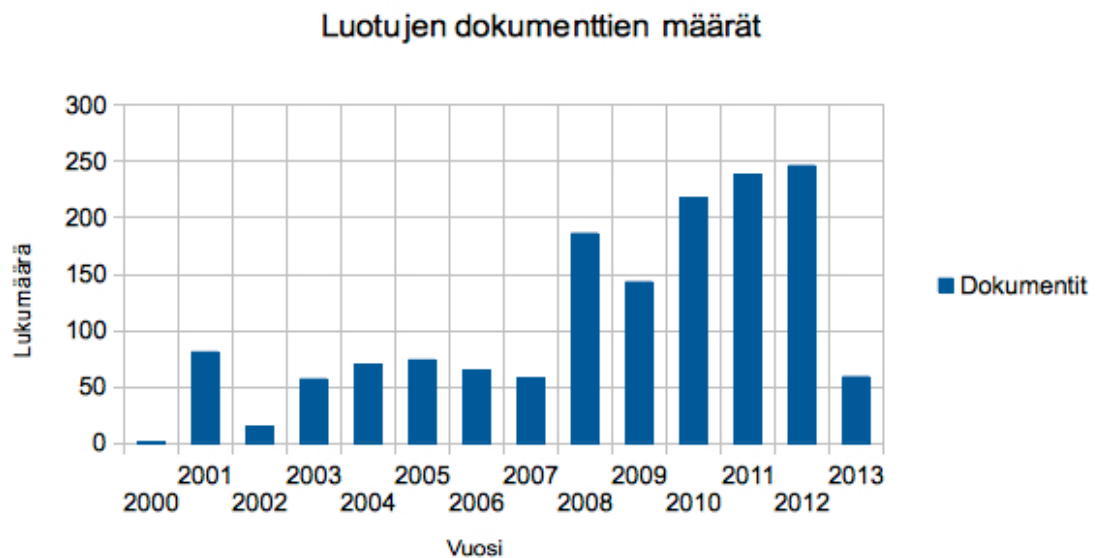
Seuraavaksi käydään läpi aikaisemmassa kohdassa käytettyjen hakutermien tuloksia. Tulokset on ryhmitelty matriisiin siten, että eri tiedostotyyppit on jaettu pystysuuntaisesti, ja niiden määrä yrityksen julkisella palvelimella voidaan havaita vaakasuorista sarakkeista.

### Kerätyt tiedostomäärät ja tiedostotyytit

	PDF	DOC/DOCX	XLS/XLXS	PPT/PPTX
Yritys K	84	0	0	1
Yritys T	318	0	0	2
Yritys V	444	2	0	0
Yritys D	345	5	13	0
Yritys S	351	1	0	1
<b>Yhteensä:</b>	<b>1542</b>	<b>8</b>	<b>13</b>	<b>4</b>

Taulukko 12. Löydettyjen tiedostojen lukumäärät yrityskohtaisesti lajiteltuna.

Taulukkoon 12 on merkitty työtä varten kerättyjen tiedostojen tiedostotyytit sekä -määrät. Kerätyistä tiedostoista 1542 oli pdf-tiedostoja, dokumenttitiedostoja oli 8, talukkolaskenta-tiedostoja 13 ja diaesitystiedostoja oli 4. Työtä varten kerättyjä tiedostoja oli yhteensä 1567. Kerätyistä tiedostosta löydettiin yhteensä 434 uniikkia käyttäjätunnusta tai käyttäjänimeä, 84 tiedoston luomiseen käytettyä ohjelmaa tai näiden versionumeroa, 16 organisaatiotunnusta sekä 50 sekalaiseksi luokiteltua tunnistetta.



Kuvio 6. Kerättyjen dokumenttien julkaisuvuodet ja -määrät metatiedon perusteella.

Kuvioon 6 on merkitty kerättyjen dokumenttien määrät, jotka on jaettu julkaisuvuosien perusteella. Kerättyjen tiedostojen luomisvuodet olivat vuosilta 2000-2013.

## Yrityskohtaiset käyttäjämäärät LinkedIn-palvelussa

LinkedIn	
Yritys K	549
Yritys T	305
Yritys V	168
Yritys D	543
Yritys S	416
<b>Yhteensä:</b>	<b>1981</b>

Taulukko 13. Työhön valittujen yritysten työntekijämäärät LinkedIn-palvelussa.

Taulukossa 13 on tulokset jotka on saatu käyttämällä taulukko 5 näkyvää hakutermiä Google-hakupalvelussa. Löydettyjä käyttäjätilejä oli yhteensä 1981.

## 5 Johtopäätökset

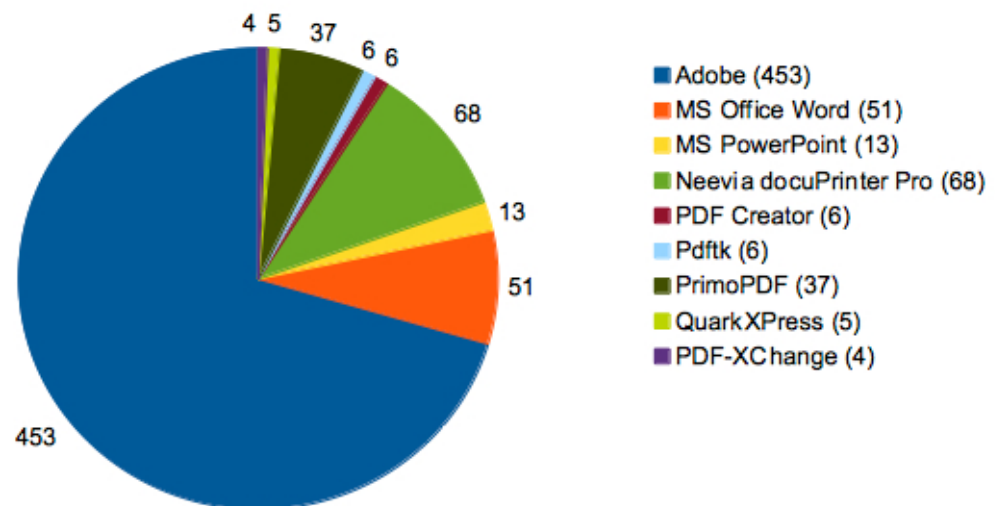
Työn lähdeaineistoksi kerätyssä lähdemateriaalissa oli vain muutamia yksittäisiä tiedostoja, joista metatiedot oli selkeästi siivottu joko osittain tai kokonaan pois. Nämä dokumentit olivat tyypiltään usein yritysten julkaisemia virallisia raportteja, ja ne koskivat usein yrityksen senhetkistä taloudellista tilannetta. Näiden kohdalla dokumenttien julkaisutietoihin oli lisätty yrityksen virallinen nimi, mutta muun muassa dokumentin muokkaushistoria oli kuitenkin jätetty pyyhkimättä. Tämän voi tulkita tarkoittavan sitä, että yritykset eivät järjestelmällisesti pyyhi metatietoja edes julkiseksi tarkoittamistaan dokumenteista. Pikemminkin kyseessä vaikuttaisi olevan julkiseksi tarkoitettujen tiedotteiden ja raporttien viimeistelyvaiheessa liitetyt yritystä koskevat tunnistet.

Kerätyistä tiedostoista valtaosa oli pdf-tiedostoja. Tämä ei toisaalta ole yllätys, sillä ne ovat vakiinnuttaneet paikkansa organisaation julkaiseman materiaalin tiedostoformaattina. Toiseksi eniten löydettyjä tiedostoja olivat taulukkolaskentatiedostot, joita löydettiin 13 kappaletta. Tulosten kannalta huomattavana seikkana on kuitenkin se, että kaikki taulukkolaskentatiedostot löydettiin saman organisaation palvelimelta. Asiakirjatiedostoja eli doc- ja docx-tiedostoja löydettiin yhteensä 8. Powerpoint-ohjelmistolla luotavia diaesitystiedostoja löydettiin yhteensä 4. Muiden tiedostotyyppien vähäisen määrän takia, kerätyt tulokset perustuvat lähes täysin pdf-tiedostojen metatiedon analysointiin. Tiedostojen luontivuodet olivat 2000-2013, joskin merkittävä määrä tiedostoista oli luotu vuoden 2007 jälkeen.

Metatiedon perusteella voidaan tunnistaa 725 tiedoston luoneen tietokoneen käyttöjärjestelmä. Näistä 659 käyttävät Windows-käyttöjärjestelmää, joista 10 voidaan tunnistaa käyttävän Windows XP-käyttöjärjestelmää ja 7 Windows XP Professional Edition Service Pack 3 käyttä-

viksi järjestelmiksi. Macintosh- tai OSX-käyttöjärjestelmää käytetään metatiedon mukaan 66 tietokoneessa. Lukujen yhteydessä on kuitenkin syytä todeta niiden edustavan parhaimmassakin tapauksessa vain hyvin varovaista arviota käyttöjärjestelmien todellisesta määrästä. Tämä johtuu siitä, ettei kerätyn tiedon pohjalta voida osoittaa, onko yrityksen julkaisemat dokumentit luotu yhden tai useamman käyttäjän tai tietokoneen toimesta. Tietoja voidaan hyödyntää kuitenkin yritysten henkilökunnan käytössä olevien eri käyttöjärjestelmätyyppien tutkimiseen. Tämän perusteella yrityksen D luomista dokumenteista 65 tapauksessa metatiedoista löytyy merkintä Windows-käyttöjärjestelmästä. Yrityksen K kohdalla luku on 19, Yrityksen S kohdalla 163, yrityksen T kohdalla 140 ja Yrityksen V kohdalla 272.

**Dokumenttitiedostojen luontiin käytettyjen ohjelmien lukumäärät**



Kuvio 7. Dokumenttitiedostojen luontiin käytetyt ohjelmat metatiedon perusteella.

Kuvasta 9 nähdään dokumenttitiedostojen luomiseen käytettyjä ohjelmistoja. Adobe'n eri tuotteita on käytetty merkittävässä osassa kerättyjä ja analysoituja tiedostoja. Tätä osaltaan voi selittää se, että Adobe on pdf-tallennusformaatin kehittäjä. Toiseksi yleisin oli Neevia docuPrinter Pro-ohjelmisto 68 tiedostolla ja kolmanneksi yleisimpänä ohjelmistona oli Microsoftin Office Word ja PowerPoint yhteensä 64 tiedostolla. Kuvasta 9 on jätetty pois ohjelmaversiot, joita löydettiin analysoidusta aineistosta vain yksi kappale.

Tiedostojen metatiedot sisälsivät runsaasti käyttäjätunnuksia sekä tiedoston luoneiden henkilöiden etu- ja sukunimiä. Käyttäjätunnuksia löydettiin analysoidusta aineistosta 162 kappaletta. Tyypillisenä käyttäjätunnusmuotona oli joko pelkkä etunimi, etunimen ensimmäisestä kirjaimesta sekä sukunimestä muodostettu tunnus tai sukunimi. Yritys D:llä käytettiin lähes yksinomaaisesti yrityksen ensimmäisen kirjaimen ja numerosarjan muodostamia tunnuksia. Numerosarjat muistuttivat työntekijälle annettavaa henkilön numeroa, joskaan tästä ei voida olla varmoja. Aineistosta löydettiin yhteensä 124 henkilön etunimi ja sukunimi. Käyttäjätunnuksiin

liittyvänä mielenkiintoisena havaintona oli se, että aineiston perusteella 14 tiedostoa oli luotu työasemalla, jossa käytettiin pääkäyttäjätunnuksia. Tällaisia tiedostoja löydettiin yrityksiltä V, T ja S. Merkittävä osa eli 8 pääkäyttäjätunnuksilla luotua tiedostoa oli kerätty yritykseltä S. Yrityksillä V ja T pääkäyttäjätunnuksilla luotuja tiedostoja oli kummallakin 3 kappaletta. Sekalaisiksi määriteltäviä käyttäjätunnisteita oli 50 kappaletta, ja niiden joukossa oli muutamia mielenkiintoisia tiettyä paikkaa tai järjestelmää määritteleviä tunnuksia. Näitä olivat muun muassa ”Lohko 3” ja ”Lohko 4”, ”THhallintokk”, ”kopiokone 2krs.”, ”KMBT\_420”, ”KMBT\_600”, ”KMBT\_C353”, ”KMBT\_C451” sekä ”KMBT\_C650”. KMBT-alkuiset tunnukset liittyvät Konica Minoltaan monitoimitulostinjärjestelmiin ja niiden eri malleihin, joka viittaa siihen, että yrityksessä käytetään kyseisen valmistajan tuotteita.

Tiedoston metatiedon otsikkokentän sisältö oli pääosin odotettua, sisältäen tiedoston varsinaisen otsikon tai alaotsikon. Analysoidussa materiaalissa oli kuitenkin 7 otsikkokenttää, joihin oli tallennettu tiedoston kokonainen tiedostopolku ja tiedostonimi. Nämä olivat tyypiltään ”Y:\hakemisto1\hakemisto2\tiedosto.pdf”-muotoisia. Kokonaisen tiedostopolun perusteella voidaan päätellä että organisaatiossa käytetään keskitettyjä tiedostopalvelimia julkaistavan materiaalin tallentamiseen. Tässä ei varsinaisesti ole mitään erikoista, sillä monissa yrityksissä käytetään keskitettyjä tiedostojen tallennusratkaisuja. Toisaalta tallennetun tiedostopolun perusteella voidaan myös päätellä se, että tiedoston metatietoihin tallennetulla käyttäjällä on todennäköisesti myös käyttöoikeus kyseiseen tiedostopalvelimeen. Tätä kautta ulkopuolisen henkilön on mahdollista selvittää tietojärjestelmien välisiä luottamussuhteita käyttäjien käyttöoikeuksien ja järjestelmien muodossa. Heikkoutena on toisaalta se, että tiedostopolusta on mahdotonta sanoa onko todellisuudessa kyseessä verkossa oleva tiedostopalvelin vai käytössä olevan tietokoneen kovalevyn osio. Tämä johtuu siitä että käyttöjärjestelmä pyrkii tekemään eri tallennuslaitteista läpinäkyviä niitä hyödyntäville ohjelmistoille, eikä näiden eroa ole siksi mahdollista havaita pelkän metatiedon perusteella.

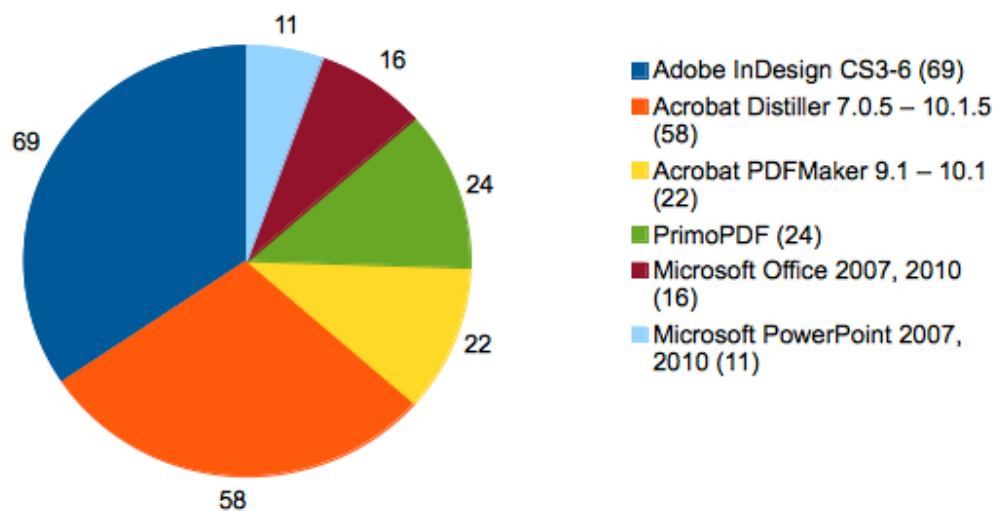
Linkedin-palvelusta kerätyt tiedot ovat korkeintaan suuntaa-antavia, sillä tuloksissa on mukana organisaatiolle aikaisemmin työskennelleitä henkilöitä. Toisena huomattavana seikkana on, että mukana ovat vain Google-hakupalvelun indeksoimat ja kaikille julkiset LinkedIn-profiilit. Piilotetut tai rajatut profiilit jäävät haun ulkopuolelle, samoin kuin henkilöt jotka eivät ole tehneet LinkedIn-profiilia. Huomattava on myös se, että koska käyttäjä itse syöttää palvelun tiedot eikä niiden todenperäisyyttä tarkasteta, saattaa profiileihin olla merkitty virheellistä tietoa joko tahattomasti tai tahallisesti. Tietoa käytetään luotettavuuteen liittyvien syiden takia vain suuntaa antavana tietona yrityksen työntekijöiden määrästä eri yhteisöpalveluissa.

Aikaisempien kohtien perusteella voidaan todeta, että ylimääräistä tietoa on saatavilla runsaasti, joskin sen aiheuttaman riskin merkitys voidaan kyseenalaistaa. Seuraavaksi käydään läpi metatiedon sekä avoimiin lähteisiin perustuvan tiedustelun hyödyntämistä tietojärjestel-

mään hyökkäävän tahon näkökulmasta. Tarkoituksena on esitellä skenaario, jota voidaan pitää mahdollisena ja joka voidaan toteuttaa hyvin yksinkertaisesti. Tarkastelun kohteeksi on valittu yritys S, jonka päätoimialaan kuuluu tietoturvallisuuteen liittyvät laitteistot ja ohjelmistot.

Alustavan tarkastelun perusteella, yritys S:tä on saatavilla runsaasti tietoa. Google-hakukoneen kautta löydetään 351 pdf-tiedostoa sekä muutamia muita dokumenttiedostoja tallennettuna yrityksen palvelimelle. Yrityksen työntekijät vaikuttavat käyttävän alustavien hakujen perusteella runsaasti LinkedIn-verkostoitumispalvelua. Käyttäjiä, jotka työskentelevät tai ovat aikaisemmin työskennelleet yritys S:n palveluksessa, löydetään 416 kappaletta. Koska dokumentteja on paljon, rajataan tutkittavat dokumentit niihin joihin luontivuodeksi on merkitty 2010, 2011, 2012 tai 2013. Rajauksen tarkoituksena on käsitellä mahdollisimman tuoreita tiedostoja, jotta niiden metatietoon tallennetut ohjelmaversiot tai käyttäjätunnukset olisivat edelleen yrityksen käytössä. Rajauksen jälkeen käsiteltäviä tiedostoja on 200.

Tiedostojen luontiin käytettyjen ohjelmistojen jakautuminen



Kuvio 8. Tiedostojen luontiin käytettyjen ohjelmistojen jakautuminen.

Kuviosta 8 voidaan nähdä käytettyjen ohjelmistojen jakautuminen niillä luotujen dokumenttien perusteella. Valtaosa tiedostoista on luotu käyttämällä joko Adobe InDesign- tai Acrobat Distiller-ohjelmiston eri versiota. Merkintä Applen Macintosh- tai OSX-käyttöjärjestelmästä löytyy 32 dokumenttiedoston metatiedoista. Microsoftin Windows-käyttöjärjestelmä on puolestaan merkitty 58 tiedoston metatietoihin. Tietojen perusteella voidaan päätellä, että yrityksessä käytetään sekä Windows- että OSX-käyttöjärjestelmiä, ja merkittävä osa julkaistuista dokumenteista luodaan joko Adoben InDesign- tai Acrobat Distiller-ohjelmistolla.



Kun yritys S:n käytössä olevat käyttöjärjestelmät sekä dokumenttien käsittelyyn tarkoitetut ohjelmistot on käsitelty, siirrytään tutkimaan metatietoihin tallennettuja käyttäjätunnuksia ja käyttäjien nimiä. Tiedostoista löydetään yhteensä 16 käyttäjätunnusta tai -nimeä, joista 5 on luotu Windows-käyttöjärjestelmän pääkäyttäjätunnuksilla. Pääkäyttäjätunnuksien käyttäminen normaalikäytössä on riski, sillä niiden väärinkäyttö mahdollistaa haittaohjelmien asen- tamisen sekä palomuurin tai virustorjuntaohjelmiston sammuttamisen (Minimising administrative privileges explained 2012). Yksi pääkäyttäjätunnuksilla luoduista tiedostoista sisältää henkilön täyden nimen tiedoston nimessä. LinkedIn-palvelun haku henkilön sekä yrityksen ni- mellä paljastaa profiilin, johon ei ole merkitty kuitenkaan nykyistä työpaikkaa. Profiilissa on kuitenkin profiilikuva. Kun henkilön sekä yrityksen nimellä suoritetaan Google-haku, yhtenä hakutuloksena on yrityksen virallinen kotisivu josta löytyy ylimpien toimihenkilöiden yleis- luontoiset profiilit ja profiilikuvat. Yksi näistä profiileista on yrityksen S nykyisen talousjohta- jan profiili, johon tallennettu kuva täsmää löydetyn LinkedIn-profiilin kuvan kanssa. Henkilön nimi on sama niin edellä mainituissa profiileissa kuin myös analyysin kohteena olevan tiedos- ton metatiedoissa. Yrityksen S-verkkosivuilla nähdään nykyisen talousjohtajan sähköposti, joka täsmää löydetyn nimen kanssa, ja joka on muotoa [etunimi.sukunimi@yritys.com](mailto:etunimi.sukunimi@yritys.com).

”Spear-phishing”-hyökkäyksellä tarkoitetaan sähköpostin kautta tehtyä hyökkäystä, jossa koh- teena olevasta tahosta kerättyjä tietoja hyödynnetään mahdollisimman yksilöidyn ja henkilö- kohtaisen sähköpostin lähettämiseen. Tällainen sähköposti sisältää tyypillisesti haittaohjel- man tai linkin haittaohjelmaan, joka vastaanottajan on tarkoitus ladata ja avata. Tiedoston avaaminen johtaa haittaohjelman suorittamiseen. Hyökkäystä käytetään runsaasti, sillä se on osoittautunut tehokkaaksi tavaksi murtautua yrityksen tietoverkkoihin. Trend Micron jul- kaiseman tutkimuksen mukaan vuoden 2012 helmikuun ja syyskuun välillä 91% tutkituista koh- distetuista hyökkäyksistä yrityksiä vastaan, tehtiin hyödyntämällä ”spear-phishing”- hyökkäystä. (Spear-Phishing Email 2012.)

Pelkän tiedoston metatiedon, sekä muutaman Google-hakupalvelussa suoritettun haun perus- teella tiedetään pääkäyttäjätunnuksilla luotu tiedosto, henkilö joka on todennäköisesti luonut kyseisen tiedoston, tiedoston luomiseen käytetty ohjelmisto, henkilön nimi, sähköpostiosoite sekä asema yrityksessä. Näitä tietoja voidaan hyödyntää ”spear-phishing”-hyökkäyksen suun- nittelussa ja kohdistamisessa. Tieto siitä, että pääkäyttäjätunnuksia käytetään tarpeettomas- ti, auttavat ensisijaisen kohteen löytämisessä. Koska pääkäyttäjätunnuksilla voidaan muuttaa lähes kaikkia käyttöjärjestelmän asetuksia, tällaisen käyttäjätilin kaappaaminen toimii hyvä- nä jalansijana yrityksen tietoverkossa. Käyttöjärjestelmän sekä tiedoston luomiseen käytetyn ohjelmiston nimen tunteminen auttaa sopivan haavoittuvuuden ja sitä hyödyntävän haittaoh- jelman valitsemisessa. Mikäli käytetyssä ohjelmistossa on tuntemattomia haavoittuvuuksia tai haavoittuvuuksia, joihin liittyyvää päivitystä ei ole vielä asennettu, tätä voidaan käyttää hy- väksi valitsemalla näiden mukaan käytettävä haittaohjelma. Henkilön asema ja tehtävät yri-

tyksessä auttavat sopivan ”spear-phishing”-sähköpostin sisällön suunnittelussa, jolla pyritään lisäämään todennäköisyyttä sille, että vastaanottaja avaa haittaohjelman sisältämän liitetiedoston. Henkilön nimen kautta voidaan selvittää tälle kuuluva sähköpostiosoite, jotta valmisteltu ”spear-phishing”-hyökkäys voidaan lähettää kohteena olevalle henkilölle.

Pyrin edellä esittämään mahdollisia seurauksia, joita voi aiheutua mikäli julkaistuista dokumenteista ei siivota niihin tallennettuja metatietoja. On hankala ymmärtää sitä, minkä takia yritys saattaisi tarkoituksella julkiseksi tietoja jotka liittyvät sen työntekijöiden käyttäjätunnuksiin, yrityksen käyttämiin ohjelmistoihin ja näiden ohjelmistojen versionumeroihin. Tällaisen tiedon tahaton levittäminen saattaa altistaa yritykset uhkille, jotka voidaan joko välttää kokonaan tai joiden toteutumisen todennäköisyyttä olisi mahdollista ehkäistä yksinkertaisilla keinoilla. Kyseessä on mielestäni täysin tarpeeton ja tiedoston loppukäyttäjälle lähes näkyvät tietovirta, joka voi aiheuttaa tiedoston luoneelle ja julkiseksi saattaneelle yritykselle vain haitallisia seurauksia.

Yritykset voivat suojautua riskiä vastaan rakentamalla dokumenttien ja tiedotteiden julkaisuprosessit sellaisiksi, että dokumentit saatetaan julkiseksi vain tietyn organisaation tahon toimesta. Jokaisen käyttäjän sijaan, yksittäisen tahon ohjeistaminen ja kouluttaminen metatiedon pyyhkimiseen ennen dokumenttien julkaisemista, on yksinkertaisempaa ja toisaalta yhden julkaisutahon käyttämisellä voidaan varmistua siitä, että metatieto pyyhitään varmasti jokaisesta dokumentista ennen näiden julkaisemista. Toisena mahdollisuutena on hyödyntää eri toimisto-ohjelmistoissa olevaa ominaisuutta, joka pyyhkii tiedoston metatiedon pois automaattisesti.

Jatkon kannalta olisi mielenkiintoista selvittää yritysten dokumenteista löydettävää metatietoa ja sen eri lajeja, käymällä läpi merkittävästi useampia yrityksiä. Tämän avulla olisi mahdollista selvittää tarkemmin eri tyyppisen metatiedon esiintymismäärää. Toisena mielenkiintoisena asiana voisi olla ajallinen vertailu, jossa yritysten dokumentteihin tallennettua metatietoa ja sen laatua verrattaisiin tietoturvastandardeihin sekä niiden julkaisuajankohtiin.

## 6 Oman työn arviointi

Opinnäytetyön keskeisimpinä haasteina olivat saatavilla olevan lähdemateriaalin vähäinen määrä sekä kerättävissä olevaan tietoon liittyvät epävarmuustekijät. Alun perin työssä oli tarkoituksena hyödyntää erityisesti verkostanalyysiä, mutta ongelmaksi kuitenkin muodostui saatavilla olevan tiedon muotoon liittyvät tekijät. LinkedIn-palvelun profiilisivut on tallennettu sellaiseen muotoon että niiden koneellinen käsittely on haasteellista. Tiedon valtavan määrän takia ei ollut mahdollista käsitellä käsin, vaan tilalle olisi pitänyt löytää koneellinen ratkaisu. Opinnäytetyön työsuunnitelma tehtiin hyvissä ajoin, mutta varsinainen työn tekeminen

tapahtui hyvin nopeassa aikataulussa joka vaikutti osaltaan työn analyttiseen sisältöön. Työsuunnitelmasta huolimatta työn aloittaminen oli hankalaa, koska tutkittavaa ongelmaa ei oltu määritelty riittävän suppeasti. Metatieto tuli työn keskeiseksi osaksi vasta sen jälkeen kun havaitsin että pelkkä avoimiin lähteisiin perustuvan tiedustelun tulosten kerääminen olisi muodostunut työn laajuuden kannalta liian suureksi. Tiukasta aikataulusta johtuen työn sisältöön jäi selkeitä puutteita, mutta katson kuitenkin että työn keskeiset teemat ja hypoteesi on pystytty toteamaan riittävän selkeästi.

## Lähteet

- Altheide, C., Carvey, H. & Davidson, R. 2011. Digital Forensics with Open Source Tools. Waltham: Syngress
- Carrier, B. 2005. File System Forensic Analysis. Indiana: Addison Wesley.
- Criminal Intelligence: Manual for Analysts. 2011. United Nations. UNODC. Viitattu 28.7.2012 [http://www.unodc.org/documents/organized-crime/Law-Enforcement/Criminal\\_Intelligence\\_for\\_Analysts.pdf](http://www.unodc.org/documents/organized-crime/Law-Enforcement/Criminal_Intelligence_for_Analysts.pdf)
- Denscombe, M. 2010. The Good Research Guide for small-scale social research projects. 4.painos. Berkshire: Open University Press.
- Duhigg, C. 2012. How Companies Learn Your Secrets. The New York Times. Viitattu 10.6.2012 <http://www.nytimes.com/2012/02/19/magazine/shopping-habits.html?pagewanted=all>
- ExifTool. 2012. ExifTool by Phil Harvey. Viitattu 10.12.2012 <http://www.sno.phy.queensu.ca/~phil/exiftool/>
- Facebook population: Status Update. 2010. The Economist. Viitattu 12.6.2012 <http://www.economist.com/node/16660401>
- Hamilton, E. 1992. JPEG File Interchange Format. 1992. Version 1.2. Viitattu 7.12.2012 <http://www.w3.org/Graphics/JPEG/jfif3.pdf>
- Heuer, Richards J. 1999. Psychology of Intelligence Analysis. Viitattu 15.1.2013 <https://www.cia.gov/library/center-for-the-study-of-intelligence/csi-publications/books-and-monographs/psychology-of-intelligence-analysis/PsychofIntelNew.pdf>
- Heuer, Richards J., Jr. & Pherson, Randolph. 2010. Structured analytic techniques for intelligence analysis. Washington: CQ Press.
- Hex Editor Definition. 2006. The Linux Information Project. Viitattu 10.12.2012 [http://www.linfo.org/hex\\_editor.html](http://www.linfo.org/hex_editor.html)
- Hirsjärvi, S., Remes, P. & Sajavaara, P. 2009. Tutki ja kirjoita. 15 uudistettu painos. Hämeenlinna: Kariston Kirjapaino Oy.
- Intelligence Collection Disciplines (INTs). 2012. The Federal Bureau of Investigation. Viitattu 1.9.2012 <http://www.fbi.gov/about-us/intelligence/disciplines>
- Intelligence Exploitation of the Internet. 2002. NATO. Viitattu 2.6.2012 [www.au.af.mil/au/awc/awcgate/nato/exploit\\_internet.pdf](http://www.au.af.mil/au/awc/awcgate/nato/exploit_internet.pdf)
- Jones, Morgan D. 2009. The Thinker's Toolkit: 14 Powerful Techniques for Problem Solving. Amazon Digital Services: Crown Business. Viitattu 14.1.2013 <http://www.amazon.com/The-Thinkers-Toolkit-Techniques-ebook/dp/B002PYFW4S/>
- JPEG File Interchange Format File Format Summary. 2012. Fileformat.info. Viitattu 10.12.2012 <http://www.fileformat.info/format/jpeg/egff.htm>
- Kaupparekisterin esittely. 2012 Patentti ja rekisterihallitus. Viitattu 14.2.2013 <http://www.prh.fi/fi/kaupparekisteri/yleista.html>
- Linkedin:About Us. 2013. Linkedin. Viitattu 14.2.2013 <http://www.linkedin.com/about-us>

MacMillan, D. 2012. Chasing Facebook's Next Billion Users. Bloomberg Businessweek. Viitattu 1.9.2012 <http://www.businessweek.com/articles/2012-07-25/chasing-facebooks-next-billion-users>

Madig, L. 2012. Girls Around Me App Is a Reminder To Be Aware What You Share. Forbes. Viitattu 4.5.2012 <http://www.forbes.com/sites/larrymagid/2012/04/09/girls-around-me-app-is-a-reminder-to-be-aware-what-you-share/>

Mikä on YTJ. 2012. Yritys- ja Yhteisötietojärjestelmä. Viitattu 14.2.2013  
<http://www.ytj.fi/mika-on-ytj>

Minimising administrative privileges explained. 2012. Australian Government. Department of Defense. Viitattu 5.3.2013  
[http://www.dsd.gov.au/publications/csocprotect/Minimising\\_Admin\\_Privileges.pdf](http://www.dsd.gov.au/publications/csocprotect/Minimising_Admin_Privileges.pdf)

Open Source Intelligence Handbook. 2001. NATO. Viitattu 4.12.2012  
[http://www.oss.net/dynamaster/file\\_archive/030201/ca5fb66734f540fbb4f8f6ef759b258c/NATO%20OSINT%20Handbook%20v1.2%20-%20Jan%202002.pdf](http://www.oss.net/dynamaster/file_archive/030201/ca5fb66734f540fbb4f8f6ef759b258c/NATO%20OSINT%20Handbook%20v1.2%20-%20Jan%202002.pdf)

Spear-Phishing Email: Most Favoured APT Attack Bait. 2012. Trend Micro. Viitattu 5.3.2013  
<http://www.trendmicro.com/cloud-content/us/pdfs/security-intelligence/white-papers/wp-spear-phishing-email-most-favored-apt-attack-bait.pdf>

TCP/IP Overview. 2005. Cisco. Viitattu 10.12.2012  
[http://www.cisco.com/en/US/tech/tk365/technologies\\_white\\_paper09186a008014f8a9.shtml](http://www.cisco.com/en/US/tech/tk365/technologies_white_paper09186a008014f8a9.shtml)

## Kuvat

Kuva 1. JPEG-kuvatiedoston sisältöä heksadesimaalieditorin kautta katsottuna

Kuva 2. JPEG-tiedoston metatietoa ExifToolin kautta katsottuna

## Kuviot

Kuvio 1. Analyysiprosessin vaiheet

Kuvio 2. Esimerkki verkostanalyysistä

Kuvio 3. Tapahtumakaavioanalyysin esimerkki

Kuvio 4. Esimerkki hyödykevirran vuoanalyysistä

Kuvio 5. Puheluanalyysin vuokaavion esimerkki

Kuvio 6. Kerättyjen dokumenttien julkaisuvuodet ja -määrät metatiedon perusteella

Kuvio 7. Dokumenttitiedostojen luontiin käytetyt ohjelmat metatiedon perusteella

Kuvio 8. Tiedostojen luontiin käytettyjen ohjelmistojen jakautuminen

## Taulukot

Taulukko 1. 4x4-järjestelmän lähteen luotettavuuden arviointikriteerit

Taulukko 2. 4x4-järjestelmän tiedon luotettavuuden arviointikriteerit

Taulukko 3. 6x6-järjestelmän lähteen luotettavuuden arviointikriteerit

Taulukko 4. 6x6-järjestelmän tiedon luotettavuuden arviointikriteerit

Taulukko 5. Google-hakutermin LinkedIn-palvelun sisällölle

Taulukko 6. Google-hakutermin pdf-tiedostojen löytämiseksi yrityksen verkkosivustolta

Taulukko 7. Google-hakutermin doc- ja docx-tiedostojen löytämiseksi yrityksen verkkosivustolta

Taulukko 8. Google-hakutermin xls- ja xlsx-tiedostojen löytämiseksi yrityksen verkkosivustolta

Taulukko 9. Google-hakutermin ppt- tai pptx-tiedostojen löytämiseksi yrityksen verkkosivustolta.

Taulukko 10. Exiftool-ohjelmassa käytetyt valinnat metatietojen tallentamiseen kerätyistä tiedostoista

Taulukko 11. Exiftool-ohjelman luoman CSV-tiedoston rakenne

Taulukko 12. Löydettyjen tiedostojen lukumäärät yrityskohtaisesti lajiteltuna

Taulukko 13. Työhön valittujen yritysten työntekijämäärät LinkedIn-palvelussa