

**Suuren datamäärän tallentaminen uusien ohjelmisto- ja
tekoälyversioiden testaustilanteisiin**



Ammattikorkeakoulututkinnon opinnäytetyö
Tieto- ja viestintätekniikka, biotalouden koulutus, Forssa
Syksy 2021
Piia Rautio

Tekijä	Piia Rautio	Vuosi 2021
Työn nimi	Suuren datamäärän tallentaminen uusien ohjelmisto- ja tekoälyversioiden testaustilanteisiin	
Ohjaaja	Anne-Mari Järvenpää	

TIIVISTELMÄ

Opinnäytetyön tavoitteena oli luoda standardijoukko metatietoja urheiluteknologian parissa toimivan Supponor-yrityksen videomateriaaleille. Tärkeimpinä osa-alueina oli metatietojen määrittäminen, metadatatallinnien toteuttaminen sekä alustan valinta näille malleille.

Tarkoituksena oli jäsenellä suuri datamäärä siten, että siitä on hyötyä yrityksen uusien ohjelmisto- ja tekoälyversioiden testauksessa. Opinnäytetyöhön toivottiin myös dokumentointia metadatatallinnien päivittämisen prosessiin, jotta yrityksen tuleva data saataisiin myös jatkossa kategorisoitua mallien mukaisesti.

Työn teoriaosuus käsittelee teknologian kehitystä, suurien datamäärien tuomia haasteita sekä sitä, mihin metatietojen kategorisointia tarvitaan datan hallinnassa.

Tutkimusmenetelminä keskeisessä roolissa olivat henkilöhaastattelut sekä toimeksiantajan teknologiaan ja dataan perehtyminen.

Tuloksena toimeksiantajalle luotiin metatietojen kategorisointi ja ohjeet näiden metatietojen päivittämisen prosessiin. Lisäksi YAML-merkintäkielen avulla toteutettiin valmiit metadatatallinnat ja opinnäytetyön viimeisessä luvussa tuotiin esiin datan automatisointiin liittyviä jatkotoimenpide-ehdotuksia. Kestävän kehityksen näkökulma nostettiin esiin omana lukunaan, sillä teknologian kehitys on tärkeässä roolissa ilmastonmuutoksen hillinnässä.

Avainsanat Metatiedot, data, YAML, teknologia

Sivut 39 sivua ja liitteitä 1 sivu

Author	Piia Rautio	Year 2021
Subject	Storing Large Amounts of Data in the Testing of New Software and Artificial Intelligence Versions	
Supervisor	Anne-Mari Järvenpää	

ABSTRACT

The aim of the thesis was to create a standard set of metadata for the video materials of the commissioner of the thesis the Research and Development Centre Supponor Oy. The most important areas were the definition of preliminary metadata, the functional implementation of metadata models and the choice of a platform for these models. It was also requested that the thesis would be documented with clear instructions for the process of updating metadata models, so that the company's future data could continue to be categorized according to these models. The purpose was to structure the company's video data so that it can be used in the future to test new versions of the software and artificial intelligence.

The theoretical part of the thesis deals with the development of technology, the challenges posed by a large amount of information and the purposes for which metadata classification is usually needed in information management. The research methods included interviews with experts and gathering knowledge about the commissioner's technology and data.

As a result, a functionally created metadata categorization was produced for Supponor, as well as instructions for the process of updating these metadata. The YAML markup language was used to implement the metadata models themselves, and the last chapter of the thesis presents suggestions for further action related to data automation. The perspective of sustainable development was raised as a separate chapter because technological development is helping to curb climate change.

Keywords Metadata, data, YAML, technology

Pages 39 pages and appendices 1 page

Sisällys

1	Johdanto	1
2	Toimeksiantaja ja yrityksen teknologia	2
3	Työn tavoitteet, menetelmät ja yhteenveto	3
3.1	Tavoitteet	3
3.2	Suunnittelu	4
3.3	Tutkimusmenetelmät ja tutkimusongelmat	5
3.3.1	Laadullinen tutkimus ja sisällönanalyysi	8
3.3.2	Teemahaastattelut ja asiantuntijahaastattelut	9
3.3.3	Tutkimusten luotettavuus	9
3.4	Tutkimusten yhteenveto	10
3.4.1	Yrityksen dataan ja teknologiaan perehtyminen	10
3.4.2	Haastattelut	11
4	Metatietojen merkitys datassa ja teknologian kehitys	13
4.1	Datan merkitys yrityksissä	13
4.2	Teknologian kehitys	14
4.3	Metatieto ja metatietojen kategorisointi	15
4.4	YAML ja JSON	17
5	Toiminnallinen toteutus	18
5.1	Metatietojen määrittäminen	18
5.2	YAML-testaus	20
5.3	Metadatatmallit	21
5.4	Ohjeet metadatatmallien päivittämiseen	25
6	Kestävän kehityksen näkökulma	28
7	Työn tulokset ja johtopäätökset	29
8	Pohdinta ja jatkotoimenpiteet	33
	Lähteet	36

Kuvat ja taulukot

Kuva 1. YAML-testaus.....	21
Kuva 2. Jääkiekkopelin metadatatamalli.....	22
Kuva 3. Koripalloperlin metadatatamalli.....	23
Kuva 4. Hevosurheilun metadatatamalli.....	24

Taulukko 1. Teemahaastattelut.....	7
Taulukko 2. Ohjeet joukkuelajien metadatatamalleille.....	26
Taulukko 3. Ohjeet yksilölajien metadatatamalleille.....	27

Liitteet

Liite 1	Aineistonhallintasuunnitelma
---------	------------------------------

1 Johdanto

Eri yrityksissä ympäri maailmaa painitaan suurien datamäärien kanssa. Tallennustilaa nykyään kyllä riittää monilla eri alustoilla datan säilyttämistä varten, mutta ongelmia ilmenee, kun datasta pitäisi löytää tiettyjä kohtia tai hyödyntää olemassa olevaa dataa yrityksen tarpeisiin. Dataa myös kertyy jatkuvasti lisää ja se vain siirretään säilöön muiden sekaan. Yritykset pitävät tallessa siis dataansa, mutta sitä ei juurikaan osata hyödyntää yrityksen käyttöön, koska sellaisenaan sen hyödyntäminen on hankalaa ja aikaa vievää. (Parikka, 2021) On työlästä vain selata isoja datamääriä ja etsiä sieltä jotakin tarpeisiin sopivaa. Toivotun tiedon löytämiseksi on osaksi datan tallennusta otettava käyttöön datan jäsentely. Datan jäsentelyssä datalle luodaan rakenne, kategorisoimalla se harkittuun muotoon. Mitä paremmin yritykset osaavat valita oleellisen ja tärkeän tiedon datastaan, epäoleellisen tiedon sijasta, sitä tehokkaammin se pystyy hyödyntämään oman datansa sisältävää tietoa. (Kleemola, 2020)

Arkinen esimerkki suurista datamääristä ja sen kategorisoinnin merkityksestä, löytyy päivittäin käytettävästä internetistä. Miten ihmeessä sieltä löytyy niin paljon tietoa? Kun haetaan esimerkiksi kuvaa tai videota tietyistä kohteista verkosta, palvelu löytää hetkessä sopivan tiedoston kriteerien mukaisesti. Tämän mahdollistaa metadata ja se, että haettavaan tiedostoon on liitetty erilaisia hakukriteereitä metatietona, tallennuksen yhteydessä. Metatiedoista saatavat hyödyt korostuvat erityisesti silloin, kun dataa on tallessa suuria määriä. Hyödyntämällä metatietoja omassa datassaan, yritykset voivat saada helposti yleiskuvan haettavasta aineistosta ja tämän jälkeen tarkentaa ja suodattaa aineistoa yksityiskohtaisempiin sisältöihin. Oleellista on, että metatiedot on kategorisoitu oikein ja tarkasti, tällöin kaikkea dataa ei useinkaan tarvitse säilyttää. (Naatula, 2018)

Data auttaa paitsi ymmärtämään, myös ennustamaan liiketoiminnan kehitystä. Dataa voi tallentaa moneen eri paikkaan, datan määrittelyä ja kategorisointia voi tehdä monella eri tavalla ja eri liiketoiminnat tarvitsevat erilaista dataa. (Parikka, 2018) Datan tallentamiseen, kategorisointiin ja hallintaan on olemassa toimijoita, jotka tarjoavat pakettiratkaisuja yrityksille ja auttavat kehittämään juuri heidän liiketoiminnalleen sopivan ratkaisun.

Tässä opinnäytetyössä keskityttiin urheiluteknologiassa toimivan yrityksen datan jäsentelyyn. Yritykselle oli kertynyt talteen suuria määriä dataa, jota hyödynnetään uusien ohjelmisto- ja tekoälyversioiden testaustilanteissa. Suurista datamääristä ei sinällään saa kovinkaan paljon irti, joten kyseisiä testauksia varten toivottiin datan kategorisointia ja kokonaisuudessaan toimivaa ratkaisua.

2 Toimeksiantaja ja yrityksen teknologia

Opinnäytetyön toimeksiantajana toimi Supponor Oy. Yritys on urheiluteollisuudessa toimiva maailmanlaajuinen yritys, joka teknologiansa ansiosta tarjoaa ratkaisuja urheilutapahtumien sisäiseen mainontaa. Yrityksen päämaja sijaitsee Lontoossa, mutta opinnäytetyöhön liittyvää teknologiaa kehitetään Suomessa, Espoon tutkimus- ja kehityskeskuksessa. Supponor on urheiluteollisuuden maailmanlaajuinen edelläkävijä pelien sisäiseen virtuaalimainontaan. Yritys tarjoaa kaupallisesti sekä teknisesti ratkaisuja, jotka tarjoavat parempaa ja monipuolisempaa mainostilaa suorissa urheilulähetyksissä. Erilaisissa tapahtumissa fyysiset mainostilat voidaan korvata lokalisoidulla sisällöllä reaaliaikaisesti ja tuotettuja videosyötteitä voidaan räätälöidä paikalliselle yleisölle tuottamaan tuotemerkin aktivointiarvoa sekä luomaan osuvampia ja kiinnostavampia käyttäjäkokemuksia live-tapahtumien aikana.

Yhteistyössä yrityksen kanssa toimivat isot organisaatiot, kuten NHL, Bundesliiga, La Liga ja NBA. Supponor on tällä hetkellä maailman ainut kyseistä teknologiaa käyttävä yritys, näin suuressa mittakaavassa. Yrityksen virtuaaliratkaisut tarjoavat urheilujärjestöille, oikeuksien omistajille, tuotemerkeille ja lähetystoiminnan harjoittajille mahdollisuuden tavoittaa paikalliset yleisöt tehokkaammalla, osuvammalla ja kohdennetulla viestillä. Jokainen urheilu-, tapahtuma- tai lähetystuotanto on erilainen ja suunnattu eri kohdeyleisölle. Yritys tarjoaa laajan valikoiman teknisiä ratkaisuja, mikä tarjoaa urheilun ja viihteen maailmalle mahdollisuuden olla henkilökohtaisempi ja osuvampi yleisölleen. (Supponor, n.d.)

Kuvattaessa testi- tai liveottelua eri tapahtumissa, käyttäen yrityksen kehittämää tekoälyä, tästä kertynyt videomateriaali tallennetaan yrityksen sisäiseen palveluun. Tallennuksessa toimii levytietovarasto, jossa on kymmeniä teratavuja tietoa tallessa. Teknologian kehittyessä ohjelmistoa tai sen käyttämää tekoälyä kehitetään eteenpäin ja silloin tulee

verifioida, miten kaikki muutokset vaikuttavat teknologian muutokseen ja toimivuuteen. Tätä säilöttyä videomateriaalia testataan silloin uusien ohjelmisto- ja tekoälyversioiden toimivuudessa. Tuloksia vertaillaan sekä vanhan, että uuden version välillä. Vertailu tapahtuu sekä numeerisesti, että visuaalisesti, parhaan lopputuloksen saamiseksi. Nämä testaukset ovat toimineet käytännössä aiemmin yrityksen testaustilanteissa jollakin tavalla, mutta datan säilöminen ja käyttäminen näihin testeihin järjestelmällisesti, vaati kehittämistä.

3 Työn tavoitteet, menetelmät ja yhteenveto

Tässä luvussa käydään läpi opinnäytetyöprosessin tavoitteet, vaiheet ja eritellään tutkimusongelmat. Tutkimuskysymykset, tutkimusmenetelmät ja työn toteutus eritellään myös tarkasti sekä yhteenveto kaikesta kerätystä tiedosta, jotta toiminnallisen osuus oli mahdollista aloittaa. Toiminnallinen osuus vaati laajoja henkilöhaastatteluja sekä yrityksen dataan ja teknologiaan perehtymisen.

3.1 Tavoitteet

Opinnäytetyön tavoitteena oli luoda standardijoukko metatietoja Supponor Oy:n nykyiselle ja tulevalle videomateriaalille. Tärkeimpinä osa-alueina olivat alustavien metatietojen määrittäminen, metadatatallien toteuttaminen ja alustan valinta näille malleille. Opinnäytetyöhön toivottiin myös dokumentointia selkein ohjein metadatatallien päivittämisen prosessiin, jotta yrityksen tuleva data saataisiin myös jatkossa kategorisoitua mallien mukaisesti. Yritykselle oli kertynyt dataa talteen suuria määriä. Tavoitteena oli toteuttaa mallit siten, että tuon suuren datamäärän läpikäymiseen ei vaadittaisi ohjelmointitaitoja, vaan ne voitaisiin toteuttaa esimerkiksi kesätyöntekijän toimesta. Kokonaisuudeksi toivottiin selkeää ja toimivaa ratkaisua, josta olisi välitöntä hyötyä yrityksen uusien ohjelmisto- ja tekoälyversioiden testaustilanteisiin.

Toimeksiantajan toivomuksena oli huomioida ja selvittää myös jatkokehitykseen liittyvät mahdollisuudet. Teknologian kehitys mahdollistaa metatietojen kategorisointiin myös automatisoituja ratkaisuja. Tavoitteena oli ottaa tämä huomioon työn edetessä, alustan valintaa tehdessä ja malleja luodessa. Toimeksiantajan kanssa sovittiin jo opinnäytetyön

alkuvaiheessa, että käytännössä metadatatallit voisi toteuttaa tietäntyyppisinä tiedostoina levyllä, esimerkiksi YAML tai JSON, sisältäen dokumentoidut metadatatiedot.

Tietoperustan tavoitteena oli tarkentaa opinnäytetyöhön liittyviä teknologian periaatteita ja kehitystä sekä tuoda esiin metadataan ja datan kategorisointiin liittyviä huomioita yritysten näkökulmasta. Myös toimeksiantajan ehdottamat YAML ja JSON, vaativat laajaa perehtymistä sekä vertailua keskenään. Tavoitteena oli kerätä tietoa datan automatisoinnista, tekoälystä ja konenäöstä, sillä niistä toivottiin jatkokehitysehdotuksia. Kestävän kehityksen näkökulma oli tarkoituksena tuoda opinnäytetyöhön selkeästi esiin, sillä toimeksiantajan teknologia ja opinnäytetyön loppuratkaisu edistää kestäväää kehitystä. Keskeisessä roolissa oli myös tavoittaa oikeat henkilöt haastatteluja varten.

Opinnäytetyö rajattiin metatietojen määrittämiseen, niiden kategorisointiin, metadata mallien luomiseen, alustan valintaan sekä ohjeiden dokumentointiin. Kuitenkin siten, että kaikkea yrityksen dataa ei ollut mahdollista läpi käydä. Tavoitteena oli koota kokonaisuus urheilulajeittain, yrityksen pientä datamäärää läpi käyden. Jatkotoimenpiteet ja datan automatisointiin liittyvät seikat, rajattiin haastattelujen perusteella tuleviin näkökulmiin ja ne päätettiin koota tämän opinnäytetyön viimeiseen lukuun, yleisellä tasolla.

3.2 Suunnittelu

Tämä opinnäytetyö määräytyi toiminnalliseksi opinnäytetyöksi. Tämä mahdollisti tietoperustan, johon kerättiin pohjatietoa tarpeellisista käsitteistä aiheeseen liittyen. Työn suunnittelussa pyrittiin selvittämään ja erittelemään selkeästi tutkimusongelmat ja lisäksi työn toteutuksessa löytämään parhaat menetelmät vastaamaan näihin tutkimusongelmiin. Tämän opinnäytetyön kohdalla tehtiin erillinen aineistohallintasuunnitelma tukemaan opinnäytetyöprosessia. Työn suunnittelu oli tärkeässä roolissa koko prosessissa, sillä opiskelijalla ei ollut ennestään tietoa metadatatallien ja datan kategorisointiin liittyvistä seikoista.

3.3 Tutkimusmenetelmät ja tutkimusongelmat

Tutkimusmenetelmien avulla ohjataan opinnäytetyöprosessia. Nämä ovat menetelmiä, joilla aineistoa sekä tietoa hankitaan ja myös analysoidaan. Tärkeää on tunnistaa mitä tietoa tutkimuksella haetaan, sillä se ohjaa tutkimusmetodien valintaa. Tutkimusongelmat koostuvat siitä, mitä opinnäytetyön onnistumiseksi on pyrittävä selvittämään. Mikäli kyseessä on hyvin laaja ongelma, aihetta on silloin rajattava. On myös mahdollista, että tutkimusongelmat muuttuvat prosessin aikana, silloin tärkeää on pitää tutkimus kasassa ja varoa lähtevänsä epäolennaisia ongelmia ratkaisemaan. (Alasuutari, 2011)

Tässä opinnäytetyössä viitekehys muodostui tietoperustaan kerätystä tiedosta. Tietoperusta koostui siitä, mitä data on yritysten näkökulmasta ja mitä uusia mahdollisuuksia teknologian kehitys on tuonut mukanaan. Näillä tietoja haluttiin tukea henkilöhaastatteluilla sekä omakohtaisia näkemyksiä toivottiin useaan tutkimusongelmaan. Laadullinen tutkimus ei voi olla teorialtonta, mutta tässä opinnäytetyössä se ei yksinään olisi tuottanut myöskään mahdollisuutta koota yhteen koko opinnäytetyöprosessia. Opinnäytetyöhön tarvittiin henkilöhaastatteluja tuomaan lisätietoa ja omakohtaisia kokemuksia, jotta tutkimusongelmiin voitiin vastata mahdollisimman kattavasti.

Tutkimusmenetelminä tässä opinnäytetyössä perehdyttiin toimeksiantajan teknologiaan sekä keskusteltiin yrityksen vastuuhenkilön kanssa opinnäytetyön tarkoituksesta ja toiveista, jotta lopputuloksesta olisi yritykselle välitöntä hyötyä. Tutkimusmenetelmänä tutustuttiin myös yrityksen videomateriaaleihin laajasti. Tärkeää oli tunnistaa mitä yrityksen data on, mitä se pitää sisällään, kuinka monipuolista data on ja mitä datasta halutaan selvittää. Myös henkilöhaastattelut olivat opinnäytetyössä suuressa roolissa.

Haastateltavien valintaan suhtauduttiin harkiten ja haastatteluihin valittiin sellaisia henkilöitä, joilla arveltiin olevan parhaiten tietoa ja taitoa tutkimusongelmiin liittyen. Haastattelujen menetelmiksi valikoituivat teemahaastattelut ja asiantuntijahaastattelut. Nämä haastattelutyylit yhdistettiin keskenään. Haastateltavaksi etsittiin ohjelmoinnin parissa työskenteleviä asiantuntijoita metatietojen tallennustekniikoihin ja ohjelmointikieliin liittyen. Tärkeää oli saada tietoa myös eri metadatatallennusmallien mahdollisuuksista, saada omakohtaisia kokemuksia YAML- ja JSON-formaateista ja toimintaperiaatteista.

Yliopistojen opettajia kutsuttiin haastatteluihin, jotta saataisiin kokonaistietoa datan kategorisointiin, teknologian kehitykseen, sekä datan automatisointiin ja jatkotoimenpiteisiin liittyen. Lisäksi haastatteluihin etsittiin tekniikanalalla toimivia yrityksistä antamaan näkemyksiä konenäköön ja datan automatisointiin liittyen.

Tutkimuskysymysten tarkoitus oli saada haastateltavilta näkemyksiä ja omakohtaisia kokemuksia, jotta niitä voisi vertailla ja tehdä johtopäätöksiä kokonaisuuden hahmottamiseksi ja oikeiden tekniikoiden valinnan tueksi. Haastattelujen tuloksena toivottiin haastateltavilta ajatuksia ja ehdotuksia, mitä itse opinnäytetyön tekijä ei ollut ajatellut tai tiennyt mahdollisesti tarvitsevensa toiminnallisen osuuden toteuttamiseksi. Tutkimusongelmat ja -kysymykset opinnäytetyön alkuvaiheessa olivat hyvin laajat, sillä aikaisempaa kokemusta tai tietoa ei aiheesta opiskelijalla ollut ennestään.

Tutkimuskysymykset muokkautuivat teemoiksi, sillä vaikka tutkimusongelmat olivat tiedossa, ei tarkkoja ja kohdennettuja kysymyksiä henkilöhaastatteluihin voitu eikä haluttu eritellä. Taulukossa 1. Teemahaastattelut. on kuvattu valitut teemat. Aiheen kokonaisuus vaati paljon hahmottamista ja teemoittain luodut aiheet mahdollistivat paremman tiedonsaannin.

Teemojen tarkoitus oli luoda pohjaa eri haastateltaville heidän osaamisalueiden mukaisesti ja tavoitteena oli avoin haastattelu, jotta haastateltavat voisivat kertoa vapaammin näkemyksistään. Teemahaastattelut toteutettiin haastateltaville sopivien aikataulujen mukaisesti ja näihin varattiin paljon aikaa. Eri henkilöiden haastattelujen välillä oli toisinaan pitkiäkin taukoja, sillä haastateltavista useilla oli aikataulut täynnä jo useammalle kuukaudelle. Haastattelut kestivät keskimäärin tunnin ja haastattelumenetelmänä käytettiin videohaastattelua, Teams-alustan avulla. Vain muutaman henkilön kohdalla haastattelut jäivät lyhyeksi, koska haastateltavat kokivat, etteivät tiedä aiheesta tarpeeksi.

Haastatteluista pyrittiin ottamaan kaikki mahdollinen tieto ylös ja kaikilta haastateltavilta kysyttiin haastattelujen lopuksi myös muiden teemojen aiheista. Näin mahdollistettiin maksimaalinen tiedonsaanti ja vaikka henkilöt kyseisellä hetkellä työskentelivätkin tietyllä alalla, ei haastateltavien taustoista voinut olla varma. Haastatteluissa pyrittiin kartoittamaan kaikki mahdollinen, jotta näkemyksiä ja kokemuksia löydettäisiin aiheisiin liittyen mahdollisimman paljon.

Taulukko 1. Teemahaastattelut.

TEEMA 1.	TEEMA 2.	TEEMA 3.
Ohjelmoinnin parissa työskentelevät henkilöt:	Yliopistojen opettajat:	Konenäön/automaation parissa työskentelevät:
Metadatatamallien merkitys	Teknologian tuomat mahdollisuudet	Konenäön mahdollisuudet
Metatietojen kategorisointi	Metadatatamallien merkitys jatkossa	Datan puoliautomatisointi
YAML ja JSON kokemukset	Datan kategorisointiin liittyvät huomiot	Datan automatisointi
Alustan valinnan merkitys		
Ohjelmointikielet		
Muut mahdolliset huomiot		

Tässä opinnäytetyössä edettiin aineistojen analyysin suhteen siten, että ensimmäiseksi kaikki kerätty aineisto luettiin huolellisesti läpi useaan kertaan. Tämän jälkeen koko aineisto jaettiin osiin, tarkoituksena saada tieto kokonaisuudessaan selkeämmäksi. Osiin jakamisen jälkeen, ylös kirjoitettiin myös omia johtopäätöksiä ja huomioita sekä valintojen perustelut, sillä tärkeää oli ymmärtää aineistojen todellinen sisältö sekä hahmottaa kokonaisuus. Litterointimenetelmää ei käytetty tässä opinnäytetyössä, sillä kaikki haastattelut toteutettiin avoimina keskusteluina. Haastattelijä kirjoitti ylös haastatteluissa ilmenneitä asioita ja lopullinen aineiston purkaminen toteutettiin koodauksen avulla. Teemoittain purettu osat käytiin läpi huolellisesti riviriviltä ja luokitteluyksikön mukaisesti aineistoon merkittiin koodeja. Tämä selvensi aineistojen kokonaisuutta ja aineistot saatiin tiivistettyä siten, että siitä voitiin koota yksi yhtenäinen kokonaisuus.

Haastatteluista myös kerättiin näkemyksiä jatkotoimenpiteet osioon. Tarkoituksena oli kartoittaa haastateltavilta omakohtaisia näkemyksiä ja ideoita, mitkä voisivat olla mahdollisia toimintatapoja datan automatisoinnin edistämiseksi. Tämän opinnäytetyön viimeisessä luvussa esiin nostetut jatkotoimenpide-ehdotukset, on luotu täysin haastattelujen avulla ja koska haastatteluissa ei käytetty yhtä tiettyä kysymysluokittelua, haastatteluaineistoa ei analysoitu tätä tarkemmin.

Lähdekritiikki huomiointiin opinnäytetyössä tarkoin. Teknologia kehittyy valtavaa vauhtia, joten opinnäytetyössä varmistettiin lähteiden sisältöjen ajankohtaisuus. Kun tietoperustaan etsittiin tietoa, tietojen ajankohtaisuuden lisäksi todenmukaisuus varmistettiin useista artikkeleista sekä lisäksi henkilöhaastatteluiden yhteydessä. Myös tiedontuottajat valittiin tarkoin lähteitä etsiessä. Huomioon otettiin kuinka luotettava ja tunnettu tiedontuottaja on, pyrittiin selvittämään lähteen laatijan tavoitteet sekä onko lähde oikeasti se, mitä se kertoo olevansa.

Tämän opinnäytetyön analyysitavaksi valittiin sisällön analyysi, sillä tätä tapaa voidaan käyttää useiden erilaisten aineistojen analyysiin. Tässä opinnäytetyössä sitä käytettiin teemahaastattelujen ja asiantuntijahaastattelujen purkamiseen. Sisällönanalyysia voitiin myös hyödyntää kirjallisen materiaalin analysoinnissa. Tällä pyrittiin saamaan kokonaiskuva tutkittavasta aiheesta ja sen kokonaisuudesta kattavassa muodossa.

Tutkimuksia ja haastatteluja tarkisteltaessa todettakoon, että reliabiliteettia olisi voitu parantaa, jos tutkimusaineistoa olisi saatu vieläkin enemmän, eritoten haastattelujen myötä. Toisaalta, koska tutkimus on laadullinen ja perustuu yksittäisiin haastatteluihin, eritoten erilaisiin henkilöihin, on reliabiliteettia myös hankala arvioida. Tutkimuksista kerätty yhteenveto vastasi kuitenkin tutkimusongelmiin niiltä osin, että toiminnallinen osuus oli mahdollista toteuttaa. Sen voidaan siis katsoa olleen luotettava, sekä tutkimuksen suorittamisia haastattelujen muodossa sekä niistä saatuja tutkimustuloksia, voidaan pitää pätevänä. Useilla haastateltavilla oli yleistietoja aiheista ja nämä henkilöt myös opettavat aiheista työkseen, joten niitä voidaan arvioida luotettavaksi. Haastatteluiden kautta löytyi myös henkilöitä, jotka työskentelevät ohjelmoinnin sekä metadatatallien parissa, joten voidaan näiden haastateltavien olevan alansa asiantuntijoita.

3.3.1 Laadullinen tutkimus ja sisällönanalyysi

Laadullinen eli kvalitatiivinen tutkimus on tieteellisen tutkimuksen menetelmäsuuntaus. Laadullisessa tutkimuksessa on tarkoituksena pyrkiä ymmärtämään tietyn kohteen laatua, ominaisuuksia ja merkitystä kokonaisuudessaan. Tämä menetelmäsuuntaus jakaa mielipiteitä, joten niiden oppaisiin tulisi suhtautua kriittisesti. (Tuomi & Sarajärvi, 2018) Sisällönanalyysi on tapa, jossa tarkastellaan tekstimuotoista aineistoa. Sisällönanalyysia

voidaan käyttää esimerkiksi haastattelujen, tekstin kuin kuvaakin sisältävien aineistojen analyysiin. (Nikander ym., 2017, s. 18)

3.3.2 Teemahaastattelut ja asiantuntijahaastattelut

Teemahaastattelu on yksi esimerkki sovelletusta puolistrukturoidusta haastattelusta ja on yksi käytetyimmistä haastattelumenetelmistä. Teemahaastattelu ei keskitä haastattelua mitenkään, kvalitatiiviseen tai kvantitatiiviseen eikä se myöskään ota kantaa haastattelujen määrään. Teemahaastattelussa oleellisinta on se, että tarkkoja määritettyjä kysymyksiä ei tarvitse olla, vaan haastattelut etenevät teemojen varassa. Tämä mahdollistaa vapaamman haastattelun ja haastateltavat saavat vapaammin kertoa näkemyksiään. (Hirsjärvi & Hurme, 2001, s. 48)

Asiantuntijahaastatteluista voi aina seurata riskinä se, että opinnäytetyön tekijällä ja asiantuntijalla on ristiriitaiset tavoitteet tai ajatus kokonaisuudesta. Asiantuntijalla ei myöskään välttämättä ole tavallisen haastateltavan halua auttaa ongelman ratkaisemisessa. Haastattelijan on osattava tällöin vaihtaa lähestymistapaansa. Toisinaan voi olla hyvä esiintyä itse asiantuntijana tai toisinaan mahdollisesti tietämättömyyden korostaminen voi auttaa saamaan tarvittavaa tietoa. (Hyvärinen ym., 2017)

3.3.3 Tutkimusten luotettavuus

Luotettavuuden arviointi on hyvin keskeinen osa tutkimusten tarkistelua. Tutkimuksella on aina asetettu joitakin tiettyjä arvoja, joihin on pyrittävä. Luotettavuuteen vaikuttaa myös se, kuinka tarkasti kerättyä aineistoa on analysoitu. (Saaranen-Kauppinen & Puusniekka, 2006) Luotettavuuskysymyksissä keskeisiä käsitteitä ovat perinteisesti reliabiliteetti ja validiteetti, arvioitaessa laadullisen tutkimuksen luotettavuutta. Validiteetti eli toisin sanoen pätevyys, tarkoittaa tutkimusmenetelmän kykyä mitata sitä mitä sen on tarkoitus mitata. Reliabiliteetti eli luotettavuus, taas tarkoittaa mittauksen toistettavuutta ja sitä, onko aineiston keruu tehty oikein. Asiantuntijahaastatteluihin on suhtauduttava myös kriittisesti, kuten kaikkiin haastattelumuotoihin. Ne antavat jokaisen haastateltavan kohdalla heidän omaa näkemystänsä tiedoksi ja näitä tietoja ei voi pitää täysin luotettavana. (Puusa & Juuri, 2020)

3.4 Tutkimusten yhteenveto

Tässä luvussa tuodaan esiin tutkimuksissa selvinneitä asioita. Esiin otetaan opinnäytetyöntekijän omat tutkimukset, yrityksen dataan ja teknologiaan liittyen sekä haastattelujen perusteella kerätyt näkökulmat tutkimusteemoihin perustuen. Haastateltavien näkökulmat eivät kuitenkaan itsessään ollut tuloksia, vaan antoivat tukea tietoperustaan kerätystä tiedosta sekä henkilökohtaisia näkemyksiä ja ehdotuksia.

3.4.1 Yrityksen dataan ja teknologiaan perehtyminen

Tärkeää tutkimuksia tehdessä, oli keskustella useasti toimeksiantajan kanssa työn tavoitteista sekä tutustua huolella yrityksen dataan. Yrityksen teknologiaa oli myös ymmärrettävä, jotta saatiin käsitys siihen, mihin metadatatamalleja jatkossa hyödynnetään ja millaisia metatietoja tulisi määrittää. Yrityksessä on toiminut ohjelmisto- ja tekoälyversioiden testaustilanteet, mutta siihen toivottiin selkeyttä ja kehittelyä sekä esiin nousi ongelmat liian suurien datamäärien kanssa. Yrityksen kehitlemä tekoäly myös helposti mallintaa ongelmatilanteissa siihen dataan, mitä löytyy eniten tallesta ja pyrkii itse korjaamaan näitä esiintyviä virheitä. Tällöin on tärkeää huomioida, että data on tasapainoista ja saman arvoista.

Vaikka ihmisen silmä ei erota näitä pieniä virheitä, mitä yrityksen teknologiassa toisinaan esiintyy, silti yritys pyrkii jatkuvasti kehittämään näitä ohjelmisto- ja tekoälyversioita. Mikäli näitä virheitä tapahtuu, ne johtuvat liveotteluissa esiintyvistä seikoista. Tekoälyn toimintaa saattaa rajoittaa sääolosuhteet, eriväriset peliasut pelaajien päällä tai esimerkiksi muut värit, mitä kentällä tai sen ulkopuolella esiintyy. Eri urheilulajien tapahtumissa kentällä ja sen ympäristössä on yllättävän paljon pieniä yksityiskohtia. Metatietojen määrittämisessä oli tärkeää huomioida kaikki mahdolliset yksityiskohdat, sillä monipuolinen data mitä toivottiin talteen, määräytyisi jatkonkin suhteen metatietojen sisällöstä.

Toimeksiantaja kertoi tarpeelliseksi, että metatietoja määritettäessä on myös tuotava esiin, mitkä joukkueet peleissä pelaa, millä kentällä se tapahtuu ja milloin peli on pelattu. Tällöin eri joukkueita voidaan verrata testaustilanteissa ja etsiä johtuuko mahdolliset virheet juuri tietyn kentän yksityiskohdista. Pelien ajankohta taas helpottaisi jatkossa oikean pelin

löytämistä. Supponorin teknologiaa käytetään tällä hetkellä useissa eri lajeissa. Lajit ovat jalkapallo, jääkiekko, koripallo, pesäpallo, tennis, rugby, kriketti, hevosurheilu ja koiraurheilu. Lisäksi yritys testaa teknologiaansa uuteen lajiin, jota ei tässä vaiheessa vielä julkaistu, mutta opinnäytetyössä määritettiin metatiedot sekä toteutettiin metadatatamalli, myös tämän lajin kohdalla. Dataan tutustuesssa selvisi nopeasti, kuinka paljon eri lajit poikkeavat toisistaan. Dataa tarkisteltaessa tiedostettiin, että kaikille lajeille olisi toteutettava omat metadatatamallit ja yksityiskohtia joka lajin kohdalla tarkennettava.

3.4.2 Haastattelut

Haastatteluihin kutsuttiin sähköisessä muodossa yhteensä seitsemäntoista henkilöä. Kaikki haastatteluihin kutsutut tavoitettiin ja jokaisen kanssa haastattelu toteutettiin.

Haastattelujen edetessä varmistui, että henkilöt olivat kohdennettu oikein, sillä usea haastateltavat mainitsi jonkin tietyn henkilön nimeltä, joka sopisi tämän opinnäytetyön tutkimusongelmiin vastaamaan ja nämä ehdotetut henkilöt löytyivät jo haastattelulistoilta.

Vain kahdella haastateltavista löytyi omakohtaista kokemusta YAML-merkintäkielestä. Haastatteluista tuli ilmi se, että JSON oli tutumpi sekä selvästi useimmin käytetty formaatti Suomessa. Näiltä kahdelta haastateltavalta saatiin kuitenkin paljon hyödyllistä tietoa. YAML-merkintäkieltä kehitettiin helppokäyttöiseksi ja tämän opinnäytetyön tarkoitusta vastaavaksi paremmin kuin JSON. Tietoa saatiin myös siitä, että YAML on helppo muuttaa JSON-muotoon tarvittaessa, jos toimeksiantaja toteaisi sen paremmaksi vaihtoehdoksi. Toiselta haastateltavista saatiin ohjeet, miten YAML-formaatin käyttöön vaadittavat työkalut ladataan koneelle, neuvoja sen käyttöönotosta ja miten pääpiireittäin YAML toimii. Lisäksi haastateltavalta saatiin hyödyllinen virheenkorjauslinkki, tukemaan metadatatamallien oikeinkirjoitusta. Esiin nousi myös, että JSON toimii hyvin silloin kun eri koneet keskustelevat keskenään, YAML taas silloin, kun ihminen toteuttaa jotakin käsin.

Haastattelujen myötä saavutettiin myös ymmärrys siihen, ettei ohjelmointikielellä malleja toteuttaessa tässä vaiheessa ollut merkitystä. Malleihin kirjataan metatietoja alustan toimivuuden mukaisesti ja valmiit mallit voidaan myöhemmin muuttaa eri ohjelmointikielille, tarpeiden mukaisesti. Tämä tieto helpotti opinnäytetyöprosessissa, eikä varsinaista ohjelmointikieltä tarvinnut pohtia sen tarkemmin. Selvisi että YAML sallii kommentit

sisällössään metadatamalleja luodessa, JSON taas ei. Ylimääräiset kommentit helpottavat ymmärtämään mallien sisältöä tarkemmin. Näiden haastattelujen perusteella tulleet asiat kirjattiin ylös ja niitä verrattiin myös yleiseen tietoperustaan.

Vaikka kaikilla ohjelmoinnin parissa työskentelevillä henkilöillä ei ollut omakohtaisia kokemuksia YAML-merkintäkielestä, niin kuitenkin metatiedoista, niiden määrittämisestä ja metadatamallien merkityksestä yleisesti, saatiin kerättyä paljon huomioita. Haastateltavat kertoivat, että tärkein osuus koko opinnäytetyönprosessissa on metatietojen oikea määrittäminen. Vain kohdentamalla metatiedot toimeksiantajan tarpeisiin sopiviksi, todella tarkasti ja harkiten, metadatamalleista olisi lopulta hyötyä. Haastateltavat neuvoivatkin käyttämään eniten aikaa metatietojen määrittämiseen, varmistamaan useaan kertaan toimeksiantajalta, että ne ovat varmasti heidän tarpeisiinsa hyödyllisiä, sekä kohdentamaan metatiedot oleelliseen tietoon, jättäen epäoleelliset huomiotta. Ohjeita saatiin myös dokumentointiin liittyen ja se ehdotettiin tehtäväksi erikseen kirjaten, eikä vain valmiisiin malleihin lisäkommentteja luoden. Kun dokumentointi tehtäisiin kerralla selkeästi, olisi metatietojen päivittäminen helppoa ulkopuolisille henkilöille, vain tarkkoja ohjeita seuraten.

Haastattelujen myötä kokonaisuutena esiin nousi datan automatisoinnin ja konenäön tärkeys jatkotoimenpiteinä. Jokainen haastateltava totesi, että datan hyödyntäminen lähtee liikkeelle aina käsityöstä ja sillä luodaan pohja jatkolle. Pidemmällä tähtäimellä tuskin kuitenkaan yhdenkään yrityksen aika tai resurssit riittäisivät käsin datanhallintaa toteuttamaan. Haastatteluista toivottiinkin näkökulmia näihin jatkomahdollisuuksiin liittyen datan käsittelyssä. Tuloksena useimmista jäi se, että mahdollisuuksia on monia ja kokonaisuus täytyy huomioida. Selkeitä käytännön esimerkkejä ei juurikaan saatu. Haastatteluista kuitenkin aukesi ymmärrys siihen, ettei ole yhtä tai kahta oikeaa vaihtoehtoa tai ratkaisua. Yrityksen teknologiaan, sen mahdolliseen hyödyntämiseen sekä kokonaisuuteen mitä toimeksiantaja toivoisi saavuttavansa, pitäisi tutustua paljon laajemmin vaihtoehtojen määrittämiseksi. Jonkin verran erilaisten rajapintojen hyödyntämisestä saatiin tietoa ja muutamia muita esimerkkejä, näitä kirjattiin tämän opinnäytetyön viimeiseen lukuun.

Yleisten eettisten periaatteiden mukaisesti haastatteluissa tulee kunnioittaa haastateltavia sekä heidän yksityisyytensä. Haastateltavista osa halusi varmistaa haastattelun

anonyymiyden ja tästä syystä henkilötietoja ei luovuteta eteenpäin tai mainita opinnäytetyössä haastateltavien nimiä. Kaikki haastateltavat vastasivat kysymyksiin asianmukaisesti sekä olivat selvästi halukkaita auttamaan parhaansa mukaan.

4 Metatietojen merkitys datassa ja teknologian kehitys

Data yleisesti on digitaaliseksi muutettua tietoa. Data voi pitää sisällään ääntä, tekstiä tai esimerkiksi liikkuvaa kuvaa. Kaikki se tieto, joka eri organisaatioiden sähköisissä järjestelmissä liikkuu, on dataa. (Muurinen, 2011) Dataa on kaikkialla ympärillämme valtavia määriä ja sitä syntyy koko ajan lisää. Jokainen myös osallistuu datan tuottamiseen jopa huomaamattaan, jakamalla kuvia sosiaalisessa mediassa, selaamalla verkon kautta uutisia, maksamalla laskuja verkkopankissa ja usealla muulla tavalla. (Andersson, 2019)

4.1 Datan merkitys yrityksissä

Eri yrityksille kertyy erilaista dataa. Se mistä data koostuu, riippuu eri toimialoista. Toiset hyödyntävät verkkosivuja yritystoiminnassa ja silloin data kertyy käyttäjien toimista verkkokauppa-alustalla. Ostettaessa esimerkiksi vaatteita verkkokaupasta, yritys pystyy hyödyntämään käyttäjän toimia, sillä siitä voidaan kerätä dataa yrityksen käyttöön. Yritys voi selvittää käyttäjän toimien perusteella, mitä vaatteita on käyty katsomassa tarkemmin, kerätä tietoa, kuinka paljon jollakin tietyllä tuotteella on kysyntää tai vain katselukertoja ja saada näistä toimista kerättyä dataa verkkokaupan ja sen sisällön kohdentamiseen. Aina käytettäessä internetiä, siitä jää jälkiä verkkoon ja tuosta datasta muodostunutta kokonaisuutta kutsutaan digijalanjäljeksi. Ihmisten toimia verkossa voidaan monella tavalla kerätä talteen ja näistä muodostuu valtavia määriä dataa.

Dataa voidaan kerätä muullakin tavalla kuin vain verkossa asioinnin avulla. Puheentunnistus on hyvin yleistä ja useinkin soitettaessa eri yritysten puhelinpalveluihin, nauhoite ilmoittaa puhelun tallentamisesta. Nauhoitteiden avulla yritykset saavat tallenteista muutettua digitaalista dataa ja hyödynnettyä sitä liiketoiminnan kehittämiseen. (Pulkkinen, 2005) Dataa syntyy siis eri lähteistä, eri yrityksille, eri tavoin. Monet yritykset ovat jo vuosikymmeniä osanneet hyödyntää dataa ainakin jollakin tavalla yritystoiminnassaan, ja teknologian kehittyessä yritysmaailma on saanut entistä enemmän mahdollisuuksia toimintoihinsa. Se

korostuu mitä enemmän opitaan datan keräämisestä sekä sen oikeanlaisesta hyödyntämisestä. (Dufva, 2020)

Dataa voi loputtomasti käsitellä, sillä se ei kulu käytössä ja uutta dataa syntyy koko ajan lisää. Se miten dataa hyödynnetään yrityksen sisällä, riippuu paljon yrityksen toimintaperiaatteista, palveluista sekä yrityksen käyttötarpeista. Merkittävää on myös datan saatavuus ja sen laatu sekä tehokas ja luotettava datan hallinta. Ihanteellista olisikin hyvin organisoitu data, joka on tarkkaa ja ymmärrettävää sekä se on tallennettu, analysoitu ja tulkittu juuri omiin tarpeisiin sopiviksi. Lisäksi tarvitaan asianmukaisia työkaluja, suunnittelua ja datan kategorisointia, ennen suurien datamassojen analysoimista. Oleellista ei ole kerätä ja tallentaa kaikkea dataa talteen, vaan tärkeää olisi saada monimuotoista dataa kerättyä kategorisoidusti. (Hirvonen, 2015, s. 10) Dataa on eroteltava toisistaan, jotta oikea tieto löytyy. Ennen kuin suuria datamääriä edes lähdetään analysoimaan, on päätettävä strategia yritykselle. Oleellista on huomioida, onko yrityksellä paljon omaa dataa käytössään ja millä tavoin sen tallennus on toteutettu. Digitaalisesti tallennettu data on aina helpompaa analysoida verrattuna esimerkiksi paperisiin arkistoihin. Paperisten arkistojen vieminen digitaaliseen muotoon vie paljon aikaa. (Lindholm, 2021)

4.2 Teknologian kehitys

Teknologian kehitys on nopeampaa kuin koskaan ennen. Uudet teknologiat ja erilaiset älylaitteet tuovat mukanaan jatkuvasti uusia mahdollisuuksia. Eri tieteenaloilla eletään kiistattomasti kulta-aikaa, sillä teknologian kehitys on jopa hämmästyttävän nopeaa. Muutosnopeutta on toisinaan jopa hankala havaita, kun erilaiset teknologiat älypuhelimista itseohjautuviin autoihin ja erilaisiin robotteihin muuttavat arkea. (Limnéll, 2019) Teknologiakehityksen myötä valtavia muutoksia tapahtuu niin monen työelämässä kuin arjessakin. Tämän kehittymisen ansiosta monen yrityksen kohdalla voidaan lisätä tuottavuutta ja siten talouskasvua sekä tehostaa monen yrityksen tuotantoa ja toimivuutta (Setälä, 2020). Erilaisten koneiden, robottien ja tekoälyjen avulla, työt voidaan toteuttaa nopeammin ja tehokkaammin, kuin mihin ihminen pelkästään yksinään kykenisi (Heikkinen & Nieminen, 2021).

Automaation seurauksena ihmisten tekemät toimet voidaan helposti korvata jopa kokonaan koneilla. Ajatellaan, että yritys on esimerkiksi osannut hyödyntää oikein keräämäänsä dataa, automaatio mahdollistaa lisäksi entistä kokonaisvaltaisempia ratkaisuja myös datan käsittelyyn. Tekoäly simuloi ihmisten älykkyyttä, joka prosessoidaan tietokoneilla. Prosessit sisältävät oppimista, merkityksen löytämistä, päättelyä ja itseoikaisua. (Copeland, 2018.) Tekoälyllä tarkoitetaan tietokonetta tai tietokoneohjelmaa, joka kykenee käyttämään kykyjä, jotka on liitetty ihmisen älykkyyteen. Tällaisia taitoja ovat mm. oppiminen, suunnitteleminen ja päättely. Tekoälyn ansiosta tekniset järjestelmät voivat havainnoida ympäristöään, käsitellä erilaisia havaintojaan ja ratkaista ongelmia. Tekoäly pyrkii saavuttamaan tietyn päämäärän näillä kyvyillä, sekä ottamaan myös vastaan erilaista tietoa niiden käsittelyä varten. Koneoppiminen on tekoälyn osa-alue, joka mahdollistaa erilaisille järjestelmille mahdollisuuden oppia kokemuksia automaattisesti. Kone oppii datasta samalla tavalla, kuin ihminen oppii kokemuksistaan. Opittuaan sovellus osaa käyttää oppimaansa päätösten tekemiseen ilman, että ihmisen tarvitsee puuttua siihen. (Euroopan parlamentti, 2020)

Monet yrittäjät ja tutkijat pyrkivät kehittämään itsenäistä tekoälyä. Tarkoituksena ettei se tarvitsisi ihmisen tukea edes todella monimutkaisten päätösten kanssa. Tämä mahdollistaisi täysin uudenlaisia liiketoimintamalleja kaikille toimialoille. Yhdistämällä automaation, koneoppimisen ja tekoälyn, tulevaisuus tuo täysin uusia mahdollisuuksia. (Vatanen, 2017)

4.3 Metatieto ja metatietojen kategorisointi

Metatieto on lyhyesti tietoa tiedosta. Metatieto voi olla vain tiedostonimi ja -pääte, mutta laajimmillaan se voi olla todella kattavasti taltioitavan aineiston sisältö. Se on tietoa, joka kulkee tiedoston mukana tai on tallennettuna tietokantaan. Moni ei arjessaan tunnista sanan merkitystä, vaikka todellisuudessa metatieto on asia, jota ilman arki- tai työpäivä ei tänä päivänä kulkisi kovinkaan helposti eteenpäin. (Ollikainen, 2017) Metatietojen tehtävänä on dokumentoida aineistojen ominaisuuksia ja näin helpottaa aineistojen käyttöä.

Metatietoja esiintyy joka puolella, vaikka niitä ei varsinaisesti havaitsekaan. Esimerkiksi internetin suoratoistopalvelut Netflix tai Viaplay, löytävät valikoimistaan toivotut elokuvat tai sarjat vain sillä, että niiden kaikki aineistohaut pohjautuvat metatietoihin. Käytettäessä googlea tai mitä tahansa hakukonetta, niiden pohjalta tieto on mahdollista löytää vain metatietojen avulla.

Pohjoismaissa metatietojen hallinta on edelleen melko alhainen. Isompia organisaatioita sekä esimerkiksi pankkeja lukuun ottamatta, yritykset eivät vielä täysin ymmärrä metatiedon merkitystä, vaikka dataa osattaisiinkin osittain hyödyntää. (Harju, 2019) Ilman metatietoja tiedonhallinta ja tiedon löytäminen olisi lähes mahdotonta. Metatiedot helpottavat ihmisten jokapäiväistä arkea ja lisäksi ne helpottavat myös työelämässä montaa asiaa. Jos jonkin tarvittavan tiedon löytäminen tuottaa yrityksessä haasteita, helpoin tapa on lähteä liikkeelle siitä, että laittaa metatiedot kuntoon. Eri organisaatioissa ja yrityksissä, joissa useat työntekijät luovat dokumentteja päivittäin, käsittelevät niitä, muokkaavat tai tallentavat, voidaan metatietojen hyödyntämisestä saada todella paljon irti. Ne helpottavat oikean tiedon löytymistä sekä tieto tulee tallennettua järjestelmällisesti. Kun oikea tieto löytyy helposti, se kuluttaa myös paljon vähemmän aikaa ja resursseja. (Forsström, 2019)

Ensimmäinen askel on määrittää yrityksen datasta tarpeellinen tieto. Tärkeää on tunnistaa mitä tiedoilla tehdään jatkossa ja siitä myös määräytyy metatiedon määrä. Turhia metatietoja tulisi välttää, hyvin suunnitellussa metatieto mallissa pärjätään parhaiten maltillisella määrällä metatietoja, toivottujen hakujen ja raportointitarpeiden toteuttamiseksi. Kaikelle erityyppiselle tiedolle voi olla eri määrä metatietoja. (Koppatz, 2020)

Metatietojen määrittäminen on luokittelua, datan tarpeellisten tietojen keräämiseksi. Metatiedot kategorisoidaan eli eritellään jonnekin talteen. Kun tarpeellinen tieto on tunnistettu ja metatiedot määritetty, niistä luodaan metadatatalleja, esimerkiksi tietyn tyyppisenä tiedostona sisäisessä tallennuksessa tai talteen tietokantaan ulkoiseen tallennukseen. Manuaalisesti metatietojen kategorisointi on työlästä ja vaatii paljon aikaa. (Suomen standardisoimisliitto SFS Ry, 2020)

Metadatasta suurin osa voidaan teknologian kehityksen ansiosta tallentaa myös puoliautomaattisesti tai automaattisesti. Edellytyksenä tälle on oikeanlaiset järjestelmät, joita on syytä harkita yrityskohtaisesti. Metatiedot ovat kuitenkin ensin itse määriteltävä ja koko prosessi lähtee liikkeelle käsityöstä, vaikka jatkossa onkin hyvä hyödyntää automaattista metatiedon tallentamista. (Ollikainen, 2017)

Dokumentointi on tärkeää metadatamalleja luodessa. Se tarjoaa tietoa, joka auttaa selventämään näitä malleja ja helpottamaan muiden osapuolten ymmärrystä siihen, mitä tietoja on määritetty. Kun metadatamalleja luodaan eri alustoille tai merkintäkielille, niiden tallennustekniikan valintaa on syytä harkita ja miettiä miten dokumentoinnin voisi toteuttaa. (Tietoarkisto, 2021)

4.4 YAML ja JSON

YAML on merkintäkieli, jonka lyhenne tulee sanoista ”YAML Ain’t Markup Language”. Tätä merkintäkieltä käytetään usein tietojen määrittämisessä, mutta sitä voidaan käyttää myös toisenlaiseen tietojen talletukseen. (Fileinfo, 2015) YAML on ihmisten luettavissa oleva tiedostomuoto, jota käytetään tietojen lukemiseen ja kirjoittamiseen. Metadatamalleja luodessa, tätä voidaan käyttää useilla ohjelmointikielillä, jotka tukevat YAML-kirjastoja, kuten C / C ++, Ruby, Python, Java, Perl, C # ja PHP. (Sharma, 2021) Maailmassa on monia erilaisia merkintäkieliä, se on kieli, joka merkitsee tekstiä siten, että tietokone kykenee manipuloimaan sitä (Tieteen termipankki).

JSON on lyhenne sanoista ”JavaScript Object Notation” ja on yksinkertainen avoimen standardin tiedostomuoto tiedonvälitykseen. Nimestään ja JavaScript-perustastaan huolimatta, JSON on JavaScriptistä riippumaton. JSON-tiedosto on tiedosto, joka tallentaa yksinkertaiset tietorakenteet ja objektit JavaScript Object Notation (JSON) -muodossa. Tämä taas on tavallinen tiedonsiirtomuoto, jota käytetään ensisijaisesti tiedonsiirtoon web-sovellusten ja palvelimien välillä. JSON-tiedostot ovat tekstipohjaisia ja ihmisen luettavissa sekä niitä voidaan muokata tekstieditorilla. JSON on myös ohjelmointikielestä riippumaton tiedonesity standardi. (Tiedostomuotojen opas)

YAML on suurissa määrin helpompi lukea kuin JSON, vaikka toimintaperiaate on enimmäkseen molemmissa sama. YAML on helpompaa kirjoittaa ihmiselle, sillä se sisältää yksinkertaisempia syntaksiin liittyviä erikoismerkkejä kuin JSON. YAML tukee myös useita kommentteja, joita JSON ei sellaisenaan tue. Esimerkiksi pilkkuja ei tarvitse rivien lopussa, lainausmerkkejä tekstikenttien arvojen ympärillä eikä hakasulkeita kentissä. YAML myös tukee omia kommentteja, mitä JSON taas ei ota vastaan. Omat kommentit helpottavat tiedon ymmärtämistä ja lukemista. (Bertram, 2021)

5 Toiminnallinen toteutus

Toiminnallinen toteutus tässä opinnäytetyössä jaettiin neljään osaan. Tärkein osuus koko prosessissa oli metatietojen määrittäminen, sillä ilman oikeanlaisia metatietoja ei yritykselle olisi lopullisista metadatatamalleista hyötyä. Metatietojen määrittämisen jälkeen päätettiin testata YAML ja harjoitella sen toimivuutta. Näiden jälkeen luotiin itse metadatatamallit ja viimeisenä dokumentoitiin ohjeet metadatatamalleille.

5.1 Metatietojen määrittäminen

Yrityksen dataan oli tutustuttu jo opinnäytetyön alkuvaiheessa, mutta toiminnallisen osuuden alkaessa, yritys lähetti lisää monipuolisempaa dataa. Tämä data koostui eri urheilulajien videomateriaaleista. Jokaisesta lajista saatiin dataa erikseen ja se koostui pienistä osista kuvattuja pelejä tai otteluja. Videoita katsottiin läpi useaan kertaan ja ylös kirjoitettiin havaintoja peleistä. Nämä havainnot olivat metatietojen määrittämistä, eli määritettiin yrityksen tarpeisiin kohdistetut tiedot datasta. Metatiedoiksi määritettiin:

- Päivä jolloin ottelu/kilpailu oli aloitettu
- Kellon aika jolloin ottelu/kilpailu oli aloitettu
- Sijainti missä ottelu/kilpailu oli käyty
- Urheilulaji mikä videolla esiintyi
- Liigan nimi
- Joukkueiden/pelaajien mahdolliset nimet
- Värit mitä esiintyi eri henkilöiden asuissa, ympäristössä, mainoksissa, muut lajikohtaiset yksityiskohdat
- Yleisön läsnäolo
- Sääolosuhteet

Metatietoja oli sovellettava lajikohtaisesti ja kaikkia määritettyjä metatietoja ei tarvittu jokaisessa lajissa. Lajeittain koottiin erikseen omat tarkennetut metatiedot, vaikka runko pyrittiin pitämään joka mallissa samana.

Kun oli määritetty mitä metatietoja tulitaisiin tarvitsemaan, niihin piti löytää myös vastaus videoista ja kirjata mikä metatieto oli kyseisessä videossa. Videoiden tallennustiedoista pystyttiin erittelemään eri pelien ja kilpailujen alkamispäiviä ja joistakin tiedoista löytyi myös suoraan näiden tarkemmat alkamisajat. Videoita tarkastellessa, vastaan tuli kuitenkin myös useita kohtia, mihin itse videon katsomisella tai tallennustiedoista ei löytynyt vastauksia. Kun metatietoihin tarvittiin esimerkiksi eri joukkueiden virallisia nimiä, pelaajien kokonimiä, paikkamerkintöjä millä kentillä pelejä pelataan tai näiden kenttien virallisia nimiä, silloin näitä tietoja etsittiin verkosta. Kun ennestään oli tiedossa, että Supponor käyttää teknologiaansa mm. koripallon osalta NBA:n peleissä, oli verkon kautta mahdollista etsiä joukkueiden nimiä ja verrata niitä videoissa näkyviin pelaajiin ja heidän peliväreihinsä. Myös monessa ottelussa näkyi selkeät logot kenttään painettuna, joiden avulla voitiin päätellä ja yhdistää joukkueita, sekä tietoja missä pelejä oli pelattu.

Esimerkki metatietojen määrittämisestä jääkiekkopelin kohdalla:

- Katso video ensin läpi kokonaisuudessaan
- Katso uudelleen ja havainnoi mitä pelissä esiintyy
- Kirjaa ylös niitä huomioita mitä videosta löytyy
- Määritä tärkeä ja oleellinen tieto ja tähän määritettyyn tietoon vastaus
- Jatka tietojen määrittämistä ja etsi määritettyyn tietoon vastaus
- Käy koko video läpi, etsi kaikki oleellinen ja tärkeä tieto, kirjaa kaikki ylös
- Päivä, jolloin peli on pelattu - 24.5.2021
- Aika, jolloin peli on aloitettu –
- Sijainti missä peli pelattu – Centre Bell, Montreal, Canada
- Olosuhde pelissä – Sisäpeli/Halli
- Mikä urheilulaji on kyseessä – jääkiekko
- Liigan nimi – NHL
- Mitkä joukkueet pelaavat – Montreal Canadiens vs. Toronto Maple Leafs
- Minkä väriset kypärät pelaajilla/tuomareilla – sininen, valkoinen, musta
- Minkä värisiä asuja pelaajilla on – sini, puna, valkoinen ja sinivalkoinen
- Tuomareiden, muiden henkilöiden asujen värit – mustavalkoinen
- Yleisön läsnäolo – ei
- Mitä värejä esiintyy ympäristössä – punainen, sininen, oranssi, vihreä, valkoinen,

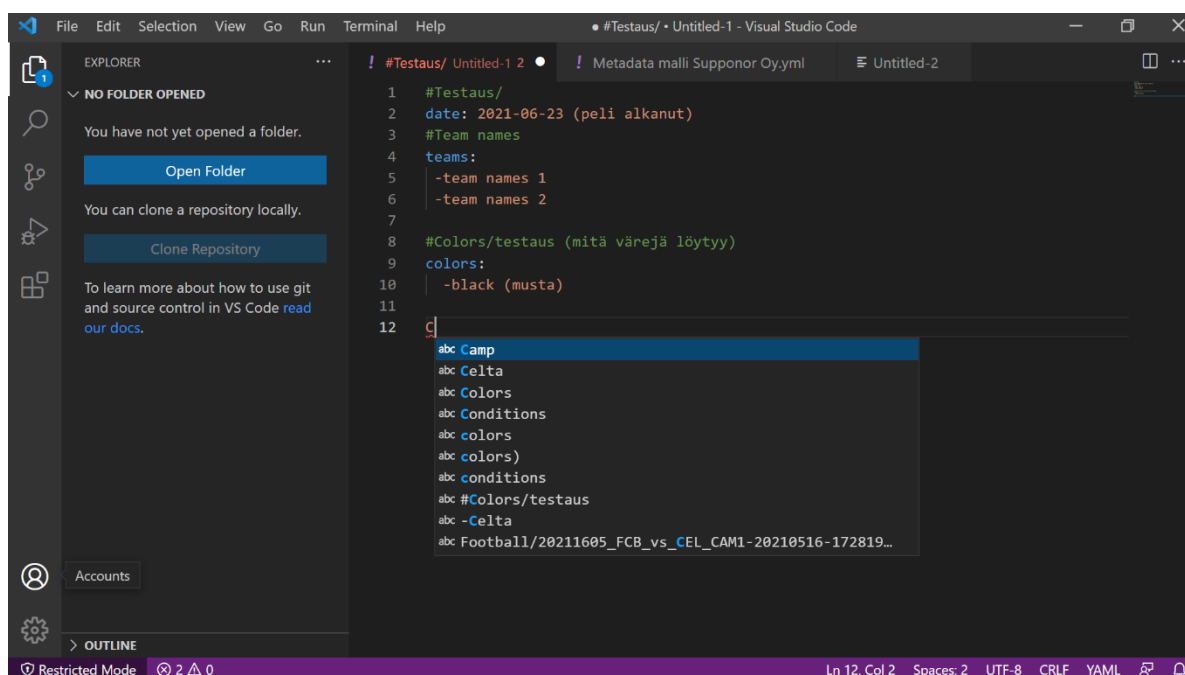
Jokaisen lajin kohdalla kaikkia tietoja ei kuitenkaan voitu löytää edes verkosta tietoja etsimällä. Jos esimerkiksi tarkasteltiin esimerkissä olevaa jääkiekkopeliä, niin videon tallennustiedoista ei löytynyt alkamisaikaa, vaikka alkamispäivä videon tallennuksen yhteydessä oli merkitty. Nostetaan toiseksi esimerkiksi tennisottelu, videoissa ei esiintynyt pelaajiin liittyviä logoja tai muuta pelaajiin liittyvää yhdistävää tekijää, joten vaihtoehtoksi jäi vain jättää kohta tyhjäksi. Tästä heräsiikin entistä vahvempi ajatus siitä, miten tärkeää metadatan automaattinen kategorisointi tulevaisuudessa yritykselle olisi, sillä tämä toimintatapa oli todella työlästä. Tämän opinnäytetyön toimeksiantajalle kertyy jatkuvasti niin suuria määriä uutta dataa, joten mahdotonta olisi jokaisen ottelun tai pelin kohdalla etsiä määritettyjä metatietoja vain videoita katsomalla.

5.2 YAML-testaus

YAML päätettiin testata ennen sen lopullista käyttöönottoa metadatatamalleja varten. YAML-formaatin testiversio ladattiin koneelle marketplace.visualstudio.com -sivuston kautta ja verkosta löytyvien ohjeiden mukaisesti harjoiteltiin sen käyttöönottoa. Kuva 1. YAML-testaus. Haastatteluista tulleet tiedot helppokäyttöisyydestä osoittautuivat oikeaksi ja formaatin toimintaperiaatteet olivat helposti ymmärrettävissä. Metatietokenttiä oli helppo luoda ja YAML itsessään jo ilmoitti sattuneista virheistä kenttiä luodessa. Yhdeltä haastateltavalta saatu virheenkorjauksen apualusta, osoittautui hyödylliseksi ja sitä testattiin myös tässä testaustilanteessa. Kyseinen virheenkorjauksen apualusta löytyy marketplace.visualstudio.com-sivun hakukentän kautta, kirjoittamalla kenttään YAML.

Periaate YAML-tiedostoja luodessa toimi siten, että #-merkillä (risuaita) luotiin kentät, jolla ilmoitettiin mitä metatietoja juuri siihen kohtaan halutaan eritellä ja millä tavalla se tieto halutaan ilmoittaa. Kentän alle eriteltiin haluttu tieto. Esimerkiksi: # Date in ISO-format (YYYY-MM-DD) kertoi, että kenttä on pelin tai ottelun päivä ja se ilmoitetaan muodossa (vuosi-kuukausi-päivä). Kentän alle eroteltiin pelin tai ottelun päivä, mikä kyseisellä videolla esiintyi, date: 2021-05-16. Näin kenttien kanssa edettiin allekkain. Team names-kentän kohdalla alku toimi samalla tavalla, mutta joukkueita ollessa kaksi, ne eroteltiin allekkain. Välilyönnillä ja – merkillä, saatiin eroteltua joukkueet allekkain, kuitenkin niin että ne pysyivät halutun joukkuekentän tietojen sisällä.

Kuva 1. YAML-testaus.



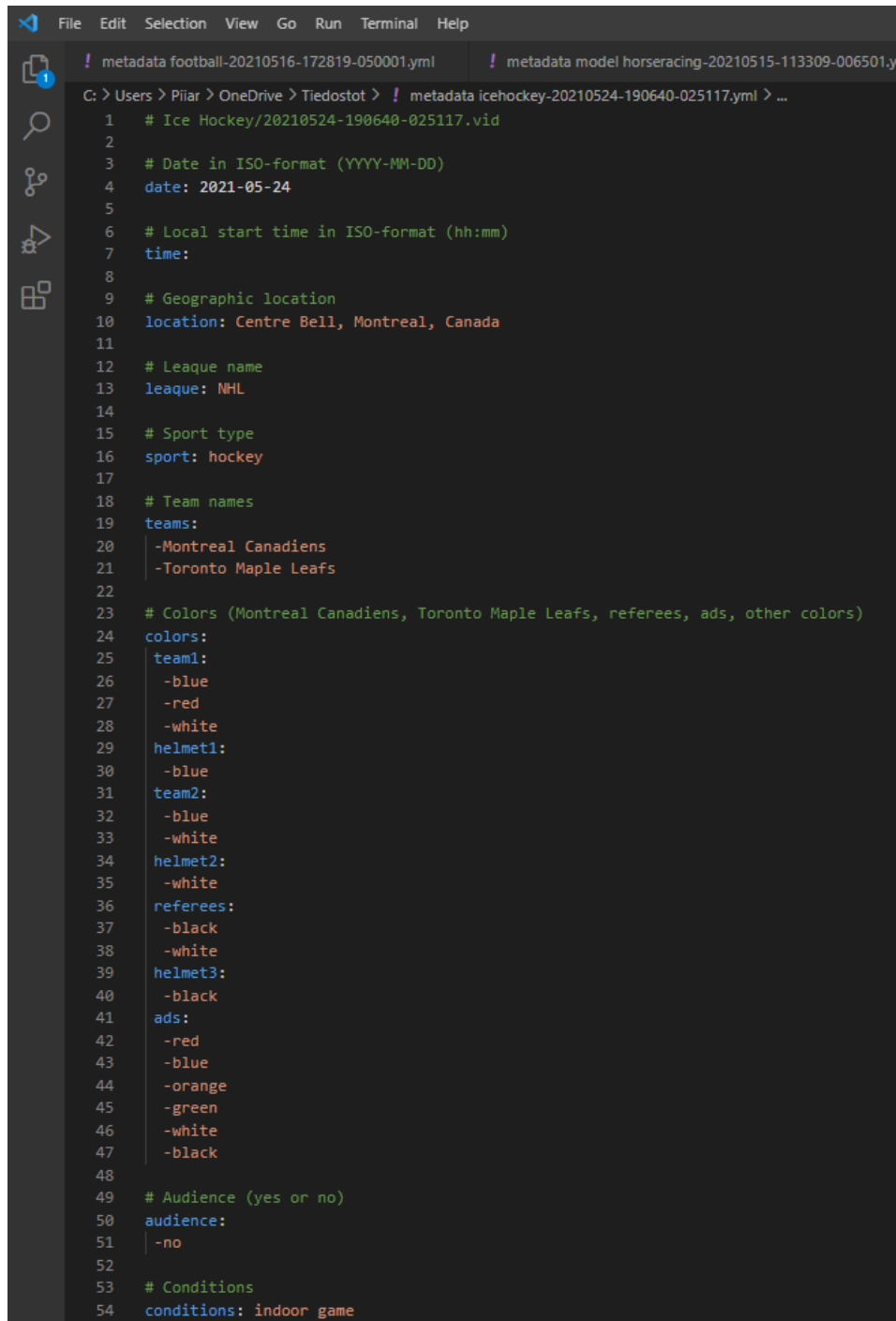
5.3 Metadatatmallit

Metadatatmallit päätettiin toteuttaa YAML-formaattia käyttäen. Aluksi hahmoteltiin alustava kaavio ja ehdotusta siihen, miten mallien runko voisi edetä ja miten metatiedot sinne määrittäisivät. Toimeksiantajalle esitellessä tätä alustavaa kaaviota ilmeni, että ajatus tekijällä itsellään oli ollut turhakin monimutkainen mallien suhteen. Ajatuksena ensimmäisessä ehdotuksessa oli, että mallissa olisi yksi pääkenttä, josta lähdettäisiin erottelemään välikenttien avulla, kohdennettuja tietoja. Toimeksiantajan kanssa tätä ehdotusta läpikäydessä selvisi, että metadatatmallit voisivat olla paljon yksinkertaisempia ja kentät niiden sisällä samanarvoisia. Vai pieni osa kohdennetuista metatiedoista vaatisi pienempiä välikenttiä. Mallien rungoissa kenttien pysyessä samanarvoisina, voitaisiin uusien ohjelmisto- ja tekoälyversioiden testaustilanteissa eri mallien tietoja vertailla helpommin keskenään. Lopputulokseksi päädyttiin selkeään ja johdonmukaiseen malliin, jossa eri metatiedot mallissa olisi mahdollisimman samanarvoisia.

Metadatatmalleja toteutettiin yhteensä kolmetoista erilaista. Eri urheilulajeja yrityksen datassa esiintyi yhteensä kymmenen, mutta jääkiekon, jalkapallon sekä koripallon osalta dataa yritykseltä löytyi eniten verrattuna muihin lajeihin. Näiden kolmen lajin kohdalla

malleja toteutettiin kaksi mallia kutakin lajia kohden, helpottamaan mallien sisällöllistä ymmärtämistä.

Kuva 2. Jääkiekkopelin metadatatamalli.



```

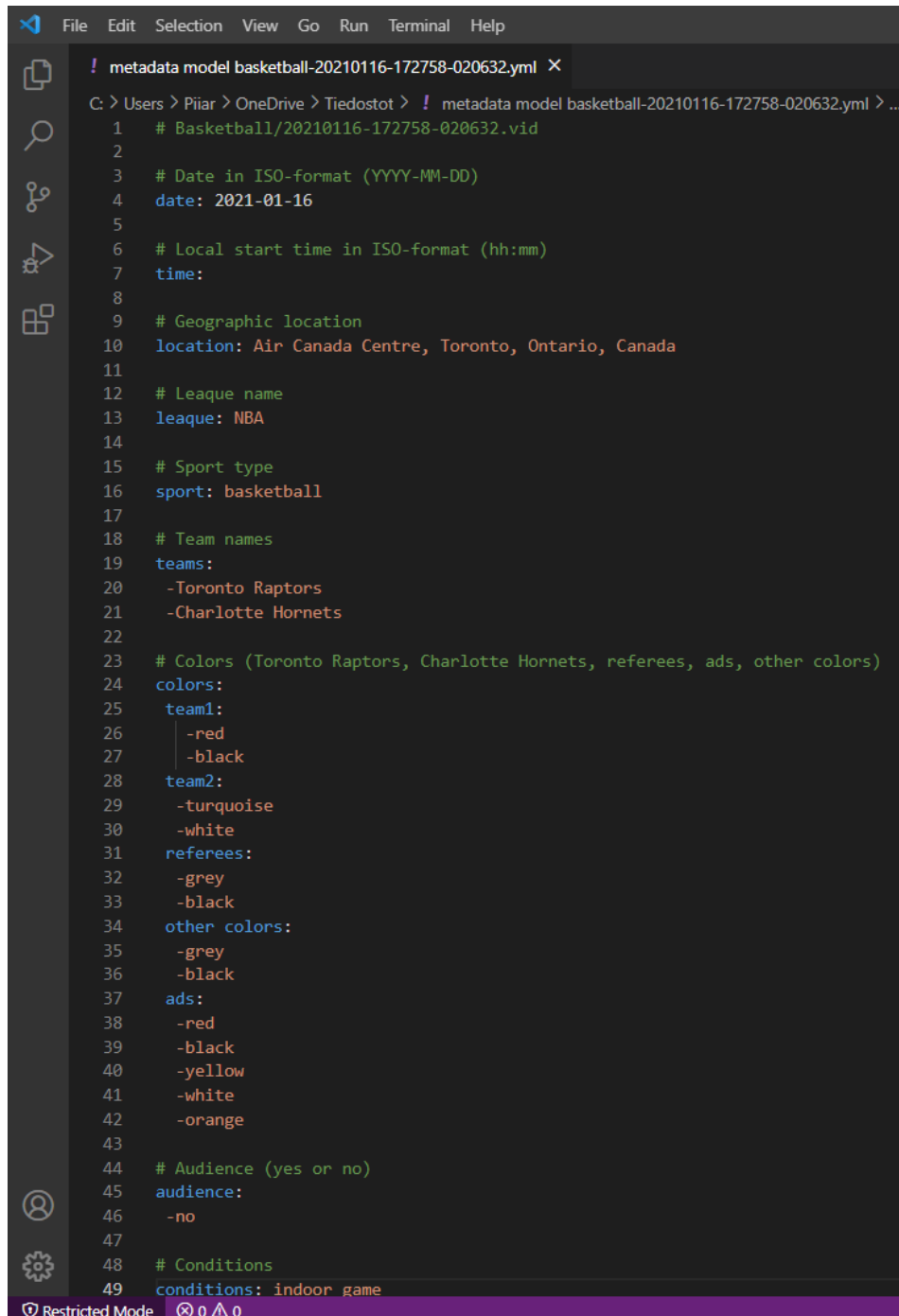
1  # Ice Hockey/20210524-190640-025117.vid
2
3  # Date in ISO-format (YYYY-MM-DD)
4  date: 2021-05-24
5
6  # Local start time in ISO-format (hh:mm)
7  time:
8
9  # Geographic location
10 location: Centre Bell, Montreal, Canada
11
12 # League name
13 league: NHL
14
15 # Sport type
16 sport: hockey
17
18 # Team names
19 teams:
20   -Montreal Canadiens
21   -Toronto Maple Leafs
22
23 # Colors (Montreal Canadiens, Toronto Maple Leafs, referees, ads, other colors)
24 colors:
25   team1:
26     -blue
27     -red
28     -white
29   helmet1:
30     -blue
31   team2:
32     -blue
33     -white
34   helmet2:
35     -white
36   referees:
37     -black
38     -white
39   helmet3:
40     -black
41   ads:
42     -red
43     -blue
44     -orange
45     -green
46     -white
47     -black
48
49 # Audience (yes or no)
50 audience:
51   -no
52
53 # Conditions
54 conditions: indoor game

```

Jääkiekkopelin metadatatamallissa, Kuva 2. Jääkiekkopelin metadatatamalli. kyseessä oli NHL-peli, vuodelta 2021 ja pelissä vastakkain pelasivat Montreal Canadiens – Toronto Maple

Leafs. Malli aloitettiin kentällä Ice Hockey ja perään lisättiin videon tallennustiedoista löydetty numerosarja. Tämä numerosarja yhdisti tiedoston oikeaan videoon.

Kuva 3. Koripallopelin metadatatamalli.

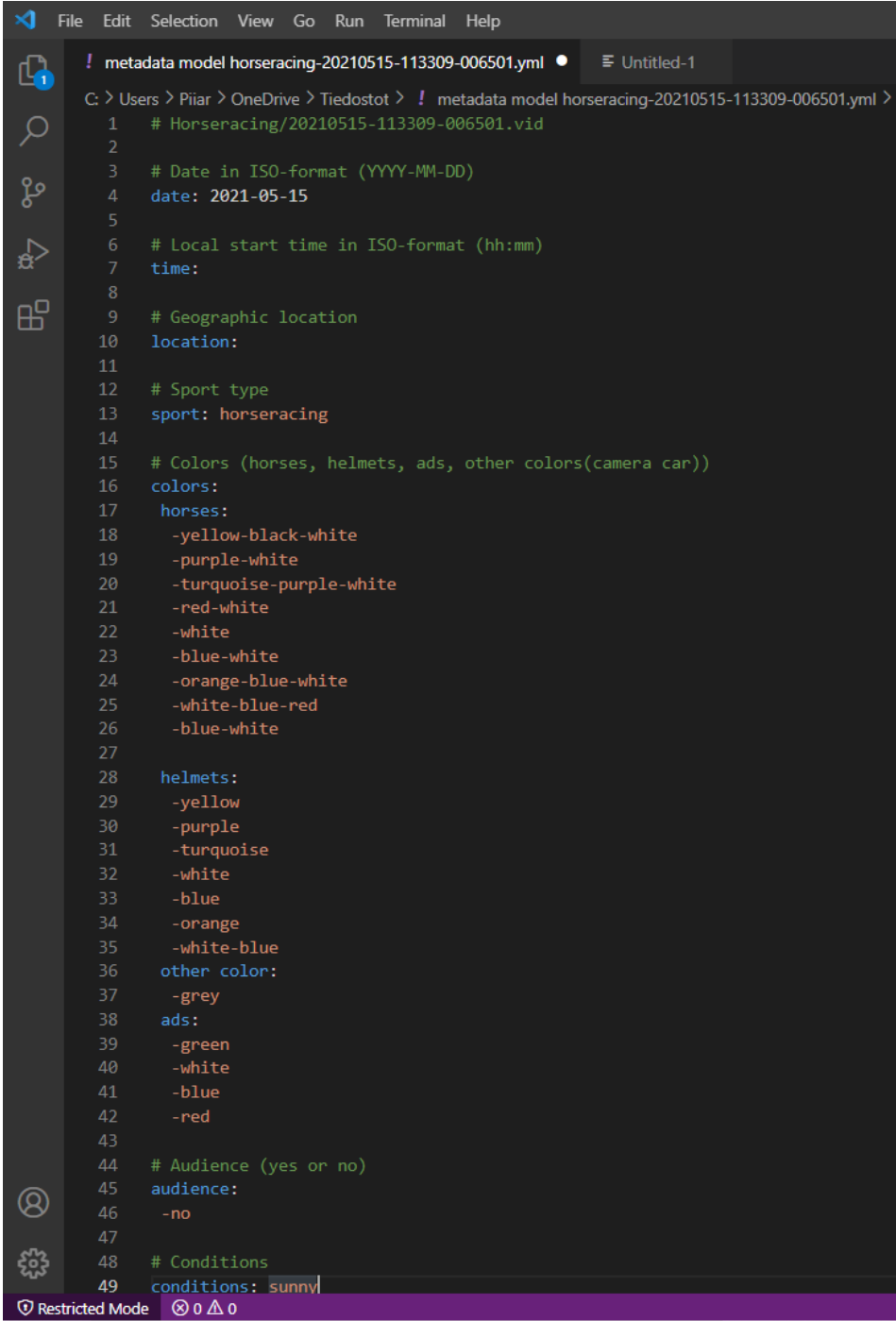


```

! metadata model basketball-20210116-172758-020632.yml X
C: > Users > Piir > OneDrive > Tiedostot > ! metadata model basketball-20210116-172758-020632.yml > ...
1  # Basketball/20210116-172758-020632.vid
2
3  # Date in ISO-format (YYYY-MM-DD)
4  date: 2021-01-16
5
6  # Local start time in ISO-format (hh:mm)
7  time:
8
9  # Geographic location
10 location: Air Canada Centre, Toronto, Ontario, Canada
11
12 # League name
13 league: NBA
14
15 # Sport type
16 sport: basketball
17
18 # Team names
19 teams:
20   -Toronto Raptors
21   -Charlotte Hornets
22
23 # Colors (Toronto Raptors, Charlotte Hornets, referees, ads, other colors)
24 colors:
25   team1:
26     -red
27     -black
28   team2:
29     -turquoise
30     -white
31   referees:
32     -grey
33     -black
34   other colors:
35     -grey
36     -black
37   ads:
38     -red
39     -black
40     -yellow
41     -white
42     -orange
43
44 # Audience (yes or no)
45 audience:
46   -no
47
48 # Conditions
49 conditions: indoor_game
  
```

Yllä oleva metadatatamalli rakentui koripallopelin pohjalta. Kuva 3. Koripallopelin metadatatamalli. Sitä verratessa jääkiekkopelin malliin, havaitsee selvästi, että rakenne malleissa on sama, vaikka tieto mallin sisällä muuttuukin.

Kuva 4. Hevosurheilun metadatatamalli.



```

1  # Horseracing/20210515-113309-006501.vid
2
3  # Date in ISO-format (YYYY-MM-DD)
4  date: 2021-05-15
5
6  # Local start time in ISO-format (hh:mm)
7  time:
8
9  # Geographic location
10 location:
11
12 # Sport type
13 sport: horseracing
14
15 # Colors (horses, helmets, ads, other colors(camera car))
16 colors:
17   horses:
18     -yellow-black-white
19     -purple-white
20     -turquoise-purple-white
21     -red-white
22     -white
23     -blue-white
24     -orange-blue-white
25     -white-blue-red
26     -blue-white
27
28   helmets:
29     -yellow
30     -purple
31     -turquoise
32     -white
33     -blue
34     -orange
35     -white-blue
36   other color:
37     -grey
38   ads:
39     -green
40     -white
41     -blue
42     -red
43
44 # Audience (yes or no)
45 audience:
46   -no
47
48 # Conditions
49 conditions: sunny
  
```

Yllä olevasta hevosurheilun metadatatamallista voidaan havaita, että se ei ole täysin verrattavissa jääkiekkopelin ja koripalloperin malleihin. Malli alkaa samalla tavalla kuin edellisestkin, mutta kaikkia kenttiä ei kyseiseen malliin tarvittu. Liiganimi puuttuu kokonaan, kuten myös joukkueiden nimet. Kentässä colors (värit), malliin yhdistettiin hevosesta löytyvät kilpailuvärit siten, että kyseessä oli hevoseen kiinnitetyn kilpailukankaan värit sekä ratsastajan asuissa esiintyvät värit. Kaikkien hevosten värit ilmoitettiin kentässä allekkain,

eikä niitä eritelty kuten aikaisimmissa malleissa, joukkueittain tai yksilöittäin. Ratsastajien kypärän värit taas eriteltiin oman otsikon alle, mutta nekin vain allekkain, eikä eroteltuna mihin hevoseen ne yhdistyvät. Kuva 4. Hevosurheilun metadatatamalli. Tämä erovaisuus edellisiin malleihin perustui siihen, että toimeksiantajalla ei ollut tarvetta erotella hevosia yksilöittäin. Malleja toteuttaessa havaittiin, että urheilulajit olisi mahdollista jakaa kahteen eri kategoriaan. Tiettyjen lajien kohdalla ei esimerkiksi otteluissa ole varsinaisia kotikenttiä tai erillisiä joukkueita, vaan näiden kohdalla kyseessä ovat kilpailijat, jotka ovat yksilöitä. Näissä lajeissa ei myöskään värit esiinny tiimeittäin tai käytetä minkäänlaisia liiganimityksiä.

5.4 Ohjeet metadatatamallien päivittämiseen

Tähän opinnäytetyöhön toivottiin erikseen metatietojen dokumentointia. Sen sijaan, että ohjeet olisi kirjattu jokaiseen lajiin erikseen, ymmärrettävämpää sekä selkeämpää oli jakaa lajit kahteen kategoriaan, joukkuelajit ja yksilölajit. Joukkuelajit pitävät sisällään jalkapallon, jääkiekon, koripallon, kriketin, rugbyn sekä pesäpallo ja poikkeuksena myös tenniksen. Yksilölajeihin luokituvat hevosurheilu ja koiraurheilu. Tennis luokitellaan todellisuudessa yksilölajiksi, mutta näiden metadatatamallien rakenteen mukaisesti, se sopi paremmin joukkuelajeihin ja on siitä syystä poikkeuksena siinä kategoriassa.

Joukkuelajien malleissa, Taulukko 2. Ohjeet joukkuelajien metadatatamalleille. metatiedot koostuvat alkamispäivästä, alkamisajasta, maantieteellisestä sijainnista, liiganimestä, urheilutyypistä, joukkueiden tai kilpailijoiden nimistä, väreistä, tuomarit, mainokset, muut yksityiskohdat, muut värit, yleisön merkitsemisestä ja olosuhteista. Joukkueet tai kilpailijat, tuomarit, mainokset, muut mahdolliset yksityiskohdat, eriteltynä malleihin allekkain.

Yksilölajien malleissa, Taulukko 3. Ohjeet yksilölajien metadatatamalleille. metatiedot koostuvat alkamispäivästä, alkamisajasta, maantieteellisestä sijainnista, urheilutyypistä, väreistä, yleisön merkitsemisestä ja olosuhteista. Ottelijat tai kilpailijat, ilman erittelyä. Malleja toteuttaessa on hyvä kuitenkin huomioida, että jokainen laji pitää sisällään toisistaan poikkeavia yksityiskohtia, mitä on syytä tarkistella ja mallintaa suoraan valmiisiin lajikohtaisiin malleihin.

Taulukko 2. Ohjeet joukkuelajien metadatatamalleille.

Mikä urheilulaji on kyseessä sekä numerosarja mihin videoon se yhdistetään
Football/20211605_FCB_vs_CEL_CAM1-20210516-172819-050001.vid
Date - Tapahtuman päiväys (vuosi-kuukausi-päivä)
Date in ISO-format (YYYY-MM-DD)
esimerkki date: 2021-05-16
Local start time - Tapahtuman alkamisaika (tunnit: minuutit)
Local start time in ISO-format (hh:mm)
esimerkki: time: 18:30
Geographic location - Tapahtuman sijainti (kentän nimi, kaupunki, maa)
Geographic location
esimerkki: location: Camp Nou, Barcelona, Spain
League name - Liigan nimi (jos kyseessä liiga peli)
League name
esimerkki: league: La Liga
Sport type - Urheilulaji
Sport type
esimerkki: sport: soccer
Teams names - Joukkueiden nimet (allekkain merkittynä)
Team names
esimerkki: teams:
FCB
Celta
Colors -Värit eroteltuna (joukkueet, tuomarit, mainokset, muut)
Colors
esimerkki: colors:
team1:
blue
red
yellow
team2:
light-blue
white
referees:
pink
black
ball:
yellow
ads:
red
Audience - Onko yleisöä paikalla
Audience
esimerkki: audience:
no
Conditions - Olosuhde tapahtumassa
Conditions
esimerkki: conditions: sunny

Taulukko 3. Ohjeet yksilölajien metadatatamalleille.

Mikä urheilulaji on kyseessä sekä numerosarja mihin videoon se yhdistetään
Greyhound/20210303-210133-005501.vid
Date - Tapahtuman päiväys (vuosi-kuukausi-päivä)
Date in ISO-format (YYYY-MM-DD)
esimerkki: date: 2021-03-04
Local start time - Tapahtuman alkamisaika (tunnit:minuutit)
Local start time in ISO-format (hh:mm)
esimerkki: time: 12:15
Geographic location -Tapahtuman sijainti (kentän/paikan nimi,kaupunki, maa)
Geographic location
esimerkki: location: Calgary, Canada
Sport type - Urheilulaji
Sport type
esimerkki: sport: Greyhound
Colors - Värit eroteltuna (kilpailijat, muut merkittävät yksityiskohdat, muut värit)
Colors (dogs, ads, other colors)
esimerkki: colors
dogs:
white
red
pink
green-pink
black-white
red-white-blue
yellow-black
other colors:
red
white
Audience - Onko yleisöä paikalla
Audience (yes or no)
esimerkki: audience:
no
Conditions - Olosuhde tapahtumassa
Conditions
esimerkki: conditions: cloudy

6 Kestävän kehityksen näkökulma

Kestävä kehitys esiintyy yrityksissä päätöksentekona ja käytännön toimina. Näiden kautta pyritään kestävämpään liiketoimintaan, ottaen huomioon yhteiskunta ja ympäristö. (Palosaari, 2020) Ilmastonmuutoksen hillinnässä tärkeitä keinoja teknologian näkökulmasta ovat energian säästäminen, energiatehokkuuden lisääminen, hiilijalanjälki sekä elinkaariajattelu (Ympäristöministeriö, 2014).

Teknologian kehittyessä digitalisaatio lisääntyy ja mahdollistaa parempia ja kustannustehokkaampia tuotteita. Kun elinkaariajattelu otetaan mukaan teknologian kehitykseen, tuotannon ja lopputuotteiden energiankulutukseen, eri tuotantotapojen ja materiaalien alkuperään sekä tarkoituksenmukaiseen käyttöön, voidaan ympäristövaikutuksia vähentää. Kun ympäristövaikutukset huomioidaan jo suunnittelussa kokonaisvaltaisesti ja valitaan ympäristöä vähemmän kuormittavat ratkaisut, voidaan saavuttaa kustannussäästöjä esimerkiksi jätteiden määrän tai materiaalin kulutuksen vähentyessä ja näin voidaan hillitä myös jätteiden turhaa syntymistä. (Teknologiateollisuus ry, 2012, s. 10)

Tämän opinnäytetyön toimeksiantajan kehrittelemää teknologiaa hyödynnetään siten, että eri tapahtumien fyysiset mainostilat voidaan korvata lokalisoidulla sisällöllä. Tämä saattaa vähentää päästöjä, vaikka tarkkoja laskelmia ei asiasta toteutettu. Olettamuksena todettakoon, että kun mainoksia ei tarvitse painattaa, kuljettaa ja vaihdattaa tapahtumien sisäiseen mainontaa, vaan tuo voidaan toteuttaa teknologian avulla, saattaa päästöjä syntyä vähemmän. Yrityksen teknologiaa voidaan myös hyödyntää monipuolisesti eri tapahtumissa, peleissä ja kilpailuissa, eikä erilasia ohjelmisto- ja tekoälyversioita tarvitse luoda jokaiseen otteluun erikseen.

Mikäli jokin muu yritys tuottaisi vastaavanlaista mainosten virtualisointia ilman tekoälyä, se vaatisi jonkin muun menetelmän pohjaksi. Tyypillisesti tällaiset menetelmät vaatisivat, että urheilutapahtumapaikoille olisi asennettu tähän käyttöön soveltuvia ratkaisuja, kuten LED-mainosaitoja. Lisäksi tapahtumissa käytetyt TV-kamerat tulisi varustaa jonkinlaisella erikoisoptiikalla, joka kykenisi näkemään edellä mainitut LED-mainosaidat esimerkiksi infrapunataajuuksina. Näiden erikoisvarusteiden tuottaminen on erittäin kallista ja

materiaalimielessä monimutkaisia. Nämä menetelmät myös vaatisivat suuria määriä logistiikkaa ja fyysisten varusteiden siirtoa paikasta toiseen. Se taas aiheuttaa matkustamista varusteiden käyttäjille sekä kuljetuksia varusteiden siirtoon. Supponorin kehittämä tekoäly mahdollistaa etätuotannon, eikä vaadi fyysistä läsnäoloa. Tekoäly mahdollistaa sen, ettei edellä mainittuja erikoisvarusteita enää tarvita, joten tämän teknologian avulla säästetään matkustamisen hiilijalanjäljessä.

Tässä opinnäytetyössä toiminnallisesti toteutetut metadatatallit edistävät energian säästämistä. Tarkennetut ja oikein kohdennetut metadatatallit auttavat yritystä kohdistamaan käytettyä tietokoneiden laskenta-aikaa paremmin, jolloin voidaan säästää sähköä. Keskittämällä vihreän sähkön käyttöön, eli sähköön, joka on tuotettu uusiutuvilla energianlähteillä, voitaisiin tukea kestävämpää energiantuotantoa.

Jatkotoimenpide-ehdotuksia on kirjattu tämän opinnäytetyön viimeiseen lukuun. Mikäli datan automatisointia saataisiin kehitettyä edes puoliautomaattitasolle, josta rajapintojen avulla tiettyjä tietoja voidaan kerätä automaattisesti, se vähentäisi entisestään tietokoneiden laskenta-aikaa sekä resursseja.

7 Työn tulokset ja johtopäätökset

Työn tuloksiksi saavutettiin tutkimus- ja kehityskeskus Supponorille, toimiva ratkaisu suuren datamäärän kategorisointiin, metadatatallien avulla. Tulokset saavutettiin etsimällä tietoa verkosta, henkilöhaastattelujen avulla, toimeksiantajan kanssa tiiviillä yhteydenpidolla sekä yrityksen teknologiaan ja dataan tarkoin perehtymällä.

Tärkeimpänä osa-alueena oli oikein kohdennettujen metatietojen määrittäminen yrityksen tarpeisiin. Metatiedot määritettiin yrityksen dataan ja teknologiaan sopivalla tavalla, ottaen huomioon eri joukkue- ja yksilölajit sekä näiden lajien yksityiskohtaiset tiedot. Yritys käyttää tällä hetkellä teknologiaansa yhdeksän eri urheilulajin kuvaamiseen ja jokaisen lajin kohdalla videomateriaalit katsottiin tarkasti läpi sekä toteutettiin niistä lajikohtaiset metatiedot, järjestelmällisesti samaa runkomallia noudattaen. Kymmenentenä lajina yritys testaa teknologiaansa uuteen lajiin ja vaikka lajia ei tässä vaiheessa vielä julkisesti mainita, tälle lajille toteutettiin myös metatietojen määrittäminen sekä lopullinen malli. Toimeksiantajan

kanssa keskusteltiin useaan otteeseen siitä, minkälaisia virheitä teknologian käytön yhteydessä toisinaan esiintyy ja näiden tietojen pohjalta metatiedot määritettiin tarpeisiin sopiviksi.

YAML- ja JSON-formaatteihin perehdyttiin tarkasti. Nämä kaksi formaattia tutkittiin tarkasti ja selvitettiin niiden erovaisuudet. Näitä vertailtiin teoriapohjaisesti, haastateltiin ohjelmoinnin parissa työskenteleviä henkilöitä ja henkilöitä kenellä omassa työssään oli konkreettisesti omakohtaisia kokemuksia kyseisistä formaateista. Lisäksi suoritettiin YAML-merkintäkielelle testaus ja näiden tietojen pohjalta valittiin YAML.

YAML-merkintäkielen avulla toteutettiin valmiit metadatatmallit, jokaiselle urheilulajille erikseen. Kolmen lajin kohdalla, malleja toteutettiin kaksi mallia lajia kohden. Tuloksena saatiin yhteensä kolmesta mallia. Mallit toteutettiin johdonmukaisesti samaa rakennetta käyttäen, kuitenkin siten, että ne sisälsivät lajikohtaiset yksityiskohdat. Kokonaisuutta harkittiin tarkoin ja yrityksen dataa käytiin läpi useaan kertaan.

Mallien lisäksi, dokumentointiin ohjeet metadatatmallien päivittämiseen. Ohjeet toteutettiin selkeästi ja ne eroteltiin kahteen kategoriaan, joukkue- ja yksilölajit. Ohjeet dokumentointiin siten, että ohjeita lukiessa ja niitä valmiisiin malleihin mallintaessa, on helppo ymmärtää koko toimintaperiaate. Ohjeet kirjattiin taulukkomaisesti, siten että ne järjestelmällisesti seuraavat valmiita malleja ja tekstit näihin taulukoihin luotiin valmiiden mallien tekstivärien mukaisesti. Kentät esiintyvät vihreällä värillä ja tarkennetut tiedot näiden kenttien sisällä sinisellä värillä. Taulukkomaiseen dokumentaatioon kirjattiin ohjeet samoja värejä käyttäen. Myös kaikki muut yksityiskohdat, mitkä malleissakin esiintyvät tietyillä väreillä, kirjattiin ohjeisiin samaa väriä käyttäen.

Tuloksia tarkastellessa opinnäytetyöprosessin päätteeksi, voidaan johtopäätöksenä todeta, että saavutettiin toivottu kokonaisuus. Tutkimusten- ja haastattelujen tulokset vastasivat kaikkiin tutkimusongelmiin ja niiden pohjalta toiminnallinen osuus saavutettiin kokonaisuudessaan. Yritykselle tuotettiin heidän tarkoitukseensa sopivat metatietojen ominaisuudet ja ne koottiin valmiiksi metadatatmalleiksi. Huomioon otettiin yrityksen teknologia sekä monipuolisuus eri urheilulajien välillä, joten kohdennettiin metatiedot tarkasti lajeittain. Voidaan myöskin olettaa, että tulevaisuudessa yrityksen teknologiaa

tullaan käyttämään vieläkin monipuolisemmin eri urheilulajeissa, joten ohjeisiin kirjatut tiedot, on siitä syystä jaettu niin joukkuelajeja kuin yksilölajeja ajatellen. Metadatamalleja tarkistellessa voidaan todeta, että alustan valinta näille malleille on onnistunut, sillä yksinkertaisempaa valintaa ei olisi voitu löytää. Näin toimeksiantajan toivomusten mukaisesti, olemassa olevan datan läpikäyminen onnistuu kesätyöntekijältä tai esimerkiksi harjoittelijalta, vain ohjeita seuraten. Myös tulevan datan talteenotossa on helppoa kategorisoida dataa, vain mallintamalla olemassa oleviin ja seuraamalla ohjeita järjestelmällisesti. Vaikka itse mallit ovat yksinkertaisia, niitä voidaan hyödyntää tulevaisuudessa monin eri tavoin. YAML-formaatti toimii siten, että sen voi tarvittaessa helposti muuttaa eri ohjelmointikielille tai JSON-formaatti muotoon. Tästä kokonaisuudesta on myös toimeksiantajalle välitöntä hyötyä ja suuria datamääriä ei tarvitse jatkossa säilyttää. Talteen otetaan vaan monimuotoista dataa, sellaista missä lajikohtaiset tiedot poikkeavat jo tallennetuista.

Kestävä kehitys ja vastuullisuus on suuressa roolissa teknologian kehittyessä. Mitä enemmän eri teknologia-alan yritykset, siinä missä jokainen muukin yritys, ottaa elinkaariajattelun ja energiatehokkuuden huomioon, saavutetaan pienillä teoilla pitkällä tähtäimellä paljon. Opinnäytetyöhön koottiin omana lukunaan kestävän kehityksen näkökulmat, tämän opinnäytetyön sisällön mukaisesti sekä myös toimeksiantajan teknologiaa ajatellen. Tuo kyseinen luku toteutettiin laajasti asiaan perehtymällä. Vastuullisuus yrityksissä koostuu siitä, että pyritään parhaisiin mahdollisiin toimintaperiaatteisiin, saavuttaaksemme ilmastonmuutoksen hillintään liittyviä tavoitteita yhdessä.

Opinnäytetyöprosessissa onnistuttiin toivomalla tavalla, vaikka aihe ei ollut ennestään tekijälle tuttu. Prosessi eteni johdonmukaisesti ja tärkeää tietoa haastateltavilta saatiin kerättyä paljon. Eri henkilöiden haastattelujen välillä oli toisinaan kesälomista johtuen pitkiäkin taukoja, joten oli tärkeää maltaa odottaa haastattelut kaikilta kutsutuilta, ennen kuin lopullisia johtopäätöksiä tutkimusteemoihin liittyen teki. Opinnäytetyötä toteutettiin opintojen ohella ja asiaa pohtiessa jälkeenpäin, prosessia olisi helpottanut vain yksin siihen keskittyminen. Uutta tietoa tuli jatkuvasti niin paljon lisää ja koska toisinaan oli keskityttävä vain opintojen sisältöihin, oli vaikeaa koota ajatukset tähän prosessiin uudelleen. Tästä johtuen materiaaleja ja haastattelujen tuloksia oli kerrattava useasti ja tämä vei myös aikaa prosessista.

Mikäli opinnäytetyöprosessin voisi toteuttaa uudelleen alusta, muuttuisi toiminnassa muutama kohta. Formaatti metadatatmallien toteuttamiselle valittaisiin suoraan sen perusteella, mikä katsottaisiin yksinkertaisemmaksi, helppolukuisimmaksi ja lisäksi se tukisi useita ohjelmointikieliä. Opinnäytetyöprosessissa käytettiin paljon aikaa formaatin valintaan ja myöhemmin vasta aukesi ymmärrys siihen, miten helposti formaatteja voidaan vaihtaa keskenään. Toteuttaessa metadatatmallit YAML-merkintäkielen avulla, mallit ovat yksinkertaisempia kuin JSON-formaattia käytettäessä, mutta todellisuudessa tuo muutos näiden kahden välillä, on vain muutaman painalluksen päässä. Formaatteja ja merkintäkieliä voidaan vaihtaa monen valmiin ohjelman avulla.

Toimeksiantajan tuesta ja avusta oli suuri apu koko opinnäytetyöprosessin ajan. Supponorin henkilökunta oli tavoitettavissa aina tarvittaessa ja monia yksityiskohtia pohdittiin yhdessä. Prosessin lopuksi todettakoon, että opinnäytetyön aihe tukee työelämän tarpeita. Koko prosessi toi tullessaan niin valtavan määrän uutta tietoa, jota työelämässä voi hyödyntää. Prosessi toi ymmärryksen metatietojen määrittämisen tarpeellisuudesta, miten metadatatmalleja luodaan sekä vertailuja eri merkintäkielten välillä. Konenäkö ja tekoäly on tulevaisuutta ja haastattelujen ansiosta näihin saavutettiin uudenlaista mielenkiintoa. Työelämässä kaikesta tämän opinnäytetyöprosessin tuottamista tiedoista, on varmasti hyötyä.

Toimeksiantajan palaute:

Tilaaajan tavoitteina työlle olivat tekoälykäyttöön tarkoitettujen metadatatmallien kartoitus ja esimerkkimallien toteutus tilaaajan toimittamaan esimerkivideoaineistoon perustuen. Tilaaajan asettamat tavoitteet saavutettiin hyvin, sekä työn valmistelun, että toteutuksen näkökulmasta. Tilaaaja voi hyödyntää opinnäytetyön tuloksia monimuotoisesti. Työn tuottamat esimerkkimallit toimivat pohjana vastaavien mallien jatkokehitykselle eri urheilulajien kontekstissa. Lisäksi työn dokumentoimat lähestymistavat kerätyn videoaineiston jatkuvaan metadatatmallintamiseen mahdollistavat tilaaajan aiheeseen liittyvien sisäisten prosessien jatkokehittämisen. Opinnäytetyön tekijä on osoittanut erinomaisia kommunikointitaitoja ja oma-aloitteisuutta työn suunnittelussa ja toteutuksessa. Tilaaajan näkökulmasta työ edistyi joustavasti ja tehokkaasti, tuottaen suunnitellun lopputuloksen.

8 Pohdinta ja jatkotoimenpiteet

Kuten jo tämän opinnäytetyön johdanto osuudessa todettiin, on metatiedoilla suuri merkitys yritysten datan hyödyntämisessä. Tärkeintä on määrittää metatiedot oikein ja kohdennetusti omiin tarpeisiin nähden. Toimeksiantajan dataan tutustuessa ja siitä kerättyjä metatietoja määrittäessä, heräsi kuitenkin muutamia huomioita. Värit olivat yksi keskeinen metatieto, mitä toimeksiantaja omasta datastaan tarvitsi ohjelmisto- ja tekoälyversioiden testaustilanteisiin. Dataa tutkiessa ja metatietoja määrittäessä, heräsikin huomio muutamasta yksityiskohdasta. Jalkapallopelien osalta, dataa ei ollut koko peliltä, joten datasta puuttui maalivahtien näkyvyys. Maalivahdeilla kuitenkin on eriväriset peliasut, kuin muilla pelaajilla, jotta ne erottuva selvästi kentältä. Metadatatamalleihin voisi eritellä jatkossa myös maalivahtien peliasut erikseen. Jääkiekkopelien osalta, malleihin kirjattiin olosuhteet, vaikka voidaan olettaa pelien tapahtuvan aina sisällä hallissa. NHL-pelejä pelataan kuitenkin myös talviklassikkona ja silloin pelit pelataan ulkona. Tämä on hyvä huomioida malleja tehdessä jatkossa. Data, mitä opinnäytetyössä käytiin läpi, oli pitkälti lähivuosien dataa. Koronarajoituksista johtuen, harvoissa tapahtumissa esiintyi yleisöä. Malleja saattaakin olla tarpeellista kohdentaa yleisönosalta paremmin, sillä pääasiassa peleissä on paljon yleisöä, erivärisissä asuissa.

Pohdintoja herätti myös toimeksiantajan videodatan pituudet yleisesti. Vaikka tässä opinnäytetyössä tutkittiin leikattua dataa, oletetaan että yrityksellä on kaikkien tapahtumien ja pelien videot tallessa kokonaisuudessa. Ajatellaan, että kyseessä on esimerkiksi jalkapallopelejä, jonka kesto on kokonaisuudessaan 90 minuuttia ja välissä on 15 minuutin tauko. Peliä kuvataan siis oletetusti 105 minuuttia. Ohjelmisto- ja tekoälyversioiden testauksiin ei kuitenkaan tarvita koko peliä alusta loppuun, vaan ainoastaan olennaisia kohtia, jotka on määritetty metatiedoissa. Videot voisikin leikata siten, että talteen tallennetaan vain se kyseisen osa, mikä on oleellista. Tällöin tallessa oleva datamäärä pienenee huomattavasti.

Opinnäytetyöhön oli helppoa löytää yleistä tietoa datan hyödyntämisestä yrityksissä ja sen tärkeydestä monien eri artikkelien ja uutisten kautta, mutta etsiessä tietoa kokonaisvaltaiseen ratkaisuun, datan jäsentelyjen toimiin ja miten se konkreettisesti tapahtuu, tietoa ei juurikaan löytynyt. Verkon hakukentistä esiin nousi vain yrityksiä, jotka

tarjoavat ratkaisuja näihin ongelmiin. Tämä herätti pohdintoja siihen, kuinka monella pienyrittäjällä todellisuudessa on mahdollisuuksia tai resursseja oman datansa kategorisointiin. Teknologia kehittyy kovemmin kuin koskaan ihmiskuntamme historiassa, mutta kuinka moni siinä kehityksessä ehtii pysyä mukana. On siis hyvin ymmärrettävää, että monella yrityksellä on haasteita oman datansa kanssa, sillä sen kategorisointi on työläs prosessi, vaatii datan huolellista läpikäymistä, johdonmukaista strategiaa edetä tuossa prosessissa tai vaihtoehtoisesti maksaa tästä palvelusta ulkopuoliselle yritykselle. Monikaan yritys taas ei mielellään maksa palveluista, jos ei ole tarkoin selvillä sen hyötyjä.

Opinnäytetyön edetessä heräsi myös ajatuksia siitä, kuinka moni henkilö tai yrittäjä, edes tietää mitä metatieto on. Mietteitä heräsi myös dataan liittyen, onko kaikille varmasti selvää se, kuinka paljon kaikkien ihmisten toimista dataa pystytään tänä päivänä keräämään ja millaisia digijalanjälkiä taakse jokainen jättää huomaamattaan. Alalla työskenteleville termit sekä niiden hyödyt ja haitat ovat tuttuja, mutta teknologian edetessä näin kovalla vauhdilla, monelta saattaa mennä paljon hyödyllistä tietoa ohitse.

Jatkotoimenpide-ehdotuksia koottiin henkilöhaastattelujen pohjalta. Tietoa kerättiin henkilöiltä, jotka työskentelevät automaation tai konenäön parissa sekä yrityksiltä, jotka itse hyödyntävät näitä toimia datan käsittelyssä. Tähän opinnäytetyöhän koottu kokonaisuus mahdollistaa välittömästi yrityksen testitilanteisiin helpotusta, mutta pidemmällä tähtäimellä se on vasta jäävuoren huippu. Teknologian kehitys mahdollistaa monia uusia toimintoja, millä datan kategorisointia voidaan nopeuttaa ja toteuttaa ilman työlästä käsin tehtävää prosessia. Jotta yritys saavuttaisi ihanteellisen ja toimivan lopputuloksen tulevan datan keräämiseen, tallentamiseen sekä kategorisoimiseen, on avuksi otettava koneoppiminen, konenäkö ja automatisoitu tiedon tuottaminen.

Konenäön avulla on mahdollista analysoida kertynyttä dataa reaaliajassa. Koneoppiminen mahdollistaa esimerkiksi videokuvan analysoimisen siten, että se seuraa liikkeitä tai ilmeitä, joiden seurauksena se kykenee tekemään ratkaisuja tai ennustamaan jotakin toivottua asiaa. Supponor on kehitellyt oman teknologian, ohjelmiston ja tekoälyn, jotka mahdollistavat virtuaalimainokset suorissa tv-lähetyksissä, voisikin ajatella, että näihin yhdistämällä oikeanlaisen sovelluksen teknologia kykenisi helposti tarkistelemaan tallennettua dataa. Kun se havaitsisi jotakin poikkeavaa, kuten erilaisia sääolosuhteita tapahtumissa, uusia määrittelemättömiä metatietoja tai muita mahdollisia poikkeavuuksia, se ilmoittaisi tästä

havainnosta ja merkitsisi kyseisen datan. Se mahdollistaisi sen, että aina siirtyy monimuotoista dataa talteen ja mikäli sovellus havaitsisi jotakin itselle tuttua, se siirtäisi sen datan syrjään. On kovin työlästä käydä suuria datamääriä läpi tai aina uuden datan kohdalla tarkastella sisältääkö se tietoja, mitkä poikkeavat jo talteen otetusta. Konenäön avulla saavutetaan tämä prosessi ainakin osittain automaattisesti.

Erilaisten rajapintojen hyödyntäminen taas mahdollistaa olemassa olevan tiedon saatavuuden. Rajapinta on määritelmä, jonka avulla eri ohjelmat voivat vaihtaa tietoja keskenään ja ns. keskustella keskenään. (Kivekäs, 2014) Rajapintojen avulla tämän opinnäytetyön toimeksiantaja voisi etsiä datan kategorisointiin monia tietoja automaattisesti. Metadatatamalleja toteuttaessa tässä opinnäytetyössä, esiin nousi, kuinka työläs prosessi oli. Mikäli videon tallennustiedoista ei löytynyt jonkin tapahtuman alkamisaikaa tai joukkueiden virallisia nimiä, niitä etsittiin hakemalla verkosta eri joukkueita ja etsimällä verkosta kyseistä peliä, jotta tietoja voisi löytää. Useinkaan metadatatamallien toteuttajan ei oleteta olevan juuri sisällöllisen aiheen asiantuntija, joten vaatii paljon etsiväntaitoja saada oikeat tiedot löydettyä jälkikäteen. Tosin, datan voisi tallentaakin aina sisällöllisten tietojen mukaisesti mutta rajapintoja hyödyntämällä kyseiset tiedot saisi kerättyä automaattisesti, sillä NHL-pelit sekä monet muut useat tapahtumat ovat löydettävissä verkosta. Se edellyttäisi rajapinnan avoimena oloa, jotta niitä voisi hyödyntää ja niiden tuottamaa tietoa.

Haastattelujen perusteella kerätyn yhteenvedon perusteella tärkeimpänä asiana pidettiin yrityksen kannalta sitä, että yrityksen videodatasta on löydetty ja määritetty juuri oikea ja hyödyllisin tieto. Katsottiin siis, että datan oikeanlainen kategorisointi on kaiken perusta ja vaikka konenäköä ja rajapintoja hyödynnettäisiin tulevaisuudessa, on datan käsittelyn suhteen liikkeelle lähdettävä aina ensimmäiseksi käsityöstä.

Lähteet

Alasuutari, P. (2011). *Laadullinen tutkimus 2.0*.

<https://www.ellibslibrary.com/reader/9789517685030>

Andersson, M. (10.12.2019). Löydä ulkoisen datan parhaat apajat. *Advian*.

<https://www.advian.fi/blogi/ulkoisen-datan-parhaat-apajat>

Bertram, A. (14.7.2021). The state of config file formats: XML vs. YAML vs. JSON vs. HCL.

<https://octopus.com/blog/state-of-config-file-formats>

Dufva, M. (29.7.2020). Teknologia tuo koronanälkeiseen aikaan paljon mahdollisuuksia – ja muutamia uhkia. *Sitra*. [https://www.sitra.fi/artikkelit/teknologia-tuo-](https://www.sitra.fi/artikkelit/teknologia-tuo-koronanjälkeiseen-aikaan-paljon-mahdollisuuksia-ja-muutamia-uhkia/)

[koronanjälkeiseen-aikaan-paljon-mahdollisuuksia-ja-muutamia-uhkia/](https://www.sitra.fi/artikkelit/teknologia-tuo-koronanjälkeiseen-aikaan-paljon-mahdollisuuksia-ja-muutamia-uhkia/)

Euroopan parlamentti. (29.3.2021). Mitä tekoäly on ja mihin sitä käytetään?

<https://www.europarl.europa.eu/news/fi/headlines/society/20200827STO85804/mit-a-tekoaly-on-ja-mihin-sita-kaytetaan>

Fileinfo. (2015). YAML tiedostopääte. <https://tiedosto.info/extension/yaml.html>

Forsström, P.-L. (23.9.2019). Metatiedot ovat viesti maailmalle. *Vastuullinen tiede*.

<https://vastuullinentiede.fi/fi/jatkokaytto/metatiedot-ovat-vesti-maailmalle>

Harju, K. (18.3.2019). Mitä metatieto on? *Suomen ympäristökeskus SYKE*.

<https://blog.edu.turku.fi/karttanyt/kartat/paikkatieto/metatieto/>

Heikkinen, S. & Nieminen, J. (17.3.2021). Tekoäly muuttaa maailman – pian se tekee jopa lääkärin ja juristin töitä. *Yle*. [https://yle.fi/aihe/artikkeli/2017/06/04/tekoaly-](https://yle.fi/aihe/artikkeli/2017/06/04/tekoaly-muuttaa-maailman-pian-se-tekee-jopa-laakaran-ja-juristin-toita)

[muuttaa-maailman-pian-se-tekee-jopa-laakaran-ja-juristin-toita](https://yle.fi/aihe/artikkeli/2017/06/04/tekoaly-muuttaa-maailman-pian-se-tekee-jopa-laakaran-ja-juristin-toita)

Hirsjärvi, S. & Hurme, H. (2001 s.48). *Tutkimushaastattelu -Teemahaastattelun teoria ja käytäntö*. Yliopistopaino Helsinki.

Hirvonen, T. (2015). *Tutkimusdatan hallinnan kehittäminen ja datan laadun arviointi asiantuntijayrityksessä*. [pro gradu -tutkielma, Tampereen teknillinen yliopisto].

<https://core.ac.uk/download/pdf/250163284.pdf>

Hyvärinen, M., Nikander, P., Ruusuvaori, J. & Aho, A.-L. (2017). *Tutkimushaastattelun käsikirja*. <https://www.ellibslibrary.com/reader/9789517686112>

Kivekäs, O. (16.6.2014). Mikä on avoin rajapinta? <http://otsokivekas.fi/2014/06/avoin-rajapinta/>

Kleemola, H. (2020). Metadatan hallinta on avain tiedon maksimaalisen arvon realisointiin.

Tivi. <https://www.tivi.fi/kumppanisisaltoa/tivi-studiovieras/metadatan-hallinta-on-avain-tiedon-maksimaalisen-arvon-realisointiin/a2887957-de71-5619-8c7e-ec97b0773d9d>

Koppatz, R. (28.9.2020). "Metadatan hallinnasta on tullut cool"!

<https://blogi.communicationpro.com/artikkelit/metadatan-hallinnasta-on-tullut-cool>

Limn  ll, J. (2019). Teknologian radikaali kehitys haastaa ihmist   ja luottamusta. *Kaleva*.

<https://www.kaleva.fi/teknologian-radikaali-kehitys-haastaa-ihmista-ja-l/1705164>

Lindholm, M. (2021). Datar  j  hdys muutti datan merkityksen yhteiskunnassa. *Tilastokeskus*.

<https://www.stat.fi/tietotrendit/blogit/2021/datarajahdys-muutti-datan-merkityksen-yhteiskunnassa-miten-siihen-pitaisi-reagoida/>

Muurinen, R. (6.9.2016). Suuret ja avoimet aineistot politiikan tutkimuksessa. *Tampereen yliopiston yhteiskuntatieteiden tiedekunnan verkkojulkaisu*.

<https://www.tuni.fi/alustalehti/2016/09/06/suuret-ja-avoimet-aineistot-politiikan-tutkimuksessa/>

Naatula, M. (2018). Tied  tk   tarpeeksi datastasi? *Tivi*.

<https://www.tivi.fi/kumppaniblogit/enfo/tiedatko-tarpeeksi-datastasi/b43f1b62-1700-38a5-b01f-cf3acce78d6d>

Nikander, P., Hyv  rinen, M. & Ruusu vuori, J. (2017 s.18). *Haastattelun analyysi*. Vastapaino Tampere.

Ollikainen, J. (12.10.2017). Mit   metatieto on ja miksi min   ?

<https://www.documenthouse.fi/blogi/mita-metatieto-on-ja-miksi-mina-kayttaisin-sita>

Palosaari, R. (14.10.2020). Kest  v  n kehityksen huomioiminen yritystoiminnassa.

MIKROYRITT  JYYSKESKUS. <https://www.microentre.fi/kestavan-kehityksen-huomioiminen-yritystoiminnassa/>

Parikka, H. (2018). Huomisen hyvinvointia datasta. *Sitra*.

<https://www.sitra.fi/julkaisut/huomisen-hyvinvointia-datasta/#1-hyvinvointia-tiedosta-visio-2025>

Parikka, H. (2021). Yhteisty  ll   reiluun datatalouteen. *Sitra*.

<https://www.sitra.fi/julkaisut/yhteistyolla-reiluun-datatalouteen/>

Pulkkinen, J. (8.3.2005). Puheluiden tallennus yleist  . *Kaleva*.

<https://www.kaleva.fi/puheluiden-tallennus-yleista/2152507>

Puusa, A. & Juuti, P. (2020). *Laadullisen tutkimuksen näkökulmat ja menetelmät*.

<https://www.ellibslibrary.com/reader/9789523456167>

Saaranen-Kauppinen, A. & Puusniekka, A. (2006). KvaliMOTV - Menetelmäopetuksen tietovaranto. *Yhteiskuntatieteellinen tietoarkisto*.

<https://www.fsd.tuni.fi/menetelmaopetus>

Sarajärvi, A. & Tuomi, J. (2018). *Laadullinen tutkimus ja sisällön analyysi*.

<https://www.ellibslibrary.com/reader/9789517685030>

Setälä, N. (2020). Raportti: Suomi ei ole hyötynyt teknologian kehityksestä kilpailijamaiden tavoin – alempi työn tuottavuus on Suomen kansantalouden suurin ongelma.

Tekniikka&Talous. <https://www.tekniikkatalous.fi/uutiset/raportti-suomi-ei-ole-hyotynyt-teknologian-kehityksesta-kilpailijamaiden-tavoin-alempi-tyon-tuottavuus-on-suomen-kansantalouden-suurin-ongelma/a84d1ebf-5e31-412b-b47a-ec325b08a0d4>

Sharma, A. (13.5.2021). What is YAML? A Beginner's Guide. [https://circleci.com/blog/what-is-yaml-a-beginner-s-](https://circleci.com/blog/what-is-yaml-a-beginner-s-guide/?utm_medium=SEM&utm_source=gmb&utm_campaign=SEM-gmb-DSA-Eng-ni&utm_content=&utm_term=dynamicSearch-&gclid=EAlaIQobChMIsv2_2uri8gIVluh3Ch3X_gTQEAMYASAAEgLiz_D_BwE)

[guide/?utm_medium=SEM&utm_source=gmb&utm_campaign=SEM-gmb-DSA-Eng-ni&utm_content=&utm_term=dynamicSearch-&gclid=EAlaIQobChMIsv2_2uri8gIVluh3Ch3X_gTQEAMYASAAEgLiz_D_BwE](https://circleci.com/blog/what-is-yaml-a-beginner-s-guide/?utm_medium=SEM&utm_source=gmb&utm_campaign=SEM-gmb-DSA-Eng-ni&utm_content=&utm_term=dynamicSearch-&gclid=EAlaIQobChMIsv2_2uri8gIVluh3Ch3X_gTQEAMYASAAEgLiz_D_BwE)

Suomen standardisoimisliitto SFS Ry. (23.9.2020). Valmistuva standardi yhtenäistää ja tehostaa metatietojen käyttöä dokumenteissa. [https://sfs.fi/valmistuva-standardi-](https://sfs.fi/valmistuva-standardi-yhtenaistaa-ja-tehostaa-metatietojen-kayttoa-dokumenteissa/)

[yhtenaistaa-ja-tehostaa-metatietojen-kayttoa-dokumenteissa/](https://sfs.fi/valmistuva-standardi-yhtenaistaa-ja-tehostaa-metatietojen-kayttoa-dokumenteissa/)

Supponor <https://supponor.com/>

Teknologiateollisuus ry. (2012). Teknologia ilmastonmuutoksen torjunnassa.

https://teknologiateollisuus.fi/sites/default/files/file_attachments/elinkeinopolitiikka_kestava_kehitys_julkaisut_teknologia_ilmastonmuutoksen_torjunnassa.pdf

Tiedostomuotojen opas. Mikä on .JSON tiedostotyyppi?

<https://www.filetypeadvisor.com/fi/extension/json>

Tieteen termipankki. (9.9.2021). Merkintäkieli.

<https://tieteentermipankki.fi/wiki/Nimitys:merkint%C3%A4kieli>

Tietoarkisto. (2021). Dokumentaatio ja metadata. <https://sites.uef.fi/rdm/dokumentointi-ja-metadata/?lang=fi>

Vatanen, P. (11.2.2017). Kasvokkain: Tekoäly-yrittäjän mielestä tarvitsemme lisää kiinnostavia ongelmia – "Niitä ei ratkaista vain insinöörien avulla". *Yle uutiset*.
<https://yle.fi/uutiset/3-9453967>

Ympäristöministeriö. (3.11.2014). IPCC: Ilmastonmuutos etenemässä kohti peruuttamattomia seurauksia; toimia tarvitaan yhteiskunnan kaikilla sektoreilla.
[https://www.ymparisto.fi/fi-FI/Ilmasto_ja_ilma/IPCC_Ilmastonmuutos_etenemassa_kohti_per\(31807\)](https://www.ymparisto.fi/fi-FI/Ilmasto_ja_ilma/IPCC_Ilmastonmuutos_etenemassa_kohti_per(31807))

Liite 1: Aineistonhallintasuunnitelma

Tämä aineistonhallintasuunnitelma on toteutettu osana opinnäytetyöprosessia. Tällä suunnitelmalla varmistetaan, että aineistojen kerääminen, käsittely, säilyttäminen sekä tuhoaminen on suunniteltu hyvin ja toteutetaan oikein.

Opinnäytetyön aikana kerättiin aineistoa verkosta luotettavista lähteistä, kirjaston kirjoista sekä haastattelujen muodossa. Hyvä tieteellinen käytäntö koskee kaikkia opinnäytetöitä, niitä noudatetaan tässäkin opinnäytetyöprosessissa. Opinnäytetyön tekijä huolehtii tutkimusluvista sekä kaikista suostumuksista. Mikäli haastateltavat haluavat pysyvä anonyymeinä, keskustelujen henkilötietoja ei julkaista eikä tallenneta minnekään talteen. Noudatetaan myöskin kaikilta osin tutkimuseettisiä periaatteita sekä henkilötietojen käsittelyyn ja tietosuojaan liittyvää lainsäädäntöä.

Toimeksiantajana opinnäytetyössä toimii Supponor Oy, yrityksen kanssa on tehty opinnäytetyöprosessin alkuvaiheessa sopimus aineiston käytöstä sekä sen julkisuudesta. Aineisto ei sisällä salassa pidettävää tietoa. Opinnäytetyön aineistoista ja tuloksista on sovittu, että sen valmistuttua opiskelija luovuttaa ne Supponor Oy:n käyttöön.

Opinnäytetyön tutkimusaineistoa säilytetään opiskelijan omassa ulkoisessa kovalevyssä. Opinnäytetyöprosessin aikana varmistetaan, että aineisto säilyy turvassa. Aineisto säilytetään opinnäytetyö ohjeiden mukaisesti 1 vuosi, tämän jälkeen aineisto tuhotaan.