

# 2

## Webin kehityssuuntia

Maailma on luotu uudelleen  
luotavaksi.

---

*Georges Duhamel (1884–1966)*

Tässä luvussa luodaan katsaus webin kehitykseen tiedon julkaisemisen ja käytön monista näkökulmista. Tämän jälkeen näkökulma tarkennetaan teoksen aiheena olevaan semanttiseen webiin.

### 2.1 Web julkaisukanavana

Web tarjoaa kanavan tiedonhankintaa ja sosiaalista kanssakäymistä varten. Fyysistä kokemusta näyttöpäätte tai mobiililaitte ei korvaa, mutta se mahdollistaa uusia tapoja tiedon saamiseen ja yhteydenpitoon. Tässä teoksessa keskitytään erityisesti webin tarjoamiin mahdollisuuksiin tietojen yhdistämisen, julkaisemisen ja haun kannalta. Näitä ovat muiden muassa:

1. *Ajasta ja paikasta riippumaton tiedonvälitys.* Tietosisällöt saadaan joustavasti tutkijoiden ja suuren yleisön käyttöön. Tietoon tutustuminen ei edellytä fyysistä läsnäoloa, esimerkiksi kirjastoon tai museoon tulemistä tiettyyn aukioloaikaan.
2. *Tietosisällön määrän lisääminen.* Laajat kokoelmat, joiden fyysinen näyttäminen ja tiedon saataville asettaminen olisi mahdotonta, voidaan avata yleisölle WWW:n välityksellä.

- vuorovaikutteisuus  
sopeutuvuus
3. *Tietosisällön laadun parantaminen.* Tietotekniikka mahdollistaa tietojen haun, yhdistämisen ja esittämisen tavoilla, jotka eivät ole teknisesti tai taloudellisesti mahdollisia fyysisissä tiloissa.
  4. *Vuorovaikutteisuuden hyödyntäminen.* Tietojärjestelmien *vuorovaikutteisuus* (interactivity) ja *sopeutuvuus* (adaptivity) tarjoavat aiempaa joustavampia ja henkilökohtaisempia tapoja tutustua aineistoihin.
  5. *Audiovisuaalisuuden hyödyntäminen.* Sisältöjä voidaan esitellä audiovisuaalisilla keinoilla, kuten kuva-, musiikki- ja videotallenteilla.
  6. *Uudet sisältökohteet ja palvelut.* Uudentyyppisten aineistojen tallentaminen ja julkaiseminen on tullut mahdolliseksi. Esimerkiksi sähköinen asiointi pankeissa ja verkkokaupoissa, sosiaalinen media ja erilaiset pelit ovat verkon käytetyimpiä sisältöjä. Internet-sivustoja on myös määrätietoisesti ryhdytty tallentamaan uudenaikaisina historiallisina dokumentteina.<sup>1</sup>

Web-sovellusten kehittämisessä on runsaasti käytännöllisiä, teknisiä haasteita. Suurimmat hankaluudet liittyvät kuitenkin tietosisältöjen esittämiseen eivätkä tekniikkaan. Avainkysymys on, miten web-sivustoista saadaan kävijöiden kannalta aidosti mielenkiintoisia, hyödyllisiä ja helpokäyttöisiä. Tähän ei pelkkä perinteinen tietojen hakumahdollisuus tietokannasta välttämättä riitä. Siinä ei hyödynnetä kuin pieni osa niistä mahdollisuuksista, joita verkon rikas sisältö voi tarjota.

Webin kehittymistä ja haasteita voidaan tarkastella monista tietoon liittyvistä näkökulmista, joihin kuhunkin liittyy omat muotiterminsä sekä poliittisia ja teknologisia kehityspolkuja:

- |           |   |
|-----------|---|
| Open Data | 1. <i>Tiedon avoimuus (Open Data).</i> Tieto on tietoyhteiskunnan tärkeimpiä tukipilareita. Mutta missä määrin sen pitäisi olla avointa (open data) ja suljettua? Erilaisia näkökohtia ja argumentteja voidaan esittää muun muassa demokratian, liiketoiminnan, tietoturvan ja yksityisyyden suojan kannalta. |
| Web 2.0   | 2. <i>Tiedon yhteisöllinen tuotanto (Web 2.0).</i> Miten webin tietoa voidaan tuottaa ja julkaista yhteistyössä sen käyttäjien kanssa mahdollisimman helposti?  |
| Big Data  | 3. <i>Tiedon määrä (Big Data).</i> Miten voidaan tuottaa ja hallita webin yhä suurempia ja nopeasti muuttuvia suuraineistoja?   |

<sup>1</sup> Ks. esim. Internet Archive <https://archive.org/> ja Kansalliskirjaston Suomalainen verkkoarkisto <http://verkkoarkisto.kansalliskirjasto.fi/va/>.

4. *Tieto palveluina (Web Services)*. Millaisina toiminnallisina palveluina tieto kannattaa julkaista toisaalta ihmiskäyttäjille, toisaalta toisten verkkopalveluiden käytettäväksi? Web Services
5. *Tiedon verkko (Web of Data, Linked Data, Semantic Web)*. Perinteiset web-sisällöt on tarkoitettu ihmisen luettaviksi toisiinsa linkitettyinä dokumentteina. Webin sisään on kuitenkin rakentumassa myös tietokoneiden käytettäväksi tarkoitettu, konetulkittavan tiedon verkko, joka yhdistää web-sivujen sijasta käsitteitä ja tietoja toisiinsa. Koneellisesti tulkittavan datan avulla voidaan rikastaa tietoja toistensa avulla, päätellä uutta tietoa ja kehittää älykkäitä sovelluksia (tekoäly). Web of Data

Seuraavaksi tarkastellaan lyhyesti näitä webin tietoon liittyviä kehityssuuntia.

## 2.2 Tiedon avoimuus – Open Data

Tietoyhteiskunnan tärkein tuote, tieto, poikkeaa perinteisistä teollisuuden tuotteista siinä, että siitä voidaan sekä valmistaa kopioita (monistaa) että jakaa näitä (sähköisesti) lähes ilmaiseksi. Teoriassa tietoyhteiskunnan kansalainen voisi siis elää tiedon osalta loputtomassa yltäkylläisyydessä ja me kaikki voisimme olla tiedon miljonäärejä. Sitä ennen on kuitenkin ratkaistava joitain perustavaa laatua olevia kysymyksiä. Kuka kustantaisi tietotuotteen ensimmäisen kopion valmistamisen, jos kaikki muut saavat sen ilmaiseksi käyttöönsä? Hyödyn korjaa tällöin jokin muu taho kuin tiedon tuottaja. Yritysten kannalta avoin ilmainen data edellyttää sitä, että varsinainen liiketoiminta syntyy datan myymisen sijasta siihen liittyvien palveluiden tai muiden tuotteiden kautta.

Julkisen tiedon tuotannon kustannamme me veronmaksajat. Ensimmäisen kopion tuottamisen ja tiedon jakamisen ei siksi pitäisi olla ongelma, jos tiedon avulla voidaan tuottaa kansalaisille hyödylliseksi koettuja palveluita. Julkinen sektori onkin ryhtynyt avaamaan tietokantojaan määrätietoisesti, vaikka julkisen sektorinkin dataan, kuten esimerkiksi paikka- ja sää tietoon, liittyy taloudellisia intressejä. Ajatuksena on, että laajemman kokonaisuuden, kuten Suomi Oy Ab:n, kannalta datan avaamisesta saatava hyöty on suurempi kuin yksittäisen julkisen organisaation menettämät myyntituotot.

Toinen perustavaa laatua oleva kysymys koskee tiedon arvoa. Eikö tiedon arvo vähene sitä levitettäessä, kun yhä harvemmalla on tarve ostaa tietoa itselleen? Avoimen datan idea haastaa tämän perinteisen ajattelutavan. Ideana on, että tiedon arvo ei vähene vaan päinvastoin kasvaa sitä avoimesti jakamalla, jos tietoa tuotetaan ja sitä voidaan hyödyntää yhteisöllisesti. Webin kehitys ja Wikipedian, YouTuben ja Facebookin kaltaiset sovellukset ovat osoittaneet, mikä voima yhteisöllisessä, avoimessa tiedontuotannossa voi piillä. Sopivasti suunnattu ilmaisuus ja avoimuus tarjoavat myös uusia mahdollisuuksia liiketoiminnallisiin innovaatioihin, kuten vaikkapa Googlen palvelut ovat osoittaneet.

Julkisen tiedon avoimeen, maksuttomaan avaamiseen sitä tarvitseville tahoille on useita perusteita.

- *Demokratian edistäminen.* Kansalaisilla on oltava demokraattinen oikeus saada ja käyttää julkisin varoin tuotettua tietoa maksutta, koska he ovat sen viime kädessä maksaneetkin esimerkiksi veroina. Tämä on jo pidempään ollut yleisperiaatteena muun muassa Yhdysvalloissa.
- *Tiedon saatavuuden parantaminen.* Saatavuutta rajoittava, yhden tahon eduntavoittelu saattaa estää tiedon hyödyntämisen laajemmassa yhteydessä. Esimerkiksi kartta- ja paikkatiedon osalta helposti saatavilla olevat aineistot, kuten Google Maps ja alun perin sotilaskäyttöön kehitetty, satelliittien kautta saatava GPS-paikkatieto, ovat käynnistäneet todellisen sovellusten ja innovaatioiden aallon. Suomessa julkisia organisaatioita on kuitenkin suoraan edellytetty muun muassa maksuperustelailta sulkemaan tietoa ja tekemään taloudellista tulosta aineistoillaan tai ainakin kattamaan datan tuotannon kuluja. Yhden organisaation pieni etu tukahduttaa kuitenkin helposti innovaatioita ja kehitystä laajemmassa yhteydessä.
- *Sisällön yhteisöllinen rikastaminen.* Yhdistämällä oma aineisto muiden toimijoiden aineistoihin jokaisen tiedontuottajan tiedon arvo kasvaa yhteisöllisesti. Esimerkiksi yksittäiset Wikipedian artikkelit erillisinä eivät ole kovin arvokkaita, vaan arvo syntyy niiden muodostamasta kokonaisuudesta.
- *Yhteentoimivuuden edistäminen.* Yhteiset avoimet pelisäännöt tiedon julkaisemisessa, yhdistämisessä ja jakamisessa edistävät tieto-

järjestelmien ja palveluiden *yhteentoimivuutta* (interoperability), jonka puute on keskeinen este muun muassa julkisen sektorin eri toimijoiden yhteistyössä.

yhteentoimivuus

- *Tuottavuuden tehostaminen.* Tietoja ja ratkaisuja jakamalla voidaan tehostaa työnjakoa, eliminoida tarpeetonta päällekkäistä työtä ja lisätä näin työn tuottavuutta. Esimerkiksi Suomen kuntien verkkopalveluilla on paljon samoja toiminnallisuuksia, mikä mahdollistaa yhteistyön tietojärjestelmien kehitystyössä. Asiakkaat ja heidän tarpeensa eri kunnissa ovat paljolti samanlaisia.
- *Uusien innovaatioiden mahdollistaminen.* Tietoa avoimesti julkaisemalla mahdollistetaan ennalta-arvaamattomia innovaatioita. Joku voi keksiä ja toteuttaa tiedolle yllättäviä sovelluksia ja hyödyntämistapoja, jos tieto vain on helposti saatavilla. Esimerkkinä mainittakoon vaikkapa asunnonhakupalvelu, jossa voidaan etsiä keskustasta tietyllä ajallisella etäisyydellä olevia kohteita julkisen liikenteen avointen aikataulutietojen avulla.

Tiedon avoimuuden edistämiseksi on perustettu useita kansallisia ja kansainvälisiä ohjelmia ja järjestöjä. Esimerkiksi Open Knowledge International<sup>1</sup> -yhteisö, jolla on aktiivinen alajärjestö Suomessakin, on toimittanut verkossa olevan monikielisen Open Data Handbook -sivuston.<sup>2</sup> Sen kautta voi tutustua tarkemmin avoimeen dataan liittyviin juridisiin, sosiaalisiin ja teknisiin kysymyksiin, avoimen datan menestystarinoihin sekä käytettävissä oleviin aineistoihin ja muihin resursseihin.

## 2.3 Tiedon yhteisöllisyys – Web 2.0

Webin voima perustuu paljolti yhteisöllisesti julkaistaviin aineistoihin. Webin alkuaikoina käyttäjien tiedontuotanto ja -jakelu, toisin sanoen web-sivustojen luominen ja omien aineistojen julkaiseminen, ei kuitenkaan ollut kovin yksinkertaista, vaan vaati teknistä osaamista. Web 2.0 -termin lanseerasivat vuonna 2004 erityisesti Tim O'Reilly ja Dale Dougherty. Sillä viitattiin uudentyyppisiin, helppokäyttöisiin ja yhteentoimiviin web-järjestelmiin, jotka perustuvat käyttäjien itse tuottamaan si-

---

<sup>1</sup> <https://okfn.org/>

<sup>2</sup> <http://opendatahandbook.org/>

sältöön.<sup>1</sup> Monet eniten käytetyistä nykyistä web-palveluista ovat tämän tyyppisiä sovelluksia, esimerkiksi YouTube ja Facebook.

## 2.4 Tiedon määrä – Big Data

**suurdata** Big Data -käsitteellä eli *suurdatalla* viitataan isoihin tietojoukkoihin, joiden tallentaminen, siirtäminen, käsitteleminen ja hallinta nykyisillä relaatiotietokanta- ja tietoteknisillä ratkaisuilla on haasteellista. Se, koetaanko tietojoukko isoksi, on siis riippuvaista kulloinkin käytettävissä olevan teknologian mahdollisuuksista ja rajoituksista käsitellä laajoja tietoaaineistoja eli *skaalautuvuudesta*.

**skaalautuvuus**

Dough Laney'n klassisen, vuonna 2001 esittämän niin sanotun kolmen V:n mallin mukaan<sup>2</sup> tiedon määrällisen kasvun haasteita ovat:

1. *Volume (määrä)*. Perinteinen tietokantatekniikka ei riitä hyvin suurien datajoukkojen tallentamiseen, siirtoon ja käsittelemiseen. Ratkaisumallina on datan jakaminen osiin.
2. *Velocity (nopeus)*. Nopeus viittaa tässä yhteydessä laskennan nopeuteen. Riittävän tehokkuuden aikaansaamiseksi laskenta tyypillisesti hajautetaan kymmenille tai jopa tuhansille rinnakkain samanaikaisesti toimiville laitteistoille, joiden tulokset kootaan lopuksi yhteen.
3. *Variety (monimuotoisuus)*. Tieto tai sen osat voidaan esittää eri tavoin, jolloin niiden käsitteleminen kokonaisuutena hankaloituu. Käytössä voi olla esimerkiksi vaihtoehtoisia tai epäyhtenäisiä tapoja nimien, osoitteiden, numeeristen arvojen ynnä muiden esittämiseksi tai näiden välisten yhteyksien kuvaamiseksi.

Mallia on sittemmin täydennetty uusilla V-ulottuvuuksilla, kuten:

4. *Veracity (totuudenmukaisuus)*. Data voi olla epätasällista, puutteellista tai suorastaan virheellistä.
5. *Variability (vaihtelevuus)*. Vaihtelevuudella viitataan datan mahdollisten arvojen lisääntymiseen isoissa datajoukoissa.
6. *Value (arvo)*. Datan arvo sovelluksissa on keskeinen datan ominaisuus.

Suurdatan käsittelyssä tarvitaan tehokkaita tietojenkäsittelyn menetel-

<sup>1</sup> Lisätietoa Web 2.0 -ilmiöstä ja tiedon yhteisöllisyydestä löytyy esimerkiksi teoksesta (Anderson, 2012).

<sup>2</sup> Ks. (Laney, 2001).