

**Relaatiotietokantojen
myötä 1970-luvulta alkaen
yksilöivät tunnisteet
vakiintuivat hakuavaimiksi
tietojärjestelmissä.**

Esa Tiainen

Yksilöivät kohdetunnisteet paikkatietoyhteisessä

TIETOTEKNIKASSA yksilöiviä tunnisteita käytetään yleisesti tiedonhakuun. Relatiiviset tietokantojen myötä 1970-luvulta alkaen yksilöivät tunnisteet vakiintuivat hakuavaimiksi tietojärjestelmissä. Tänä päivänä paikkatietoyhteisessä ja alakohdittaisissa tai eri tietolähteitä hyödyntävissä sovelluksissa tarvitaan sovellusalaan riippumattomia yksilöiviä tunnisteita mm. sen varmistamiseksi, että tietoaineistoa voidaan ylläpitää vain sen kohteita koskevien muutosten osalta ilman koko aineiston koostamista uudelleen. Sijaintiperusteiset operaatiot yksinään

ovat osoittautuneet riittämättömiksi kohteiden luotettavaan ja nopeaan identifiointiin paikkatiedon laajassa hyödyntämisessä.

IT-standardoinnissa yksilöivien tunnisteiden voidaan sanoa olevan ajankohdittaisesti ”kuuma peruna”, ja Suomessakin on käynnistynyt ISO-OID-koodin kansallisen SFS-standardin valmistelu.

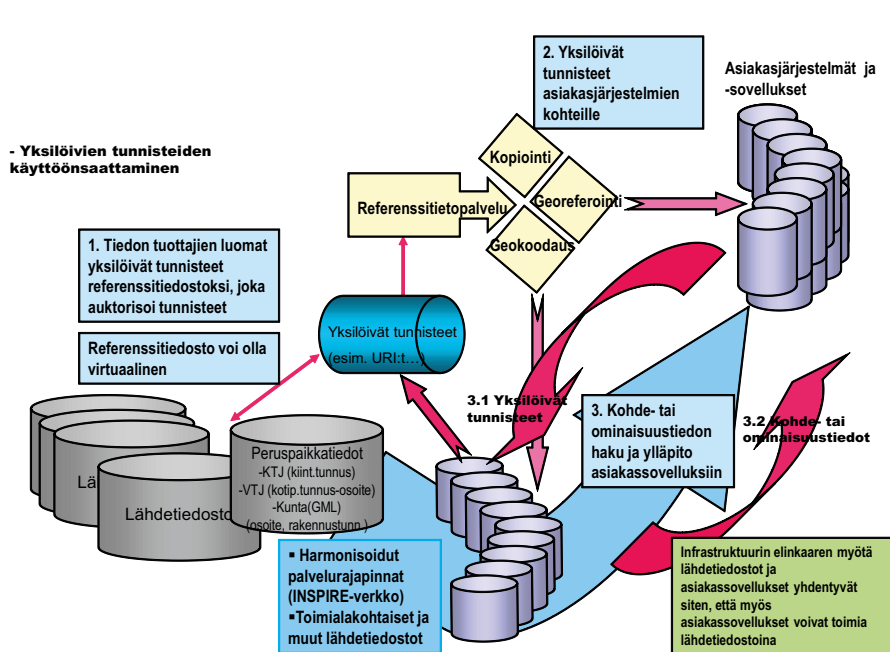
Paikkatietoalalla OGC:n GeoLinking Service (GLS) Interface -standardiluonnos on juuri ollut lausuntovaiheessa, jo pitkään odoteltuna tarkennuksena OGC-dokumenttiin Abstract Specification,

Topic 14: Semantics and Information Communities (www.opengeospatial.org), jossa lähtökohtana on jäsentää systemaattisesti eri käyttäjäyhteisöjen näkemyksiä informaation merkityksiin. Tämä näkökulma on yhä tärkeämpi, koska useimmissa tietoaisteistoissa todellisuuden kohteita on yksilöity sovellusalueittain eri tarkoituksiin, mihin yksilöinti on perustunut. Niinpä samaa todellisuuden kohteita koskevaa tietoa on yksilöitynä erilaisin perustein ja tunnistein, mikä vaikeuttaa oleellisesti tiedon luotettavaa yhdistelyä.

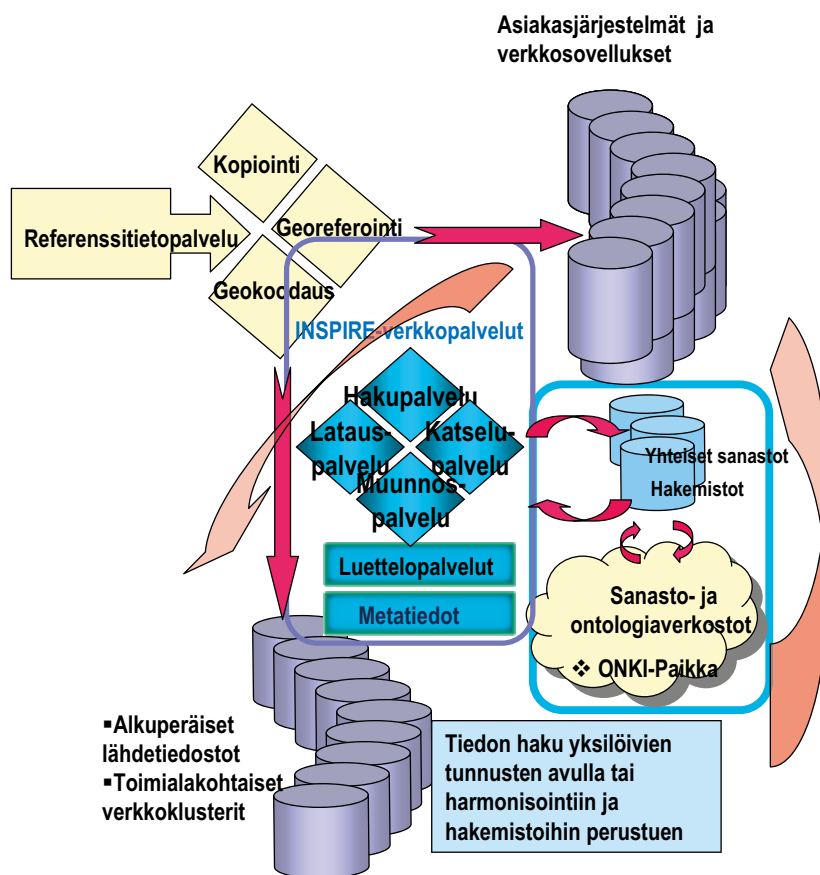


Antero Aaltonen

Yksilöiviä kohdetunnisteita ovat esimerkiksi rakennustunnus, kiinteistötunnus ja henkilötunnus. Tunnistekäytäntö on laajenemassa koskemaan muitakin tiedon käyttäjille merkityksellisiä paikkatietokohteita.



Kuva 1. Referenssietietopalvelun toiminnallinen rooli.



Kuva 2: Referenssietietopalvelu verkko-ympäristössä.

Yksikäsitteiset yksilöivät tunnisteet mahdollistavat tehokkaan hajautetun tiedonkeruun. Kun tieto on kertaalleen tallennettu, se on kaikkien käytettävissä paikkatietoinfrastruktuurin puitteissa. Tiedon keruussa voidaan siten tehokkaasti hyödyntää kaikkia toimintoja, jotka ovat tekemisissä kohteen kanssa. Viranomaisten ja yritysten ohella myös kansalaisten ja kolmannen sektorin kytkeminen tehokkaaseen tiedonhallintaan mahdollistuu, Open Street Map-sovelluksen tapaan. Tätä varten tarvitaan yleinen tietopalvelu yksilöivien tunnisteiden jakeluun kaikkien saataville. Siitä voitaisiin käyttää esim. työnimeä *referenssietietopalvelu*.

Yksilöinti, identiteetti ja versiohallinta

Yksilöiviin tunnisteisiin liittyy paikkatiedossa oleellisesti niiden versiohallinta, joka määrittelee kohteiden identiteetin ja version suhteessa ajallisiin, sijainnillisiin, geometrisiin, topologisiin yms. muutoksiin. Tätä varten paikkatietokohteille tarvitaan elinkaarisäännöt, jotka tarvittaessa ovat yksilölliset eri kohdetyypeille. Paikkatiedon toinen ominaispiirre on horisontaalisuus yli aihe- ja sovellusalueiden sekä hallinnonalojen.

Kaksijakoinen kehityskuva

Paikkatietonäkökulmasta tunnistestandardointiin liittyvä kehitys näyttää kaksijakoiselta:

Pelkkää tunnistetta käsittelevät standardit eivät hahmota niiden laajemmassa käytössä tarvittavaa palveluinfrastruktuuria, sen rakenteita ja kustannuksia (UUID, OID, ehkä myös INSPIRE). Nämä kustannukset eivät kuitenkaan automaattisesti tai yleisesti sisälly nykyisiin infrastruktuureihin. Tämän ilmeisen kompastuskiven takia aiheita on toisaalla lähestytty järjestelmä-näkökulmasta (ISO TC 211/19112 Spatial referencing by geographic identifiers, ISO-TC211/19155 Place Identifier Architecture). OGC:n GLS puolestaan rakenteistaa paikkatiedon merkityksellisiä (semanttisia) suhteita relaatiarakenteina.

AIHEESEEN LIITTYVÄÄ STANDARDINTIA TMS:

- Spatial referencing by geographic identifiers (ISO-TC211/19112 (2003 E))
- UUID – Universally Unique Identifier (ITU-T Rec. X.667 | ISO/IEC 9834-8:2005)
- OID – Object identifier (ASN.1, ISO/IEC, ITU-T 1984; CCITT X.208 1988; ISO/IEC 8824)
- INSPIRE (DT Data Specifications, 2008/ 2009)
- Place Identifier Architecture (ISO-TC211/19155 N2493 (NWIP 2008))
- OGC – Geographic Linkage Service (GLS, OGC 08-006r2; 2009-01-26 Draft)
- OGC – Abstract Specification, Topic 14: Semantics and Information Communities (99-114.doc, 1999; www.opengeospatial.org)

INSPIRE:n kehitystilanne

Yksilöivien tunnisteiden suhteen INSPIRE-direktiivin 8. artiklan mukaan täytännönpanosäännöissä on käsiteltävä

- (a) yhteiset puitteet paikkatietokohteiden yksilöivää tunnistamista varten, joihin voidaan rinnastaa kansallisten järjestelmien mukaiset tunnisteet niiden välisen yhteentoimivuuden varmistamiseksi; sekä lisäksi yksilöiviin tunnisteisiin liittyen
- (d) tietojen ajallista ulottuvuutta kuvaava tieto;
- (e) tietojen päivitykset.

Yksilöivät tunnisteet tulee olla aineistoliitteiden I ja II teemoilla, sekä liitteen III teemoilla, mikäli niitä käytetään referenssitietona. Tunnisteet ovat teema-kohtaisia.

Tunnisteiden roolina on

- helpottaa pelkkien muutostietojen päivytystä
- tunnistaa samat kohteet eri aineistoissa tai mittakaavoissa (yleistys)
- mahdollistaa inkrementaaliset päivitykset esim. tietyltä aikaväliltä.

Tunnisteen rakenteena on

- tietolähteen tai organisaation yksilöivä osa
- tiedon tuottajan antama kohteen lokaalitunnusosa
- versio-id.

Tunnisteiden elinkaarisääntöjen paikallisuus on vielä käsittelyssä, mutta tarkoitus on että ne olisi dokumentoitu metatiedoissa. Elinkaarisäännöt on määritelty teemakohtaisissa valmisteluryhmissä (TWG) aineistoliitteen I kohdalla toistaiseksi vain osoitteille, joskin INSPIRE-valmistelun tavoitteena on ollut, että elinkaarisäännöt määriteltäisiin em. rajauksella kaikille aineistoteemoille.

Taksonomioita, ontologioita...

Semanttisessa webissä (Web 2.0) tiedon merkityksiä eri käyttäjyhteyksissä kuvataan ontologioilla (~sääntö- tai suhdarakenteina) esim. RDF-pohjaisesti, tai jopa heuristisin menetelmin. GLS:n kohdalla vastaava rakenteistaminen herättää kysymyksen, edellyttääkö se lopulta esim. taksonomioita, jotka jäsentävät systemaattisesti ilmiömaailman aihealueita. Taksonomiat ovat osin sukua ontologioille, ja soveltuvat parhaiten eksakteihin ilmiöihin. Ontologiat puolestaan eivät rajaudu systemaattisiin ryhmittelyihin, ja niissä voi olla sääntöpohjaisuutta.

Semanttisen verkon malli tarkentuu?

Yksilöinnin problematiikka tiivistyy kohteiden identiteetin ja semantiikan hallin-

taan. W3C edustaa edellä mainittuihin kehityskulkuihin verrattuna kolmatta tietä: www:ssä on jo olemassa valmis nimipalvelu – infrastruktuuri (DNS), ja W3C:n Semanttisen Webin standardit kattavat yksilöintitunnusten määrittelyn sekä tunnusten käytön ohjeistuksen. Lisäksi W3C:n Semanttisen Webin standardit on pyritty määrittelemään siten, että alemmilla teknisillä kerroksilla voi tapahtua kehitystä ilman että sovelluksia tarvitsee jatkuvasti päivittää. Mm. nimeämiskysymykset ja URI-tunnisteiden valinta sekä jakautuminen osiinsa puhuttavat myös www-kehittäjiä.

STANDARDIEN SISÄLLÖISTÄ

• **Spatial referencing by geographic identifiers – ISO 19112**

ISO 19112:n lähtökohtana on määritellä sijainnillinen vertausjärjestelmä (spatial reference system) sijaintia ilmaiseville tunnisteille maantieteellisten hakemistojen (gazetteer) muodostamiseksi yhtenäisellä ja johdonmukaisella tavalla, sekä sijaintiviittausten kuten geokoodauksen ja -referoinnin tueksi.

• **UUID**

(Universally Unique Identifier)

UUID:n tarkoituksena on, että hajautetuissa järjestelmissä informaatio voidaan yksilöidä yksikäsitteisesti ilman merkittävää keskitettyä koordinaatiota. Tausta-ajatuksena näyttää olevan, että myöhemmin tunnisteet voitaisiin koota yhteen tietokantaan ilman että jouduttaisiin ratkomaan nimeämisiongelmiä. UUID suuntautuisi siten lopulta rekisterien (registry) ja rekisterijärjestelmien perustamiseen tiedonhallinnan vakauttamiseksi, ja mahdollisesti URI:ien käyttämiseen yksilöivinä tunnisteina.

• **OID (Object identifier)**

OID:a käytetään kohteiden nimeämiseen muodostaen solmun hierarkkiseen nimiavaruuteen (ASN.1). Uudet solmut varataan valtuutetulta rekisteröintiorganilta (esim. SFS). Suomessa OID on vakiintunut lähinnä terveysalalla. OID ei sisällä tiedon merkityksiä, vaan on pelkkä merkkijono. Solmut ilmaisevat aihealueen, mutta niiden perusteella ei vielä voida löytää tai yhdistellä dataa.

• **PI (Place Identifier Architecture)**

Place Identifier Architecture kuvaa arkkitehtuurin ja palvelurajapinnan "paikkatunnisteiden" (PI) hyödyntämiseen. PI-alustan (platform) avulla käyttäjyhteisöt voivat linkittää paikkatunnisteisiin liittyvää tietoa. Auktorisoitujen paikannimihakemistojen avulla kohteet (paikat) voitaisiin tunnistaa. Tämä voi toimia jossakin selvärajaisissa tai hallinnollisesti säädellyissä yhteisöissä, mutta ei mitenkään kattavasti. Paikannimien ja -tunnisteiden semantiikkaan työaihe ei ota kantaa. Kolmansille osapuolille esitetty arkkitehtuuri tarjoaa apuvälineitä informaation yhdistelyyn. Voidaan ajatella, että INSPIRE olisi tällainen hallinnollisesti säädelty kokonaisuus, jossa "annotointi" hakemistoon voidaan hoitaa. Sitä varten PI-arkkitehtuuri on kovin massiivinen, ja osin rinnakkainen INSPIRE:lle.

• **GLS**

GLS-standardiehdokas määrittelee rakenteen palvelurajapinnalle, joka mahdollistaa paikkatietoa sisältävän datan yhdistelyn. Rajapinta on esitetty massiivisena taulurakenteena (yht. 63 taulua!), joissa kohteet, niiden ominaisuudet ja yksilöinti verkottuvat taulurakenteen relaatioiden kautta URI-rakenteita ja XML:ä hyödyntäen. URI esiintyy kuitenkin vain FrameworkDataset- ja Dataset-tasoilla. GLS esittelee kaksi "omaa" operaatiota: GetData ja JoinData, operoi XML-tiedostoilla ja tukee versiointia. GLS:n oleellinen lisäarvo on sen tavoittelema attribuuttitiedon hallinta. Voisiko tämän tehdä yksinkertaisemmin, ja käyttäen URI:a myös kohdetasolla?

Konkreettisesti semantiikka-näkökulma tuo terminologiaan semanttiset tunnisteet, joilla tarkoitetaan sitä että yksilöivät tunnisteet voivat olla myös semanttisia tunnisteita. Semanttisen tunnisteiden komponentteina olisivat yksinkertaisimmillaan sen muodostavien geometrinen osien tunnisteet taulurakenteena.

EU ja eurooppalaiset toimijat

SEMIC.EU kokoaa sanastorakenteiden raaka-ainemateriaalia jatkotyöhön, jota parhaillaan suunnitellaan. Lisäksi EU:ssa semanttiseen nimipalveluun liittyy OCCAM-hanke. Tässä suhteessa SEMIC.EU on relevantti paikkatietoalalle ja INSPIRE:n toimeenpanossa.

INSPIRE:ssä yksilöivien tunnisteiden referensseinä ovat vahvasti UK:n karttalaitoksen Ordnance Survey'n topografinen tunniste (TOID, <http://en.wikipedia.org/wiki/TOID>) ja Hollannin ID-käytäntö, jossa kohteille annetaan kaikissa muutostilanteissa uusi tunniste lukuun ottamatta mittaustarkkuuden tarkennuksia. Hollannin mallissa tarvitaan kohtalaisen laaja versiohallinta.

UK:n The digital national framework (DNF) -konseptissa lokaalien tunnisteiden yksilöivä organisaatiokohtainen etuliite on tallennettu DNF-rekisteriksi. Kohteiden muutokset ylläpidetään muutosattribuutteina (change attribute). Organisaatiot määrittelevät kohteilleen elinkaarisäännöt.

Yhteys Ordnance Survey'n (OS) kohteisiin (TOID) muodostuu 16 merkin tunnisteella, joiden avulla OS hallitsee kaikkien organisaatioiden tuottamat kohteet. Jos käyttäjän ja alkuperäisen tiedontuottajan elinkaari kohteelle on erilainen, tarvitaan kohde- ja versiohistoriaa.

Ordnance Surveyssa yksilöivät tunnisteet on koettu käyttöönoton jälkeen oleellisiksi tietojen ylläpidossa ja aineistojen yleistystarpeissa. Niiden avulla myös varmistetaan, että tiedot on oikein mallinnettu ja kohteiden elinkaari hallittu, ja tiedot voidaan toimittaa asiakkaille joko kokonaispäivityksenä taikka vain muuttuneiden tietojen osalta asiakkaan todellisen tietotarpeen mukaan. Seurauksena on myös selkeitä laatuhyötyjä; kun asiakkaat määrittelevät tietotarpeensa tarkemmin heidän prosessinsa tehostuvat, ja toisaalta myös OS voi tehostaa resurssiansa kohdentamista asiakastarpeiden mukaan. Molemmat hyötyvät.

On ilmeistä, että kansallisella ja EU-tasolla tarvitaan DNF:ä vastaava viite- ja säädöskehys, jotta yksikäsitteisten yksilöivien tunnisteiden laaja käyttöönotto ja sen tuomat hyödyt ovat saavutettavissa.

Tällaista ei toistaiseksi ole määritelty ja spesifioitu INSPIRE-valmistelussa, vaan aihe on jäänyt virallisten INSPIRE-valmisteluryhmien katvealueeseen, eikä mikään ryhmä ole syventynyt siihen toiminnallisista tai infrastruktuurin lähtökohdista.

ESDIN (European Spatial Data Network)

ESDIN-projekti (www.esdin.eu) määrittelee INSPIRE-direktiivin jalkauttamista karttalaitosten näkökulmasta niiden eurooppalaisten yhteistuotteiden suhteen. Eräänä työaiheena on myös eurooppalaisen yksikäsitteisen yksilöivän tunnisteiden (EUID) ja sitä koskevien suuntaviivojen täsmäntäminen. Maanmittauslaitos on mukana hankkeessa ja tässä työaiheessa, jota koskevat ehdotukset valmistuvat ensi vuoden puolella.

Standardoinnin jakolinjat

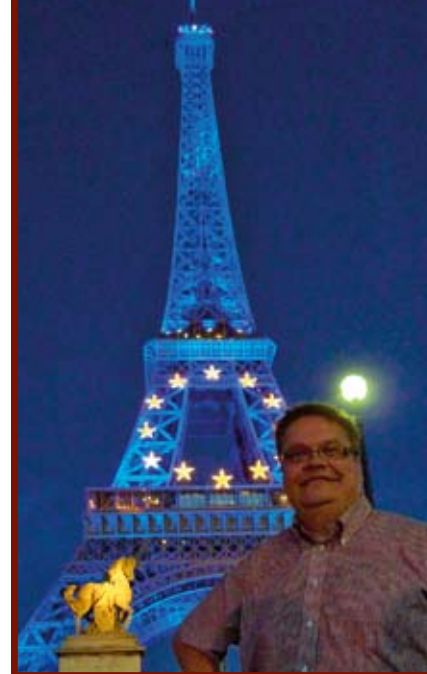
ID-standardoinnin kuvattu tilanne lienee vain eduksi kestäväälle kehitykselle ja toivottavasti toimii kannusteena OGC:n ja W3C:n vastaiselle kumppanuudelle. INSPIRE:llä on rooli päänavaajana Euroopassa.

Epäilemättä olisi eduksi, että paikkatieto- ja Web 2.0 -yhteisöt kohtaisivat tämän aiheen vaiheilla merkityksellisen tiedon haltuunoton läpimurtona. Web 2.0 ja GLS edellyttävät molemmat sanastotyötä, jonka viitekehys tulee ensin vankasti kiinnittämään. Sanastotyö on vahvasti tulollaan paikkatietomarkkinoille.

Aiheeseen liittyvää Suomessa:

- Pääjohtaja Jarmo Ratia; Maastotietokanta, *Maankäyttö* 3/2008; s. 17. www.maankaytto.fi/arkisto/mk308/mk308_1173_ratia.pdf
- FinnOnto 2.0 (Semantic web 2.0) www.seco.tkk.fi/projects/sw20/.

**Kirjoittaja toimii johtavana
asiantuntijana Maanmittaus-
laitoksessa, sähköposti
esa.tiainen@maanmittauslaitos.fi.**



KUKA ENSIMMÄISENÄ tiesi Amerikan olemassa olosta Euroopassa? Vaikka Kolumbus löysikin Amerikan olivat viikingit käyneet Kanadan pohjoisosissa jo aiemmin. Kartografian kannalta asiaan liittyy ns. *Vinlandin kartta*, joka löytyi 1950-luvulla ja on nykyisin Yalen yliopistossa. Kyseinen kartta on ajoitettu vuodelle 1440, siis n. 50 vuotta ennen Kolumbusen kuuluisaa matkaa 1492. Löydön jälkeen karttaa on tutkittu ahkerasti ja epäilyjä sen aitoudesta on esitetty. Osallistuin heinäkuussa kansainväliseen kartografian historian konferenssiin Kööpenhaminassa, jossa esiteltiin tanskalaisen tutkijaryhmän tuloksia. Heidän johtopäätöksensä erilaisten analyysien jälkeen oli, että karttaa on hyvin vaikea pitää väärennöksenä ainakaan teknisen analyysin perusteella. He olivat mm. tutkineet mustetta, väitetyjä tuplaviihoja ja sitä olivatko paperissa olevat madonreiät olleet ennen paperille piirtämistä ja todeneet ettei näin ollut. Kyseinen konferenssi pidetään Suomessa vuonna 2013.

Syksyllä on mukava muistella vietettyä kesää. Tänä vuonna kesäloma jakaantui kahteen osaan. Ensimmäisen osan vietimme Euroopassa ja Ranskassa ja toisen osan Suomessa. Juhannuksen jälkeen pakkasimme uuden auton ja suuntasimme kohti Vuosaaren uutta satamaa. Yhden laivalla nukutun yön jälkeen hurautimme suoraan ilman tulli- tai rajamuodollisuuksia laivasta Lyypekkiin, jossa yövyimme. Lyypekin jälkeen suuntasimme Hampurin kautta kohti Bremeniä jossa ihailimme Grimmin veljesten sadun soittoniekkvoja. Yön jälkeen siirryimme Saksasta Hollantiin jossa yllättäen rajalla moottoripyöräpoliisi pysäytti meidät. Passien tarkastuksen ja puhalluksen jälkeen vierailimme mm. Leidenissa, Haagissa ja Delftissä.