

Luonnontieteellisten museokokoelmien digitointi

Strategia ja toimintasuunnitelma 2010-2015

Toimittaneet

Veli-Pekka Pelkonen, Hannu Saarenmaa ja Nina Laurenne

Helsingin yliopisto, Luonnontieteellinen keskusmuseo

31.12.2009

Sisällys

1 Johdanto.....	5
1.1 Toimeksianto.....	5
1.2 Digitoinnin tarkoitus: Saatavuus ja pitkäaikaissäilytys.....	6
1.3 Digitoinnin tarve biodiversiteetin suojelun kannalta.....	6
1.4 Kansalliset ja kansainväliset sitoumukset.....	7
1.5 Digitoinnissa käytettäviä määritelmiä.....	8
2 Katsaus nykytilanteeseen.....	9
2.1 Tietojärjestelmät ja aineistojen saatavuus.....	9
2.2 Digitointityö meillä ja muualla.....	10
2.3 Kokoelmia hallinnoivat organisaatiot ja niihin tehty aineistokysely.....	11
2.4 Digitoinnin tilanne kokoelmittain.....	12
2.5 Käyttäjien tarpeet digitoidulle aineistolle.....	17
2.5.1 Tutkijoiden käyttämät tietoaineistot.....	18
2.5.2 Tietoaineistojen kartuttaminen.....	18
2.5.3 Digitaalisten aineistojen merkitys tutkimuksessa.....	18
2.6 Vaihtoehdot.....	19
3 Visio.....	20
4 Tavoitetila.....	22
4.1 Digitoitavan aineiston priorisoinnin periaatteet.....	22
4.1.1 Käyttäjälähtöinen priorisointi.....	23
4.1.2 Aineistolähtöinen priorisointi.....	23
4.2 Museokohtaiset toimenpiteet	26
4.3 Tietojärjestelmien kehittäminen.....	29
4.3.1 Tietoarkkitehtuuri.....	29
4.3.2 Saatavuus.....	30
4.3.3 Pitkäaikaissäilytys.....	31
4.4 Osaamisen kehittäminen ja laadun varmistaminen.....	31

4.4.1 Makroekologian ja biodiversiteetti-informatiikan opetus.....	31
4.4.2 Laadun varmistaminen.....	32
4.4.3 Tiedostointikeskuksen käynnistyshanke Joensuussa.....	34
4.5 Tilakysymykset ja voimavarojen kohdentaminen.....	35
4.6 Digitointityön voimavarat ja niiden hallinnointi.....	36
4.6.1 Rahoituslähteet.....	37
4.6.2 Rahoituksen kohdentaminen.....	38
4.7 Yhteistyön kehittäminen.....	39
5 Toimenpideohjelma.....	41
5.1 Rahoitusmekanismi.....	42
5.2 Henkilöstökustannukset.....	43
5.3 Infrastrukturi ja kokonaiskustannukset.....	44
6 Lähdeluettelo.....	45

Tiivistelmä ja alkusanat

Luonnontieteellinen keskusmuseo (LTKM) yhdessä muiden luonnontieteellisten museoiden kanssa sopi keväällä 2009 kansallisen luonnontieteellisen materiaalin digitointityön suunnitteluhankkeen aloittamisesta. Vuoden 2009 alussa opetusministeriö myönsi rahoituksen suunnitteluhanketta varten, jonka tavoitteena oli kartoittaa nykyinen digitoinnin tilanne sekä luoda strategia ja toimintasuunnitelma valtakunnallisen digitointityön organisointia ja käynnistystä varten vuosille 2010 – 2015.

Suunnitteluryhmä on tarkastellut vaihtoehtoja ja esittää vision, jonka mukaan kaikki maamme tärkeimmät kokoelmat olisivat digitoitu vuoteen 2035 mennessä. Tavoitetaan pääsemiseksi tehdään 15 suositusta ja esitetään noin yhden miljoonan euron vuotuista budjettia, joka olemassa olevien voimavarojen uudelleen kohdentamisella sekä yhteistoiminnan ja osaamisen lisäämisen avulla hankittaisiin laajasta joukosta erilaisia rahoituslähteitä. Kokoelmat ja niiden digitaalinen saatavuus tulee nähdä tärkeänä osana kansallista ja kansainvälistä tutkimusinfrastruktuuria.

Tässä esitetty strategia ja toimintasuunnitelma, joka toimii valtakunnallisen digitointityön ohjenuorana, on suunnitteluhankkeen ohjausryhmän käsittelemä, ja se luovutetaan opetusministeriölle vuoden 2009 lopussa.

Suunnitteluryhmässä ovat työskennelleet Anders Albrecht (LTKM Eläinmuseo), Susanne Heiska (JOSEK Oy), Jukka Kettunen (Kuopion luonnontieteellinen museo), Janne Kotiaho (Jyväskylän yliopisto), Raino Lampinen (LTKM Kasvimuseo), Atte Moilanen (Helsingin yliopisto), Ilari Sääksjärvi (Turun yliopisto), Liisa Tuominen-Roto (Suomen ympäristökeskus), Risto Virtanen (Oulun yliopisto). Sen puheenjohtajana on toiminut Hannu Saarenmaa, sihteerinä Veli-Pekka Pelkonen ja avustajana Nina Laurene.

1 Johdanto

1.1 Toimeksianto

Vuoden 2009 alussa opetusministeriö, Helsingin yliopiston Luonnontieteellinen keskusmuseo (LTKM) ja muut luonnonhistorialliset museot sopivat yksityiskohtaisen strategian ja toimintasuunnitelman laatimisesta museoiden **K**Okoelmien ja tietovarantojen **D**igitoinnille, sähköiselle saatavuudelle ja pitkäaikaissäilytykselle vuosille 2010 – 2015. Hankkeelle otettiin nimeksi **DIKO**-suunnitteluhanke. Strategiasta ilmenee minkä vuoksi digitointia tehdään, nykytilan ongelmat, luonnontieteellisten museoiden nykyinen ja tavoiteltava rooli tietoyhteiskuntakehityksessä, museoiden lähtökohdat digitoinnin edistämiseksi sekä yhteisesti sovittava tavoitetila. Hankkeessa mukana ovat kaikki merkittävät Suomessa olevat yksityiset sekä julkishallinnon alaiset luonnontieteelliset museot, puutarhat sekä kokoelmat. Toimintasuunnitelma sisältää kuvauksen yhteistyöstä ja työnjaosta, toteutettavista toimenpiteistä, henkilö- ja muista voimavaroista sekä museoiden ja niiden omistajayhteisöjen sitoutumisesta suunnitelman toteuttamiseen. Toimintasuunnitelma laaditaan siten, että se on toteutettavissa varoin, joita organisaatiot ovat omien toimintamäärärahojensa ja muun tiedossa olevan rahoituksen puitteissa sitoutuneet suuntaamaan digitointiin, digitoitujen aineistojen säilyttämiseen sekä sähköisten aineistojen verkkosaatavuuteen. Tämän lisäksi suunnitelmasta käy ilmi, mihin mahdollisesti haettava lisärahoitus 2010 – 2015 suunnitellaan kohdennettavaksi.

Luonnonsuojeluhallinnon LAJI- ja SETI-työryhmät esittivät 2008 suuntalinjat aineistojen saatavuudelle. Luonnontieteellisen keskusmuseon (LTKM) asemaa selvitti opetusministeriön asettama työryhmä (2007), jonka keskeiset ehdotukset olivat:

- Luonnontieteelliset museot ja Suomen ympäristökeskus (SYKE) tiivistävät yhteistyötään. Tästä solmitaan sopimukset Helsingin, Jyväskylän, Oulun ja Turun yliopistojen ja Kuopion kaupungin kesken sekä em. yliopistojen ja Suomen ympäristökeskuksen kesken.
- Luonnontieteelliset museot perustavat harvoin käytetyille aineistoille yhteisen säilytystilan, jonka esikuvana on kirjastojen yhteinen Kuopiossa toimiva varastokirjasto.
- Luonnontieteellinen keskusmuseo vastaa luonnontieteellisten museoiden tietokantapalveluiden ja tietoteknisen osaamisen sekä tähän liittyvän työnjaon koordinoinnista.
- Pitemmällä aikavälillä tavoitteena on Helsingin, Jyväskylän, Oulun ja Turun yliopistojen sekä Suomen ympäristökeskuksen ja Kuopion kaupungin yhteinen Luonnontieteellinen museo, jolla on toimipisteitä eri puolilla Suomea.

Yksikään em. tavoitteista ei ole toistaiseksi edennyt. Tässä strategiassa otetaan kantaa näiden ehdotusten toteuttamiseen käytännössä sikäli kuin ne sivuavat digitointia. LTKM:aan on 2009 nimitetty kokoelmajohtoryhmä, jossa valmistellaan päätöksiä koskien mm. kokoelmien asemaa, sijoitusta ja digitointia.

1.2 Digitoinnin tarkoitus: Saatavuus ja pitkäaikaissäilytys

Luonnontieteellisten museokokoelmien pääasiallisena tarkoituksena on turvata aineiston tietojen säilyminen tuleville sukupolville sekä parantaa aineiston saatavuutta nykyisessä tietoyhteiskunnassa. Luonnontieteellisissä kokoelmilla ja niiden keräämiseen liittyvillä tiedoilla on sekä tieteellistä että kulttuurihistoriallista merkitystä.

Aineiston verkkosaatavuuden mahdollistamiseksi se on saatettava digitaaliseen muotoon, eli digitoitava. Digitoitavaa aineistoa kertyy jatkuvasti museoiden oman tutkimustyön tuloksena ja lahjoituskokoelmista. Kokoelmien aineistoa hoidetaan ja säilytetään hamaan tulevaisuuteen. Digitoidut aineistot vaativat ylläpitämistä. Tietotekniikkaa sinänsä ei voida ennakoida 50-100 vuoden tarpeisiin. Tiedostomuotojen pitäminen mahdollisimman yhtenäisenä ja tietojen dokumentoiminen mahdollisimman tarkasti metatiedoin sekä termistön kuvaaminen ja standardoiminen sanastoilla ja ontologioilla mahdollistaa kuitenkin tiedon tulkitsemisen ko. aikavälillä. Aineistojen aktiivinen ylläpito ja hyvä dokumentaatio mahdollistavat myös uusien tekniikoiden hyödyntämisen varhaisessa vaiheessa.

Digitoinnin yhtenä tavoitteena on integroida kansalliset kokoelmat osaksi kansainvälisiä tietojärjestelmiä. Aineiston maailmanlaajuinen saatavuus mahdollistaa tehokkaamman biodiversiteetin tutkimuksen. Global Biodiversity Information Facility (GBIF) on kansainvälisesti merkittävin organisaatio, joka säilyttää ja jakaa biodiversiteettitietoa ja myös suomalaiset aineistot on tarkoitus tehdä saataville tätä kautta. GBIF on todettu sopivaksi kanavaksi myös kansalliseen tiedonvaihtoon (SETI, 2008). Koska toimeksianto koski erityisesti biodiversiteettitutkimusaineistoa, on tässä vaiheessa keskitytty ainoastaan biologisiin näytekokoelmiin, ja geologiset kokoelmat on jätetty tarkastelun ulkopuolelle. Niihin on syytä palata erillisessä tarkastelussa myöhemmin.

Aineistojen saatavuus digitaalisessa muodossa vähentää näytteiden käsittelyä ja museoiden välisiä lainoja, mikä alentaa kustannuksia ja säästää henkilöstöresursseja. Ainutlaatuisten alkuperäisnäytteiden käyttö vähenee ja jo tämä pidentää niiden ikää ja säilyvyyttä. Aineiston sähköinen saatavuus tuo logistisia ja taloudellisia etuja, ja aineisto on saatavilla sen fyysisestä kunnosta ja paikasta riippumatta.

1.3 Digitoinnin tarve biodiversiteetin suojelun kannalta

Maa-ilmankannan ja YK:n rahoittamassa tutkimuksessa vuonna 2005 todettiin maailmassa olevan 16 000 uhanalaista eläinlajia ja 60 000 kasvilajia. Biodiversiteetin suurimmat uhat ovat ilmastonmuutos, vieraslajit ja kestämatön ympäristön käyttö. Luonnonympäristöjen häviäminen ja pirstoutuminen johtaa eliölaajien sukupuuttoon. Noin 60% tärkeistä ekosysteemeistä on heikentyneitä tai niitä käytetään kestämatmättöinä.

Ihmisen toiminnasta aiheutuneesta muutoksesta ollaan oltu tietoisia pitkään, mutta sen nopeus on yllättänyt tutkijat (Ympäristöministeriö, 2007).

Ihmisen suoran vaikutuksen lisäksi viime vuosina on herännyt huoli ihmisen toiminnasta välillisesti johtuvan ilmastonmuutoksen seurauksista. Ilmastonmuutoksen seurauksia on yhä aktiivisemmin seurattu mm. eri eliölajien levinneisyystietojen perusteella. Suuret GBIF-tietokannat ovat tässä eräs tärkeimmistä apuvälineistä (Peterson ja Soberon, 2006).

Luonnon monimuotoisuutta koskevat tiedot tulisi olla helposti saatavilla riippumatta siitä, missä ja kenen toimesta tiedot ovat kerättyjä. Digitoitu aineisto vuorostaan auttaa kohdistamaan seurantoja ja suojelutoimia tehokkaammin. Luonnontieteellisten kokoelmien digitaaliset aineistot ovat keskeisessä asemassa luonnonsuojeluviranomaisille mm. uhanalaisuusarvioinneissa.

1.4 Kansalliset ja kansainväliset sitoumukset

Euroopan unioni on kansainvälisen yhteisön mukana sitoutui vuonna 2002 toimillaan hidastamaan merkittävästi luonnon monimuotoisuuden häviämistä niin, että luonnon monimuotoisuuden köyhtyminen saadaan pysähtymään vuoteen 2010 mennessä (Convention of Biological Diversity 1994, Millenium Ecosystem Assesment 2005). Tästä tavoitteesta ollaan jo myöhässä. Tähän työhön liittyy myös erityisesti laji- ja havaintotietomateriaalin kattava tallentaminen ja saatavuus. Näitä tietoja voidaan käyttää erilaisten seurantatutkimusten materiaalina, joilla seurataan ympäristön muutoksia.

Kattavia ja pitkäaikaisia seuranta-aineistoja tarvitaan EU:n lainsäädännön sekä kansainvälisten sopimusveloitteiden toteutumista koskevaa raportointia varten. Tällaisia myös Suomen ratifioimia sopimuksia ovat mm. biologista monimuotoisuutta koskeva YK:n yleissopimus (Rio de Janeiro 1994), Eurooppaneuvoston päätös biodiversiteetin köyhtymisen pysäyttämistä (Göteborg 2001) ja YK:n päätös biodiversiteettikadon merkittävästä hidastamisesta maailmanlaajuisesti (Johannesburg 2002).

Kattavia ja pitkäaikaisia seuranta-aineistoja tarvitaan EU:n lainsäädännön sekä kansainvälisten sopimusveloitteiden toteutumista koskevaa raportointia varten. Tällaisia myös Suomen ratifioimia sopimuksia ovat mm. biologista monimuotoisuutta koskeva YK:n yleissopimus (Rio de Janeiro 1994), Eurooppaneuvoston päätös biodiversiteetin köyhtymisen pysäyttämistä (Göteborg 2001) ja YK:n päätös biodiversiteettikadon merkittävästä hidastamisesta maailmanlaajuisesti (Johannesburg 2002).

Valtioneuvosto hyväksyi vuonna 2006 Suomen luonnon monimuotoisuuden suojelun ja kestävän käytön strategian vuosille 2006 – 2015 (Ympäristöministeriö, 2007). Strategiassa edellytetään mm. että luonnontieteellisen keskusmuseon ja tärkeimpien maakunnallisten museoiden kokoelmat siirretään digitaaliseen muotoon. Strategiaa toteutetaan valtionhallinnon yleisten taloudellisten ehtojen puitteissa, ts. strategia ei merkitse merkittävää valtion kokonaismenojen lisäystä, mutta sen toteutuminen saattaa edellyttää

huomattavaakin uudelleen kohdentamista. Tähän liittyen seurantajärjestelmien kehittämistä on tarkastellut useampikin luonnonsuojeluhallinnon projektiryhmä, viimeisimpänä luonnonsuojeluhallinnon tuottavuusohjelman työryhmät SETI ja LAJI (2008).

Luonnontieteellisen aineiston digitoiminen noudattaa kansallisen tietoyhteiskuntastrategian 2007 – 2015 periaatteita edistämällä kansalaisten sekä yhteisöjen tietoyhteiskuntavalmiuksia ja tietoisuuden kehittämistä sekä parantamalla julkishallinnon palvelurakenteiden uudistamista asiakas- ja prosessilähtöisesti teknologiaa hyödyntäen (Valtioneuvosto, 2006).

INSPIRE-direktiivin soveltaminen on yksi Suomen tekemistä EU:n laajuisista sitoumuksista. Vaikka direktiivi puuttuu alkuvaiheessa vain paikkatietojen metatiedon julkaisemiseen, direktiivillä on toteuttamisvaiheessa vaikutusta paikkatietojen saatavuuteen. LTKM:n vastuulla oleviksi *paikkatietotuotteiksi* on määritelty lajien levinneisyyskartat sekä biomaantieteelliset alueet. Kokoelmatietojen paikkatiedon laatu on heikkoa ja lajien levinneisyyskarttojen tuottaminen vaatii tietojen mittavaa prosessointia. Karttojen tuottamiseksi tarvitaan mm. luotettava ja saatavilla oleva paikannimistö, joka kattaa myös historialliset paikannimien muutokset.

Luonnontieteellisen kokoelma- ja havaintoaineiston digitointi edistää seurantajärjestelmien käytettävyyttä sekä ns. e-sciencen kehitystä. Euroopan tutkimusinfrastruktuuristrategiafoorumi on laatinut 2006 tiekartan Euroopan tutkimusinfrastruktuurien rakentamis- ja päivitystarpeista (ESFRI, 2008). Luonnontieteellinen keskusmuseo mainitaan yhtenä kansallisen tason tutkimusinfrastruktuurina Suomen kansallisen tason tutkimusinfrastruktuurikartoituksessa (Tiekartta, 2009) ja aineistojen digitointi ja avoin saatavuus on nostettu tiekartassa esiin ensimmäisenä yleisenä suosituksena. Biodiversiteetin alalla tärkeäksi uudeksi infrastruktuuriksi on valittu LIFEWATCH/FinLTSER, minkä hankkeen toteutus nojautuu suurelta osin GBIF:iin.

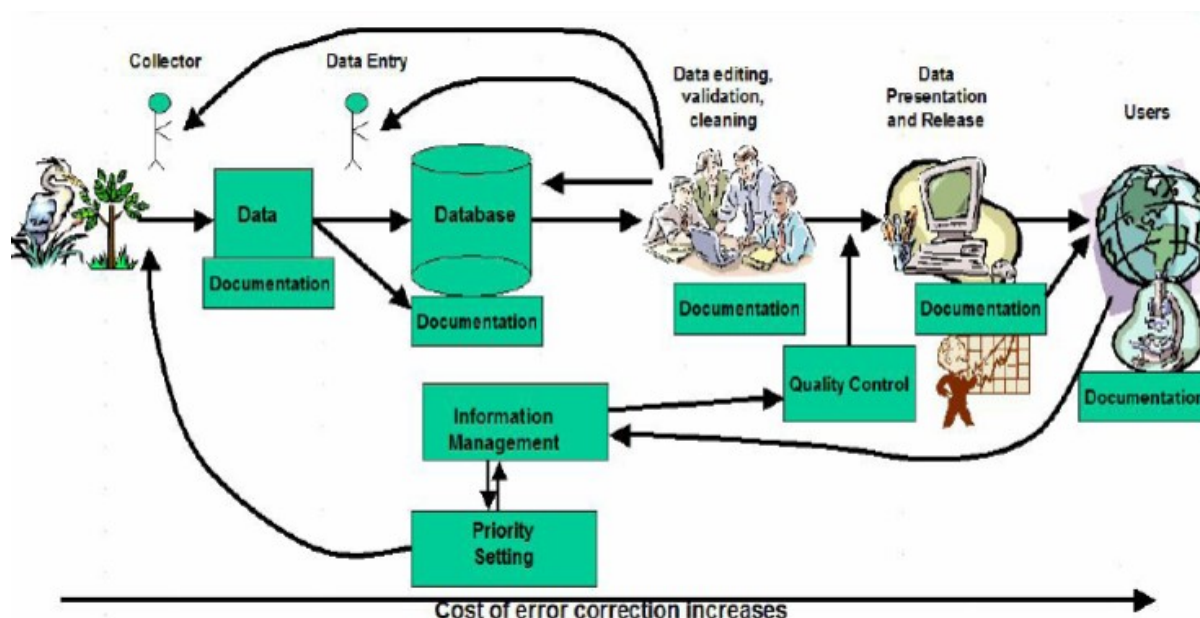
1.5 Digitoinnissa käytettäviä määritelmiä

Digitointiprosessi pitää sisällään hyvin monenlaisen aineiston käsittelyä ja erilaisia työvaiheita (ks. kuva 1). Työn ensimmäinen vaihe on näytteiden valinta, joka tehdään kokoelmakohtaisten priorisointikriteerien pohjalta. Näytteiden kunto tarkistetaan, ja etiketit ym. informaatio tallennetaan. Tarvittaessa etiketit korvataan uusilla, ja näytteet merkitään myös viivakooditunnuksilla. Mahdollinen kuvantaminen tehdään seuraavaksi kullekin aineistolle sovitulla tekniikalla. Lopuksi näytteiden säilytysastiat, nesteet, arkit, kuoret ym. vaihdetaan tarvittaessa uusiin, ja näytteet palautetaan säilytystiloihin.

Seuraavassa on selitetty näitä digitointiin liittyviä aineistotyypppejä, prosesseja sekä terminologiaa. Aineiston tallentaminen voi olla ei-sähköistä tai digitaalista luettelointia. Edellisessä tapauksessa aineistot ovat kokoelmayksilöiden etikettejä, käsin kirjoitettuja kortistoja tai muistiinpanovihkoja, joista viimeksi mainittuja on erityisesti vanhassa aineistossa. Tällaisessa muodossa materiaalista on n. 88 %, loput 12 % on sähköisessä muodossa luuteloitua eli digitoituna.

Kuvantaminen (engl. imaging) voi olla osa digitointiprosessia, mutta läheskään aina se ei ole tarpeen tai mahdollista. Yleisimmin kuvannetaan (skannataan tai valokuvataan) putkilokasvi- tai eläinnäytteitä, joiden kuvamateriaali tuo lisäinformaatiota näytetietoihin tai parantaa näytteiden käytettävyyttä. Kuvamateriaalista voidaan tulkita myös havaintoa koskevat tiedot, mikä useimmiten tapahtuu manuaalisesti kuten luettelointitietojen tallentaminen etiketeistä. Kuvannettua aineistoa on maassamme hyvin vähän.

Digitoinnilla käsitetään usein myös GIS-paikkatietoihin perustuvaa tallentamista, joka sisältää esim. löytöpaikkojen tai havaintoaluetta kuvaavien karttojen tallennuksen. Lisäksi digitaalisesti tallennettua materiaalia ovat mm. DNA-sekvenssitiedot.



Kuva 1. Digitointi kattaa keskeiset osat eliölajeja koskevan tiedon keruusta. Chapmanin (2005) mukaan.

2 Katsaus nykytilanteeseen

2.1 Tietojärjestelmät ja aineistojen saatavuus

Luonnontieteellisten museoiden kokoelma-aineistoille ei ole olemassa mitään yhtä tietojärjestelmää. Naapurimaistamme Ruotsissa ja Tanskassa tilanne on sama. Kummassakin arvioidaan parhaillaan valmiita sovelluksia (Specify ja KEmu). Norjassa on käytössä koko museosektorilla yhteinen järjestelmä (Musit, Oracle). Siitä ollaan tekemässä luonnontieteelliseen käyttöön uutta versiota. Tietojärjestelmäyhteistyöstä on keskusteltu maiden GBIF-solmujen kesken.

Suomalaisista biologisista näyte- ja havaintoaineistoista on olemassa LTKM:n hoitama BioCASE NoDITmetatietokanta (PHP, Microsoft SQL-Server). Siihen on tallennettu metatiedot jo noin 150 aineistosta, jotka sisältävät yhteensä noin 50 miljoonaa näyte- ja havaintotietoa.

Useat luontoseurannat tallentavat tietonsa Hatikka-tietokantaan (PHP, Microsoft SQL-Server), joka periaatteessa voi ottaa vastaan kaikkia biologisia havaintoja ympäri maailman. Kasvitieteen havainto- ja kokoelma-aineistoa on puolestaan tallennettu Kastikka-tietokantaan (Oracle), joka on museoiden yhteiskäytössä. Luonnontieteellisten harrastajayhdistysten aktiivisuus erityisesti havaintotietojen osalta on aina ollut merkittävää. Useilla yhdistyksillä on omat havainto- ja aineistotietokantansa (Liite 1). Hyönteistieteellisten seurojen Hyönteistietokantaan (Java, MySQL) on tallennettu jopa LTKM:n kokoelmatietoja. Maakunnallisista museoista ainoana Kuopion luonnonhistoriallisella museolla on oma kokoelmatietokantansa (PHP, MySQL). Ympäristöhallinnon koordinoimat eliöryhmät ylläpitävät lähes kaikki omia tietokantojaan, joissa on sekä kokoelma- että havaintotietoa. Aineistoja on siis saatavilla, mutta hajanaisesti ja heterogeenisessä muodossa.

Etenkin havaintotietokantojen tietoja on jo saatavilla myös GBIF-verkoston kautta, jonne tällä hetkellä on viety tiedot 33 tietokannasta ja niiden noin 1,8 miljoonasta näyte- tai havaintotiedosta. Maassa on tällä hetkellä neljä GBIF-palvelinta. Pohjoismaisessa vertailussa Suomi pitää tällä hetkellä perää aineiston GBIF-saatavuudessa.

Yksi tärkeimpiä suomalaisia digitaalisen aineiston hankkeita on opetusministeriön ylläpitämä Kansallinen digitaalinen kirjasto (KDK). KDK:ssa on kehitteillä asiakasliittymä ja pitkäaikaissäilytyksen järjestelmä ja se on osa eurooppalaista Europeana-järjestelmää. Jatkossa KDK:oon on tarkoituksena liittää tähän aineistoon myös luonnontieteelliset aineistot.

2.2 Digitointityö meillä ja muualla

Suomessa kansallista materiaalia on digitoitu erityisesti humanistisella ja kulttuurihistorian alalla. Mikkelissä on toiminut 50 henkeä työllistävä kansalliskirjaston mikrokuvaus- ja arkistointilaitos (DIMIKO) jo noin 20 vuotta.

Luonnontieteellisiä digitointihankkeita on käynnissä useissa eri maissa ympäri maailmaa. Erityisesti on mainittava alan pioneereina Costa Rican INBio, Meksikon CONABIO ja Kolumbian Humboldt Institute. Näissä biodiversiteetiltään rikkaissa maissa biologisen tiedon arvo kansallisena voimavarana ymmärrettiin aikaisin. Euroopassa on digitointiohjelmia ainakin Alankomaissa, Espanjassa, Ranskassa ja Sveitsissä. Pohjoismaissa mm. Norjassa ja Ruotsissa on aloitettu laaja-alainen kansallisten kokoelmien, arkistojen ja museoiden aineiston digitointi. Ruotsin Naturhistoriska Riksmuseetilla on tavoitteena saada kaikki materiaali digitaalisesti tallennettua 15 vuoden kuluessa. Ruotsissa investoidaan noin 18 miljoonaa kruunua vuodessa digitointiin ja aineistojen saatavuuteen. Norjassa toiminta on organisoitu niin, että julkishallinto ostaa digitointipalvelut Oslon yliopiston palveluita tuottavalta yritykseltä (DigForsk A/S). Toimintaan työllistetään Pohjois-Norjan kehitys- ja haja-asutusalueiden työttömiä (Matkaraportti, 2009).

Yliopistojen ja muiden julkishallinnollisten yksikköjen hallinnoimista näytekokoelmista on vasta pieni osa digitaalisessa muodossa. Työ on ollut aina vuoteen 2009 asti yksiköiden omien päätösten ja resursoinnin varassa. Rutiiniksi digitointi on muodostunut vain LTKM:n kasvimuseolla ja Jyväskylässä. Yhteisistä standardeista ei ole sovittu, joten aineisto on useimmiten hyvin vaihtelevaa niin tallennusmuotonsa kuin hallintaohjelmistonsa puolesta. Tämä hankaloittaa käyttöä ja on yhtenä haasteena tietokantojen integroitua käyttöä ajatellen. Myös aineistojen vaihteleva saatavuus ja mahdolliset käyttöoikeudet aiheuttavat ongelmia.

Harrastajien merkitys kokoelmiensa digitoinnissa on erittäin tärkeä. On tarve ohjeistukselle erilaisen aineiston luettelointia ja digitointia varten, jotta harrastajat voisivat saattaa aineistonsa digitaalseksi jo aktiivisena aikanaan. Tällaista ohjeistusta ollaan laatimassa mm. hyönteisten suojeluneuvottelukunnassa, jossa laaditaan ohjeita näyte- ja keräyskoodien sekä paikkamerkintöjen osalta. Vastaavanlainen ohjeistus tulisi saada luotua kaikkia eliöryhmiä sekä kaikenlaisia näytetyyppejä varten.

Kokoelmien näytetietokantojen osalta digitointitilanne on hyvin erilainen. Suurimman työosan muodostavat vanhat, digitoimattomat kokoelmat, joiden määrä museoissa lisääntyy jatkuvasti lahjoituksina saatavien kokoelmien myötä.

2.3 Kokoelmia hallinnoivat organisaatiot ja niihin tehty aineistokysely

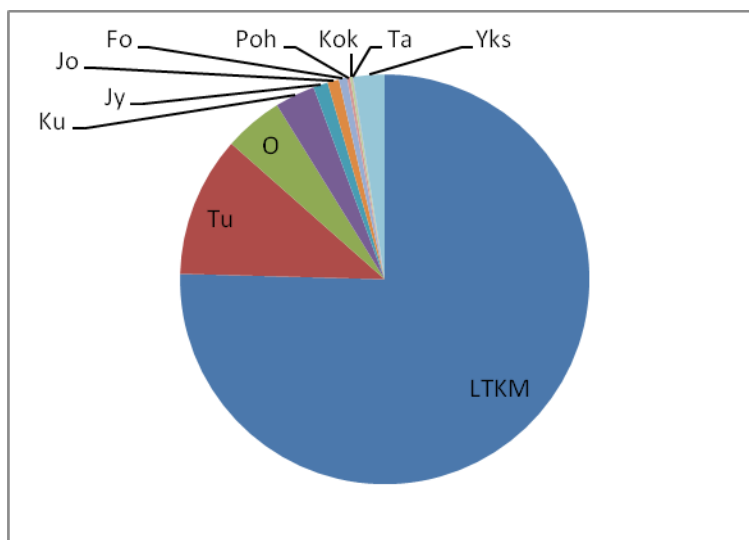
Suomessa julkishallinnollisia (yliopistot, kunnat, tutkimuslaitokset) luonnontieteellisiä kokoelmia on yhteensä noin 20 hallinto-organisaatiossa sekä lukuisia yksityishenkilöiden ja yhdistysten hallussa (Taulukko 1). Näytemäärältään suurimmat kokoelmat (Taulukko 2) ovat LTKM:n jälkeen Turussa, Oulussa, Kuopiossa, Jyväskylässä ja Joensuussa. Lisäksi on maakunnallisia luontokeskuksia ja museoita mm. Tampereella, Forssassa, Vaasassa, Kokkolassa, Porvoossa ja Rovaniemellä (Liite 2).

Taulukko 1. Kokoelmamäärät (arvio) organisaatioittain (* = koostuu erillisistä lahjoistukokoelmista).

Yliopistot	Kunnat	Muut
LTKM 8	Forssan museo 2	SYKE 24
Joensuu 3	Pohjanmaan museo 3 *	(Merentutkimuslaitos 9)
Jyväskylä 6	Tampereen museo 1	Metsähallitus 12
Kuopio 5	Luontokeskus Kieppi 3 *	RKTL 2
Oulu 7	Porvoon museo 3 *	
Turku 7	Lapin maakuntamuseo 3	yksityiset n. 1000

Suomen kokoelma-aineistoista tehtiin selvitys kesän 2009 aikana käyttäen hyväksi NODIT-metatietokannassa olevia tietoja sekä kokoelmakohtaisten kyselyjen avulla (Liite 3). Kysely koski kokoelmien näytemääriä, digitointiastetta sekä priorisointikriteereitä. Havainto- ja seuranta-aineistot jätettiin kyselyn ulkopuolelle, mikäli niihin ei sisältynyt myös näytetietoja. Kyselyyn saatiin sekä kirjallisia että suullisia vastauksia yhteensä 20 kokoelmaa hallinnoivalta organisaatiolta. Tästä selvityksestä tehtiin yhteenveto, jota käytettiin toimintasuunnitelmassa nykytilanteen kartoituksessa sekä kokoelmakohtaisten digitointisuunnitelmien laadinnassa (Liite 4).

Taulukossa 2 on listattuna kokoelmittain arviot kokonaisnäytemääristä sekä digitoitavien näytteiden määristä. Tätä yhteenvetotaulukkoa käytettiin yhdessä aineistokyselyn vastauksissa saatujen yksityiskohtaisempien tietojen kanssa lähtökohtana kokoelmakohtaisia digitoimissuunnitelmia laadittaessa. Lukumääräisten arvioiden lisäksi käytettiin myös tietoa kunkin kokoelman sisältämien näytteiden laadusta, digitointitarpeesta, priorisoinneista ja käytöstä.



Kuva 2. Hallintaorganisaatioiden osuudet koko digitoitavien näytteiden kokonaismäärästä (ks. Taulukko 2). LTKM - keskusmuseo, Tu - Turku, O - Oulu, Ku - Kuopio, Jy - Jyväskylä, Jo - Joensuu, Poh - Pohjanmaan museo, Fo - Forssan luontomuseo, Ta - Tampereen museo, Kok - Luontokeskus Kieppi (Kokkola), Yks - yksityiskokoelmat.

2.4 Digitoinnin tilanne kokoelmittain

Suomessa olevan luonnontieteellisen aineistojen määräksi voidaan eri aineistojen sekä kyselytutkimusten perusteella arvioida noin 50 miljoonaa tietoa, josta noin 20 miljoonaa on kokoelmanäytteitä ja loput havainto- ja seurantatietoa. Kokoelmanäytteistä on digitaalisesti luetteloitu vain 12 %, mutta havainto- ja seurantatieto on pääosin digitaalista, koska se on uudempaa, ja jo syntyyään sähköisessä muodossa. Jatkotarkastelussa olevat luvut koskevat näytekokoelmia.

Aineistoyhteenvedosta (Liite 4) sekä yhteenvetotaulukosta (Liite 2) voidaan todeta, että pienimpien kokoelmien osalta digitointityön loppuun vieminen on mahdollista jopa lähimpien vuosien aikana (Kuva 2). Suurimpien kokoelmien osalta tarvitaan pitkäjänteisempää suunnittelua, ja erityisesti keskusmuseon osalta työn voidaan odottaa kestävän jopa useita kymmeniä vuosia. Lisäksi varsinkin keskusmuseon digitoitavan ns. ”vanhan aineiston” määrä myös kasvaa johtuen jatkuvasti museolle lahjoitettavista kokoelmista.

Taulukko 2. Näytteiden määrä kokoelmittain sekä näyteyksilöiden täydellisen digitaalisen luetteloinnin vaatimat henkilöresurssit.

YKSIKKÖ	NÄYTTEITÄ YHT.	DIGITOIMATTOMIA NÄYTTEITÄ	HENKILÖTYÖVUOSIA	
			YHT.	/ 25 v.
LTKM	n. 14 milj.	n. 13 milj.	555	22
-hyönteiset	9,5 milj.	9,5 milj.	444	18
-muut selkärangattomat	550 000	550 000	28	1,1
-selkärangattomat	130 000	80 000	4	0,1
-putkilokasvit	1,7 milj.	1 milj.	44	1,8
-itiökasvit	n. 1,1 milj.	n. 1 milj.	44	1,8
-sienet (ml. jäkälät)	960 000	930 000	41	1,6
Turku	2 milj.	1 800 000	80	3
Oulu	1,2 milj.	770 000	34	1
Kuopio	650 000	520 000	23	0,8
Jyväskylä	290 000	190 000	8	0,3
Joensuu	150 000	150 000	7	0,2
Forssa	100 000	100 000	5	0,2
Pohjanmaa	90 000	42 000	2	0,1
Kokkola	30 000	30 000	1	>0,1
Porvoo	20 000	20 000	1	>0,1

Lappi	20 000	20 000	1	>0, 1
Tampere	10 000	10 000	0, 4	>0, 1
YHTEENSÄ	18 620 000	16 880 000	748	30

Helsingin yliopiston Luonnontieteellisen keskuksen (LTKM) hallinnoimia erillisiä kokoelmia on noin 20, joissa on yhteensä lähes 30 milj. näyte-, havainto- tai seurantatietoa. Kokoelmatietoja näistä on vajaat puolet.

Eläinmuseossa suurimman yksittäisen kokoelma-aineiston muodostavat selkärangattomat ja näistä erityisesti hyönteiskokoelmat, joissa on yhteensä noin 9,5 miljoonaa näytettä. Näistä on tiedostoituna vain n. 1 %. Digitointia on tehty viime vuosina erityisesti uhanlaisluokituksien aineistoja kootessa. Selkärangaskokoelmissa näytteitä on kaikkiaan n. 100 000, joista noin puolet on tietokannoissa.

Kasvimuseossa suurimmat kokoelmat ovat itäfennoskandiset ja pohjoismaiset kokoelmat, maailman putkilokasvikokoelmat sekä itiökasvi- ja sienikokoelmat. Näistä vain ensiksi mainittu on merkittävässä määrin digitoitu (41 %). Kokoelmanäytteistä erityisesti tyyppinäytteet ovat ahkerassa käytössä lajimääritystyössä sekä muussa tutkimuksessa. Kokoelmien suurta arvoa kuvaa se, että niistä tyyppinäytteitä on arviolta noin 50 000 kappaletta.

Kasvimuseolla onkin jo 1980-luvulla aloitettu putkilokasvien tyyppinäytteiden digitointi ja vuonna 2006 putkilokasvien tyyppinäytteiden kuvantaminen. Laajempi digitointihanke (kuvantaminen) alkoi 2009 jäkälänäytteiden tyypeistä amerikkalaisen Andrew W. Mellon -säätiön rahoituksella. Tavoitteena on digitoida noin 50 000 kasvi- ja sieninäytettä.

Keskuksella, kuten muillakin luonnontieteellisillä museoilla, on eläinnäytteiden lisäksi myös muuta arvokasta materiaalia, kuten karttoja tai valokuvia, joiden digitoinnin tarve tulisi arvioida muiden kokoelma-aineistojen ohessa. Erillisiä, kulttuurihistoriallisesti merkittäviä erilliskokoelmia on kolme, joissa on yhteensä reilut 100 000 näytettä.

Opetusministeriö myönsi valtion elvytyslisäbudjetista vuonna 2009 määrärahan LTKM:n vuoden 2009 digitointisuunnitelman toteuttamiseksi. Tämä oli ensimmäinen määräraha laatuaan. Sen avulla on digitoitu kasvimuseolla luovutetun Karjalan alueelta kasvi- ja sieninäytteitä sekä eläinmuseon äyriäis- ja monijalkaiskokoelma. Eläinmuseolla laajempaa digitointityötä on aloitettu historiallisista keräyspäiväkirjoista ja kokoelmaluetteloista ("kladikirjat"), joita on kuvattu sekä tulkittu tietokantoihin. Vuoden 2009 loppuun mennessä kladikirja-aineisto tulee myös yleisön saataville, kun siitä julkaistaan käyttöliittymäsovellutus. Tätä laajennetaan myöhemmin kattamaan myös muuta digitointihankkeessa digitoitavaa aineistoa. Hankkeessa

kehitettävät tietotekniset sovellutukset sekä käytettävät standardit sovitetaan yhteen KDK- ja GBIF-verkostojen kanssa.

Helsingin yliopistolla on kaksi GBIF-palvelinta, joista toinen keskusmuseolla. Näistä on toistaiseksi ollut saatavissa vain osa Hatikan aineistoja, Virtalan hyönteistietokanta sekä joitakin pieniä kokoelma-aineistoja.

Turun yliopiston hallinnoimien kokoelmien kokonaisnäytemäärä on noin 2 miljoonaa, joista on digitoituna noin 12 %. Turussa valtaosa digitoimattomista näytteistä on putkilokasveja (yli 1 milj.), joista tietokannoissa on vajaat 10 %. Hyönteisistä digitoimattomia on noin 600 000 näytettä, tietokannoissa neljäsosa. Turun maailmanlaajuisestikin merkittävistä hämähäkkieläinkokoelmista on tietokannoissa reilu kolmasosa. Turussa on ollut GBIF-palvelin vuodesta 2002, mutta siinä olevien aineistojen määrä on pieni, eikä sitä ole juuri päivitetty perustamisen jälkeen.

Lisäksi hallussa on useita historiallisesti arvokkaita Turun palolta pelastuneita kokoelmia Turun akatemian ajalta. Åbo Akademin kokoelmat sijaitsevat samassa paikassa kuin Turun yliopiston kokoelmat.

Oulun yliopiston hallinnoimat kokoelmat koostuvat kasvi- ja eläinmuseon sekä kasvitieteellisen puutarhan kokoelmista. Kokonaisaineistomäärä on arviolta reilut miljoona näytettä. Valtaosa myös Oulun kokoelmista koostuu hyönteisistä (n. 600 000) sekä putkilokasveista (n. 300 000). Digitointiaste on kotimaisten putkilokasvien osalta verraten korkea (35 %), mutta sammalten ja sienten (ml. jäkälät) osalta alhainen (3 %). Uhanalaisten lajien tietoja on tallennettu ympäristöhallinnon tietojärjestelmään. Digitoimattomia näytteitä on n. 750 000, joista suurin osa on selkärangattomia ja putkilokasveja.

Jyväskylän yliopiston hallinnoimia kokoelmia on 4, joissa yhteensä on noin 290 000 näytettä, näistä tietokannoissa kolmannes. Valtaosa digitoimattomista aineistosta on selkärangattomia, joita on arviolta 140 000 sekä putkilokasveja, joita on noin 60 000. Korkea digitointiaste on mahdollistanut kokoelmien viemisen GBIF-verkoston yleisesti saataviksi, missä suhteessa Jyväskylässä ollaan johtavassa asemassa Suomessa. Digitoitavaa materiaalia Jyväskylässä on noin 200 000 näytettä, joista valtaosa hyönteisiä. Arvioitu työvoimatarve on 9 henkilötyövuotta. Jyväskylässä on myös Keski-Suomen luontomuseon kokoelma.

Joensuun yliopiston opetus- sekä tutkimuskokoelmat koostuvat eläin- ja kasvinäytteistä. Opetuskokoelmissa on noin 100 000 näytettä eri eliöryhmistä. Erillisissä putkilokasvi- ja sienikokoelmissa on noin 55 000 näytettä.

Joensuussa on aloitettu luonnontieteellisen aineiston tiedostointikeskuksen perustamista varten kehittämishanke. Tavoitteena on luoda digitointityössä tarvittavan asiantuntemuksen sekä infrastruktuurin osaamiskeskus (ks. luku 4.4).

Kuopion luonnontieteellinen museo on suurena museona poikkeuksellisesti kaupungin omistama. Sen kokoelmat koostuvat eläin- ja kasvinäytteistä, joista selkärangattomia on 400 000, selkärankaisia neljässä erillisessä osakokoelmassa vajaat 40 000 sekä putkilokasveja 150 000.

Muista museokokoelmista voidaan mainita Forssan luonnonhistoriallinen museo, joka on kaupungin ylläpitämä. Sen kokoelmat koostuvat pääasiassa eläinnäytteistä, jotka jakaantuvat tasan selkärankaisiin ja selkärangattomiin. Forssan kokoelmia ei ole digitoitu ja kokonaisnäytemäärä on reilut 200 000.

Vaasassa sijaitsevan Pohjanmaan museon hallinnoima luonnontieteellinen kokoelma, Terranova – Merenkurkun luontokeskus, sisältää selkärankaisia ja selkärangattomia sekä putkilo- ja itiökasvinäytteitä. Pohjanmaan museon erikoisuutena on suuri historiallisten kokoelmien määrä, mikä lisää materiaalin kulttuurihistoriallista arvoa. Kaikkiaan näyteaineisto koostuu parista kymmenestä pienemmästä erilliskokoelmasta. Eliöiden lisäksi Pohjanmaan museossa on myös geologinen kokoelma, joka sisältää vajaat 700 näytettä. Aineisto on jaettu opetus- ja näyttelykokoelmiksi, ja yhteensä näytteitä on noin 90 000, joista tietokannoissa on noin puolet.

Kokkolassa toimivan luontokeskus Kiepin kokoelmiin kuuluu kolme lahjoituskokoelmaa, jotka koostuvat yhteensä 31 000 näytteestä.

Lapin maakuntamuseon yhteydessä on myös luonnontieteellinen kokoelma, joka sisältää reilut 20 000 näytettä. Näistä valtaosa on hyönteisiä, mutta mukana on myös selkärankais- ja kasvinäytteitä.

Tampereen luonnontieteellinen museo toimii osana kaupungin ylläpitämää museokeskus Vapriikkia. Luonnontieteellisiä näytteitä on noin 10 000, jotka ovat pääasiassa digitoimattomia. Porvoon museolla on myös luonnontieteellisiä kokoelmia, jotka on saatu lahjoituksina. Kokonaisnäytemäärä on 15 000 – 20 000 ja koostuu selkärangattomista, selkärankaisista sekä itiö- ja putkilokasveista. Kokoelmien näytteitä ei ole digitoitu.

Metsähallituksen kokoelmat ovat eri tutkimushankkeissa syntyneitä seurantatietoja. Aineiston yleinen saatavuus vaihtelee suuresti: osa on pelkästään tutkimusorganisaation käytössä, osa on julkaistu GBIF:ssä Jyväskylän yliopiston palvelimella. Kyselyyn vastattiin kahden kokoelman osalta, jotka kumpikin ovat selkärangattomien seurantatutkimuksia. Kumpikin on myös yleisissä tietokannoissa.

Suomen ympäristökeskuksen (SYKE) aineistoja on erillisinä tutkimusaineistoina sekä osana yliopistojen tutkimuskokoelmia. SYKE hallinnoimissa kokoelmissa näytetietoja on n. 10 000, jotka ovat pääosin tietokannoissa. Mikäli SYKE:n aineisto on osana jonkun tiedeorganisaation tutkimusta, on aineisto yhdistetty ko. yliopistokokoelmaan. Hertta-tietokannassa on reilut 100 000 havaintotietoa uhanalaisista lajeista. NODIT:issa on tiedot parista kymmenestä SYKE:n hallinnoimista kokoelmasta. Kokoelmat ovat pääosin omien organisaatioidensa käytössä, tosin osa on julkaistu myös yleisissä tietokantaportaaleissa sekä GBIF:ssä. Tiedostoimatonta materiaalia on vain vähän, ja sekin siirtyy tietokantoihin pääasiassa tutkimusprosessin myötä.

SYKE:n kokoelmissa ovat mukana myös aikaisemmin erillisenä organisaationa toimineen Merentutkimuslaitoksen aineisto.

Metsähallituksen hallinnoimia kokoelmia NODIT:issa on kaksitoista. Aineisto koostuu pelkästään havainnoista, jotka koskevat mm. uhanalaisten lajien seurantatutkimuksia. Tiedot ovat joko yksityisissä tai yleisissä tietokannoissa. Aineiston yleinen saatavuus vaihtelee suuresti: osa on pelkästään tutkimusorganisaation käytössä, osa on julkaistu GBIF:ssä Jyväskylän yliopiston palvelimella.

Riista- ja kalatalouden tutkimuskeskuksen (RKTL) tilanne lienee sama kuin edellisten tutkimusorganisaatioiden. Kyselyvastauksia tuli yksi, joka koski pitkäaikaisseuranta-aineistoa, joka sisältää 120 000 tietuetta 50 lajista. Digitoituna siitä on n. 90 %. Aineistoa käytetään biodiversiteetin seurantaan Suomessa ja EU:n alueella.

Yksityisistä tietokannoista suurimmat aineistot ovat harrastajien ylläpitämiä havaintotietokantoja (Birdlife, Virtalan Hyönteistietokanta) sekä yksityisiä hyönteis- ja kasvikokoelmia. NODIT-metatietokannassa olevien kokoelmatietojen perusteella yksityiskokoelmissa näytteitä on kaikkiaan noin 400 000, joista digitoituna on alle 10 %. Tiedot ovat kuitenkin monelta osin vanhentuneita, joten ne eivät anna kuvaa yksityiskokoelmien todellisesta tilanteesta.

Maassamme on noin 1000 – 2000 omaa kokoelmaa keräävää hyönteis-, kasvi- tai muuta luontoharrastajaa ja kussakin kokoelmassa voidaan arvioida olevan muutama tuhat näytettä. Näistä kokoelmista valtaosa päättyy tämän tarkastelun kattamana aikana julkisiin kokoelmiin, mikä nostaa digitoitavien yksilöiden määrää useilla miljoonilla. Tämän vuoksi olisikin tärkeää ohjeistaa yksityisiä kerääjiä itse luetteloimaan kokoelmansa, mikä tulevaisuudessa vähentäisi digitointityön määrää.

2.5 Käyttäjien tarpeet digitoidulle aineistolle

Biologisten aineistojen digitoinnin tarpeita selvitettiin kyselyllä, joka suunnattiin sellaista aineistoa tarvitseville tahoille. Näistä tärkeimmiksi todettiin eliötyöryhmien jäsenet ja makroekologian tutkijat. Kyselyä varten haastateltiin 13 henkilöä, joista 8 kuului johonkin ympäristöministeriön asettamaan eliötyöryhmään ja lisäksi haastateltiin 4 yliopistotutkijaa. Haastateltujen tutkimuskohteet kattoivat seuraavat eliöryhmät: hyönteiset (8), hämähäkkieläimet (2) ja sammaleet (2). Kirjallisuudesta etsittiin suomalaista tietoaineistoa laajasti hyödyntäviä tutkimuksia, jotka on julkaistu vuoden 2000 jälkeen (ks. kirjallisuus). Niitä löytyi viisi ja ne kaikki käyttivät samaa lähdeaineistoa (Huldén et al. 2000) ja olivat lähes saman tutkimusryhmän julkaisemia.

Kyselyssä selvitettiin, kuinka nykyiset tietoaineistot palvelevat heidän tarpeitaan ja kuinka niitä voitaisiin tulevaisuudessa kehittää. Nykyinen tilanne todettiin ongelmalliseksi, sillä digitoidut aineistot ovat erittäin puutteellisia ja hajallaan. Aineiston digitointi itsessään ei takaa sen saatavuutta, sillä läheskään kaikki olemassa olevat aineistot eivät ole yleisesti saatavilla vaan suppean ryhmän käytössä. Tärkeimmiksi syiksi käyttä

digitoituja aineistoja nousivat lajistoselvitykset, lajiston muutokset, eliölajien ajallinen seuranta ja elinympäristön mallintamiseen liittyvä tutkimus.

2.5.1 Tutkijoiden käyttämät tietoaineistot

Käytetyimpiä tietokantoja olivat yhteiseurooppalaiset tietokannat (mm. Fauna Europaea, 2004) ja eurooppalaisten museoiden ylläpitämät omat tietokannat (mm. The Natural History Museum, Lontoo). Lisäksi saatavilla on varsin laadukkaita yksityishenkilöiden ylläpitämiä tietokantoja, joiden tietosisältö on luotettavaa, mutta usein niiden rajoituksena on alueellisuus ja sen vuoksi suomalaisesta lajistosta ei juuri ole tietoa saatavilla. Tarjolla olevan tiedon käyttöä hankaloittaa myös mm. käytetyn nimistön erilaisuus, jolloin eri lähteistä saatavan tietoa-aineiston yhdistäminen on ongelmallista. Tietoa-aineiston käyttäjäkyselyssä korostuivat tiedon jäljitettävyyden ja luotettavuuden. Erityisesti vaikeasti määritettävien eliöryhmien osalta tiedolla on merkitystä vain, jos se on jäljitettävissä eliöyksilöihin.

Kotimaisista tietokannoista eniten käytetään Kasviatlasta (Lampinen & Lahti, 2009) ja Hyönteistietokantaa (ks. kirjallisuus), joista jälkimmäisen todettiin olevan liian puutteellinen ollakseen hyödyllinen. Lisäksi tiedon laatua pidettiin epäluotettavana. Kotimaisilla eläintietokannoilla katsottiin olevan merkitystä lähinnä vain harrastusmielessä. Luotettavimpia tietokantoja olivat ympäristöministeriön eliöryhmien ylläpitämät kannat, mutta osa näistä on suljettuja eikä siten tutkijoiden saatavilla.

Yhtenä tärkeänä osana digitointityötä on ajanmukaisen ja päivitetyn nimistötiedon saatavuus. Tätä varten on varsinkin ympäristöhallinnon taholta jatkuvasti toivottu ajantasaisesti ylläpidettävää nimistötietokantaa, josta saisi myös taksonelle nimistömuutoksista riippuvia tunnisteita, ja jota voidaan käyttää sekä digitointityössä että laajemminkin nimistöreferenssinä.

2.5.2 Tietoa-aineistojen kartuttaminen

Tietokantoihin tallennus on pääsääntöisesti yksittäisten tutkijoiden vastuulla ja tietoja säilytetään omalla työasemalla. Nämä aineistot ovat asiantuntijoiden keräämiä ja luotettavuudeltaan korkeatasoisia. Laji- ja levinneisyystiedon lisäksi tarjolla on lisääntyvässä määrin kuva-aineistoa ja muuta multimedia-aineistoa. Ongelmana nähdään infrastruktuurin puute ja tutkijat priorisoivat tutkimustyön ensisijaiseksi katsoen, ettei heillä ole tarvittavaa tietotaitoa aineistojensa julkaisuun. Tutkijat olivat valmiita luovuttamaan omaa aineistoa saataville, mikäli siihen on helppokäyttöiset työkalut tarjolla. Biologisten aineistojen digitaalisen saatavuuden todettiin nimenomaan helpottavan tutkimusta.

2.5.3 Digitaalisten aineistojen merkitys tutkimuksessa

Museokokoelmien tyyppiyksilöiden digitointia pidettiin ensisijaisen tärkeänä, sillä tieteelle uutena kuvatun lajin on oltava yksilöitävissä kyseiseen eliöyksilöön. Kansainvälisen käytännön mukaan tyyppiyksilöiden on oltava

tutkijoiden saatavilla. Tieteelle uuden lajin se yksilö, johon lajikuvaus sidotaan, on holotyyppi, ja holotyyppiä on vain yksi kappale. Kaikki sen kaksoiskappaleet ovat isotyyppiä. Kaikkia lajeja ei kuitenkaan ole kunnolla tyypitelty, mikä aiheuttaa sekaannuksia. Tilanne paranisi huomattavasti jo siitä, että tyyppiyksilöitä koskevat tiedot olisivat yleisesti saatavilla. Tyyppiyksilöiden etikettitietojen digitoiminen vähentää tieteellisesti arvokkaiden yksilöiden fyysistä käsittelyä, johon sisältyy rikkoutumisen vaara. Kuvantaminen toisi yksilöt kaikkien tutkijoiden ulottuville ja vähentäisi museoiden välisiä lainoja.

Tiedon digitaalista tallentamista pidettiin välttämättömänä, sillä sen avulla selvitetään lajien esiintyvyyttä ja tehdään eliömaantieteellisiä analyysejä. Ympäristötekijöiden vaikutus eliöihin on tärkeä tutkimusala, mutta toistaiseksi aineistojen riittämättömyyden vuoksi, tutkimus on keskittynyt maantieteellisesti pienille alueille. Makroekologia on tutkimuksen ala, jossa selvitetään eliön ja sen elinympäristön suhdetta laajalla maantieteellisellä alueella. Biodiversiteetin tutkimus perustuu paljolti mallintamiseen ja yhä laajenevassa määrin on ymmärretty lajitiedon osuus tutkimuksessa. Mm. ilmastonmuutoksen vaikutuksia eliöeläimistöön voidaan mallintaa vain, jos lajitietoa on riittävästi saatavilla. Toistaiseksi laajamittaiseen mallinnukseen soveltuvaa aineistoa on saatavilla vain muutamista eliöryhmistä, ja lisäksi selvitykset aineistojen riittävydestä makroekologiseen tutkimukseen ovat vasta karttumassa (Komonen et al. 2009). Suomalaista koko maan kattavaa perhosaineistoa (Huldén et al. 2000) on käytetty mm. tutkittaessa, kuinka tietyt biologiset ominaisuudet vaikuttavat lajien levinneisyyteen ja mikä on niiden osuus lajiston säilymisen kannalta (Mattila et al. 2009, 2008; Päivinen et al. 2005). Eteläisten lajien levittäytymistä pohjoiseen (Mitikka et al. 2008) on niin ikään tutkittu perhosfaunalla. Aikatiedon osuus korostuu tutkittaessa muutoksia eliöeläimien vuosittaisessa kierrossa. Kattavaa fenologia-tietoa on olemassa, mutta se ei ole yleisesti saatavilla. Museoaineistot ovat tärkeä resurssi myös invasiivisten lajien tutkimuksessa ja seurannassa.

Museoaineistojen digitoinnin avulla moninkertaistetaan saatavilla olevan tiedon määrä useista eri eliöryhmistä ja tällöin voidaan selvittää myös aineistojen maantieteellistä keskittymistä ja suunnata tutkimusta vähemmän kartoitetuille alueille. Museoaineistot ovat kansallisuusomaisuutta ja yleisesti pidettiin tärkeänä, että myös luonnontieteellisiä aineistoja digitoidaan kaikkien saataville kuten muutakin kansallisuusomaisuutta.

2.6 Vaihtoehdot

Digitoinnin toteuttamiselle voidaan nähdä useita eri **skenaarioita**, joilla on erilaiset kustannukset ja seuraukset:

1. Jatketaan nykyisellä tavalla ilman erityistä resursointia tai uudelleenjärjestelyitä. Digitointi otetaan hieman laajemmin osaksi normaalia toimintaa kaikissa museoissa. Sitä tehdään tilaustöinä, esim. ympäristöhallinnon uhanalaisuusarvioinneissa. Nämä sekä tyyppinäytteet saadaan digitoitua ensi vuosikymmenellä. Pienempien museoiden aineistot saadaan digitoitua ehkä muutaman vuosikymmenen sisällä. LTKM:n aineistoja, varsinkaan eläinmuseolla ei koskaan. Sen vuoksi tätä vaihtoehtoa ei voida pitää hyväksyttävänä.

2. Kuten edellä, mutta siirretään henkilökuntaa muista tehtävistä digitointiin, niin että kaikki tärkeät kokoelmat saadaan digitoitua esim. 25 vuoden sisällä. Laskelmien mukaan tämä edellyttäisi noin 30 henkilön tehtävien vaihtamista, mikä koskisi 20 % museoiden nykyisestä henkilökunnasta. Museoiden henkilövoimavaroja on muutenkin vähennetty ja tätä ei voida pitää mahdollisena. Yleistä 20 % tasokorotusta museoiden voimavaroihin ei myöskään pidetä mahdollisena ainakaan yliopistobudjetista. Kyse on kuitenkin valtakunnallisesta tutkimusinfrastruktuurista ja kansalliskokoelmasta, jonka ylläpito kuuluu yleiseen sivistystehtävään eikä yliopistojen perustehtäviin.
3. Edellisten välimuoto, jossa tavoitteeksi asetetaan, että vain noin tarkkaan valittu kolmasosa tärkeimmistä kokoelmista on digitoitu 25 vuoden sisällä. Tässä vaihtoehdossa digitointi on otettava laajemmin osaksi koko museohenkilöstön toimenkuvaa. Yksittäisen museohenkilökunnan jäsenen tutkimukseen käytettävissä oleva aika vähenee, mutta kokoelmiin perustuvan tutkimustyön tehokkuus kokonaisuutena lisääntyy. Kokoelmatyöstä saatava tieteellinen meriitti on otettava huomioon, jotta tämä vaihtoehto voisi toimia. Lisäksi on uudelleen kohdennettava voimavaroja digitointiin esimerkiksi tilavuokrista, joissa säästöt tulevat mahdollisiksi kokoelmien etäkäytön tullessa mahdolliseksi.
4. Luodaan hankeperustainen rahoitusmekanismi digitoinnille, mikä mahdollistaa kokoelmien digitoinnin hajautetusti eri museoissa esimerkiksi 25 vuoden aikajänteellä. Tässä mallissa nojaututaan olemassa olevaan fyysiseen infrastruktuuriin. Tämä maksaisi noin 1 miljoonaa euroa vuodessa ja kaikkiaan siis noin 25 miljoonaa euroa.
5. Luodaan erillinen fyysinen tutkimusinfrastruktuuri digitoinnille, joka huolehtisi pääosasta digitoinnista 25 vuoden aikajänteellä, mutta keskitetysti. Digitoinnin osalta tämä maksaisi saman kuin edellinen, eli noin 1 miljoonaa € vuodessa ja kaikkiaan siis noin 25 miljoonaa euroa. Sen lisäksi tulisi fyysisen infrastruktuurin kustannus.

Digitoinnin toteutuminen edellyttää, että museoiden perustoiminnan resurssit turvataan joten vaihtoehdot 1 ja 2 eivät ole hyväksyttäviä. Digitointi voidaan siis toteuttaa vain kohtien 3 – 5 mukaisesti tai niiden yhdistelmänä.

3 Visio

Tämän strategian tavoitteena on, että vuoteen 2015 mennessä on käynnistetty prosessit ja sen tutkimusinfrastruktuurin rakentaminen, joilla kaikki tärkeimmät kokoelmat saadaan digitoitua 25 vuodessa, eli vuoteen 2035 mennessä. Tämän aikajänteen arvioidaan olevan vielä riittävän lyhyt, jotta digitointia voidaan suunnitella kokonaisuutena, mutta kuitenkin riittävän pitkä urakan realistiseksi toteuttamiseksi.

Tämän kehityksen osana vuoteen 2015 mennessä digitoinnin ja saatavuuden rahoitusmekanismi on saatu toimintaan. Kokoelmien digitoinnille voi vuosittain hakea määrärahoja eri lähteistä, jotka on mitoitettu riittäväksi tavoitteen saavuttamiseksi em. 25 vuoden aikajänteellä.

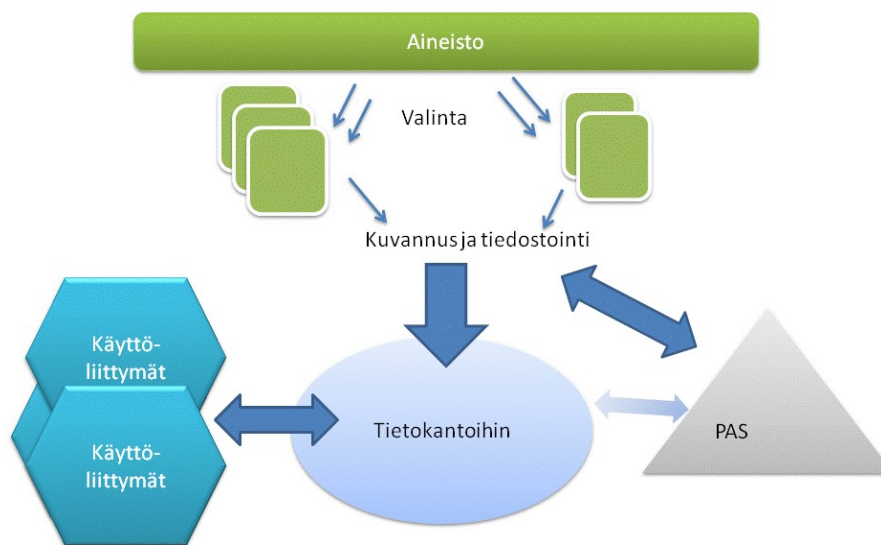
Julkisen tutkimusrahoituksen saajille on määritelty velvoite saattaa uusi tutkimusaineistonsa yleisesti saataville soveltuvien osin. Mikäli kokoelma-aineistoja käytetään, hankerahoitukseen varataan oma digitointiosuus osana kokonaisrahoitusta.

Osatavoitteena vuoteen 2015 mennessä on luotu toimintatavat, joilla digitointityö etenee. Siihen mennessä on myös saatu digitoitua ensimmäisiksi priorisoidut näytteet (eniten käytetyt, tyyppinäytteet, kladikirjat, uhanalaiset sekä pienet yksittäiskokoelmat). Uusi aineisto digitoidaan osana rutiininomaista tutkimus- tai tiedonhankintaprosessia eikä se enää lisää kokonaisurakkaa.

Museoiden yhteistyö toimii nimenomaan osaamisen vaihdossa ja tietojärjestelmien yhteiskäytössä (Kuva 3). Museoilla, puutarhoilla ja yhdistyksillä on hyvät asiantuntijavalmiudet aineiston digitointia ajatellen. Tärkeimmistä lisätarpeista liittyen resursseihin sekä yhteisten toimintaperiaatteiden luomiseen on huolehdittu yhteisissä keskusmuseon kanssa sovituissa kokoelmakohtaisissa digitointisuunnitelmissa.

Biodiversiteetti-informatiikka on tullut osaksi yliopistollista biotieteiden opetusta. Digitointia varten rekrytoidaan henkilöstöä, jolla on kutakin työvaihetta ajatellen riittävät perustiedot. Henkilöstö koulutetaan erikseen tehtäviinsä. Erillisien, koulutetun hankehenkilöstön etuna on toiminnan suunnittelun ja ennakkoinnin helppous sekä aineiston käsittelyyn ja tallennukseen liittyvien riskien minimointi. Maassa toimii ainakin yksi huomattava digitoinnin osaamiskeskus.

Museoiden käytössä on tietotekniset sovellutukset ym. palvelut, joiden avulla voidaan joustavasti hallita kokoelmien tietoa-aineistoja. Mitään yhtä järjestelmää ei edellytetä, mutta halukkaat museot voivat halutessaan käyttää muiden museoiden järjestelmiä. Kaikista tietokannoista aineisto on avoimesti saatavissa GBIF- ja KDK-verkoston kautta ja muodostaa näin yhteisen virtuaalisen järjestelmän. Museot osallistuvat merkittävästi biodiversiteetin alan tiedonvaihtoon Suomessa ja kansainvälisesti.



Kuva 3. Digitointityön eri työvaiheet: pienet siniset nuolet kuvaavat näytteitä tai havaintoja, paksut nuolet osoittavat digitaalista dataa. (PAS = pitkäaikaissäilytys).

4 Tavoitela

4.1 Digitoitavan aineiston priorisoinnin periaatteet

Luonnontieteellinen kokoelma-aineisto voidaan jakaa vanhaan ja uuteen materiaaliin. Vanha materiaali i) on tyypillisesti vuosikymmenten aikana kertynyttä näytemassaa, johon liittyy usein myös kulttuurihistoriallisia tekijöitä. Tämä aineisto on pääosin ei-digitaalista. Huomattava osa kartunnasta on "uusvanhaa" eli ulkopuolisten lahjoittajien museolle antamia tai muista museoista vaihdossa saatuja "vanhoja" näytteitä. Uusi aineisto ii) syntyy taas tutkimustyön tuloksena ja on jo syntyjään digitaalista tai digitoidaan tutkimuksen osana. On tärkeää tarkastella näitä kahta erillisenä. Edellisen digitoimiseksi ei vielä ole olemassa rahoitusmekanismia, mutta jälkimmäisen digitointi on jo periaatteessa rahoitettu osana normaalia tutkimustoimintaa. Maamme museot ja tutkimuslaitokset ovat resursoituja nykyään vain kohdan ii) työtä varten. Jotta uusi syntyvä aineisto ei enää kartuttaisi digitointiurakkaa, näiden uusien aineistojen saatavuusjärjestelyt ja arkistointi olisi hoidettava ensi tilassa.

- **Suositus 1: Jo digitaalisen tutkimuksissa käytetyn kokoelma-aineiston saatavuuden varmistamiseksi laaditaan tietoarkisto. Tämä tietoarkisto voi olla hajautettu eri museoihin ja siihen talletettujen aineistojen tulee olla saatavilla GBIF-mekanismien kautta. Museot ja muut laitokset toteuttavat tämän arkistoinnin omilla toimintamenoillaan.**

- **Suositus 2: Vanhan kokoelma-aineiston digitointia varten perustetaan erityinen rahoitusmekanismi, koska sen digitointia ei katsota mahdolliseksi nykyisillä resursseilla.**

4.1.1 Käyttäjälähtöinen priorisointi

GBIF-järjestön työryhmässä ”Global Strategy and Action Plan for Mobilisation of Natural History Collections Data” (wiki.gbif.org/gbif/wikka.php?wakka=DiscussionDocument) on aineiston priorisoinnista ja valinnasta mainittu seuraavasti:

”... yksityiskohtaisen digitointityön lähtökohtana on käyttäjien tarve. Tästä johtuen digitoinnin priorisoinnissa on huomioitava menossa olevien tai suunniteltujen tutkimushankkeiden tarpeet sekä yhteiskunnalliset ja poliittiset velvoitteet (ratifioidut sopimukset jne). Tämän johdosta digitointia voikin tarkastella palvelujärjestelmänä, jonne asiakkaat voivat tehdä tilauksia.”

Tämä on suositeltavin tapa valita digitoitava aineisto, koska se sisältää myös potentiaalisesti digitoinnin rahoitusmekanismi. Se edellyttää, että käyttäjät tietävät mitä aineistoja on kokoelmista periaatteessa saatavilla ja millä kustannuksilla.

- **Suositus 3: Museot ylläpitävät julkisesti yhdenmukaisella tavalla saatavilla olevia metatietoja kokoelmistaan. Tietojen tulee olla riittävän tarkkoja paikan, ajan ja eliöryhmän suhteen, jotta käyttäjät voivat sen perustella päätellä mitä aineistoa on saatavilla. Museoiden tulee kuvata nettisivuillaan mahdollisesti tarjoamansa digitointipalvelut, digitoinnin tarvelähtöinen rahoitusmekanismi ja myös millä ehdoilla ja käytännöillä käyttäjät itse voivat digitoida kokoelmia. Näistä verkkopalveluista tehdään yksi yhteinen portaali.**

4.1.2 Aineistolähtöinen priorisointi

Toinen tapa valita digitoitava aineisto perustuu kokoelmien ominaisuuksiin. Tässä tapauksessa aineistoa digitoidaan ikään kuin varastoon, josta se (ehkä) joskus otetaan käyttöön. Tämän tavan huonoja puolia ovat se, että ei ole varmuutta että aineistoa joskus tarvitaan, ja että se vaatii erityisen rahoitusmekanismi. Hyvä puoli on se, että kun aineisto on valmiiksi digitoitu, sitä voidaan käyttää välittömästi ja pienempiinkin tarpeisiin.

Digitoitavan materiaalin priorisointia varten on aineistolle sovittava arviointikriteerit. Pienempien kokoelmien kohdalla tämä ei välttämättä ole tarpeen, sillä koko aineisto voidaan digitoida kerralla, kun taas erityisesti keskusmuseon aineisto vaatii kokonaisvaltaisempia suunnitelmia aikatauluineen.

Priorisointikriteereihin vaikuttavat aineiston sisällöllinen arvo, ikä, käyttö ja saatavuus. Sisällöllisiä kriteereitä ovat ainutlaatuisuus (esim. kotimaiset tai harvinaiset ja uhanalaiset lajit), tieteellinen merkittävyys (lajityypinäytteet sekä lajin sisäiset tyypit- ja populaationäytteet) sekä kulttuurihistoriallinen arvo. Käyttöön liittyen merkittävimpiä kriteereitä ovat käytön tiheys sekä rinnakkaisten tai verrokinäytteiden määrä.

Käyttökriteerien perusteella voidaan digitointityötä tehdä myös tarvelähtöisesti eli kysytyt näytteet digitoidaan käytön yhteydessä. Sähköisen aineiston osalta on arvioitava nykyisen formaatin käytettävyyden, yhteensopivuuden sekä päivitysaste.

Kokoelmien hallussa on myös ulkomailta lainassa olevia näytteitä. Ulkomaisten näytteiden käyttöön saattaa liittyä rajoituksia tai erityisvaatimuksia. Tällaisia ovat mm. tyyppinäytteet, joiden käyttö on rajoitettu ainoastaan esim. lajikuvausta varten. Erikoisnäytteitä varten on luotava erillinen ohjeisto, joka koskee näytteiden sekä niistä saadun tietomateriaalin käyttöä.

Erillisissä kokoelmissa olevan identtisen materiaalin osalta on arvioitava mahdollisuus aineiston yhdistämiseen. Priorisoinnissa on myös mahdollista toimia tarveharkintaisen periaatteen mukaisesti eli aineistoa voidaan digitoida kysynnän mukaan. Tämä koskisi lähinnä ns. toisarvoista aineistoa. Digitointi tulisi aloittaa kaikkein vanhimmasta sekä arvokkaimmasta aineistosta. Luonnontieteellisessä materiaalissa on otettava huomioon myös aineiston mahdollinen kulttuurihistoriallinen arvo.

Esimerkiksi LTKM:ssa on noudatettu tähän mennessä tehdyssä digitointityössä seuraavia yleisiä suuntalinjoja aineiston priorisoinnin suhteen:

- Työnkulkujen vaatima looginen järjestys
- Tämän kriteerin perusteella on ensimmäisenä aloitettu kortistojen ja keruupäiväkirjojen ("kladikirjat") digitointi.
- Museolle tulevat tietopyynnöt ja lainauspyynnöt
- Osa aineistosta on hyvin ahkerassa käytössä. Kokoelmien käyttäjiä ovat museoiden oman henkilöstön lisäksi ulkopuoliset tutkijat ja muut käyttäjät. Sekä aineiston saatavuuden että näytteiden säilyvyyden kannalta on digitoinnin priorisoinnissa huomioitava näytteiden odotettu käyttötarve.
- Museolle tulevien vierailijoiden tarpeet
- Näytteiden saatavuus sähköisesti verkossa mahdollistaa useinkin ainakin näytetietojen helpon tarkastelun, esimerkiksi ei tarvitse matkustaa lukemaan tietoja kokoelmista. Verkkosaatavuus lisää eri kokoelmien kiinnostavuutta ja tunnettavuutta ympäri maailmaa. Tällä perusteella sekä runsaalla käytöllä on priorisoitu ns. luovutettujen alueiden kasvi- ja sieninäytteiden digitointi.
- Museon oman asiantuntemuksen ylläpito
- LTKM henkilökunnasta merkittävä osa jää lähivuosina eläkkeelle. On varmistettava, että heidän hoitamansa kokoelmat, niiden luettelot, tutkimusaineisto ja muistiinpanot saadaan dokumentoitua, arkistoitua ja digitoitua asianmukaisesti.

- Edustavuus ja ainutlaatuisuus
- Erityisesti tyyppilajinäytteet sekä näytteet uhanalaisista lajeista muodostavat useiden kokoelmien merkittävimmän osan. Myös kulttuurihistorialliset näytteet ovat ainutlaatuisuudessaan korvaamattomia. Digitaalinen tallennus parantaa kysytyjen ja vaikeasti lainattavien näytteiden saatavuutta.
- **Suositus 4: Kukin museo ylläpitää kokoelmiensa ominaisuuksiin sovitettuja digitoinnin priorisointikriteerejä, pitkän tähtäimen digitointisuunnitelmaa ja hakee sen perusteella rahoitusta digitointityölle.**

Museoilla tulisi siis olla omat, kokoelmakohtaiset priorisointikriteerinsä, jotka perustuvat sekä käyttäjä- että sisältölähtöisiin seikkoihin. Mikäli tällaisia kriteereitä ei kokoelmilla ole, digitoinnin tulisi edetä seuraavissa vaiheissa:

1-vaiheessa digitoidaan a) tyyppinäytteet, b) uhanalaiset tai muuten harvinaiset lajit, c) ulkopuolisten hankkeiden tarvitsemat käytetyimmät näytteet sekä lainat, joiden digitoinnista saadaan selvää logistista etua. Kokoelmaluettelot ("kladikirjat") digitoidaan, koska ne ovat avain monien etenkin vanhempien näytteiden etikettien tulkintaan. Kokoelmien metatiedot tehdään yleisesti saataviksi.

2-vaiheessa digitoidaan museon omassa tutkimustyössä aktiivisessa käytössä olevat kotimaiset ja ulkomaiset näytteet sekä kulttuurihistoriallisesti arvokkaat kokoelmat.

3-vaiheessa digitoidaan loput tärkeät kokoelmat, joiden digitoiminen nähdään tulevan tutkimustyön kannalta tarpeelliseksi.

Näihin priorisointiluokkiin kuuluvat aineistot ja luvussa 2.6 kuvatut vaihtoehdot on taulukossa 3 esitetty ristikkäin. Tämä valottaa sitä, miten resursointi vaikuttaa digitointiin.

Taulukko 3. Eri skenarioissa ja kiireellisyysluokissa digitoittavien näytteiden arvidut määrät (tuhansia) vuosina 2010–2015. Vrt. liite 2.

	Skenario 1	Skenario 3	Skenario 5	Yhteensä, noin
1-vaihe	Tyypit 100, uhanalaiset 50, hankkeet 50, lainat 10	Tyypit 100, uhanalaiset 50, hankkeet 50, lainat 10, kladit 100	Tyypit 100, uhanalaiset 50, hankkeet 50, lainat 10, kladit 500	200–700

2-vaihe	Aktiivinen tutkimus 200, kultturi-historia 0	Aktiivinen tutkimus 500, kultturi-historia 200	Aktiivinen tutkimus 500, kultturi-historia 500	200-1000
3-vaihe	Muut tärkeät 0	Muut tärkeät 0	Muut tärkeät 4000	0-4000
Yhteensä, noin	400	900	5000	

4.2 Museokohtaiset toimenpiteet

Tarpeiden kartoittamiseksi on suunnitteluhankkeen aikana tehty aineisto- ja tarvekyselyt osapuolilta. Kyselyissä pyrittiin mahdollisimman kattavasti selvittämään kunkin yksikön aineiston määrä ja laatu, digitointiaste, digitointitarve, nykyisten mahdollisten tietojärjestelmien tila (formaatti, yhteensopivuus) sekä digitointiin mahdollisesti käytettävissä olevat resurssit.

Seuraavassa tarkastellaan kokoelmittain digitoitavien aineistojen järjestystä.

Luonnontieteellisen keskusmuseon kaikista kokoelmista digitoimattomien näytteiden määrä on noin 12 miljoonaa, näistä hyönteisiä on valtaosa (n. 10 milj.). Eläinmuseolla ensisijaisesti digitoitavia ovat tyyppinäytteet (n. 10 000) sekä uhanalaiset lajit ja kotimaiset näytteet, joita on noin puolet kaikista digitoitavista näytteistä. Näistä kaikista hyönteisnäytteiden kuvantamista ei nähdä ensiarvoisena tai edes nykYTEKNIKALLA mahdollisena, vaan oleellista on etikettitietojen sähköinen tallentaminen. Selkärankaisnäytteiden kuvantamisen tarve puolestaan vaihtelee eläinryhmittäin. Digitoinnin alkuvaiheessa tallennetaan kokoelmaluettelot.

Kasvimuseon aineistosta digitoitavia näytteitä on yhteensä arviolta 3 miljoonaa, joista reilut puolet ovat itiökasveja tai sieniä ja loput putkilokasveja. Ensimmäisessä vaiheessa digitoidaan tyyppinäytteet kaikista ryhmistä, alkaen niistä joita pyydetään lainaksi. Toisessa vaiheessa digitoidaan uhanalaisten lajien näytteet. Kolmannessa vaiheessa digitoidaan ne kotimaiset ja ulkomaiset näytteet, joiden digitointi on käytön ja saatavuuden vuoksi tarpeellista. Aluksi tallennetaan yleensä vain etikeissä olevat tiedot. Kasvinäytteistä kuvannettaviksi sopivia ovat lähinnä kasviarvit, joista tärkeimpiä tallennettavia tietoja ovat etiketit laji- ja paikkatietoineen. Itse kasvinäytteistä kuvannettavia olisivat mahdollisesti suuremmat, erikoisemmat ja arvokkaimmat näytteet, joita kaikista näytteistä on arviolta 10 %.

Digitoitavien näytemäärien arviot eri skenarioissa on esitetty liitteessä 2. Henkilöresurssitarve on kaikkiaan 555 henkilötyövuotta. Jos tämä jaetaan vision mukaisesti 25 vuoden ajalle, vaatii se vuotta kohden 22 henkilön suorittavan työpanoksen. Skenarion 5 mukaan vuosina 2010–2015 digitoidaan kaikista näytteistä noin 10 % eli n. 2 miljoonaa. Mikäli käyttöön ei saada merkittävää lisärahoitusta (ja joudutaan toimimaan luvun 2.6 skenarion 3 mukaisesti), vain noin 200 000 näytettä voidaan digitoida. Tällöin LTKM:n hyönteiskokoelman pääosa jää toistaiseksi digitoimatta.

LTKM ylläpitää myös luetteloita maassa tavattujen eliölajien nimistöistä. Niihin kohdistuu suuri kysyntä. Nämäkin ovat ensi vaiheessa digitoitavaa aineistoa ja niiden käsittely ja olemassaolevan tietokannan viimeistely vie arviolta kaksi henkilötyövuotta.

Turun yliopiston kokoelmissa on yhteensä digitoitavia näytteitä lähes 2 miljoonaa, joista puolet selkärangattomia eläimiä ja puolet kasveja. Lisäksi kokoelmien erikoisuutena ovat Åbo Akademin kokoelmat (n. 22 000 näytettä) sekä historiallisesti arvokkaat kokoelmat. Henkilöresurssitarve on 80 henkilötyövuotta, mikä jaettuna 25 vuodelle edellyttäisi 3 henkilön vuotuista työpanosta. Turussa eläinkokoelmista ensimmäisenä digitoidaan hämähäkkikokoelmat sekä hyönteisistä tyyppinäytteet (yhteensä n. 270 000 näytettä). Kasvinäytteistä ensimmäiseen vaiheeseen valitaan tyyppinäytteet jäkälistä (n. 30 000) sekä putkilokasveista tyyppinäytteet sekä historialliset näytteet (n. 150 000). Toisessa vaiheessa digitoitavia ovat kulttuurihistoriallisesti arvokkaat hyönteiskokoelmat, loput putkilokasveista sekä soveltuvin osin selkärangattomat näytteet. Pääosa hyönteisistä jää kolmanteen vaiheeseen ja digitoimatta ellei edetä skenaarion 5 mukaisesti.

Oulun kokoelmien digitoitavien kokonaismäärä on reilu miljoona, ja henkilöstötarve 53 henkilötyövuotta, eli noin 2 henkilön vuotuista työpanosta 25 vuodelle jaettuna. Ensisijaisia digitoitavia on hyönteisissä (n. 200 000) sekä itiö- ja putkilokasveissa (130 000). Lisäksi Oulangan tutkimusaseman näytteet tulisi digitoida soveltuvin osin (n. 6 000). Oulun näytteistä digitoidaan ensivaiheessa hyönteisiä (200 000) sekä kasveja (130 000) ja Oulangan vesihyönteisnäytteitä (n. 6 000). Toisessa vaiheessa digitoidaan loput selkärangattomat, selkärangattomat soveltuvin osin sekä kasvimuseon kuva-arkisto (n. 4 000).

Jyväskylässä digitoitavien näytteiden määrä on 190 000, joista valtaosa hyönteisiä. Korkeasta digitointiasteesta johtuen loput näytteistä saataneen digitoitua jo muutamassa vuodessa yhden henkilön työpanoksella. Jyväskylässä digitoidaan 140 000 selkärangattomia näytettä, 19 000 prässätyä putkilokasvinäytettä sekä kuivattuja itiökasvi- ja sieninäytteitä 12 000. Lisäksi digitoidaan kasvitieteellisen puutarhan kuva-arkisto (n. 1000 kuvaa). Henkilöresurssitarve on 8 henkilötyövuotta.

Kuopiossa digitoitavia näytteitä on yhteensä n. 500 000. Henkilötarve on 22 henkilötyövuotta, eli 25 vuodelle jaksettuna noin yksi henkilö tarvittaisiin digitointitehtäviin. Kuopion näytteistä valitaan ensivaiheessa

hyönteisnäytteitä 100 000, muita selkärangattomia 20 000, putkilokasveja 100 000 sekä itiökasveja ja sieniä 10 000. Toisessa vaiheessa digitoidaan soveltuvien osien loput.

Joensuun kokoelmista digitoitavia näytteitä ovat ns. opetusnäytteet, jotka sisältävät selkärangattomia, selkärankaisia ja kasveja sekä norppa-aineisto, joka koostuu lähinnä luustonäytteistä. Kokonaisnäytemäärä on n. 150 000 näytettä, mikä vaatii 7 henkilötyövuotta. Joensuun kokoelmista digitoidaan ensimmäisessä vaiheessa kasvi- ja sieninäytteet (n. 50 000) sekä opetuskokoelman eläinnäytteet (100 000). Toisessa vaiheessa digitoidaan norppa-aineisto soveltuvien osien.

Muista kokoelmista suurin yksittäinen digitoitavan aineistomäärän osalta on Forssan luontomuseo, jossa digitoitavia näytteitä on n. 100 000. Lopuissa 4 museossa (Terranova, Kieppi, Lappi, Porvoo ja Tampere) digitoitavien näytteiden määrä on yhteensä saman verran, n. 100 000. Henkilöresurssitarve pienimmille kokoelmille yhteensä on 9 henkilötyövuotta. Näiden pienempien kokoelmien osalta digitoidaan ensivaiheessa kaikista kokoelmista n. puolet (100 000), ja toisessa vaiheessa loput. Koska näissä kokoelmissa on erityisen paljon pieniä erilliskokoelmia, edetään digitoinnissa kokoelma kerrallaan tapauskohtaisesti.

Kokoelmien digitointiin kaikkiaan tarvitaan rahoitus, joka vastaa noin 32 hengen vuotuista suorittavaa työpanosta, eli 1 miljoona €/vuosi.

Edellä kuvattujen, muiden kuin LTKM:n kokoelmien neljän miljoonan näytteen digitointi saadaan järjestettyä 25 vuoden aikajänteellä jos käytettävissä on noin 11 henkilön vuotuinen suorittava työpanos. Tämä työ voidaan järkevästi hajauttaa maakunnallisiin museoihin, joissa kokoelmat ovat, ja tehtäväksi hanke pohjalla esimerkiksi opiskelijoiden osa-aikaisesti tehtäväksi ja kesätöinä. Tästä syntyy maahan 20 – 30 hajautetun työntekijän virtuaalinen digitointitiimi. Sen tukemiseksi ja siis hanke pohjaisen suorittavan työn lisäksi tarvitaan kuitenkin pitempijänteisiä tukitoimia osaamisen ja tietojärjestelmien osalta. Nämä on tarpeellista toteuttaa museoiden välisenä yhteistyönä ja keskitetysti.

- **Suositus 5: Digitoinnin rahoitusta tulisi voida hakea museoiden digitointisuunnitelmien ja valittujen tutkimushankkeiden vaihtuvien tarpeiden perusteella suosituksen 2 mukaisen rahoitusmekanismin kautta. Museot tekevät vuoden 2010 puoliväliin mennessä yksityiskohtaiset digitointisuunnitelmansa, jotka toimivat myös rahoitushakemusten perusteina.**

Kaikki digitointi ei kuitenkaan voi olla hanke pohjaista, vaan erityisesti LTKM:n kohdalla kyse on pitkän tähtäimen suururakasta, jota voidaan käytännössä pitää jatkuvana toimintona. Kokoelmien jatkuva kartunta mm. lahjoitusten kautta tekee myös tyhjäksi ajatuksen, että urakka voitaisiin joskus katsoa päättyväksi. Arvokkaimpien näytteiden käsittelyä ei voida uskoa hankkeisiin palkattavan vaihtuvan tilapäisen henkilökunnan käsiin. Laajassa mittakaavassa tehtävä hanke pohjainen lyhyen tähtäimen rahoitukseen perustuva työ johtaisi myös väistämättä ns. pätkätöiden teettämiseen sellaisessa mitassa, että sitä ei voida pitää hyväksyttävänä.

LTKM:n kohdalla tarvitaan siis muita järjestelyitä. Ennen näitä koskevia suosituksia on kuitenkin tarpeen tarkastella muita asiaan liittyviä näkökohtia.

4.3 Tietojärjestelmien kehittäminen

Kuten luvussa 2.1 todettiin, luonnontieteellisten museoiden kokoelma-aineistoille ei ole olemassa mitään yhtä tietojärjestelmää. Useissa mietinnöissä on kuitenkin pidetty tätä sekä tietojärjestelmäyhteistyötä toivottavana.

Luonnontieteellisten museoiden yhteistyösopimuksen luonnoksen (Opetusministeriö, 2007) mukaan tavoitteena on edistää museoiden yhteistyötä sekä edistää tietojärjestelmien keskinäistä yhteensopivuutta, kehittää eri museoiden kokoelmille yhteinen kaikkien käyttöön sopiva sähköinen tietojärjestelmä sekä varmistaa järjestelmään tallennettavan tiedon oikeellisuus ja korkea laatu. Lisäksi tavoitteena on edistää kaikkien Suomessa tavattavia eliöitä koskevien tietoaineistojen ajantasaista ylläpitoa ja saatavuutta lajitietokannan avulla. Tällainen järjestelmä vaatii myös tukipalveluita. Yhtenä tärkeimpänä osana on nimistötietokannan ylläpito. Tarvitaan myös pysyvää historiallisten paikkatietojen ylläpitoa (esim. paikkatieto-ontologia), kansallisella tasolla ylläpidetty muuttujien määrittelyn käsikirja sekä palveluita koordinaatti- ym. muunnosten ja tarkistusten tekemiseen.

4.3.1 Tietoarkkitehtuuri

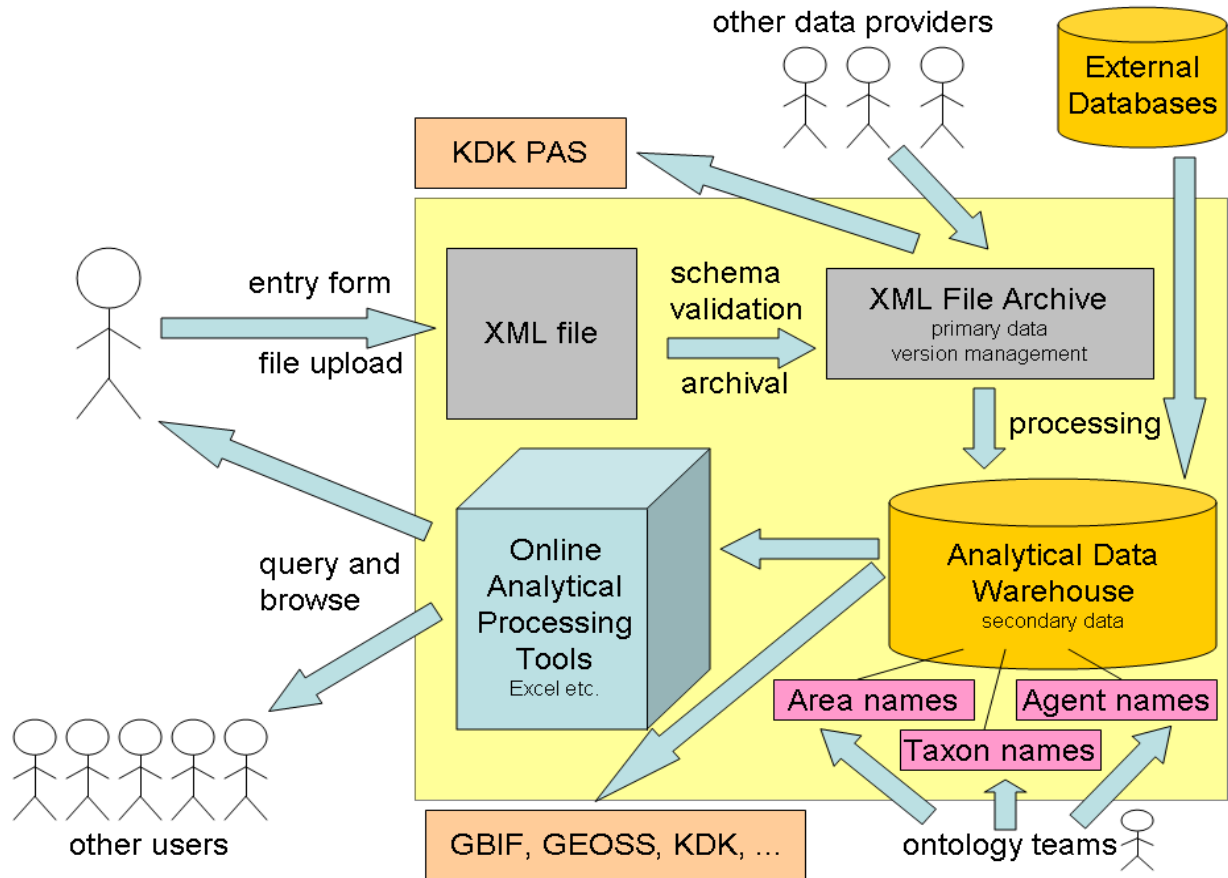
Luonnonsuojeluhallinnon LAJI-työryhmä ehdotti vuonna 2008 yhteistä lajitiedon tietojärjestelmää. Samanaikaisesti SETI-työryhmä ehdotti GBIF-mekanismiin perustuvaa tiedonvaihtoverkkoa, jossa eri organisaatioiden hallinnoiman alkuperäisaineiston kopiot olisi koottu yhteen portaaliin avoimesti saataviksi.

LTKM:ssa on meneillään tietoarkkitehtuurin uudistaminen, jonka osana tehdään myös Hatikka-järjestelmästä uusi versio. Arkkitehtuuri perustuu XML-pohjaiseen alkuperäistiedon arkistoon. Tämän yhtenäisen alkuperäistiedon varaston varaan voidaan laatia erilaisia sovelluksia, kuten relaatiotietokannat i) havainto- ja seuranta-aineiston tallettamiseen selaamiseen (Hatikka/ fieldjournal.org), ii) kokoelmatiedon digitointiin ja muokkaamiseen, tai iii) lajisuojelunäkökohdat huomioon ottava suodatettu näkymä GBIF-yhteiskäyttöä varten (Kuva 4). Liite 5 kuvaa LTKM:n tietojärjestelmiä yksityiskohtaisemmin.

Työhön käytettävissä olevien henkilöresurssien niukkuuden vuoksi tulee kuitenkin kestävänsä aikansa, ennen kuin arkkitehtuurin kaikki osat ovat paikallaan. Kaikkien sen osien laatiminen itse ei edes ole järkevää nykyisten resurssien puitteissa, vaan esimerkiksi kokoelmasovellus (ii) voitaisiin hyvin laatia pohjoismaisena yhteistyönä, perustuen esimerkiksi Norjassa jo tehtyyn työhön.

- **Suositus 6: Laaditaan yksityiskohtainen toteutussuunnitelma museoiden yhteisen kokoelmatietojärjestelmän rakentamiseksi. Suunnitelmassa tulee ottaa huomioon mahdollinen pohjoismainen yhteistyö ja ulkomailta mahdollisesti valmiina saatavat komponentit. Sen avulla tulee voida tukea**

laajamittaista koko maahan hajautettua digitointia. Järjestelmän tulee olla käyttökunnossa vuoden 2011 loppuun mennessä ja aineistot siirrettyinä järjestelmään vuoden 2012 loppuun mennessä.



Kuva 4. Luonnontieteellisen keskusmuseon uusi, rakenteilla oleva tietoaarkkitehtuuri (Tapani Lahden mukaan).

4.3.2 Saatavuus

Tavoitteena on, että museot voivat hyödyntää toistensa tietorekistereitä, kokoelmia ja asiantuntijapalveluita museo-, tutkimus-, seuranta-, opastus- ja koulutustarkoituksiin. Pääperiaatteena on tiedon mahdollisimman vapaa liikkuminen. Tätä tavoitetta edistää niin kansallisesti kuin kansainvälisestikin osallistuminen GBIF:n ja KDK:n toimintaan.

Aineistojen yhteiskäyttö ei voi odottaa mahdollisesti tulevaisuudessa Suomeen syntyvää yhtenäistä järjestelmää. GBIF-jäsenmaat ovat sitoutuneet huomattavasti nopeampaan aineistojen mobilisointiin. KDK:n asiakasliittymäkin, jonne parhaillaan digitoitavat aineistot on tarkoitus liittää, valmistuu vuoden 2011 alussa.

Olemassa olevat digitoidut aineistot, joita hallitaan taulukkolaskimessa tai henkilökohtaisissa tietokannoissa voidaan jo nyt laittaa suoraan näistä alkuperäisjärjestelmistä GBIF-palvelimiin ja saada näin yleisesti saataville. Tämä ei vaadi kuin aineiston muuttamisen Darwin Core -standardin mukaiseksi taulukoksi ja asettamisen palvelimeen (käytettävät standardit kuten Dublin Core, Darwin Core, Ecological Metadata Language on kuvattu tarkemmin liitteessä 5). Metatiedot tulee tehdä saataville OAI-PMH-standardin kautta, mitä sekä GBIF että KDK käyttävät (GBIF-raportti 2009). Tämä onnistuu helpoimmin ottamalla käyttöön GBIF:n uusi Integrated Publishing Toolkit -ohjelmisto (IPT). LTKM:ssa toimiva Suomen GBIF-solmu antaa näissä tehtävissä tarvittaessa apua.

- **Suositus 7: Kaikki museot asentavat GBIF IPT-palvelimen tai päivittävät olemassa olevat palvelimet vuoden 2010 aikana. Siihen kopioidaan kaikki olemassa olevat kokoelma-aineistot yleisesti saataville.**

4.3.3 Pitkäaikaissäilytys

Opetusministeriön asettamassa Kansallinen digitaalinen kirjasto -hankkeessa laaditaan suunnitelma sähköisten kulttuuriperintöaineistojen pitkäaikaissäilytysjärjestelmästä. Suunnitelma valmistuu kesällä 2010. Mikäli pitkäaikaissäilytysjärjestelmän toteuttamisessa voitaisiin edetä suunnitellusti, sen käyttöönotto olisi mahdollista vuonna 2016. Valitut aineistot kootaan ja dokumentoidaan ns. luovutuspaketiksi, joka jää PAS:n hoitoon ja hallintaan. Luovuttava organisaatio on itse vastuussa tietoaaineistojensa tiedostomuotojen päivittämisestä. LTKM:ssä on tämän suunniteltu tapahtuvan XML-muodossa aineistojen versionhallintajärjestelmässä, josta ne päivitetään tietovarastokantaan (relaatiotietokanta) ja samalla XML-muodossa PAS-järjestelmään. Luovuttava taho voi määritellä aineiston saatavuudelle rajoitteita, mitkä ovat tärkeitä ainakin uhanalaisten lajien tietojen kohdalla. On selvää että tällainen järjestelmä on hyödyllinen myös luonnontieteellisten museoiden näkökulmasta. Se on suorastaan välttämätön jotta LTKM voi toteuttaa uudessa yliopistolaissa sille annetun velvoitteen.

LTKM:n uusi tietoarkkitehtuuri mahdollistaa luovutuspakettien suoraviivaisen tuottamisen alkuperäistiedon XML-pohjaisesta tietovarastosta.

- **Suositus 8: Kopiot museoiden valituista aineistoista annetaan PAS-järjestelmän säilytettäväksi. Tämä tapahtuu keskitetyn tietovaraston kautta.**

4.4 Osaamisen kehittäminen ja laadun varmistaminen

4.4.1 Makroekologian ja biodiversiteetti-informatiikan opetus

Luonnontieteellisten kokoelmien digitointia ei varsinaisesti opeteta missään. Kokoelmien hoitokaan ei sisälly nykyisiin tutkintoihin Suomessa. Taksonomia ja systematiikka, jotka ovat kokoelmiin perustuvia tieteenaloja jääneet vähälle huomiolle biologian opetuksessa jo pitkään. 1930-luvulta alkaen biologian opetuksen

painopiste on kohdistunut evoluutioon ja ekologiaan. Lajituntemus ja eliömaantiede ovat vasta viime vuosina alkaneet saada uudelleen arvostusta, koska niihin perustuu globaalin ympäristömuutoksen ymmärtäminen. Makroekologia on uusi, suuriin biodiversiteettiaineistoihin nojaava tieteenala.

Koska digitointia tehdään jatkossa enenevässä määrin, yliopistollisen opetuksen tulisi ottaa nämä asiat huomioon. Akateeminen työttömyys on ongelma erityisesti biologien joukossa. Tämä kertoo opetuksen ja työmarkkinoiden kohtaamattomuudesta. Opetetaan väärää asioita. Kaikki ympäristö- ja luonnonvarahallinnon palvelukseen tulevat henkilöt joutuvat nykyään tietokantojen kanssa tekemisiin, mutta opetuksessa nämä asiat eivät tule esiin. On tarpeellista lisätä tutkijoiden tietoisuutta tuotettavan aineiston digitaalisen tallentamisen merkityksestä sekä suositeltavista standardeista, ohjelmista ja tallennusfoorumeista. Kun biologit osaavat nämä asiat, tulevaisuudessa enää synny nykyisenlaista järjestämätöntä vanhan tietoaineiston massaa, vaan tieto siirtyisi sitä mukaa digitaalisesti saataville, kun sitä synnytetään.

Biodiversiteetti-informatiikka ("lajitiedon käsittely") on uusi tieteenala, joka käsittelee näitä kysymyksiä, ja jolle on maailmalla noin parikymmentä oppituliola. Olisikin syytä harkita sellaisen perustamista maahamme. Opetuksen osaksi voidaan liittää myös digitointityötä, joka tehtäisiin valvotusti opiskelijatyönä. Tästä on jo kokemusta Turun eläinmuseolla, jossa tiettyjä osien kokoelmista on tällä tavalla saatu digitoitua osana tietotaidon ja osaamisen säilyttämisen kehittämistä.

- **Suositus 9: Jotta digitointiin saataisiin osaavaa työvoimaa, biologien opetuksessa tulisi ottaa nykyistä enemmän huomioon kysyntä joka kohdistuu lajintuntemukseen, kokoelmien hoitoon ja digitointiin ja tietoaineistojen käsittelyyn. Olisi harkittava biodiversiteetti-informatiikan opetuksen käynnistämistä maassamme.**

4.4.2 Laadun varmistaminen

Digitointi on hyvin työvoimaintensiivistä. Tiedostointi on tarkkuutta vaativaa työtä ja sen tekeminen täydellä teholla koko työpäivän läpi ei ehkä ole mahdollista eikä näytteiden rikkoontumisvaaran takia väsyneenä suotavaakaan. Sen vuoksi henkilökunnalle on suunniteltava vaihtelevia ja kehittäviä työtehtäviä. On selvää, että osaamista on jatkuvasti kehitettävä, jotta tehokkuus kasvaa. Automatisointiin ja menetelmien kehittämiseen on kiinnitettävä jatkuvaa huomiota.

Digitointityön laadun valvonta on ratkaisevan tärkeää. Esimerkiksi erään maakunnallisen museon 5 000 kasvinäytteen digitoinnin tuloksena oli vain yksi täysin virheetön näyte! Tarkistamatonta tietoa ei voida laittaa yleiseen jakeluun. Jos vaikka suorittava työ voidaan antaa sopivasti valitulle kouluttamattomalle henkilöstölle, laadun valvonta edellyttää kokemusta ja koulutusta. Tehty laadun valvonta on myös dokumentoitava. Lajimääritysten varmistaminen on kokonaan oma lukunsa. Pitkään jatkuneen taksonomian ja systematiikan aliarvostuksen tuloksena ei asiantuntijoita ole aina museoissakaan. Sen vuoksi joudutaan

laittamaan tietokantoihin ja jakeluun aineistoa, joiden lajimääritykset voivat olla epävarmoja. Tämä on kuitenkin hyväksyttävää, jos epävarmuus on dokumentoitu ja metatieto siitä saatavissa aineiston ohella. Epävarmat tai karkealla tasolla tehdyt määritykset eivät haittaa taksonomian ja systematiikan tutkijoita jotka tarkistavat ja oikaisevat tieteellisissä eliöryhmää koskevissa revisioissaan lajimääritykset.

Digitoinnin suorittaminen ei sinänsä vaadi laajaa koulutusta. Norjan digitointikeskuksen työntekijöillä ei ole mitään alan koulutusta, vaan he oppivat työn ohessa. Yksinkertaisimmillaan digitoinnin tarvitsee vain kirjata ylös kokoelmayksilön etiketissä olevat tiedot. Nämä ovat useimmiten kirjoitettu käsin ja sisältävät erilaisia lyhennettyjä paikannimiä, ym. Niissä on lyhenteitä, joiden selitykset ovat erillisissä kokoelmaluetteloissa. Jopa tekstien alleviivaukset ja etikettipaperin väri ovat merkityksellisiä. On kirjattava ylös myös lajimääritykset, jotka usein riippuvat kokoelmayksilön sijoittamisesta kokoelmaan. Kokoelmayksilöt ovat poikkeuksetta helposti vahingoittuvia. Siksi äärimmäinen tarkkuus ja huolellisuus ovat välttämättömiä. Voidaankin todeta, että suorittavan työvoiman henkilökohtaiset ominaisuudet ovat tärkeimpiä. Muu voidaan vähitellen oppia työn ohessa.

Tallennustyö on myös mahdollista tehdä kuvista. Kuvantaminen kameralla on 2-ulotteisten näytteiden kohdalla (kasviarkit, kladikirjat) niin nopeaa, että yksi kuvannusyksikkö voi pitää kymmenen tallentajaa kiireisenä. Tämä myös mahdollistaa etätönn.

Luonnontieteellisiin museoaineistoihin liittyy myös kulttuurihistoriallisia arvoja. Monet kokoelmat kuvaavat luonnontutkijoiden elämää, matkoja, ym. ja heijastavat aikansa olosuhteita. Nämä tiedot on usein kirjattu havaintopäiväkirjoihin. Tällainen tieto tulee ymmärtää osaksi kokoelmien aineistoa. Sen olemassaolo tulee dokumentoida kokoelmien metatietoihin. Se tulee soveltuvin osin myös digitoida.

Paikkatiedon keruu on tärkeä osa työtä, sillä digitaalinen tieto usein päätyy makroekologian tutkimuksiin ja kartoille. Varsinkin vanhemmissa näytteissä ei aina ole maantieteellisiä koordinaatteja. Näitä voidaan etsiä eräiden automaattisten palveluiden avulla, mutta kyseessä on työtä vaativa selkeästi erillinen työvaihe, johon on syytä varata resursseja. Liittämällä työprosessin yhteydessä haetut vanhat paikannimet nimistötietokantaan tai paikkatieto-ontologiaan voitaisiin säästää aikaa tulevilta haulilta. Nyt samaa työtä tehdään uudelleen eri kokoelmien parissa. Tarvitaan paikkatiedon ylläpitoon tietojärjestelmä, joka on pysyvästi ylläpidetty ja kansallisesti eri toimijoiden käytössä.

Edellä esitetystä voidaan päätellä muutamia asioita: Digitointityö on organisoitava työnkuluiksi, joissa eri ihmiset tarkistavat eri vaiheita. Tallentaminen ja tulkinta on pidettävä erillään: Kaikissa tapauksissa alkuperäinen tieto, joka on näytteen etiketissä on tallennettava sellaisenaan, mikä työ ei välttämättä vaadi muuta kuin käsialan lukutaidon ja muita merkintöjä koskevan ohjeistuksen. Koska eri eliöryhmien näytteet poikkeavat toisistaan, voi olla hyvä erikoistua tiettyihin ryhmiin, ja muodostaa erikoistuvia digitointitiimejä.

Työntekijät oppivat työssään ja voivat ajan mittaan siirtyä vaativampiin tehtäviin. Tiimin vetäjältä voidaan odottaa kykyä dokumentoida lajimäärityksiä ja muita laatunäkökohtia.

Digitoinnin kokonaiskustannus on jotain muuta kuin suorittavan työn ensi vaiheen kustannus. Jokaista suorittavaa henkilöä kohti on laskettava tietty määrä laadun varmistamiseen, työn ohjaukseen, tietojen päivitykseen, hallintoon ja tietojärjestelmien ylläpitoon tarvittavaa henkilökuntaa. Tämä henkilökunta on myös palkkakustannuksiltaan kalliimpaa. Tähänastisten rajoitettujen kokemusten mukaan tarvitaan vähintään noin 50 % lisäys budjettiin, joka on laskettu vain näytemäärien perusteella.

Jos digitointitehtävät suunnitellaan monipuolisiksi, henkilökunta voi osaamisensa perusteella jatkossa työllistyä esim. tallennuspalveluissa, kuvankäsittelyssä, ohjelmoijana, nettisivujen toimittajana, ym. Digitointi on maailmanlaajuinen tehtävä ja kehitettäviin menetelmiin voi olla laajaa kiinnostusta. On harkittava ulkopuolisille tarjottavan koulutuksen ja vierailuohjelmien järjestämistä osana digitointia.

- **Suositus 10: Kokoelmiin liittyvä kuvailutieto, mukaan lukien kulttuurihistoriallinen tieto ja paikkatieto tulee ottaa järjestelmällisesti huomioon digitoinnissa. Siihen liittyvä osaaminen tulee varmistaa digitointihankkeissa.**

4.4.3 Tiedostointikeskuksen käynnistys hanke Joensuussa

Joensuussa on aloitettu luonnontieteellisen aineiston tiedostointikeskuksen perustamista varten hanke. Tavoitteena on luoda digitointityössä tarvittavan asiantuntemuksen sekä infrastruktuurin osaamiskeskus. Eräänä mallina hankkeessa on kulttuurihistoriallisen museoaineiston digitoimiseen erikoistunut Mikkelin DIMIKO. Hanketta varten haetaan parhaillaan rahoitusta mm. EU:n aluekehitysrahastosta.

Joensuun tiedostointikeskus olisi yhteistyössä Helsingin yliopiston sekä Joensuun alueen eri toimijoiden kanssa, ja tarjoaisi keskitetysti palveluita luonnontieteellisen materiaalin digitointiin, tietoaineistojen arkistointiin ja pitkäaikaiseen säilyttämiseen. Yksikkö toimisi myös osaamiskeskuksena ja tarjoaisi mm. opetus- ja suunnittelupalveluita koko maassa ja kansainvälisesti (Luonnontieteellisten kokoelmien tiedostointikeskus: Käynnistysuunnitelma, 2009). Siellä järjestetään koulutusta, tiedotustoimintaa ja näyttelyitä sekä henkilökunnalle että yleisölle. Tiedotus- ja koulutustoiminnan yhtenä tavoitteena on lisätä tutkijoiden sekä harrastajien tietoisuutta aineiston digitaalisesta tallentamisesta. Näin voidaan edistää uuden tiedon digitaalista tallentamista ja sitä kautta myöhemmin tehtävän digitointityön tarpeen vähenemistä.

On ajateltu, että yksikkö voisi mahdollisesti toimia myös jo digitoidun tai muuten vähän käytettävän aineiston varastopaikkana. Tiedostointikeskuksen perustaminen suurten kasvukeskusten ulkopuolelle on perusteltua sekä kustannus- että rahoitussyistä. Tila- ja henkilöstökustannukset tulevat edullisemmiksi, toisaalta rahoitusta on mahdollista hankkia erityisistä aluekehitysrahoista sekä hajasijoitusrahoituksista.

4.5 Tilakysymykset ja voimavarojen kohdentaminen

Luonnontieteellisen keskusmuseon (LTKM) asemaa selvittänyt työryhmä ehdotti mm. että Luonnontieteelliset museot perustaisivat harvoin käytetyille aineistoille yhteisen säilytystilan, jonka esikuvana on kirjastojen yhteinen Kuopiossa toimiva varastokirjasto. Tässä yhteydessä sopii käsitellä sitä, miten tämä ehdotus mahdollisesti liittyisi kokoelmien digitointiin. Tämän selvityksen asia ei ollut ottaa kantaa kokoelmien sijoittumiseen sinänsä. Toisaalta toimeksianto edellyttää, että digitointiin tarvittavat voimavarat tulisi ensisijaisesti löytää olemassa olevia voimavaroja uudelleen kohdentamalla. Tilavuokrat ovat yksi tällainen potentiaalisesti merkittävä voimavara. Kokoelmien digitointi periaatteessa mahdollistaa niiden etäkäytön, jolloin ne voidaan sijoittaa kustannuksiltaan edulliseen paikkaan.

Vähän käytettyjä kokoelmia on paljon, mutta niiden digitointi ei ole tärkeysjärjestyksessä ensimmäisenä, vaan päinvastoin. Toisaalta vähäinen käyttö voi johtua myös puutteellisesta saatavuudesta. Kun kokoelma on digitoitu, sen virtuaalinen käyttö lisääntyy, mutta fyysinen käyttö vähenee. Voidaan siis ajatella, että digitoitu aineisto voitaisiin jopa siirtää varastoon. Etäkäyttäjät voi toisaalta kiinnostua jostain alustavasti digitoidusta näytteestä niin paljon, että haluaa siitä kuvia, preparaatteja ja DNA-näytteitä tai koko näytteen lähetettäväksi lainaan. Nämä on syytä kyetä tuottamaan digitaalisen tiedon varaan rakennetun palvelujärjestelmän kautta. Silloin kokoelman tulisi olla lähellä henkilöitä, jotka kykenevät nämä palvelut tuottamaan.

Asiantuntijoita sekä teknistä apua voidaan tarjota myös ns. liikkuvan yksikön avulla. Siinä digitointihenkilökunta voi siirtyä kokoelmasta toiseen ja suorittaa aineiston käsittelyn paikan päällä. Liikkuva digitointiryhmää voidaan käyttää myös paikallisen työvoiman perehdyttämiseen digitointityötä varten. Erityinen mahdollisuus on nopea kuvantaminen keskitetysti ja kuvien etätulkinta, eli etikettitietojen tallentaminen kuvista.

Kaikki LTKM:n kokoelmat eivät nytkään ole varsinaisissa kokoelmatiloissa, vaan tilanpuutteen takia erillisissä varastoissa eri puolilla kaupunkia on noin 1/3 näytteistä ja sen vuoksi osin huonosti käytettävissä. Vuonna 2009 Helsingin yliopisto vuokrasi 1450 m² kellaritiloja eri puolilla kaupunkia, mistä koitui 291 000 € kustannus. Vastaavat tilat olisi saatu alle puoleen hintaan esimerkiksi Joensuussa. Erotuksella olisi saatu kolme henkilötyövuotta digitointiin. Tehtyjen selvitysten tuloksena suunnitteluryhmä pitää kuitenkin mahdottomana siirtää aktiivisessa käytössä oleva kokoelma toiselle paikkakunnalle, jossa sitä voi käyttää vain netin välityksellä. Taksonominen tutkimustyö vaatii lähes aina näytteen avaamista, preparointia ja mikroskopointia. Tähän tarvitaan po. eliöryhmään erikoistuneita tutkijoita paikan päällä. Mikäli käytettävissä on nopea digitointipalvelu ja kokoelma on karkealla tasolla luetteloitu, niin että sen perusteella voidaan tehdä digitointipyyntöjä, asia voidaan järjestää joissakin käyttötapauksissa.

Joensuun tiedostointikeskuksen on ajateltu toimivan vähän käytettyjen aineistojen sijoituspaikkana ja digitoida näitä ja valittuja muualta lähetettyjä aineistoja, jotka digitoinnin jälkeen yleensä palautetaan. Samalla yksikkö joutuu kehittämään tekniikkaa ja voi myös toimia valtakunnallisena luonnontieteellisen aineiston digitointityön osaamiskeskuksena. Avainkysymys Joensuun tiedostointikeskuksen perustamisessa on se, onnistutaanko sen

avulla mobilisoimaan sellaista rahoitusta digitointiin ja muuhun siihen liittyvään toimintaan, jota muuten ei olisi saatavissa. Edullisten tilojen, saatavissa olevan työvoiman ja kehittyvän digitointiosaamisen yhdistelmän kautta tämä näyttäisi olevan mahdollista.

Yhteenvetäen, pysyvän etäsijoittamisen ongelmaa ja resurssien uudelleenkohdistamista sen avulla on harkittu perinpohjaisesti, mutta hyvin toimivaa järjestelyä ei ole ollut mahdollista löytää. Asia vaatisi paljon kallista kehitystyötä (vrt. telemedicine) ja laajempaa osaavan henkilökunnan siirtoa kuin mitä kuuluu tämän digitointistrategian toimeksiantoon. Vähän käytettyjen aineistojen (passiivikokoelmien, massanäytteiden) siirtoa voidaan harkita, jos sellaiset aineistot voidaan helposti erottaa muista.

- **Suositus 11. Perustettavaan Joensuun tiedostointikeskukseen kerätään digitointityön ja biodiversiteetti-informatiikan osaaminen sekä tekniikka, jotta se voi tukea digitointia koko maassa. Se kehittää menetelmiä kokoelmien etäkäytön vähittäiseksi mahdollistamiseksi soveltuvin osin. Sen päätehtävä on riittävän osaamispohjan luominen rahoituksen hankkimiseksi koko maan digitointityölle. Yksikkö tarjoaa digitointipalveluita, tietotekniikkapalveluita ja edullisen sijoituspaikan vähän käytetyille kokoelmille.**

4.6 Digitointityön voimavarat ja niiden hallinnointi

Toimeksiannon mukaan suunnitelma on laadittava siten, että "se on toteutettavissa varoin, joita organisaatiot ovat omien toimintamäärärahojensa ja muun tiedossa olevan rahoituksen puitteissa sitoutuneet suuntaamaan digitointiin, digitoitujen aineistojen säilyttämiseen sekä sähköisten aineistojen verkkosaatavuuteen. Tämän lisäksi suunnitelmasta käy ilmi, mihin mahdollisesti haettava lisärahoitus 2010 – 2015 suunnitellaan kohdennettavaksi."

Edellä on jo käynyt selväksi, että organisaatiot voivat sitoutua vain seuraaviin tehtäviin nykyisten voimavarojensa avulla:

- Uuden kerättävän aineiston digitointi siten että se ei enää kasvata periytyvää vanhaa aineistoa.
- Digitointityön ohjaus sekä työnjohdon ja laadunvalvonnan ohjaus.
- Aineistojen verkkosaatavuuden järjestäminen.
- Voimavarojen hankinnan edistäminen ulkopuolisista lähteistä sekä tarkoituksenmukaisen työnjaon avulla.

Historiallisen jo kertyneen aineiston digitointia ei katsota mahdolliseksi olemassa olevien voimavarojen kautta, vaan tarvitaan noin noin 1 miljoona € vuodessa ulkopuolisista lähteistä, jotta urakasta selvittää näköpiirissä olevassa ajassa. Tietojärjestelmän kehittämiseen tarvitaan myös lisää voimavaroja, jotka ovat arviolta 5 – 10 % em. kokonaiskuluista.

4.6.1 Rahoituslähteet

Rahoituksen järjestämiseksi katseet kohdistuvat ensisijaisesti opetusministeriöön. Tämä on perusteltua, koska kyseessä on kansallisen tutkimusinfrastruktuurin osa, jonka rahoittamista ei voi pitää yliopistosektorin tehtävänä. Kansallisen tutkimusinfrastruktuurin tiekartalla nämä asiat on nostettu esiin kaikkein tärkeimpinä yleisesti edistettävänä asioina.

Suunnitteluryhmä on tietoinen, että myös muita rahoituslähteitä opetusministeriön lisäksi tarvitaan. Tiedossa olevia mahdollisia lähteitä ovat ainakin seuraavat:

- EU:n rakennerahastot, erityisesti Euroopan Sosiaalirahasto (ESR) ovat ilmeinen rahoituslähde vuoteen 2013 asti. Kaksi rahoitusesitystä niihin on jo tekeillä. Digitoinnin kehittäminen vaatii osaamisen kehittämistä ja biologian alan täydennyskoulutusta (erityisesti akateemisen työttömyyden vähentämiseksi), mikä vastaa hyvin ESR:n tarkoitusta.
- Myös EU:n aluekehitysrahasto ja rajat ylittävä ENPI-yhteistyö ovat mahdollisia, mutta vain jos digitointi tehdään kohdealueilla. Tätä varten on Joensuun tiedostointikeskuksen hanke jo aloitettu. Valtion alueellistamistoimet voivat tässä tulla myös kysymykseen rahoituslähteenä.
- Ympäristöhallinto on digitoidun tiedon yksi tärkeimmistä käyttäjistä. LAJI- ja SETI-selvityksissä ehdotettiin voimavaroja tiedon saatavuuden järjestämiseksi. Ympäristöhallinto rahoittaa jo nyt noin 250 000 eurolla vuosittain museoiden toteuttamia luontoseurantoja. Nämä varat menevät lähes kokonaan kuitenkin uuden tiedon keräämiseen maastossa. Pitää kuitenkin muistaa, että kokoelma-aineisto on tärkeä tietolähde erityisesti puutteellisesti tunnettujen eliöiden kohdalla.
- Tilaustyöt palvelujärjestelmän kautta olisivat ideaalinen rahoituslähde. Esimerkiksi Mikkelin DIMIKO saa merkittävän osan rahoituksestaan näin. Tilaustöiden tuottaminen voi toimia mielekkäästi vain riittävän suurissa puitteissa.
- Kehitysyhteistyörahoja voitaisiin ehkä käyttää trooppisista maista olevien aineistojen digitointiin, tietojen palauttamiseen kohdemaihin ja sikäläisten laitosten kapasiteetin ja osaamisen kehittämiseen.
- EU:n tutkimusinfrastruktuuria kehitetään voimallisesti ja Suomessakin on kansalliseen tutkimusinfrastruktuuriin tulossa uudelleen rahoitusta. LIFEWATCH/FinLTSER on kansalliselle tutkimusinfrastruktuurien tiekartalle valittu uusi tärkeä infrastruktuuri. Uusimmissa suunnitelmissa LIFEWATCH rakentuu yhtäläisesti LTER ja GBIF-verkoston varaan. Sen vuoksi museoiden täysimittainen osallistuminen GBIF-verkoston on tärkeää ja avaa rahoitusmahdollisuuksia.
- EU:n tutkimusrahoitus ja innovaatorahoitus tietoyhteiskunnan sisällöntuotantoon ovat myös varteen-otettavia rahoituslähteitä. EU ei kuitenkaan rahoita varsinaista digitointia, joka katsotaan kansalliseksi

velvoitteeksi. Digitaaliseen tietoon liittyvien palveluiden ja tutkimuksen kehittäminen, mukaanlukien infrastruktuurin kehittäminen sopivat tähän kuitenkin erittäin hyvin. Toisin sanoen, kotimaisin varoin tehty riittävän korkealla profiililla ja osaamisella tehty innovatiivinen digitointi ja tietoaisteiden käyttö mahdollistaa suurenkin EU:n tutkimusrahoituksen hakemisen. Tätä voidaan pitää ehkä kaikkein merkittävämpänä rahoituslähteenä pitkällä tähtäimellä. Sen toteuttaminen onnistuu vain, jos osaamista kehitetään riittävälle alkuasteelle tietoisesti. Suomessa on kansainvälisessä vertailussa maan kokoon nähden poikkeuksellisen suuret biodiversiteetin tietovarannot, niitä kartuttavat verkostot ja vahva ekologian perustutkimus. Niiden varaan voidaan perustaa makroekologian tutkimuksen huipulle yltävä tutkimus, jos mahdollisuudet vain osataan käyttää hyväksi.

Yhteenvetona, mikään edellä kuvattu rahoituslähde ei yksinään riitä 1 miljoonan euron rahoituksen saamiseen vuodessa kokoelmien digitointiin, mutta 2 – 3 sellaista hanketta samanaikaisesti opetusministeriöltä saatavan pitkän tähtäimen tutkimusinfrastruktuuria ylläpitävän rahoituksen kanssa ja museoiden yhteisin toimin koordinoimana voi nousta tälle tasolle ja ylikin.

On tähdennettävä, että monet em. rahoituslähteistä edellyttävät riittävää omaa kapasiteettia eli uskottavuutta rahoittajan silmissä. On keskitettävä osaamista ja rahoituksen hankinnassa ja hallinnoimisessa pätevää henkilökuntaa. Samanaikaisesti on koordinoitava yhteistyötä koko maan puitteissa.

- **Suositus 12: Digitoinnin rahoitus koostuu opetusministeriöltä saatavasta tutkimusinfrastruktuurin pitkän tähtäimen perusrahoituksesta sekä museoiden itse hankkimasta, kilpailukykyiseen digitointiosaamiseen perustuvasta hankerahoituksesta.**

4.6.2 Rahoituksen kohdentaminen

Edellä kuvattujen suuntaviivojen toteuttamista varten tarvitaan erityinen ohjausmekanismi, jota voidaan kutsua työnimellä luonnontieteellisten kokoelmien ja aineistojen tietoyhteiskuntaohjelma (LKAT).

Tälle ohjelmalle perustetaan ohjausryhmä, joka vastaa tässä esitetyn strategian ja toimintaohjelman toteuttamisesta. Ryhmässä ovat edustettuna kaikki yhteistyösopuudet sekä rahoittajien edustajat. Ohjausryhmä kokoontuu säännöllisesti useamman kerran vuodessa ja päättää ohjelman rahoituksesta ja seuraan sen tuloksellisuutta.

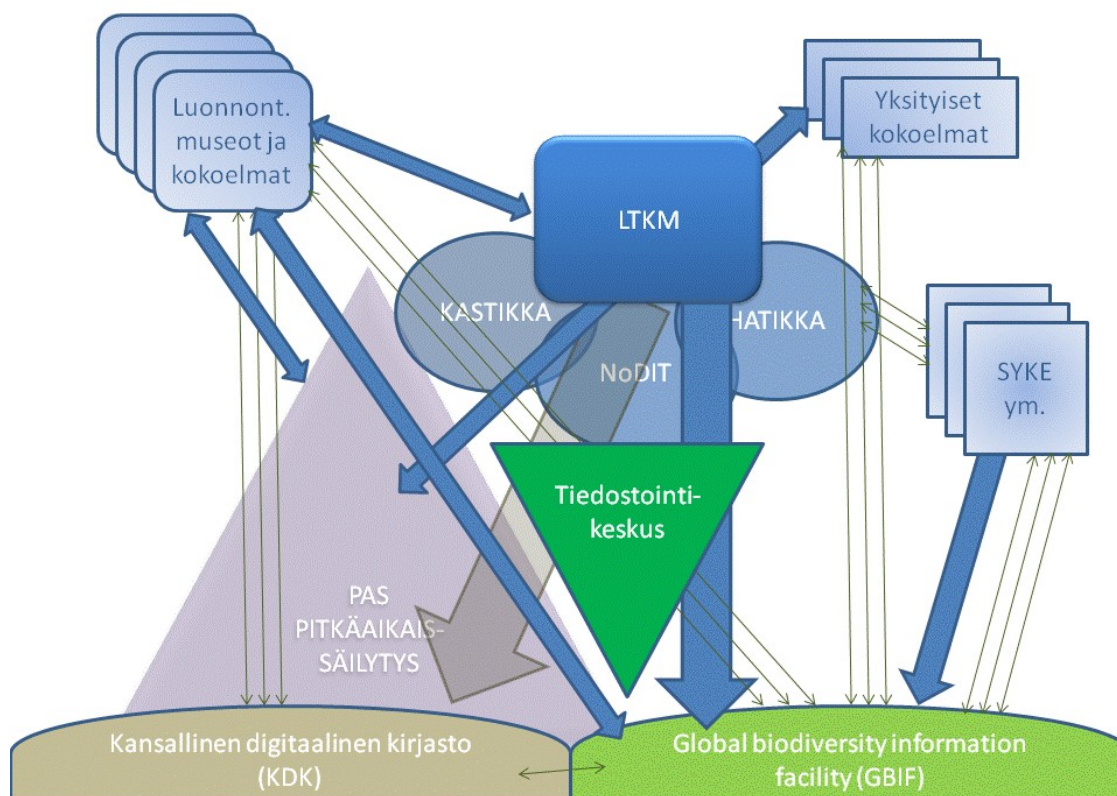
Opetusministeriö osoittaa vuosittain määrärahan ohjelman toteuttamiseksi. Tämä jaetaan museoihin ja laitoksiin hakemusten perusteella niiden ajantasaisten digitointisuunnitelmien mukaisesti. Pääpiirtein määräraha jakautuu kolmeen osaan maakunnallisten museoiden, LTKM:n ja Joensuun tiedostointikeskuksen kesken. Viimeksi mainittu käyttää osuuttaan digitoinnin ja yhteisten tukitoimien lisäksi osaamisen nostamiseksi sille tasolle, että ulkopuolisen rahoituksen hakeminen on menestyksellistä.

Esitetty malli on samankaltainen Ruotsissa noudatetulle hyvin menestykselliselle mallille. Siellä Forskningsrådet Formas osoittaa määrärahan (vuonna 2009 noin 20 miljoonaa kruunua) Artdatabankenille, joka jakaa 90 % edelleen museoille ja käyttää osan itse aineistojen mm. verkkosaatavuuden edistämiseen.

- **Suositus 13. Perustetaan luonnontieteellisten kokoelmien ja aineistojen tietoyhteiskuntaohjelma ja sille ohjausryhmä. Ryhmä päättää museoiden suunnitelmien perusteella digitoinnista, huolehtii seurannasta sekä antaa neuvoja voimavarojen hankkimiseksi.**

4.7 Yhteistyön kehittäminen

Luonnontieteellisillä museopäivillä vuonna 2009 todettiin, että aika on kypsä museoiden yhteistyön tiivistämiseksi digitoinnin ja tietojärjestelmäyhteistyön avulla. Edellä on kuvatut toimet johtavat toivottuun yhteistyön tiivistymiseen laajalla rintamalla ja käytännön tasolla. Tarvitaanko em. tietoyhteiskuntaohjelman perustamiseksi muita tai laajempi yhteistyösopimus ei kuulu tämän selvityksen tehtäväksi. Esitys tällaiseksi yhteistyösopimukseksi on jo LTKM:n asemaa käsittelevässä raportissa (Opetusministeriö, 2007). Kuvassa 5 on hahmotettu yhteistyön eri osapuolien suhteita toisiinsa sekä tiedonkulun pääsuuntia organisaatioiden välillä.



Kuva 5. Kaaviomainen esitys digitoinnin ja museoiden sekä muiden tietoyhteiskuntaohjelman osapuolten roolista ja yhteistyöstä. Siniset viivat kuvaavat digitaalisen tiedon kulkua, ohuet viivat tiedonhakuja.

Luonnonsuojeluhallinnon tuottavuusohjelman lajistonseurannan kehittämisen suunnitelmassa (LAI, 2008) on esitetty suuntaviivoja kansalliselle luonnontieteellisiä museoitakin koskevalle yhteistyölle sekä työnjaolle luonnon monimuotoisuutta koskevan tiedon hallinnan parantamiseksi. Luonnontieteellisten museoiden sekä SYKE:n välisestä yhteistyöstä on sopimusluonnos LTKM:n asemaa käsittelevässä raportissa (Opetusministeriö, 2007). Sopimuksen toteutusta varten esitetään perustettavaksi yhteistyöryhmä, jossa osapuolet voivat erikseen sovittavalla tavalla osallistua yhteistyön rahoitukseen ja hakea yhteishankkeisiin ulkopuolista rahoitusta sekä hyödyntää toistensa tietorekistereitä, kokoelmia ja asiantuntijapalveluita. Tietorekisterien hyödyntämiseen ratkaisun ovat jo esittäneet LAJI- ja SETI-ryhmät.

Edellä ovat tulleet esiin osaamisen kehittämisestä saatavat edut, mitkä ulottuvat myös museosektorin ulkopuolelle. Osittain näistä syistä naapurimaissamme Ruotsissa ja Norjassa on perustettu "Artdatabanken" nimellä kulkevat laitokset Uppsalaan ja Trondheimiin. Näistä laitoksista saadut kokemukset ovat olleet hyviä, kuten LAJI-työryhmä on jo todennut. Kumpikin maa on kaukana Suomen edellä aineistojen saatavuudessa ja digitoinnissa. Näille laitoksille on uskottu myös joitakin ympäristöhallinnon tehtäviä, kuten uhanalaisten lajien arvointi ja luetteloiden ylläpito.

- **Suositus 14: Olisi tutkittava Artdatabanken-tyyppisen tietokeskuksen perustamista Suomeen naapurimaiden esimerkin mukaisesti. Tämä selvitys tulisi tehdä yhteistyössä luonnonvarat ja ympäristö -konsortion osapuolien kanssa ja siinä tulisi ottaa huomioon Joensuun tiedostointikeskuksen toiminnoista jo tehdyt suunnitelmat.**

Yhteistyö ympäristöhallinnon kanssa tiivistyy edelleen, kun valtakunnallisen tutkimusinfrastruktuurin tiekartan toteutus edistyy. Museoiden tulee GBIF-verkoston kautta olla mukana LIFEWATCH-infrastruktuurin toteuttamisessa.

Kansallisen digitaalisen kirjaston (KDK) toiminta on ollut mallina myös luonnontieteellisten kokoelmien digitoinnin valtakunnallisessa toteutuksessa. KDK:n arkkitehtuuri ja standardit ovat yhteneviä GBIF-verkoston verkostojen kanssa. Pilottihankkeessa on jo tehty järjestelyitä digitoidun materiaalin (kladikirjat) liittämiseksi KDK:n digitaaliseen aineistoon. Kyseisen aineiston tyyppinen materiaali soveltuu muutenkin jatkossa helpommin KDK:n tietokantoihin integroitavaksi. Tämä kaikki on omiaan lähentämään luonnonhistorian ja kulttuurihistorian verkostoja.

- **Suositus 15. Yhteistyö KDK:n kanssa jatkuu, ja aineistoa siirretään soveltuvin osin KDK:n tietojärjestelmiin. Pitkäaikaisäilytyksestä tehdään sopimus KDK:n kanssa.**

Yhteistyö niin pohjoismaisten toimijoiden kuin myös globaalien organisaatioiden (GBIF) kanssa jatkuu edelleen. Tulevaisuudessa pyritään entistä aktiivisemmin vaikuttamaan myös kansainvälisissä foorumeissa sekä järjestöissä. Erityisesti pohjoismaista yhteistyötä pidetään jatkossakin erityisen tärkeänä kansallisen työn

edistymistä ja ohjausta ajatellen. Yhtenä yhteistyömuotona on mm. yhteisen nimistötietokannan perustaminen ja ylläpito.

5 Toimenpideohjelma

Edellä esitetyn jälkeen on tarve koota suositukset, toimenpiteiden ajoitus ja kustannukset yhteen tarkasteluun, jota voidaan kutsua toimenpideohjelmaksi. Taulukossa 4 on yhteenveto edellä esitetyistä toimenpidesuosituksista sekä resursseista.

Taulukko 4. Yhteenveto suunnitteluhankkeen toimenpide-esityksistä ja niiden ajoittumisesta.

Suositu s	Luku	Toimenpiteet	2010 - 2011	2012 - 2013	2014 - 2015
1	4.1	Tietoarkiston perustaminen tutkimusaineistoille.	X	X	
2	4.1	Rahoitusmekanismin perustaminen vanhan kokoelma-aineiston digitointia varten.	X		
3	4.1.1	Museoiden metatietoarkistojen ylläpito, digitointipalvelujen sekä käyttäjäehtojen kuvaukset.		X	
4	4.1.2	Museokohtaisten digitoinnin priorisointikriteerien määrittäminen ja ajantasaisten digitointisuunnitelmien ylläpito.	X	X	X
5	4.2	Rahoituksen haku vuosittain museoiden digitointisuunnitelmien perusteella suosituksen 2 mukaisen rahoitusmekanismin kautta.	X	X	X
6	4.4	Toteutussuunnitelma museoiden yhteisen kokoelmatietojärjestelmän rakentamiseksi.		X	
7	4.4.2	GBIF IPT-palvelimen asentaminen tai olemassa olevan palvelimen päivittäminen.	X	X	
8	4.4.3	Museoiden valittujen aineistojen siirtäminen PAS-järjestelmän säilytettäväksi.			X
9	4.4	Biodiversiteetti-informatiikan opetus aloitetaan.	X	X	X
10	4.4	Kokoelmien kuvailutiedot huomiotava järjestelmällisesti ja digitoidun aineiston laatu dokumentoitava.	X	X	X

11	4.5	Digitointityön sekä biodiversiteetti-informatiikan osaamisen kerääminen perustettavaan Joensuun tiedostointikeskukseen.	X	X	X
12	4.6.1	Digitoinnin rahoitus osaksi kansallista tutkimusinfrastruktuuria. Täydentävä rahoitus osaamisen avulla saaduista hankkeista.	X	X	
13	4.6.2	Luonnontieteellisten kokoelmien ja aineistojen tietoyhteiskuntaohjelma ja sille ohjausryhmä päättävät voimavarojen kohdistamisesta.	X		
14	4.7	Selvitys Artdatabanken-tyyppisen tietokeskuksen perustamisesta Suomeen naapurimaiden esimerkin mukaisesti.		X	
15	4.7	KDK-yhteistyötä tiivistetään.	X	X	X

5.1 Rahoitusmekanismi

A) Uudelleenkohdentaminen: Kohdennetaan voimavaroja tilavuokrista digitointiin siinä laajuudessa kuin kokoelmien etäkäyttö todetaan mahdolliseksi pilottihankkeissa. Tällä tavalla pyritään vapauttamaan noin 3 henkilötyövuoden panos.

B) Opetusministeriön rahoitus: Perustetaan luonnontieteellisten kokoelmien ja aineistojen tietoyhteiskuntaohjelma (LKAT). Sille myönnettävä perusrahoitus, joka alkuvaiheessa voi olla vain noin 300 000 € vuodessa, toimii haettavan hankerahoituksen siemenenä. Opetusministeriö myöntää tämän perusrahoituksen ohjelmalle toimintavuosittain. Rahoitus jakaantuu pääpiirtein seuraaviin kolmeen yhtä suureen osaan:

1. LTKM: Digitoinnin perusrahoitus koostuu LKAT:n vuosittain jakamasta perusrahoituksesta ja hankerahoituksesta. Rahoitusta kohdennetaan myös tilavuokrista. LTKM nojautuu soveltuvin osin Joensuun tiedostointikeskuksen palveluihin kuljetusta sietävän aineiston kohdalla ja hajauttaa maanlaajuisesti digitointia etätöiden mahdollistavan digitoitavan aineiston keskitetyn kuvantamisen avulla.
2. Muut kokoelmat: Digitoinnin perusrahoitus koostuu LKAT:n vuosittain jakamasta perusrahoituksesta ja hankerahoituksesta. Digitointi tapahtuu pääosin paikallisesti esim. opiskelijoiden työvoimalla.
3. Joensuun tiedostointikeskus digitoi pääosin LTKM:n kokoelmia ja säilyttää niitä soveltuvin osin.

C) Ulkopuolinen hankerahoitus: Tiedostointikeskus hakee täydentävää rahoitusta yhdessä yhteistyökumppanien eli Helsingin yliopiston sekä Joensuun alueen kehittämishankeorganisaatioiden kanssa. Yksikkö toimii Helsingin yliopiston erityistilanteen korjaamisessa sekä Joensuun alueen kehittämishankkeissa. Tarvittaessa sieltä voidaan hankkia ostopalveluina digitointityötä sekä asiantuntija-apua myös muihin yksiköihin

ja tutkimushankkeisiin. Myös muut museot toteuttavat digitointipalveluita ulkopuolisten hankkeiden tilausten mukaan.

5.2 Henkilöstökustannukset

Henkilöstöresursseja arvioitaessa (Taulukko 5) voidaan käyttää nyrkkisääntönä, että yksi henkilö digitoi 100 näytettä päivässä. (Tämä luku vaihtelee paljon, ks. aineistoselvitystä ja varsinkin korkealle priorisoidut vanhat tyyppinäytteet ovat hitaita käsitellä.) Työpäiviä on keskimäärin 225 vuodessa, joten yhden henkilötyövuoden panos on keskimäärin 22 500 näytettä. Henkilötyövuoden hinnan arvioinnissa käytetään keskimääräistä bruttopalkkaa 2 400 € / kk. Sivukuluneen (x 1,533) henkilötyövuoden hinnaksi saadaan 44 150 € (näytteen digitoinnin hinnaksi tulee näin laskettuna 1,96 €).

Taulukko 5. Digitointihankkeen arvioitu henkilöresurssitarve eri toimintaskenarioiden mukaan (ks. luku 2.6 ja liite 2)

Henkilötyövuosien määrä yhteensä toimintakaudella 2010 - 2015						
	Vaihtoehto 1		Vaihtoehto 3		Vaihtoehto 5	
Yksikkö	lkm	€	lkm	€	lkm	€
LTKM	4	176 600	9	392 400	92	4 061 800
Turku	4	166 800	8	374 800	47	2 066 000
Oulu	3	149 100	7	309 100	32	1 424 600
Jyväskylä	2	96 100	6	269 400	9	391 100
Kuopio	1	43 200	3	117 700	21	28 100
Joensuu	1	39 200	3	117 700	7	296 300
Muut	4	160 700	5	235 300	5	235 300
Yhteensä	19	832 00	41	1 810 000	213	9 403 200

Tämän varsinaisen digitointityön henkilökustannusten lisäksi tulevat työn vaatimat muut hallinnolliset henkilöstökulut. Tällaisia ovat henkilöstön työsuhteisiin, ohjaukseen ja palkanmaksuun sekä itse hankeorganisaation hallintoon, kuten rahoituksen hakemiseen ja hallinnointiin liittyvät henkilöstökustannukset (sis. sivukuluihin 53,3 %). Näiden voidaan arvioida kattavan n. 50 % henkilöstön kokonaiskuluista. Tällöin kokonaishenkilöstökustannuksiksi saadaan edellisen taulukon perusteella:

Toimintakaudelle 2010 – 2015 vuotta kohden yhteensä 2 400 000 € ja kaikkiaan viidelle vuodelle 14 400 000 €. Kaikkien noin 20 miljoonan näytteen digitointi vuoteen 2035 mennessä olisi henkilöstökustannusten osalta vastaavasti noin 60 000 000 €.

5.3 Infrastrukturi ja kokonaiskustannukset

Tarvittava infrastrukturi sisältää laitteet, tilat sekä logistiikan (Taulukko 6). Laitteiden osalta voidaan arvioida työpisteiden tarpeeksi yksi/henkilötyövuosi. Lisäksi tarvitaan kuvantamislaitteet (kamerat, kuvauspöydät, valaisimet sekä tallennusvälineet). Tiloja tarvitaan sekä digitointityötä varten, henkilökohtaista työpöytätilaa sekä varastotiloja näyttöä varten. Tilavuokrat vaihtelevat suuresti eri yksiköissä, sillä esim. toimistotilan neliövuokra LTKM:ssa keskusta-alueella on 203 € / vuosi, kun se Joensuussa on lähes puolet tästä (n. 120 € / vuosi). Logistiikkakustannuksiin lasketaan näyttöiden kuljetukset, digitointiyksikön ja laitteiden kuljetukset sekä muu asiantuntijoiden liikkuminen.

Investoinnissa on otettava huomioon niiden ylläpito pitkällä tähtäimellä. Ennen merkittäviä investointeja haltijaorganisaatioiden sitoutumisesta ylläpitoon on sovittava. Tämä koskee erityisesti Joensuun tiedostointikeskusta, joka on uusi perustettava infrastrukturi.

Kokonaiskustannukset saadaan, kun kerrotaan henkilökustannusten sekä materiaalikulujen summa yleiskustannuskertoimella (overhead), joka LTKM:ssa on 1,22. Näin saadaan vaihtoehdon 5 mukaan:

- Kokonaiskustannukset / vuosi: 2 250 900 €
- Kokonaiskustannukset 2010 – 2015: 11 255 720 €

Digitoinnin vaatima kokonaisvaltainen tietotekninen infrastrukturi vaatii erityispanostusta mm. tietokantaan, levytilaan, web-palvelimiin ulkopuolisia hakuja varten ja ehkä erillisiä edusta- ja hakupalvelimiäkin. Tämän kokoisesta aineistosta puhuttaessa taustajärjestelmän rakentamisessa voidaan puhua omasta erillisestä projektista. Tämän vuoksi suunnitelman tässä vaiheessa arviot tietotekniikan vaatimista investoinnista (5-10 % kaikesta) ovat vain suuntaa antavia. Toisaalta yleiseen saatavuuteen liittyen edes keskikorkean käytettävyyden järjestelmän osalta on otettava huomioon ylläpitoon liittyvät seikat, joista tulevia kustannuksia on vaikea arvioida.

Taulukko 6. Arviot vuotuisista infrastruktuurikustannuksista (2010 – 2015) vaihtoehdon 5 mukaan.

Kustannuserä	Yksikköä	á €	YHTEENSÄ €
laitteet työpiste	35	500	17 500
kamera	5	1 100	5 500

	kuvauspöytä	5	150	750
	valaisimet	10	23	230
	muu			9 020
tilat	toimistotila	350 m2	200	70 000
	kuvaustila	120 m2	150	18 000
	varastotila (ei sisällä varsinaisia museotiloja)	800 m2	80	64 000
logistiikka	näytteet	10 000 km	1,5	15 000
	digit.yksikkö	10 000 km	6,5	65 000
	muu	10 000 km	3,5	35 000
YHTEENSÄ				300 000

6 Lähdeluettelo

Fauna Europaea Web Service (2004) Fauna Europaea version 1.1, Available online at <http://www.faunaeur.org>

GBIF-raportti, 2009. Global Biodiversity Information Facility, Preliminary Analyses of Participants' Reports 2009. GB16, 6 and 8 October, Copenhagen, Denmark.

Göteborg 2001. Göteborgin biodiversiteettikokouksen päätöslauselmat. Göteborg 2001;
http://europa.eu.int/eur-lex/en/com/cnc/2001/com2001_0264en01.pdf

Huldén, L., Albrecht, A., Itämies J., Malinen, P. & Wettenhovi, J. 2000. Atlas of Finnish Macrolepidoptera. Lepidopterological Society of Finland, Finnish Museum of Natural History, Helsinki.

Hyönteistietokanta (<http://www.fmnh.helsinki.fi/insects/index>). Luonnontieteellinen keskusmuseo, Helsinki.

Kansallisen tason tutkimusinfrastruktuurit: Nykytila ja tiekartta. Tiivistelmä ja suositukset. Opetusministeriön julkaisu (2009:3) .

Kokoelmien digitointisuunnitelma 2009. Luonnontieteellinen keskusmuseo (LTKM), 3.4.2009. Hannu Saarenmaa, Pertti Uotila, Olof Biström, Anders Albrecht, Mikko Heikkinen & Juhani Lokki.

Johannesburgin biodiversiteettikokouksen päätöslauselmat 2002; <http://www.un.org/jsummit/html/brochure/brochure12.pdf>

Komonen, A., Päivinen, J. & Kotiaho, J. 2009. Missing the rarest: is the positive interspecific abundance–distribution relationship a truly general macroecological pattern? *Biology Letters* 23:492–494.

LAJI 2008. Luonnonsuojeluhallinnon tuottavuushanke. Esitys eliölajien suojelun tuottavuuden parantamiseksi. Luonnonsuojeluhallinnon eliölajien suojelun tuottavuutta parantavan projektiryhmän (LAJI) loppuraportti, 25 s. Ympäristöministeriö.

Lampinen, R. & Lahti, T. 2009: Kasviatlas 2008. Helsingin Yliopisto, Luonnontieteellinen keskusmuseo, Kasvimuseo, Helsinki. Levinneisyyskartat osoitteessa <http://www.luomus.fi/kasviatlas>

Luonnon puolesta – ihmisen hyväksi. Suomen luonnon monimuotoisuuden suojelun ja kestävä käytön strategia ja toimintaohjelma 2006-2016.

Luonnontieteellisen keskusmuseon asemaa selvittäneen työryhmän muistio. Opetusministeriön työryhmämuistioita ja selvityksiä 2007:20, Opetusministeriö.

Luonnontieteellisten kokoelmien tiedostointikeskuksen käynnistäminen. Aiesopimus 23.6.2008.

Luonnontieteellisten yliopistomuseoiden yhteinen suunnitteluhanke vuonna 2009. National Digitisation Centre / Centre for Microfilming and Conservation. Mikkeli, Finland.

Mattila, N., Kotiaho, J., Kaitala, V. & Komonen, A. 2008. The use of ecological traits in extinction risk assessments: A case study on geometrid moths. *Biological Conservation* 141:2322–2328.

Mattila, N., Kotiaho, J., Kaitala, V., Komonen, A. & Päivinen, J. 2009. Interactions between Ecological Traits and Host Plant Type Explain Distribution Change in Noctuid Moths. *Conservation Biology* 23:703–709.

Mitikka, V., Heikkinen, R.K., Luoto, M., Araújo, M.B., Saarinen, K., Pöyry, J. & Fronzek, S. 2008. Predicting range expansion of the map butterfly in Northern Europe using bioclimatic models. *Biodiversity Conservation* 17:623–641.

Peterson, T. & Soberon, J. 2006. Development of indicators of compliance with the 2010 target of the Convention on Biological Diversity using primary biodiversity data provided by GBIF . 33 p. Report for the GBIF Governing Board 13, Biodiversity Research Center , University of Kansas.

Saarenmaa, H. 2008. Biologisten kokoelmien tiedostointipalveluiden kehittäminen. Muistio, 22 s. 10.3.2008. Luonnontieteellinen keskusmuseo.

SETI 2008. Luonnonsuojeluhallinnon tuottavuushanke. Ehdotus luonnon monimuotoisuuden seurantajärjestelmästä ja siihen liittyvien tietojärjestelmien kehittämisestä. Luonnonsuojeluhallinnon luonnon monimuotoisuuden seurantaa ja tietojen hallintaa koordinoivan projektiryhmän (SETI) loppuraportti. 59 s. Ympäristöministeriö.

Liitteet

- Liite 1. Yhteenveto NoDIT-aineistosta
- Liite 2. Kokoelmayhteenveto aineistokyselyn pohjalta
- Liite 3. Tietoaineistokysely
- Liite 4. Tietoaineistoraportti
- Liite 5. Luonnontieteellisten museoiden tietojärjestelmäkuvaus

Versiohistoria

Numero	Tiedoston nimi	Muutos-pvm.	Kommentoineet /Muokanneet
1	Strategiatoimin-v1-Wp-131009	13.10.2009	HS
4	Strategiatoimin-v4-Wp-151109	16.11.2009	HS, RV, LT-R, SH, WP
6	Strategiatoimin-v6-HS-201109	22.11.2009	HS, NL
7	strate-v7-Wp-211109	25.11.2009	HS, AM, LT-R, RV, RL, SH, JL
8	strate-v8-091209	8.12.2009	OHJAUSRYHMÄ, SUUNNITTELURYHMÄ
10	strate-v10-101209	10.12.2009	HS, WP
11	strate-v11-111209	11.12.09	MK, HS
12	strate-v12-141209	14.12.09	HS, WP, LM
16	diko-v16-161209-PenUltimateDraft	15.12.09	HS
18	Diko-v18-301209	30.12.09	PU, SS, MP, AA, LS, HV, DT, HK, TL, HS, WP