

6. ARK

Sisältöä päivitetty viimeksi maaliskuussa 2023.

6.1 ARK-järjestelmän tausta ja tukiorganisaatio

Archival Resource Key eli ARK-tunnistejärjestelmän ja sen resolveriohjelmiston kehitti alun perin John Kunze University of Californian California Digital Library (CDL, <https://cdlib.org/>) -yksiköstä. ARK julkistettiin hieman muiden PID-järjestelmien jälkeen vuonna 2001, ja siinä on pyritty korjaamaan muiden järjestelmien puutteita. ARK-tunnisteissa ei esimerkiksi saa käyttää muita kuin tulostettavia ASCII-merkkejä, ja ARK-resolverille voi tallentaa periaatteessa mitä tahansa metatietoa.

ARK-järjestelmän kehittämiseen osallistuu nykyään CDL:n ohella useita muitakin tunnistetta käyttäviä organisaatioita, mutta CDL vastaa edelleen muun muassa yhteisen resolveriympäristön ylläpidosta. Jokaisella ARK-tunnisteen käyttäjällä on oltava oma resolveri-instanssi, mutta sille tallennetut metatiedot linkityksineen siirtyvät automaattisesti yhteiselle resolverille.

6.2 ARK-järjestelmän soveltamisalat ja keskeiset käyttäjät

ARK-tunnisteella voidaan identifioida periaatteessa mitä tahansa. John Kunzen ja Emmanuelle Bermésin laatima ARK-määritys^[1] kuvaa asian seuraavasti:

ARK is well suited to long-term access and identification of any information resources that accommodate reasonably regular electronic description. This includes digital documents, databases, software, and websites, as well as physical objects (books, bones, statues, etc.) and intangible objects (chemicals, diseases, vocabulary terms, performances).

ARK-tunnisteella on yli tuhat käyttäjäorganisaatiota. Viimeisimmän 10 vuoden aikana käyttäjämäärä on kasvanut nopeasti (ks. <https://arks.org/community/>). Merkittäviä ARK-tunnuksia hyödyntäviä organisaatioita ovat esimerkiksi Louvren taidemuseo, Ranskan kansallisarkisto sekä Ranskan ja Ison-Britannian kansalliskirjastot. Ranskassa ARK on erittäin suosittu keskeisten muistiorganisaatioiden tarjoaman vetoavun ansiosta. Louvre on kieltämättä vaikuttava esimerkki: sen digitoiduille kokoelmille annettiin 2021 483 000 ARK-tunnusta^[2]. Suomessa käyttäjiä ei vielä ole, minkä takia tunnistetta käsitellään tässä tekstikokonaisuudessa varsin lyhyesti. Jatkossa tilanne voi muuttua, koska Kansallisarkisto selvittää kevään 2023 mittaan PID-tunnistejärjestelmän käyttöönottoa. Keskeiset vaihtoehdot ovat ARK ja URN:NAN (kansallisarkistoille rekisteröitävä uusi URN-nimialue, joka vastaisi rakenteeltaan URN:NBN-nimialuetta).

6.3 ARK-tunnuksen rakenne

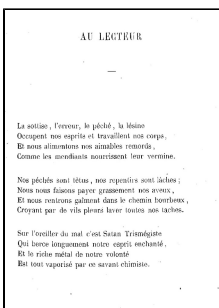
ARK koostuu seuraavista osista, jotka erotetaan kauttaviivalla (/):

- Name Mapping Authority (NMA), ARK-resolverin osoite
- ARK Label "ARK:", ARK-tunnuksen nimiö
- Name Assigning Authority Number (NAAN), ARK-tunnuksen rekisteröijän tunnus
- Name, identifioitun objektin tunnus
- Qualifier, objektin nimen tarkenne

Resolverin osoite on vapaavalintainen osa tunnusta. ARK-resolveereita on eri osoitteissa satoja, mutta niiden linkitystiedot kootaan CDL:n ylläpitämään yhteiseen ARK-resolveriin osoitteessa <http://n2t.net/>. URN-tunnisteen tavoin ARK kestää sen että resolverin osoite jätetään pois, koska ARK:nimiön ansiosta tunnus (esimerkiksi <ark:/12148/bpt6k5834013m>) on yhä tunnistettavissa tekstin keskeltä. Hyperlinkki siitä tulee vain resolverin osoitteen kera.

ARK-tarkenne on piirre, jota muissa PID-tunnisteissa ei ole. Tarkenne erotetaan varsinaisesta tunnuksesta joko kauttaviivalla, jos kyse on identifioitun objektin osakohteesta, tai pisteellä, kun identifioidaan alkuperäisen objektin versio. Esimerkiksi Baudelairen Pahan kukkia -teoksen ensipainoksen digitoidun version ARK on <https://gallica.bnf.fr/ark:/12148/bpt6k5834013m>.

Runolla Au Lecteur (Lukijalle, kuva 1) on ARK <https://gallica.bnf.fr/ark:/12148/bpt6k5834013m/f19>, ja tämän runon alla olevan sivukuvan ARK on <https://gallica.bnf.fr/ark:/12148/bpt6k5834013m/f19.highres>.



Kuva 1. Runolla Au Lecteur (Lukijalle) on ARK <https://gallica.bnf.fr/ark:/12148/bpt6k5834013m/f19>, ja runon tämän oppaan kuvitukseksi valitun sivukuvan ARK on <https://gallica.bnf.fr/ark:/12148/bpt6k5834013m/f19.highres>.

Lähde: gallica.bnf.fr / Médiathèque de la communauté urbaine d'Alençon. Klikkaa kuvaa, jos haluat nähdä sen suurempana.

Tunnukseen voidaan lisätä myös resolverille suunnattuja pyyntöjä. Vanhoissa ARK-standardiluonnoksissa "?" vaati resolveria toimittamaan identifioidun objektin metatiedot; "???" säilytystiedot. Nykyinen ARK-versio on tukee myös URI-syntaksin mukaista kyselyä "?info", joka tuottaa identifioidun objektin metatiedot. URL-muodossa esitettyyn tunnukseen voidaan lisätä myös muuta informaatiota, kuten metadataformaatin määre.

ARK <http://n2t.net/ark:67531/metadc107835?info> tuottaa Larry Austinin vuonna 1952 valmistuneen opinnäytteen laskeutumissivun, joka sisältää julkaisun metatiedot. ARK-tunnuksen versiolla <http://n2t.net/ark:67531/metadc107835??> selviää, että University of North Texas Libraries on vuonna 2008 sitoutunut tallentamaan tämän julkaisun pysyvästi. Ja ARK <http://n2t.net/ark:67531/metadc107835/metadata.dc.xml> tuottaa julkaisun Dublin Core -metatiedot XML-muodossa

6.4 ARK julkaisujen tunnisteena

Perinteiset julkaisujen tunnukset, kuten ISBN, voidaan esittää ARK-tunnuksen osana. ARK-määrittäminen ei kielä olemassa olevasta tunnuksesta poikkeavan ARK-tunnuksen antamista. ARK-yhteisö on kuitenkin päättänyt, että jos julkaisulla on jokin standarditunnus, sille ei anneta tuosta tunnuksesta poikkeavaa ARK-tunnusta.

On käytännössä mahdotonta olla varma siitä, onko ARK-yhteisön linjaus pitänyt, koska ARK-tunnusten käyttöä ei valvota keskitetysti. Järjestelmä ei myöskään vaadi tunnuksen antajaa korjaamaan tilannetta, jos hänen hallinnoimansa ARK lakkaa resoluutiovasta. Tässä suhteessa ARK muistuttaa Handlea ja sellaisia URN-nimialueita, joissa tunnusten antamista ei valvota. Hallinnoinnin puute on ongelma kaikille näille PID-tunnisteille, koska keskitetysti valvottu PID-tunnus on lähes varmasti pitkäikäisempi kuin valvomaton. Hyvin hallinnoidussa järjestelmässä tunnuksen antaja saa reklamaation resoluution tuottamasta 404-virheestä, ja hänen pitää korjata tilanne. ARK-järjestelmässä ongelmaa ei välttämättä edes huomata, elleivät käyttäjät valita siitä. Niinpä ARK-standardin vanhoissa luonnoksissa oli pitkään ARK-tunnus <https://444.berkeley.edu/ark:28722/x9t38rk45c> joka tuotti 404-virheen myös yhteisellä <http://n2t.net-resolverilla>.

6.5 ARK-järjestelmän tulevaisuuden näkymät

ARK on teknisesti vakaa järjestelmä, jolla on yhteinen ja keskitetysti ylläpidetty resolverisovellus. Tunnisteen standardisointi on kuitenkin yhä kesken. Prosessi alkoi Internet Engineering Task Force (IETF) jo vuonna 2001, ja maaliskuuhun 2023 mennessä ARK-määrittäksen Internet Draft -luonnoksia oli ilmestynyt jo 36^[ii]. Uusia versioita voi olla edelleen tulossa, mutta ellei tavoitetta vaihdeta, on mahdollista ettei sitkeys tuota toivottua tulosta. IETF on hyväksynyt URN-tunnisteen Internet-standardiksi, eikä se yleensä siunaa päällekkäisiä ratkaisuja samaan tarkoitukseen. Nykymuodossaan ARK on URN-tunnisteen suora kilpailija. Lisäksi ARK-määrittästä ei ole laatinut IETF:n asettama virallinen työryhmä. Kuka tahansa voi laatia Internet-standardiluonnoksen eli Internet draftin, mutta yksityishenkilöiden aloitteita ei aina edes käsitellä IETF:n johtelmissä, ja vaikka käsiteltäisiinkin, niistä ei yleensä tule varsinaisia standardeja. Parhaassa tapauksessa ARK-määrittäminen julkaistaan Handle-tunnisteen tapaan Informational-statuksella, jolloin siitä tulee RFC-julkaisu, mutta ei standardia.

IETF-standardisoinnin pullonkaulaan on olemassa yksinkertainen ratkaisu: ARK-tunnistelle rekisteröidään URN-nimialue, ja julkaistaan sen jälkeen RFC, jossa tunnistejärjestelmä kuvataan. Vastaavaa menettelyä on jo sovellettu NBN-tunnisteeseen, eikä ole mitään syytä, miksi sitä ei voitaisi soveltaa myös ARK-tunnisteeseen. Kansalliskirjasto on esittänyt tätä toimintamallia ARK-kehittäjille maaliskuussa 2023, ja asia on tätä kirjoitettaessa (30.3) yhteisön pohdinnassa. Nimialuekisteröinnin pitäisi olla teknisesti yksinkertainen prosessi varsinkin sen jälkeen kun URN-syntaksi salli "/"-merkin käytön. ARK-tunnuksista tulee URN lisäämällä tunnuksen alkuun merkkijono "URN:".

ARK-yhteisö voi myös luopua 22 kestäneestä IETF-standardisointiprosessista, ja kääntyä jonkin muun standardisointijärjestön puoleen. Tällöin tulisivat kyseeseen ainakin ANSI/NISO ja Ranskan kansallinen standardisointijärjestö AFNOR, koska suurin osa tunnisteen käyttäjistä on Ranskassa ja Yhdysvalloissa.

ARK-järjestelmällä oli maaliskuussa 2023 jo yli 1000 käyttäjäorganisaatiota, joilla on yhteistyötä edistävä allianssi^[iv]. Käyttäjäkuntaan kuuluu kirjastoja, arkistoja ja museoita, mutta myös datakeskuksia ja valtionhallinnon laitoksia. Tunnukset ovat maksuttomia, ja niitä on luotu peräti 8.2 miljardia, eli todennäköisesti enemmän kuin mitään muuta PID-tunnistetta. Suuren suosion ansiosta ARK-resolverisovelluksen ja muiden ARK-ohjelmistojen sekä kehittäminen sekä N2T.net-resolverin ylläpito lienevät vakaalla pohjalla. Varmuutta tästä on vaikeaa saada, koska esimerkiksi teknisen infrastruktuurin kehittämiseen ja ylläpitoon tarkoitetut rahoitusjärjestelyt eivät ole julkisia. Tähän sisältyy kenties merkittävin tunnistetta koskeva epävarmuustekijä: jos sen suosio kasvaa merkittävästi, kuka maksaa käytön kustannukset?

ARK-tunniste ja perinteiset tunnistejärjestelmät ovat tähän asti täydentäneet toisiaan, ja ARK-tunnuksia annettaneen jatkossakin aineistoille, joilla ei ole niille kehitettyä tunnistestandardia (kuten taideteokset) tai joita kyseisille aineistoille luotu tunnistestandardi ei kata (ennen ISBN-järjestelmän käyttöönottoa julkaistut monografiat, opinnäytteet). DOI- ja Handle-tunnisteiden tavoin ARK hyötyy valmiista ohjelmistoympäristöstä, mutta niistä molemmista sen erottaa maksuttomuus, joka taas on yhteinen piirre URN-tunnisteen kanssa. Tarkenteen käyttömahdollisuus antaa tunnusjakeluun joustavuutta, joka muista järjestelmistä puuttuu ainakin toistaiseksi. Näillä eväillä ARK säilynee PID-kartalla pitkälle tulevaisuuteen, toivottavasti niin että tunnistelle rekisteröidään URN-nimialue.

[i] <https://datatracker.ietf.org/doc/draft-kunze-ark/>

[ii] <https://arks.org/blog/the-louvre-collection-goes-online-with-483000-arks/>

[iii] <https://datatracker.ietf.org/doc/html/draft-kunze-ark>

[iv] <https://arks.org/>