



Software Heritagen hyödyntäminen julkaisutiedonkeruun i2-julkaisuille

17.5.2024
Minna Ala-Mantila, CSC



Mikä on Software Heritage?

- Software Heritage ("SWH") on avoin ja voittoa tavoittelematon aloite, jonka tavoite on kerätä yhteen, säilyttää ja tuoda saataville kaikki verkossa avoimesti saatavilla olevat lähdekoodit versiotietoineen.
- Hanke on aloitettu vuonna 2016. Unesco mukana hankkeessa.
- Tarjoaa mm. näkymän saatavilla oleviin lähdekoodeihin, lähdekoodin arkistoinnin ja pysyvän tunnisteen (SWHID)
- Haravoi säännöllisesti 27 repositoriota (+ useita muita tarvittaessa)
 - Mm. GitHub, GitLab ja Bitbucket, Python, R
- Haku: <https://archive.softwareheritage.org>



Software Heritage

SWH:n hyödyntäminen julkaisutiedonkeruussa?

1/2: Organisaatiot

- Organisaatiot voisivat etsiä SWH:n arkistosta omia lähdekoodejaan raportoivaksi osana OKM:n julkaisutiedonkeruuta (I2)
- Mahdollista hakea lähdekoodeja esim. organisaation nimellä
- SWH:n tarjoama pysyvä tunniste (SWHID) täyttäisi tiedonkeruun ehdon *"Sovelluksen lähdekoodiin tai algoritmiin tulee voida viitata pysyvällä tunnisteella, esim. DOI"*
 - Huom. Myös ehdon *"Tieto julkaisun tekijäistä/tekijöistä löydyttävä itse julkaisusta tai julkaisualustalta. Yhteys raportoivaan organisaatioon oltava todennettavissa."* toteutuminen tarkistettava

SWH:n hyödyntäminen julkaisutiedonkeruussa?

2/2: VIRTA-julkaisutietopalvelu

- Virta voisi jatkossa haravoida SWH:n arkistoa, ja etsiä vastaavuuksia Virrasta jo löytyville aineistoille, ja hakea näille SWH:n tarjoaman pysyvän tunnisteiden lisätiedoksi
 - Mahdollisesti myös muita SWH:n tarjoamia lähdekoodia kuvailevia metatietoja
- Tietojen näyttäminen tiedejatutkimus.fi:ssä

Miten SWH voisi hyötyä TTV:n sisältämistä tiedoista?

- Yksi mahdollinen jatkokehityskohde olisi myös haravointi toiseen suuntaan, eli SWH haravoisi TTV:n lähdekoodeja omaan arkistoonsa.
 - Tekijän ORCID-id yksi keskeinen tieto jolla SWH voisi rikastaa heiltä jo löytyviä tietoja.