

Löytösampo: kansalaisten arkeologiset löydöt semanttisessa webissä

Eero Hyvönen

Aalto-yliopisto, tietotekniikan laitos

Helsingin yliopisto, Digitaalisten ihmistieteiden keskus HELDIG

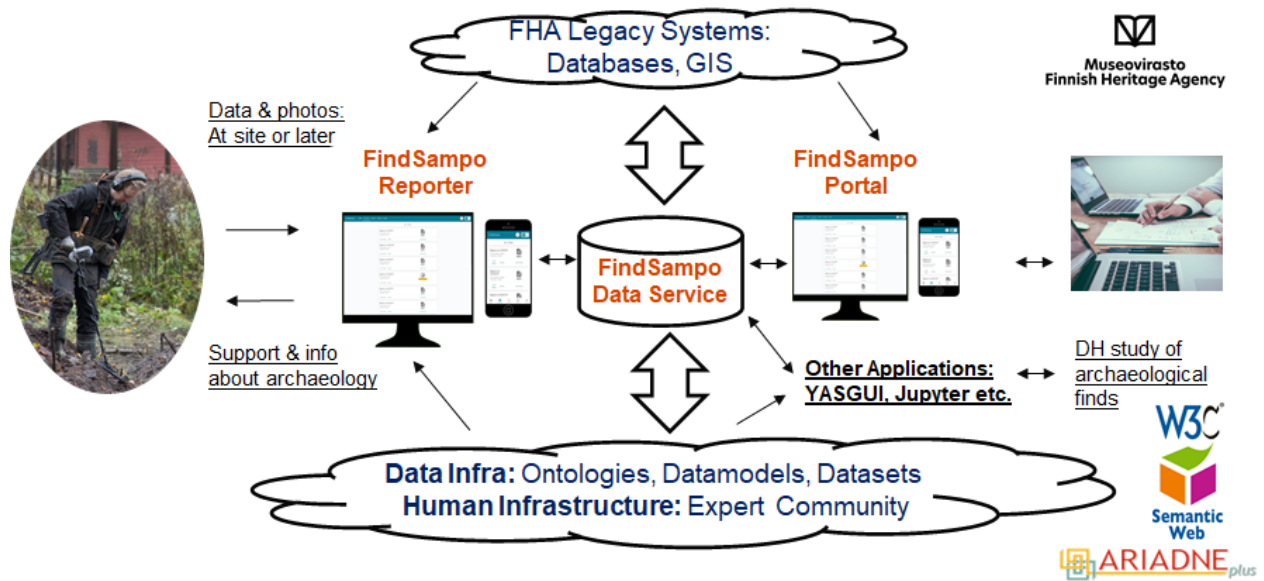
<https://seco.cs.aalto.fi/u/eahyvone>

*Löytösampo - Arkeologiset löydöt semanttisessa webissä*¹ (Hyvönen et al., 2021) on Aalto-yliopiston, Helsingin yliopiston ja Museoviraston SuALT-tutkimushankkeessa² (2017–2020) (Thomas et al., 2018, 2019) kehitetty prototyyppi arkeologisesta tietoinfrastruktuurista ja verkkopalvelukokonaisuudesta. Se sisältää tietoa arkeologisista löydöistä Suomessa, erityisesti metallinetsinnän harrastajien tekemistä. Löytösampo tarjoaa arkeologian harrastajille, museoiden henkilöstölle ja digitaalisten ihmistieteiden tutkijoille uudenlaisia välineitä arkeologisten kansalaisaineistojen keräämiseen (Wessman et al., 2019ab), julkaisemiseen ja tutkimiseen verkossa (citizen science) digitaalisten ihmistieteiden (Digital Humanities, DH) menetelmillä (Gardiner & Musto, 2015).

Järjestelmän innovaationa on tarkastella arkeologisen kansalaistiedon tuotanto- ja hyödyntämisprosessia kokonaisvaltaisesti. Keskeisenä haasteena prosessissa – mutta samalla mahdollisuutena – on metallinetsijöiden tuottaman tiedon nopea lisääntyminen 2010-luvulta alkaen, jonka tarkastaminen, luetteloiminen ja hallitseminen tietoa valtakunnallisesti keräävässä Museovirastossa on muodostunut käytännössä hankalaksi: viiveet tietojen rekisteröinnissä ovat pidentyneet merkittävästi eikä kokoelmätietojen julkaisemiseen ole ollut olemassa sopivia välineitä. Nopeasti karttuvan kokoelmätiedon mahdollisimman ajantasainen julkaiseminen metallinetsijöiden työtä helpottamaan ja ohjaamaan, mukaan lukien ohjeistukset esimerkiksi suojelluista alueista, olisi kuitenkin tarpeen. Löytöaineiston julkaiseminen käyttökelpoisessa muodossa avoimena datana mahdollistaisi myös uudenlaista digitaalisten ihmistieteiden (digital humanities) laskennallista tutkimusta sekä kokoelmien aineistoista, niiden karttumisesta että metallinetsinnästä harrastuksena.

¹ Löytösampo-järjestelmän hankesivu: <https://seco.cs.aalto.fi/projects/sualt>

² Koko SUALT-projektin kotisivu: <https://blogs.helsinki.fi/sualt-project/>



Kuva 1. Löytösammon komponentit ja linkitetyn avoimen datan tietoinfrastruktuuri

Kuva 1 esittää metallinetsintään liittyvän prosessin eri osia ja näihin SuALT-projektissa kehitettyjä ratkaisumalleja ja prototyypijärjestelmiä: Löytösampo-ilmoittaja (FindSampo Reporter), Löytösampo-datapalvelu (FindSampo Data Service) ja Löytösampo-portaali (FindSampo Portal). Perusajatuksena on hyödyntää prosessin eri vaiheissa semanttisen webin teknologioita yhteisen infrastruktuurin toimiessa kokonaisuuden “semanttisena liimana” ja perustana. Vasemmalla ovat tiedon alkuperäiset tuottajat, jotka raportoivat löydöistään Museovirastolle (Finnish Heritage Agency, FHA) valokuvilla ja löytöön liittyvillä tiedoilla, kuten löytöpaikka, löydön syvyys maassa, esineen tyyppi ja materiaali. Museoviraston arkeologit tarkastavat tiedot ja luetteloivat löydöt tietokantoihinsa. Vuoteen 2019 saakka raportointi tapahtui paperisen lomakkeen avulla, mutta nyttemmin käyttöön on otettu Museovirastossa Ilppari-niminen sähköinen raportointijärjestelmä³. Ilppari on klassinen lomakeperustainen tietojen syöttöjärjestelmä, jonka haasteena on mm. se, ettei järjestelmä sovellu joustavasti mobiiliin käyttöön löytöpaikalla luonnossa.

Löytösampo-hankeessa kehitettiin moderniin mobiiliin web-teknologiaan perustuva, uudenlainen raportointikonsepti ja sen prototyyppi *FindSampo Reporter* metallinetsijöiden raportointityötä helpottamaan (Hassanzadeh et al., 2019). Sen ja koko Löytösampo-järjestelmän tiedolliseksi perustaksi on luotu ontologiaperustainen infrastruktuuri arkeologiaan eri tavoin liittyvistä käsitteistä kuten esinetyypeistä, materiaaleista ja aikakausista (kuvassa pohjalla). Ideana on, että uuden arkeologisen tietoinfrastruktuurin avulla löytöihin liittyvää tietoa voidaan rikastaa linkittämällä sitä eri museoiden muihin arkeologisiin kokoelmiin ja palveluihin, esimerkiksi arkeologian termien selityksiä julkaiseviin sanastoihin, kuten Tieteen termipankkiin⁴, ja Wikipediaan. Järjestelmä hyödyntää myös tarjolla olevia kansallisia paikkatietopalveluita (kuvassa ylhäällä), kuten Maanmittauslaitoksen karttapalveluita ja Museoviraston

³ <https://www.museovirasto.fi/fi/ajankohtaista/ilppari>

⁴ <https://tieteentermipankki.fi/wiki/Termipankki:Etusivu/en>

muinaisjäännösrekisteriä⁵; näistä selviää mm. lailla suojellut alueet, joilla metallinetsintä on kiellettyä, ja erilaisten karttojen avulla on mahdollista tarkastella ja visualisoida jo tehtyjä löytöjä kartoilla ja suunnitella uusia etsintäretkiä luontoon. Suomalaista tietoinfrastruktuuria on yhdistetty yleiseurooppalaisessa AriadnePlus-hankkeessa⁶ käytettyihin kansainvälisiin ontologioihin, kuten Getty Säätiön AAT-sanastoon⁷ esinetyyppien osalta ja aikakausien esittämiseen, linkittämiseen ja visualisointiin tarkoitettuun Periodo-ontologiaan⁸. Näin suomalaisia löytöjä voidaan jatkossa linkittää vastaaviin löytöihin muualla Euroopassa, vaikkapa Viikinki-ajan esineistöön tai kolikkolöytöihin liittyen. SuALT-hankkeessa on tähän ajatukseen liittyen meneillään kokeilu suomalaisen löytödatan rikastamiseksi British Museumin Portable Antiquities Scheme -tietokannan⁹ aineistolla, johon sisältyy lähes miljoona raportoitua löytöä (tietuetta) Englannin ja Walesin alueelta.

Löytöihin liittyvät tiedot ja kuvat talletetaan Löytösampo-järjestelmässä semanttisen webin linkitetyn data tietomalleihin¹⁰ perustuvaan ns. tietämysgraafiin (knowledge graph) (Noy et al., 2019), joka julkaistaan W3C:n linkitetyn datan standardien, julkaisumallin ja parhaiden käytäntöjen mukaisella tavalla SPAQRL-palvelupisteessä (Heath & Bizer, 2011). Data on silloin käytettävissä näppärästi SPARQL-rajapinnan kautta, mikä mahdollistaa erilaisten sovellusten kehittämisen sekä datan tutkimisen data-analyttisin välinein, esimerkiksi YASGUI-editorin (Rietveld & Hoekstra, 2017) avulla tai Google Colab -työkalulla¹¹ Jupyter dokumentteina Python-kieltä käyttäen. Esimerkkinä sovelluskäytöstä SuALT-projektissa on kehitetty Löytösampo-portaali¹² (kuvassa oikealla), joka sisältää ”älykkäitä” välineitä löytöjen hakuja ja selailua varten sekä joukon data-analyttisiä työkalua löytöjen visualisointiin ja tutkimiseen. Löytösampo on uusi jäsen nopeasti kasvavassa Sampo-portaalien sarjassa (Hyvönen, 2021) – tällä kertaa sovellusalueena on arkeologia ja kansalaistiede. Löytösampo-portaalin demonstraattori avataan verkossa avoimeen koekäyttöön 17.5.2021 pidettävän Löytösampo-seminaarin yhteydessä.

Löytösampo-järjestelmän keskeinen osa on linkitetyn avoimen datan palvelu (kuvassa 1 keskellä), joka on julkaistu Linked Data Finland -alustalla¹³ (Hyvönen et al., 2014). Sen avoimen datan ja rajapintojen varaan voi kuka tahansa kehittää sovelluksia. Tästä esimerkkinä toimii semanttinen Löytösampo-portaali data-analyttisine työkaluineen ja visualisointineen¹⁴. Datapalvelun toinen käyttötapa on hyödyntää sen SPARQL-rajapintaa tutkimuskäytössä erilaisilla työkaluilla, kuten YASGUI-editorilla ja Google Colabilla. Datan voi myös ladata

⁵ https://www.kyppi.fi/palveluikkuna/mjreki/read/asp/r_default.aspx

⁶ <https://ariadne-infrastructure.eu/>

⁷ <http://www.getty.edu/research/tools/vocabularies/aat>

⁸ <https://perio.do/en/>

⁹ British Museum PAS-hankkeen esittely: <https://finds.org.uk>

¹⁰ <https://www.w3.org/standards/semanticweb/data>

¹¹ <https://colab.research.google.com/notebooks/intro.ipyn>

¹² Löytösampo-portaali on käytettävissä osoitteessa <https://loytosampo.fi> 17.5.2021 alkaen.

¹³ <https://ldf.fi>

¹⁴ Ks. esim. Löytösammon visulisoinnit artikkeleissa (Hassanzadeh et al., 2020) ja (Anafi et al., 2021)

palvelusta käytettäväksi ulkoisissa työkaluissa, kuten taulukkolaskimissa tai R-ympäristössä¹⁵ tilastollisia analyysejä varten.

Löytösampo on kehitetty osana Suomen Akatemian rahoittamaa SuALT-tutkimusprojektia (2017–2021). Se on uusi jäsen kulttuurialan Sampo-järjestelmien sarjassa¹⁶. Sampo-järjestelmien perustana oleva Sampo-malli (Hyvönen, 2021) ja siihen liittyvä, käyttöliittymien kehittämiseen tarkoitettu Sampo-UI ohjelmointikehys (Ikkala et al. 2021) ovat osoittautuneet joustavaksi tavaksi kehittää erilaisia ja eri aloihin liittyviä sovelluksia semanttisessa webissä. Tällä kertaa sovellusalueina ovat kansalaistiede ja arkeologia.

Projektiryhmä

Löytösammon tekninen ratkaisumalli on kehitetty Aalto-yliopiston tietotekniikan laitoksessa, yhteistyössä Helsingin yliopiston kulttuurien laitoksen tutkijoiden ja Museoviraston arkeologien kanssa. Yhteistyössä on ollut mukana myös Digitaalisten ihmistieteiden keskus HELDIG. Kehitystiimiin ovat Aallossa kuuluneet Eero Hyvösen johdolla Heikki Rantala, Esko Ikkala, Mikko Koho ja Jouni Tuominen sekä aiemmin Babatunde Anafi ja Pejman Hassanzadeh. Kiitokset yhteistyöstä Helsingin yliopiston tutkijoille (Suzie Thomas, Anna Wessman, Sisko Pajari ja Eljas Oksanen) ja Museoviraston edustajille (Ville Rohiola, Jutta Kuitunen, Minna Ryyppö, Ulla Rantala ja Vesa Hongisto). Projektin on rahoittanut Suomen Akatemia ja työssä on hyödynnetty CSC – Tieteen tietotekniikan keskus Oy:n palvelininfrastruktuuria.

Kirjallisuutta

1. Anafi, B. Mikko Koho, M. & Hyvönen E. (2020). Temporal Visualization and Data Analysis of Archaeological Finds: Case FindSampo. *Conference on Cultural Heritage and New Technologies (CHNT 25)*, Museum Stadt Archäologie Wien. Posters. <https://seco.cs.aalto.fi/publications/2020/anafi-et-al-findsampo-2020.pdf>
2. Gardiner, E. & Musto, R. G. (2015). *The Digital Humanities: A Primer for Students and Scholars*. Cambridge University Press, New York, NY, USA. <https://doi.org/10.1017/CBO9781139003865>
3. Hassanzadeh, P., Hyvönen, E., Ikkala, E., Tuominen, J., Thomas, S., Wessman, A. & Rohiola, V. (2020). FindSampo Platform for Reporting and Studying Archaeological Finds Using Citizen Science. *3rd Workshop on Humanities in the Semantic Web (WHiSe 2020)*, CEUR Workshop Proceedings, Vol. 2695, 33–40. <http://ceur-ws.org/Vol-2695>
4. Heath, T. & Bizer, C. (2011). Linked Data: Evolving the Web into a Global Data Space. *Synthesis Lectures on the Semantic Web: Theory and Technology*, Morgan & Claypool, Palo-Alto. <https://doi.org/10.2200/S00334ED1V01Y201102WBE001>

¹⁵ <https://www.r-project.org/>

¹⁶ Sampo-järjestelmistä tarkemmin sivulla <https://seco.cs.aalto.fi/applications/sampo/>

5. Hyvönen, E. (2021). Digital Humanities on the Semantic Web: Sampo Model and Portal Series. Submitted for peer-review. Pre-print:
<https://seco.cs.aalto.fi/publications/2021/hyvonen-sampo-model-2021.pdf>
6. Hyvönen, E., Tuominen, J., Alonen, M. & Mäkelä, E. (2014). Linked Data Finland: A 7-star Model and Platform for Publishing and Re-using Linked Datasets. *The Semantic Web: ESWC 2014 Satellite Events, Revised Selected Papers*, Springer-Verlag, 226–230.
https://doi.org/10.1007/978-3-319-11955-7_24
7. Hyvönen, E., Rantala, H., Ikkala, E., Koho, M., Tuominen, J., Anafi, B., Thomas, S., Wessman, A., Oksanen, E., Rohiola, V., Kuitunen, J. & Ryyppö, M. (2021). Citizen Science Archaeological Finds on the Semantic Web: The FindSampo Framework. *Antiquity, A Review of World Archaeology*. Accepted. Pre-print:
<https://seco.cs.aalto.fi/publications/2021/hyvonen-et-al-findsampo-2021.pdf>
8. Ikkala, E., Hyvönen, E., Rantala, H. & Koho, M. (2021). M. Sampo-UI: A Full Stack JavaScript Framework for Developing Semantic Portal User Interfaces. *Semantic Web*. Accepted. <http://www.semantic-web-journal.net/>
9. Noy, N., Gao, Y., Jain, A., Narayanan, A., Patterson, A., & Taylor, J. (2019). Industry-scale knowledge graphs: lessons and challenges. *Commun. ACM* 62(8), 36–43.
10. Rietveld, L. & Hoekstra, R. (2017). The YASGUI Family of SPARQL Clients, *Semantic Web* 8(3), 373–383. <https://doi.org/10.3233/SW-150197>
11. Thomas, S., Wessman, A., Ikkala, E., Tuominen, J., Koho, M. & Hyvönen, E. (2019). (co-)Creating a Sustainable Platform for Finland’s Archaeological Chance Finds: The Story of SuALT. *Digital Heritage and Archaeology in Practice* (Ethan Watrall and Lynne Goldstein (eds.)), University Press of Florida, December.
12. Thomas, S., Wessman, A., Tuominen, J., Koho, M., Ikkala, E., Hyvönen, E., Rohiola, V. & Salmela, U. (2018). SuALT: Collaborative Research Infrastructure for Archaeological Finds and Public Engagement through Linked Open Data. *Digital Humanities in the Nordic Countries 3rd Conference (DHN 2018), Book of Abstracts*.
13. Wessman, A., Thomas, S., Rohiola, V., Koho, M., Ikkala, E., Tuominen, J., Hyvönen, E., Kuitunen, J., Parviainen, H. & Niukkanen, M. (2019a). Citizen Science in Archaeology: Developing a Collaborative Web Service for Archaeological Finds in Finland. *Transforming Heritage Practice in the 21st Century: Contributions from Community Archaeology* (John Jameson and Sergiu Musteață (eds.)), Springer, 337-352.
14. Wessman, A., Thomas, S., Rohiola, V., Kuitunen, J., Ikkala, E., Tuominen, J., Koho, M. & Hyvönen, E. (2019b). A Citizen Science Approach to Archaeology: Finnish Archaeological Finds Recording Linked Open Database (SuALT). *DHN 2019 Digital Humanities in Nordic Countries. Proceedings of the Digital Humanities in the Nordic Countries 4th Conference*, CEUR Workshop Proceedings, Vol. 2364, 469-478.
<http://ceur-ws.org/Vol-2364>