

## Kuidas jõuda looduse digitaalse teisikuni?

*Urmas Kõljalg*

Digitaalne teisik on lihtsalt öeldes arvutimudel, mis simuleerib reaalse maailma objekti või mingit protsessi. Avalikkusele on ilmselt üks kuulsamaid digitaalse teisiku näiteid NASA 1970. aastal kiiruga loodud mudel, mis simuleeris rikkis Apollo 13 kosmoselaeva. Eesmärk oli tuua selle meeskond turvaliselt tagasi maa peale. Termin Digital twin tuli esimest korda kasutusele küll alles 2011. aastal, aga teoreetiline ja praktiline töö oli kestnud juba vähemalt üle 50 aasta – alustades näiteks Alan Turingi uurimistööst.

Millised küsimused on olulised looduse digitaalse teisiku puhul? Looduse eripära võrreldes tehnoloogiaga on mudeli loomine millelegi, mida me ise loonud ei ole. Mistahes masin ehitatakse tänapäeval üldjuhul esmalt digitaalse teisikuna, mida saab testida ja uurida. Looduse puhul peame aga kõigepealt välja selgitama selle ehituse alates rakkudest kuni ökosüsteemideni. Teisisõnu on masinatega võrreldes tegevus ümber pööratud.

Lisaks on looduse ülesehitus äärmiselt keeruline. Selle põhielemendi ehk raku ülesehituse ja protsesside keerukuses saab igaüks veenduda, lugedes värskelt ilmunud omakeelset Sulev Kuuse ja Toivo Maimetsa rakubioloogia õpikut. Kui rakule lisada organism, populatsioon, liik ja ökosüsteem, saame süsteemi, mille üksikosadele puuduvad põhjalikud digitaalsed teisikud. Rääkimata nende elementide ühiselt töötavast mudelist ehk kogu looduse digitaalsest teisikust. Aga vajadus selle järele on oluline nii inimese kui ka muu eluslooduse tervise kaitsmiseks ning tulevikku puudutavate tõendus põhiste otsuste tegemiseks.

Looduse digitaalse teisiku aluseks on seega andmed selle ülesehituse ja seal toimuvate protsesside kohta. Eluslooduse andmeid koguvad juba sajandeid nii teadlased kui ka harrastusuurijad. Selleks, et nende andmete põhjal mudel ehitada, peavad need olema kõigile ligipääsetavad ehk avaandmed. Samuti on oluline, et andmed oleks omavahel seostatud ja masinloetavad.

Maailmas on viimase 40 aasta jooksul loodud mõned masinloetavate ja avaandmete nõuetele vastavad globaalsed andmearhiivid. Kõige tuntumad on kolm juhtivat geenipanka USAs, Euroopas ja Jaapanis ning OECD soovitusel loodud Global Biodiversity Information Facility (GBIF) keskusega Kopenhaagenis. Uurijatelt nõutakse üha sagedamini andmete talletamist just nendes digiarhiivides. Seetõttu on neist kujunenud kõige olulisemad andmebaasid maailma eluslooduse digitaalse teisiku loomiseks.

Paraku on neis talletatud info maht ja seostatus nagu tilk meres. Nende baasilt saab mere elik ökosüsteemide ehituse kohta küll teoreetilisi mudeleid luua, aga reaalsest objektist jäävad need tõenäoliselt väga kaugeks. Looduse digitaalse teisiku loomine on seega alles lapsekingades ja vajab korrastatud suurandmeid nii objektide (rakk, organism, liik, ökosüsteem) kui ka looduslike protsesside kohta.

Ilmselt ei ole kohe mõistlik tormata kogu maailma eluslooduse digitaalse teisiku ehitamise juurde. Ehkki teadlastel, poliitikutel, ametnikel ja ettevõtjatel on

ahvatlev anda lubadus luua loetud aastate jooksul eluslooduse digitaalne teisik, mis võimaldab teha reaalajas õigeid ja säästlikke otsuseid mistahes paiga kohta maailmas. Mõistlik on olla kannatlik ja võtta suund, kus teadlased arendavad esmalt väiksemate objektide või protsesside digitaalseid teisikuid. Näiteks kaitset vajavad liigid, lindude ränne, mikroorganismide ja taimede suhted põllumuldades, kaevandamise mõju elurikkusele, üksiku looduskaitseala digitaalne teisik jne. Sellised uurimustööd annavad baasteadmuse, kuidas järkjärgult liikuda suurema ja keerukama digitaalse teisiku arendamise juurde. Selleks võib olla linna, regiooni või ka ühe riigi digitaalne teisik.

Eesti on kindlasti koht, kus saaks tema väiksuse ja kõrge digikultuuri tõttu tekkida üks esimesi riigiüleseid eluslooduse digitaalseid teisikuid. Aga siin on mitmeid probleeme, kus head lahendused veel puuduvad.

Üks on seotud tarkvara ja teine andmetega. Rahvusvaheliselt puuduvad standardid, kuidas eluslooduse digitaalse teisiku tarkvara arendada viisil, mis võimaldab tulevikus luua olemasolevate teisikute põhjal globaalse digiteisiku. Teine probleem on, kuidas hallata ja analüüsida eluslooduse keerukaid suurandmeid. Siin on ilmne vajadus tehisarul põhinevate meetodikate järele.

Tehisaru peab ilmselt tagama nii andmete kvaliteedi kontrolli kui ka indiviidide liigipõhise määramise ning tegema need andmed kõigile kättesaadavaks. See eeldab kõigi looduse andmeid koguvate osapoolte – ülikoolide, riigiasutuste ja ettevõtete – ühistööd andmete kättesaadavaks tegemisel. Arvan, et me oleme selleks koostöök valmis.

[Ilmunud ajalehes Postimees 4. novembril 2023](#)