Georef - Linked Data Deployment for Spatial Data; Finnish Initiative

Esa TIAINEN, Finland

Key words: Linked Data, Semantics, Ontology, Web, RDF, Big Data

SUMMARY

'Georef' is an initiative for a service and application development platform that employs httpURIs of place names for geocoding different data assets to enable and improve data combinations of spatial data and any other data using linked data technology.

The scope of Georef is including wide application areas, like re-use of scientific and research data across disciplines, re-use in media (local-national), public and personalized services (health, education, professional training, immigrants.) as well as commercial services. As for societal aspects Georef facilitates inclusion of community members e.g. in municipal decision making by providing sufficient information base where crowdsourcing is envisioned for updating current and historical contents related to places and areas to provide and create novel viewpoints and information in land and city development (citizen science). Here also privacy protection shall be considered as well as in any commercial use. In short, it is about geocoding by place names, which makes local to global.

Place names can be used for bridging different information and data assets as, a lot or even most of information carry place names but most information do not carry direct location data. This applies to data assets of any format: detailed contents in scientific research reports, different types of documents; factual and fiction, textual, images, photos, movies, music etc. To link or combine spatial data with these other data types, first we need place names bound to coordinates. But place names are tricky too: many places have the same names (Paris, Texas), there are endonymes, exonyms and different conjugations in different languages etc. Secondly to be deployable, they need unique identifiers i.e. httpURIs. To be live data links the simplest solution is to implement a RDF database which enables search by SPARQL as well as transformations to different popular formats like geoJSON.

To combine spatial data with all other data is to boost and verify the mighty potential of spatial data and to release the dead capital of information resources; by adding value, by making visible phenomena and reasons behind them, visualizing and verifying processes of long term. Furthermore place names are fundamental reference data in INSPIRE. Globally other place name sources such as Open Street map may be employed.

Georef - Linked Data Deployment for Spatial Data; Finnish Initiative

Esa TIAINEN, Finland

1. NATIONAL URI-RECOMMENDATION ON SPATIAL DATA (2015)

1.1 Scope of the national recommendation "Unique identifiers of geographic data" (JHS 193)

The recommendation for public administration regarding unique identifiers of spatial data was approved September 1st 2015. It defines the structural model of URIs and practices for redirection in data retrieval, and offers instructions for the implementation of identifiers and lifecycle rules for geographic features. In addition, this recommendation presents the utilisation of URIs and examples of use cases. The structural model of URIs in accordance with this recommendation is intended to the identification and linking of not only spatial data but all data within the entire field of public administration, which forms the basis linked data infrastructure for spatial data.

Similarly the *INSPIRE Directive* requires that the seven data themes of its Annex I, the four data system-independent identifiers in accordance with the implementation rules of the Directive. In the preparation of the implementation rules of the Directive and in the data product specifications, it has been recommended that they also be applied to other geographic features that are required by several user groups.

The objective of the recommendation is to establish and harmonise the standard structure of unique geographic information identifiers and their release in machine-readable format, as well as consistent procedures for redirection when using identifiers. Another objective of this recommendation is to adapt the INSPIRE requirements to existing identifier practices and related JHS recommendations.

1.2 Towards linked data ecosystems

To verify interoperability, the INSPIRE Directive demands that unique identifiers be released in HTTP URI format. Accordingly in the national recommendation, "URI" refers to identifiers in HTTP URI format, which makes the identifiers data system-independent identifiers. The purpose of the *INSPIRE Directive (2007/2/EC)* and its implementing rules is to standardise the shared use of data in European infrastructure for spatial information and to improve the shared use of registers in public administration.

Unique identifiers and their management methods (realisation of the linked data concept) form central parts of cross-administrative and more extensive interoperability. To this end the national recommendation for unique identifiers is taking a step further in interoperability and data integration introducing a framework for linked data infrastructure for spatial data and any other data, and within a scope of application wider than that of the INSPIRE Directive.

Furthermore, unique URIs are necessary for open data application developers, particularly in the maintenance of reference data, and, therefore, they improve the openness of data and the benefits obtained from it in a concrete way. Electronic transactions and the crowdsourcing of data collection and maintenance offer diverse ways to utilise them.

In addition to the national infrastructure for spatial data, unique identifiers are also included in the geographic information reference architecture in the overall public administration architecture to enable linked data infrastructure and ecosystems regarding spatial data.

2. DATA LINKING PRINCIPLES

2.1 Identification based on data modelling

A data model presents and defines a description of a real-world object in accordance with data needs within its scope of application. A single real-world object can be described by several data features based on different data models. Therefore, geographic features based on different data models describe a single real-world object. Unique identifiers describing a single real-world object can be accompanied by a reference to this real-world object, allowing the properties (attributes) of features in accordance with different data models to be linked.

In the national recommendation, real-world objects are referred to using URIs. The placeholder procedure is applied to identifiers of real-world objects so that the URI of the geographic feature modelling it, in which the /so/ component is replaced by the /id/ component, represents a real-world object. This also answers the philosophical question of what is a real-world object or its identity.

2.2 Structure of a unique identifier

The URI consists of a domain name, a URI type, a dataset identifier which identifies the data source of a geographic feature and a local identifier of a geographic feature. In addition, the URI may include a version identifier (not mandatory) which is generated in accordance with the lifecycle rules of the geographic feature.

The URI format is as follows:
http://{domain name}/{URI type}/{dataset identifier}/{local identifier}[/{version identifier}]

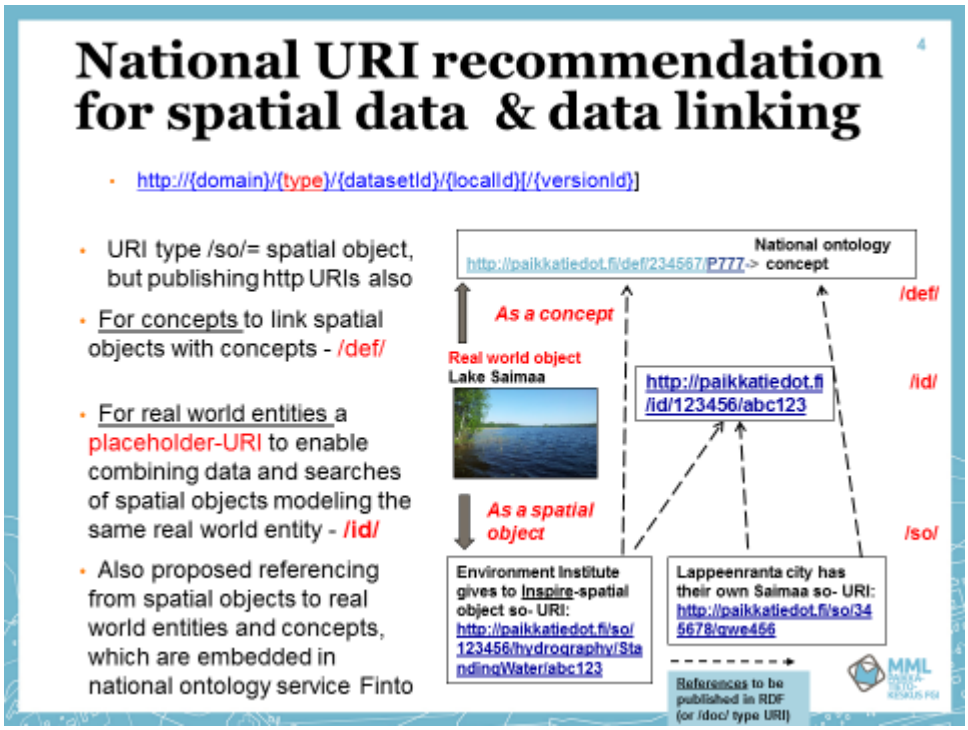
The general domain name of URIs for geographic information is http://paikkatiedot.fi which acts as a central redirection service. This redirection service directs all references to URIs via a central service to services of data producers that provide users with responses in accordance with this recommendation. The redirection service is managed by the National Land Survey of Finland.

The path component {URI type} included in the URI indicates the type of the data resource being referred to:

'id' - a real-world object or phenomenon
'so' - a geographic feature
'def' - a concept represented by the geographic feature

'doc' - documentation representing an object, for example, its different forms of portrayal

The dataset identifier identifies the data source of a geographic feature. For geographic information, the dataset identifier used is the dataset identifier (a seven-digit numerical set) of the Finnish discovery service Paikkatietohakemisto (GeoNetwork) and minted from metatieto@maanmittauslaitos.fi by the National Land Survey .
A dataset may be a geographic dataset, a dataset series, a vocabulary or a code list.



Data producers can also refer to the identifier of a real-world object released under the paikkatiedot.fi domain name in the RDF description linked to the documentation identifier of their geographic feature representing the same real-world object.

2.3 Linking concepts

The release of URIs for geographic feature types, i.e. concepts, in addition to geographic features enables an interface service in the infrastructure for spatial information to search for ontological concepts and their interrelations and, using concepts, for related geographic information.

The concept source may be any vocabulary used by the data producer. For example, the vocabulary may be an ontology, data specification, schema, code list, taxonomy or thesaurus.

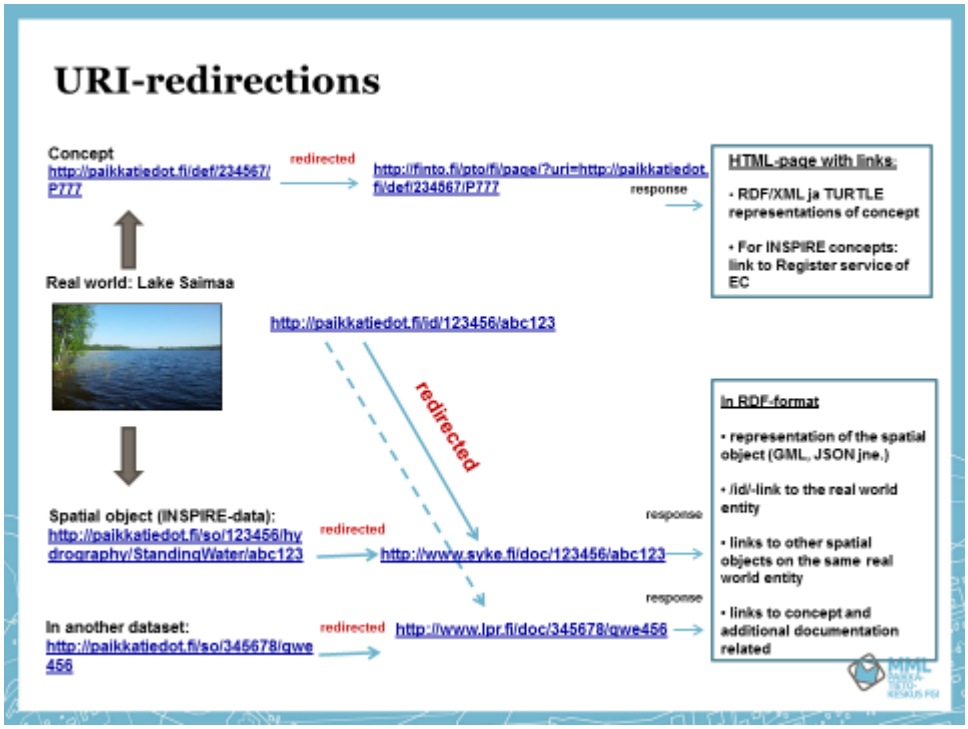
The dataset identifier of a vocabulary acting as the concept source can be obtained from metatieto@maanmittauslaitos.fi. For example, the dataset identifier of the Finnish Geospatial Domain Ontology is 1001000.

In order to be able to refer to concepts in a unique way, each concept included in the vocabulary must be provided with a unique URI. A redirection can be made from the URI of a concept to, for example, the (national) Finnish Thesaurus and Ontology Service (finto.fi) or a catalogue service including schemas.

2.4 Practices of redirection and linked data

Linked data requires that the namespaces and the URI structure be agreed and arranged at a national level. As a result, the data resources of society can be used effectively to develop public administration services and new business operations and to use data extensively within society.

Redirection practice for linked data are represented in the following diagram:

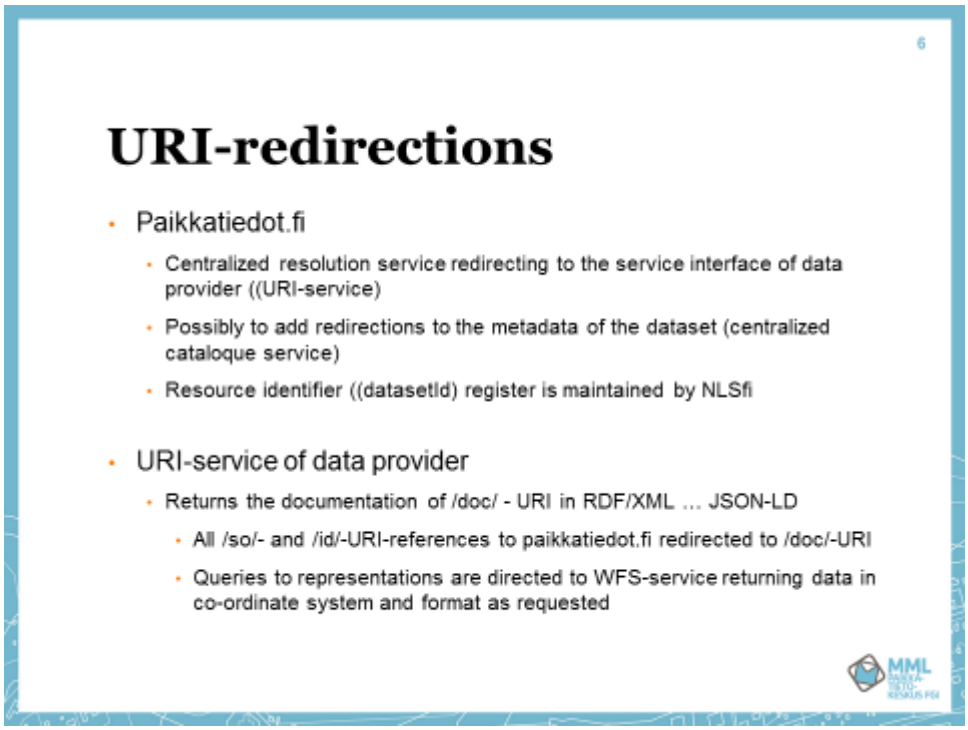


The identifiers of geographic features, concepts and real-world objects start with http://paikkatiedot.fi (spatialdata.fi), which is a centralized resolution service redirecting to the service interface of data provider called URI-service offered by the data producer. The documentation identifier returns the instance linked to it or the RDF description of the real-world object.

Using the data included in the RDF description of the geographic feature linked to the documentation identifier, users of data can link the geographic feature to the real-world object modelled by it, to the concept representing the feature and to other information related to the feature.

Using the identifier of the real-world object, geographic features representing the same real-world object and other documentation that has been presented using documentation identifiers can be linked to geographic features using redirection.

Linking the URI of a concept to a geographic feature enables the traceability of geographic features using, for example, ontology searches in the Finto service, in which case the efficiency of data enrichment and the use of existing data improves. Other benefits can be obtained from crowdsourcing applications, such as the collection of data and the allocation of feedback, and from the improved efficiency of application development.



2.5 Responding practices

The recommended responding practices of different URI types are presented in the following.

Identifier of a real-world object (id)

Content of the response: Redirection to the documentation identifier

Identifier of a geographic feature (so)

Content of the response: Redirection to the documentation identifier

- If the URI is given without any local identifier, redirection to metadata of the dataset
- Different forms of portrayal of geographic features are presented in a separate document.

Concept identifier (def)

Content of the response: Redirection to the documentation identifier

Documentation identifier (doc)

Content of the response: The data content of the data feature and references to optional forms of portrayal of the data content and to other data features and real-world objects

Recommended forms of portrayal: RDF/XML, JSON-LD, Turtle

2.5.1
2.5.2
2.5.3 Subtitle, level 3

Type text here …

3. IMPLEMENTATION APPROACH

The implementation strategy proposes that the data providers populate also /id/- and /def/- URIs with same as /so/-URI.

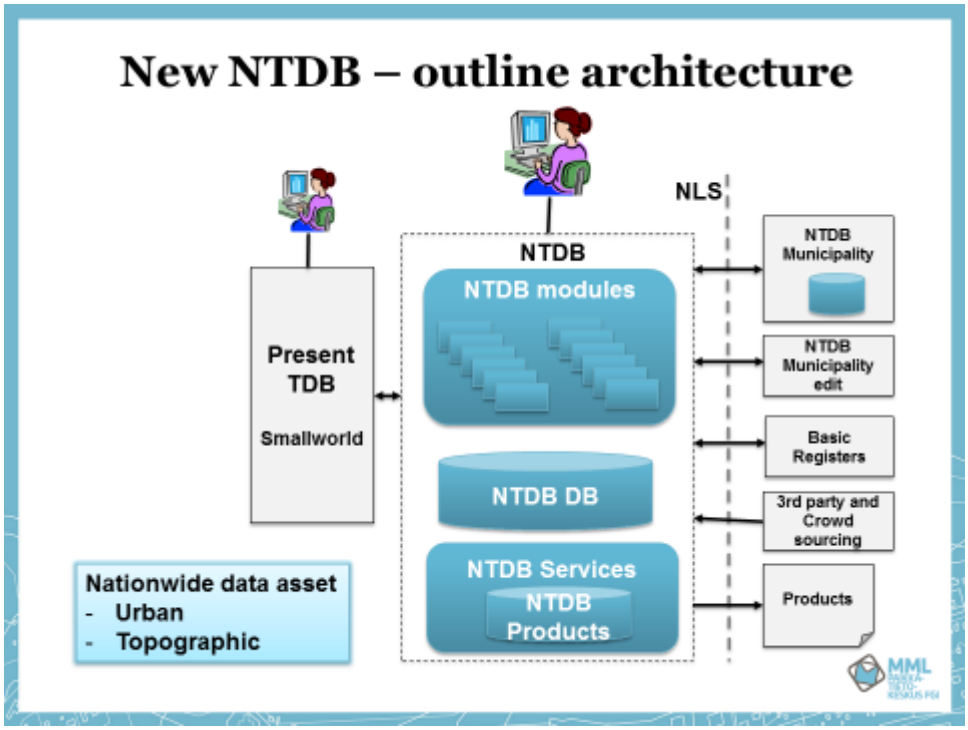
The URIs for spatial data are all minted in nationally centralized domain with redirections to URI-services of data providers. As a first stage the infrastructure is established with INSPIRE data - a critical mass and stepping stone. URIs for spatial objects shall be delivered through national geoportal. Unique ids are employed to establish a URI-based production of national core location data (NTDB).

Using URIs service providers and data consumers can

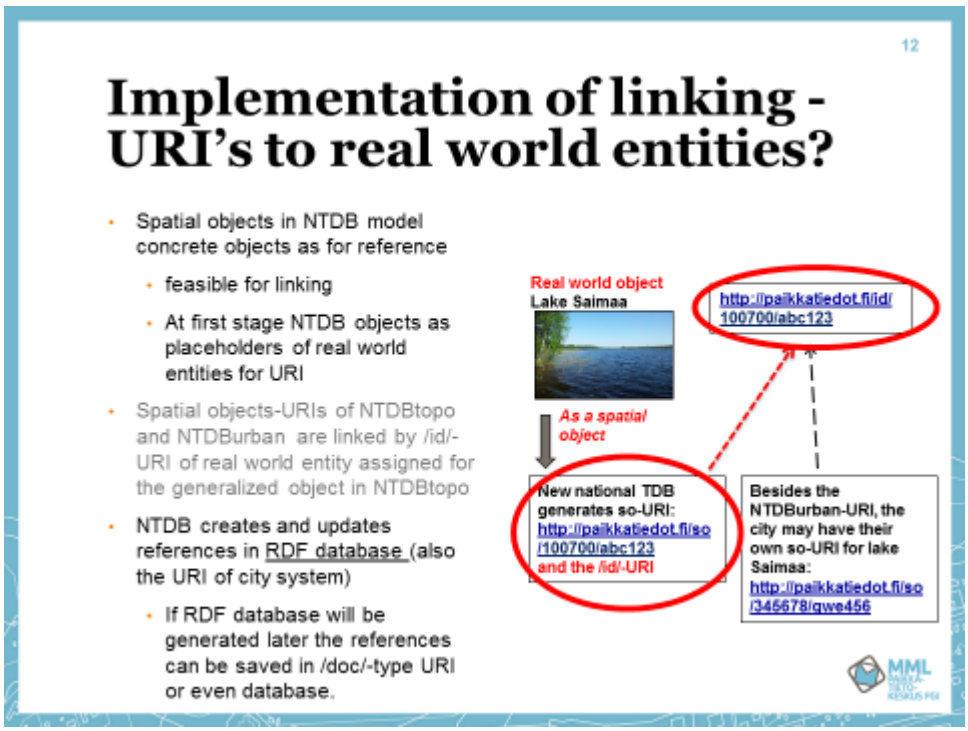
- combine additional data from different data sources (added value)
- consume data different sources with better quality or coverage and from local to national
- have more flexibility in enriching data and delivery (RDF, GeoJSON)

Data integration through URI's brings great benefits to the society at large

- Finnish Government Open Data Program (2013-2015): "The greatest benefits result from linking different type of data"
- URIs are glue integrating processes using the same data and as such Digitalisation enabler
- Crowdsourcing enabler – to identify the very spatial object.



NTDB generates and submits spatial object-URIs to the spatial objects of city system for updates or reuse.



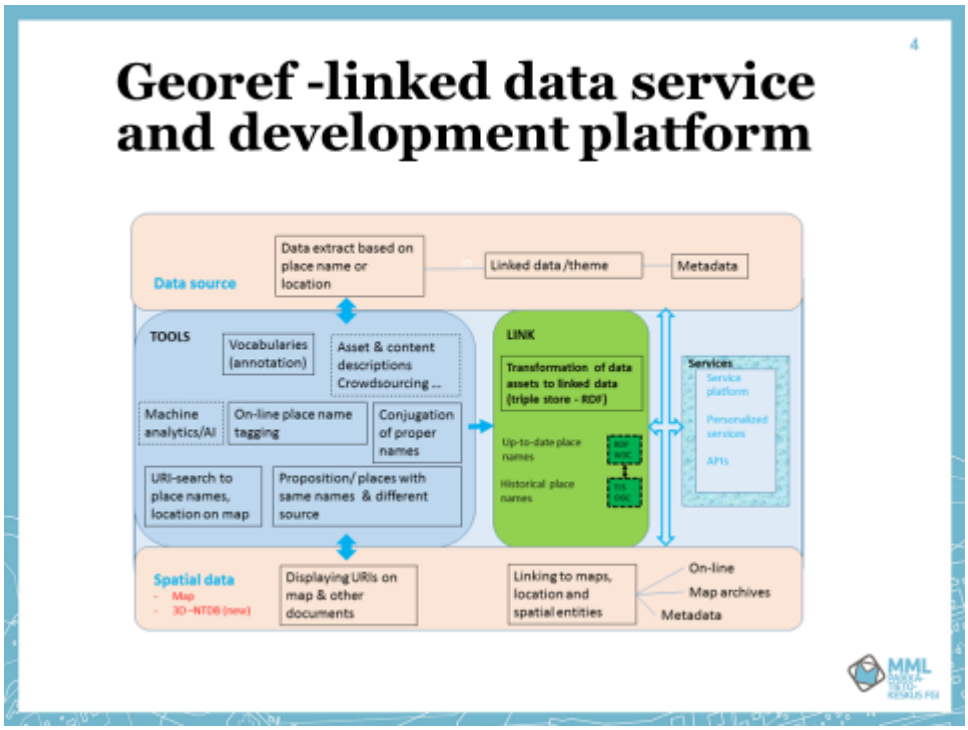
4. GEOREF – GEOCODING BY PLACE NAME

Georef is an initiative targeting to enable and improve data combinations of spatial data and any other data deploying URIs of place names and towards nationwide linked data SDI.

Place names can be used for bridging different information and data assets as, well known, a lot or even most of information carry place names but most information do not carry direct location data. This applies to data assets of any format: scientific research reports, different types of documents; factual and fiction, textual, images, photos, movies, music etc. To link or combine spatial data with these other data types, first we need place names bound to coordinates.

But place names are tricky too: many places have the same names (Paris, Texas), there are endonymes, eksonymes and different conjugations in different languages etc. To be deployable, secondly, they need unique identifiers i.e. httpURIs. To be live links for data the simplest solution is to implement a RDF database which enables search by SPARQL as well as transformations to different popular formats like geoJSON and other.

Place names for bridging spatial data to different information and data assets:



REFERENCES

INSPIRE Directive (2007/2/EC) and its implementation rules

Commission Regulation 1089 (2010) and its amendment 1253 (2013), cf. literature references

Spatial Data Infrastructure Act (421/2009) and Spatial Data Infrastructure Decree (725/2009)
Vocabulary of Geoinformatics, 3rd edition, Finnish Terminology Centre TSK, 2014. ISBN 978-952-9794-34-8

JHS 158 Metadata for the geographic information (http://www.jhs-suositukset.fi/)
JHS 162 Modelling of the geographic information for data transfer (http://www.jhs-suositukset.fi/)
JHS 177 Specification of a geographic data product (http://www.jhs-suositukset.fi/)
JHS 180 Content services for geographic information (http://www.jhs-suositukset.fi/)
JHS 193 **Unique identifiers of geographic data** (http://www.jhs-suositukset.fi/)

RFC 4122 A Universally Unique IDentifier (UUID) URN Namespace

ISO 8601 Data elements and interchange formats – Information interchange – Representation of dates and times

ISO/IEC 8824-1:2008 Information technology – Abstract Syntax Notation One (ASN.1): Specification of basic notation

ISO 19148 Linear Referencing

BIOGRAPHICAL NOTES

Esa Tiainen, M.Sc. Senior advisor at National Land Survey of Finland (NLS), has extensive experience GIS, semantics and data management. Currently intensive on Linked Data and Geospatial Data, recently completed chairing a national recommendation on unique identifiers for spatial data. Served from 1982 City of Helsinki, removed 1987 to Helsinki Metropolitan Area Information Management. From 1992 served NLS since 2000 as senior advisor at NLS in strategic projects such as Inspire based SDI and EU-projects i.e. several INSPIRE WGs, E.L.F. (European Location Framework (2014-) ESDIN (2008-2011); EULIS since 2002, as WP leader in EULIS 2.0/LINE-project 2010-2012. Beyond EU: 2015- UN GGIM Europe WG B, 2007-2013 leading a co-op project of NMCAs for developing a SDI to Russian Federation; since 2011 Arctic SDI project.

CONTACTS

Senior advisor Esa Tiainen
Finnish Geospatial Research Institute FGI, National Land Survey
Opastinsilta 12
Helsinki
FINLAND
Tel. +358 40 5570512
Email: esa.tiainen@nls.fi
Web site: www.nls.fi