8. The Handle System

Sisältöä päivitetty viimeksi maaliskuussa 2023.

8.1 Handle-järjestelmän tausta ja tukiorganisaatio

1990-luvun alussa Internet oli kasvanut niin, että alettiin kaivata verkkoon tallennettujen digitaalisten objektien yleistä hallinnointiratkaisua. Alan toimijoista tärkeimpiin kuulunut Defence Advanced Research Project (DARPA) rahoitti aikavälillä 1992–1996 Corporation for National Research Initiativesin (CNRI, https://www.cnri.reston.va.us/) tähän päämäärään tähdännyttä projektia. Vuonna 1994 hanke julkisti The Handle System -tunnistejärjestelmän ja sen taustalla olevan digitaalisten objektien hallintamallin. Niiden sijaan digitaalisten aineistojen yleiseksi hallinnointijärjestelmäksi vakiintui kuitenkin Tim Berners-Leen CERN:issä kehittämä World Wide Web (WWW).

2020-luvulla WWW on yhä selvästi tunnetuin ja käytetyin digitaalisten aineistojen hallinnointijärjestelmä, mutta myös CNRI;n kehittämällä vaihtoehtoisella ratkaisulla on omat kannattajansa. CNRI-mallin uusimmat kuvaukset ovat Digital Object Interface Protocol (DOIP, versio 2 vuonna 2018) ja Handle Systemin seuraaja, kesällä 2022 julkaistu Digital Object Identifier Resolution Protocol (DO-IRP), jotka ovat vapaasti poimittavissa osoitteesta https://www.dona.net/specsandsoftware.

Vastuuorganisaatio on Sveitsiin rekisteröity The DONA Foundation, joka kertoo itsestään seuraavaa[i]:

> *In the late 2000s, following the many years of discussion about "Internet Governance," it was clear that a system for managing information in digital form that was rooted in the U.S. would have limited global acceptance. Specifically, many countries and organizations interested in developing and deploying components of the Digital Object Architecture (DOA), including in particular the identifier/resolution mechanism were basically reluctant to do so.*

Edellä viitataan Internetin nimipalveluun. Se on elektronisten julkaisujen hallinnoinnin kannalta ongelmallinen poliittisesti Yhdysvaltain vaikutusvallan takia. Tämä lienee ainakin yksi syy siihen, että Handle system on tiettävästi erittäin suosittu Kiinassa. Mutta nimipalvelu on ongelma myös teknisesti, koska domain-nimiä voi omistaa. Niitä voi vain vuokrata määräajaksi. Handle Systemin digitaalisten aineistojen hallintamalli nojaa siihen, että jokainen digitaalinen objekti saa järjestelmässä heti pysyvän tunnuksen, joka ei muutu, vaikka aineistoa hallinnoiva organisaatio ja/tai sen domain-nimi muuttuisivat. Tunnuksen toimijakohtaista etuliitettä ei koskaan anneta jollekin toiselle toimijalle.

Handle-tunnisteesta ja sen tekniikkaan perustuvasta DOI-järjestelmästä on tullut suosittuja eritoten tieteellisten julkaisujen ja data-aineistojen tunnistejärjestelminä, mutta niitä voidaan soveltaa paljon laajemminkin, periaatteessa mihin tahansa objekteihin. Kiina on tästä hyvä esimerkki, kuten alempaa ilmenee. Handle ja DOI ovat molemmat objektien digitaalisia tunnistejärjestelmiä, eivät vain digitaalisten objektien tunnisteita.

The DONA Foundation on vastannut Handle Systemiin liittyvistä standardeista vuodesta 2015. Sen tukiorganisaatioita ovat tunnisteen alkuperäinen kehittäjä CNRI ja DOI-tunnuksen vastuutaho DOI Foundation. Tulevaisuudessa näitä tukiorganisaatioita saattaa järjestelmän käytön laajetessa tulla lisää. The DONA Foundation tekee tiivistä yhteistyötä myös International Telecommunications Unionin eli ITU:n[ii] kanssa. ITU on esimerkiksi julkaissut DOIP-standardin version 1.0 standardina ITU-T X.1255 vuonna 2013.

The DONA Foundationin välit Internet Engineering Task Forceen (IETF) ja W3C:hen ovat etäiset. Jälkimmäisen osalta tämä on helppo ymmärtää, koska Handle-toimintamalli on perinteisen Webin kilpailija. IETF:n kohdalla asiaa selittää historia. CNRI yritti julkaista The Handle Systemin Internet-standardina, mutta IETF ei siihen suostunut, koska Handle System oli Internetin nimipalvelulle ja myös tuolloin kehitteillä olleelle URN-tunnisteelle vaihtoehtoinen tekninen ratkaisu. Koska IETF:llä ei ole tapana hyväksyä toistensa kanssa kilpailevia määrityksiä, Handlesta ei tullut Internet-standardia.

Handle-määritykset ilmestyivät alun perin RFC-julkaisuina RFC 3650-3652 vuonna 2003. Niiden status on Informational, ei Internet-standardin Standards Track[iii]. IETF:n johtoelin Internet Engineering Steering Group (IESG) alleviivasi varautuneisuuttaan lisäämällä Handlen RFC-julkaisuihin seuraavan poikkeuksellisen huomautuksen (alleviivaus tekijän):

> *Several groups within the IETF and IRTF have discussed the Handle System and its relationship to existing systems of identifiers. The IESG wishes to point out that these discussions have not resulted in IETF consensus on the described Handle System, nor on how it might fit into the IETF architecture for identifiers. Though there has been discussion of handles as a form of URI, specifically as a URN, these documents describe an alternate view of how namespaces and identifiers might work on the Internet and include characterizations of existing systems which may not match the IETF consensus view.*

"Alternate view" edellä viittaa siihen, että Handle-järjestelmässä Internetin nimipalvelun domain-nimen korvaa Handle-tunnuksen etuliite. Tunnuksissa oleva osoite (http://hdl.handle.net, https://doi.org) on vain välityspalvelimen eli proxyn osoite, Handle-järjestelmän sisällä oikea resolveri löytyy Handle-tunnuksen etuliitteen avulla. Tällä ratkaisulla on se merkittävä etu, että toisin kuin domain-nimet, Handle-etuliitteet ovat pysyviä ja ainutkertaisia.

Kalsea vastaanotto IETF:ssä selittänee sen, miksi The Handle Systemin määritykset valmistellaan nykyään The DONA Foundationissa ja julkaistaan International Telecommunication Unionin standardeina. ITU:n kannalta merkitsevää oli se, että Handle on hyvin toimiva tekninen järjestelmä, ei sen suhde IETF:n URN-tunnisteeseen tai Internetin nimipalveluun.

Pesäero IETF:ään ja W3C:hen mahdollistaa Handle-standardien itsenäisen kehittämisen irrallaan muusta Internetin standardisoinnista, mutta voi pahimmassa tapauksessa johtaa marginalisoitumiseen. Riippuvuus yhdestä teknisestä määrityksestä ja sen ainoasta teknisestä toteutuksesta on ongelma, jos Handle-määritystä ja resolveria ei pidetä yllä. Tunnisteen laajan käytön vuoksi tämä riski vaikuttaa kuitenkin hyvin pieneltä.

8.2 Handlen soveltamisalat ja keskeiset käyttäjät

The Handle System -järjestelmällä on yksi keskitetysti ylläpidetty resolverisovellus, Handle.Net, jota käyttää myös DOI. Tätä kirjoitettaessa ohjelmiston viimeisin versio on helmikuussa 2023 julkaistu 9.3.1[v]. Java-pohjaisena resolveri toimii monissa käyttöjärjestelmäympäristöissä. Muiden PID-sovellusten tapaan ohjelmisto on tehokas eikä vaadi paljoa laitteistokapasiteettia. Ohjelmiston käyttöönotto edellyttää Handle-etuliitteen eli rekisteröijätunnuksen hankkimista CNRI:ltä. On hyvä pitää mielessä, että kaikki mikä näyttää Handlelta, ei ole sitä – DSpace-julkaisuarkistosovellus luo tallennettaville julkaisuille Handlelta näyttäviä tunnuksia, mutta niistä tulee virallisia ja toiminnallisia vasta, kun tunnusten Handle-etuliite on rekisteröity CNRI:n ylläpitämään kansainväliseen Global Handle Registryyn[vi].

CNRI ylläpitää Global Handle Registry -tietokantaa, mutta se ei seuraa annettujen tunnusten määrää. The Handle System lienee käyttäjien määrällä mitaten maailmanlaajuisesti suosituin PID-järjestelmä. Tunnusten määrässä se kilpailee ARK-tunnisteen kanssa. Kilvan voittajaa emme tiedä, koska Handle-tunnuksien kokonaismäärästä ei ole tilastoja.

Tunnisteiden jakelun edellyttämä Handle-etuliitteen rekisteröinti[vii] maksaa vain 50 Yhdysvaltain dollaria, ja etuliitteen vuotuinen ylläpitomaksu on sama. Itse tunnukset ovat maksuttomia. CNRI:n ohella kahdeksalla muulla organisaatiolla on oikeus Handle-etuliitteiden antamiseen[viii].

The DONA Foundationin Executive Director Christophe Blanchi kertoo haastattelussa (https://podcast.polyneme.xyz/episodes/christophe-blanchi) Handlen historiasta seuraavaa:

> *…by the mid 1990s, already, it was clear that the institutions that were interested in the handle system and what it offered, which is a level of dereference, were the publishers. Because the publishers were new to the web and they realized very quickly that URIs were breaking a lot – for all sorts of reasons. It's not because DNS doesn't work, it's not because URIs don't work, but it's mostly the social-technical fabric of what you put in a URL. And, so, the publishers realized that they needed an identification system that was independent of the underlying technology and the naming conventions that were part of the typical url. And so they were the first ones to say "all right, this is the right approach for identifying journal articles or things that we need to identify over long term."*

DOI-tunnisteen kehittäminen Handle-tekniikan varaan perustui pitkälti kustantajien aktiivisuuteen. DOI on edelleen ensi sijassa julkaisujen tunnistejärjestelmä, mutta Handlen soveltamisala on Blanchin mukaan olennaisesti laajempi etenkin Kiinassa:

> *…the Chinese are the biggest users of the Handle system – potentially by orders of magnitude – but we would not know it here because we don't see what happens in China. But in China, they are using the Handle system to do tracing and tracking of pretty much everything and anything under the sun. And they have used the Handle system as ways to program sensors and read from sensors. So they're sort of the leader when it comes to thinking outside the box and using the technology.*

Kiinassa Handle-tunnisteelle on siis jo tullut tai on tulossa samankaltainen rooli kuin IETF on tarkoittanut URN-tunnisteelle kehittäessään Internet of Things -tunnusten URN:DEV-nimialueen.

8.3 Handle-tunnuksen rakenne

Tässä luvussa kuvatut Handle-tunnuksen ominaisuudet koskevat myös DOI-tunnuksia.

Handle-tunnus koostuu kahdesta osasta: etuliitteestä (prefix) sekä takaliitteestä (suffix), joita erottaa kauttaviiva. Etuliite on pysyvä tunnuksen rekisteröijän tunnus ja samalla organisaation Handle-resolverin osoite. Takaliite on identifioidun objektin tunnus, jonka on oltava ainutkertainen etuliitekohtaisesti.

Tunnus esitetään yleensä HTTP URI -muodossa, jolloin sen eteen lisätään välityspalvelimen osoite. Esimerkiksi Handle-tunnuksessa https://hdl.handle.net/10138/300023 https://hdl.handle.net on Handlen proxy- eli välityspalvelin, etuliite 10138 on Helsingin yliopiston tunnuksen rekisteröijän tunnus ja sen resolverin osoite Handle-järjestelmän sisällä, ja 300023 julkaisun tunnus. Koska DOI ja Handle jakavat saman infrastruktuurin, Handle-proxy on korvattavissa DOI-tunnusten proxy-osoitteella https://doi.org. Mutta tämä ei muuta Handle-tunnuksia DOI-tunnuksiksi, vaikka maallikosta siltä näyttäisi. Vastaava proxy-osoitteen korvattavuus toimii myös toisinpäin.

Handle-tunnusten on oltava pysyviä ja ainutkertaisia, minkä lisäksi niiltä edellytetään semanttisuuden karsimista. Tunnus ei siis saisi kertoa identifioidusta objektista mitään. Käytännön kokemus onkin osoittanut, että ei-semanttisia tunnuksia on yleensä helpompi luoda ja ylläpitää kuin semanttisia tunnuksia. Handle-tunnuksien pitäisi olla myös sekä koneiden että ihmisten luettavissa ja niiden tulisi noudattaa URI-syntaksia. Handle-määritykset julkistettiin kuitenkin jo ennen RFC 3986 –standardia, minkä vuoksi URI-vaatimusten ja Handlen välillä on esimerkiksi merkkivalikoimaa koskevia eroja. Tämä rajoittaa koneluettavuutta: Handle-tunnuksen sisältävän HTTP URI:n voi varmuudella tulkita oikein vain Handle-resolverisovellus.

DO-IRP -standardin mukaan Handle-tunnuksissa voi käyttää periaatteessa mitä tahansa Unicode-merkkejä, kuten vaikkapa emojeja. Siksi myös kaikki ne merkit, joilla on vuonna 2005 julkistetun URI-syntaksin mukaan erikoismerkitys, ovat olleet vapaasti käytettävissä. Tällainen on esimerkiksi HTTP-palvelimelle tehtävää hakupyyntöä eli Queryä merkitsevä ?, jota ei voi soveltaa Handle-tunnuksissa RFC 3986 -standardin edellyttämällä tavalla. Ja koska on tai voi olla olemassa Handle-tunnuksia joissa ?-merkkiä käytetään ”väärin”, DO-IRP-standardia ei voi takautuvasti korjata RFC 3986:n mukaiseksi.

Handle-tunnukset voivat olla etuliitekohtaisesti joko merkkikokoriippuvaisia tai merkkikokoriippumattomia. Tämä voi rajoittaa Handle-tunnuksien muuttamista DOI-tunnuksiksi, koska DOI on aina merkkikokoriippumaton.

8.4 Handle julkaisujen tunnisteena

Koska Handle-tunnisteen käyttöä ei valvota keskitetysti, kellään ei ole kattavaa tietoa siitä, miten sitä sovelletaan. Suomessa esimerkiksi Helsingin yliopisto käyttää Handlea opinnäytteiden ja muiden Helda-julkaisuarkistoon tallennettujen aineistojen tunnisteena. Handle-tunnistetta käytetään paljon myös kansainvälisesti, esimerkiksi Kongressin kirjastossa ja monissa muissa organisaatioissa, joilla on laajat digitoitujen tai digitaalisena syntyneiden aineistojen kokoelmat.

Handle-määritys ei kiellä perinteisten standarditunnisteiden käyttämistä Handle-tunnuksen osana. Käytännössä esimerkkejä tästä on vaikea löytää. Kun ISBN-tunnuksen omaavalle e-kirjalle annetaan Handle-tunnus, se ei yleensä perustu kirjan ISBN-tunnukseen (ks. esim. http://hdl.handle.net/10138/152270) . Jos Handle linkittyy laskeutumissivulle kuten yllä olevassa esimerkissä, se on tulkittavissa teoksen tunnukseksi, jos laskeutumissivuun linkittyy useampia manifestaatioita ja niillä on edelleen omat ISBN-tunnuksensa. ISBN-tunnistejärjestelmän kannalta tämäkin käytäntö on silti haasteellinen, koska e-kirjaan viitattaessa on houkuttelevampaa käyttää toiminnallista Handlea kuin ei-toiminnalista ISBN-tunnusta. Jos e-kirjan manifestaatiot eivät ole identtisiä, teokseen viitattaessa nykyisiä viittauskäytänteitä voi olla tarpeen muuttaa. Viitteessä ei pitäisi käyttää sivunumeroa, vaan se pitää kohdistaa tiettyyn kohtaan tekstissä (esimerkiksi luku 7, kappale 3).

8.5 Handlen tulevaisuuden näkymät

On varmaa, että The Handle Systemiä käytetään vielä pitkään. Standardi on juuri päivitetty, resolverisovellusta kehitetään aktiivisesti ja palvelun tekninen infrastruktuuri vaikuttaa vakaalta.

PID-tunnisteen valintaa harkitsevalle resolverisovelluksen saatavuus sekä järjestelmän levinneisyys ja tuki ovat valinnan kannalta tärkeitä asioita. Mutta jos niiden lisäksi tarvitaan hyvin hallinnoitu järjestelmä, Handlea parempi vaihtoehto on samaa teknologiaa käyttävä Digital Object Identifier eli DOI. Näiden järjestelmien suhdetta voi kuvata siten, että jokainen DOI on myös Handle, mutta kaikki Handlet eivät ole DOI-tunnuksia. DOI-standardin vuonna 2022 julkaistun painoksen mukaan Handle-tunnuksista on kuitenkin mahdollista tehdä DOI-tunnuksia takautuvasti. International DOI Foundation voi pyynnöstä rekisteröidä Handle-etuliitteen DOI-etuliitteeksi, jolloin kaikki kyseisen etuliitteen omaavat Handle-tunnukset muuttuvat takautuvasti DOI-tunnuksiksi. Muutoksen teko edellyttää kuitenkin sitä, että jokin DOI Registration Agency ottaa tunnukset vastuulleen, ja niillä voi olla lisävaatimuksia koskien esimerkiksi identifioidun aineiston luonnetta ja metatietojen sisältöä.

[i] https://www.dona.net/aboutus

[ii] https://www.itu.int/

[iii] https://en.wikipedia.org/wiki/Request_for_Comments#Status

[iv] https://www.dona.net/specsandsoftware#block-identifierandresolutionprotocolspec

[v] http://handle.net/download_hnr.htm

[vi] https://www.dona.net/prefix/resolve

[vii] https://www.handle.net/prefix.html

[viii] https://www.dona.net/mpas