WeTransform!

GeoPackage for INSPIRE

Thorsten Reitz, Co-Founder/CEO

April 13th, 2021

Agenda

1. Introduction
2. Data usability and usefulness
3. An overview of encodings: GML, GeoJSON, GeoPackage
4. Model Transformations
5. The END use case (including demo)
6. INSPIRE and Geopackage: Next steps and Timeline, getting involved
7. Summary, Questions & Answers

wetransform in a Nutshell

Who are we and for whom do we work?

- Based in Darmstadt, Germany
- Mission: Building Green Data Ecosystems
- More than 100 customers, over 4,000 users
- hale»studio: OS software for data transformation
- hale»connect: Integrated & automated (INSPIRE) data platform


Simon


Thorsten


Sven


Claudia


Livia


Nicolas


Akshat


Johanna


Kate


Florian


Franziska


Christopher

Miljø- og Fødevareministeriet
Miljøstyrelsen

Rijkswaterstaat
Ministerie van Verkeer en Waterstaat

Europäische Umweltagentur

TNO

GEONOVUM

BEUTH HOCHSCHULE FÜR TECHNIK BERLIN
University of Applied Sciences

LANDKREIS NIENBURG / WESER

M.O.S.S.
Computer Grafik Systeme Geoinformationssysteme

Arbeitsgemeinschaft der Vermessungsverwaltungen der Länder der Bundesrepublik Deutschland (AdV)

RheinlandPfalz
LANDESAMT FÜR VERMESSUNG UND GEOBASISINFORMATION

LGL

umweltbundesamt

irCELine + + +

swisstopo

PRISMA solutions

ForstBW

MINERVA INTELLIGENCE

JRC
EUROPEAN COMMISSION

Komm.ONE

GOBIERNO DE ESPAÑA
MINISTERIO DE AGRICULTURA, ALIMENTACIÓN Y MEDIO AMBIENTE

EPSILON Italia

Making Open Standards Work!

Agenda

1. Introduction
2. **Data usability and usefulness**
3. An overview of encodings: GML, GeoJSON, GeoPackage
4. Model Transformations
5. The END use case (including demo)
6. INSPIRE and Geopackage: Next steps and Timeline, getting involved
7. Summary, Questions & Answers

Three Steps to Useful Data

Why are we discussing GeoPackage et al.?

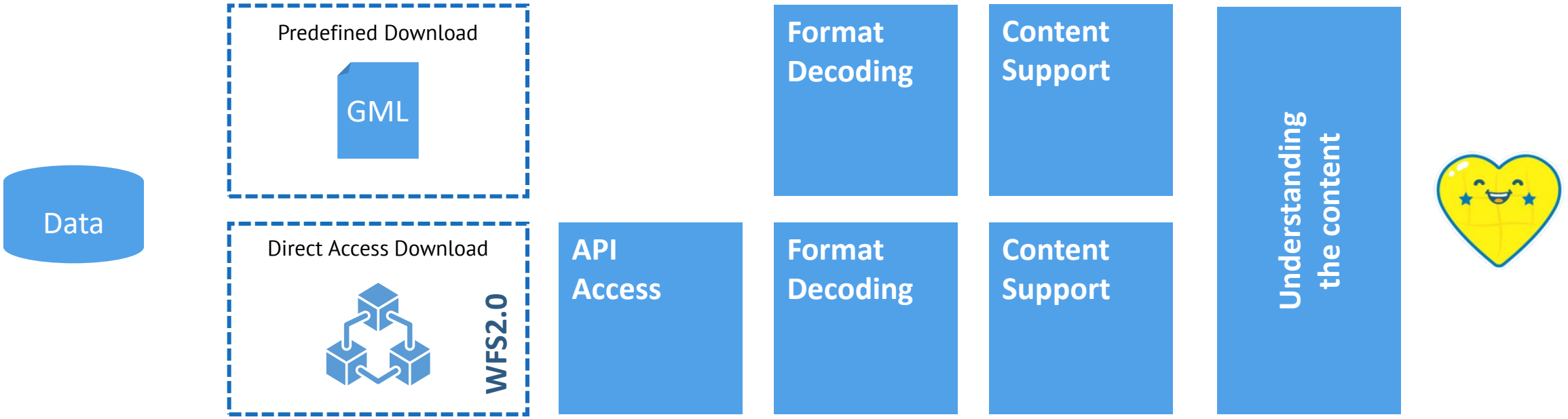
- Can I solve my problem?
- Can I find data to solve my problem?
- Can I access the data?

- Can I use the data in my environment?
- Do I need to spend a lot of effort to integrate it?

- Can I use the data to solve my problem?
- Can I share the solution with others?

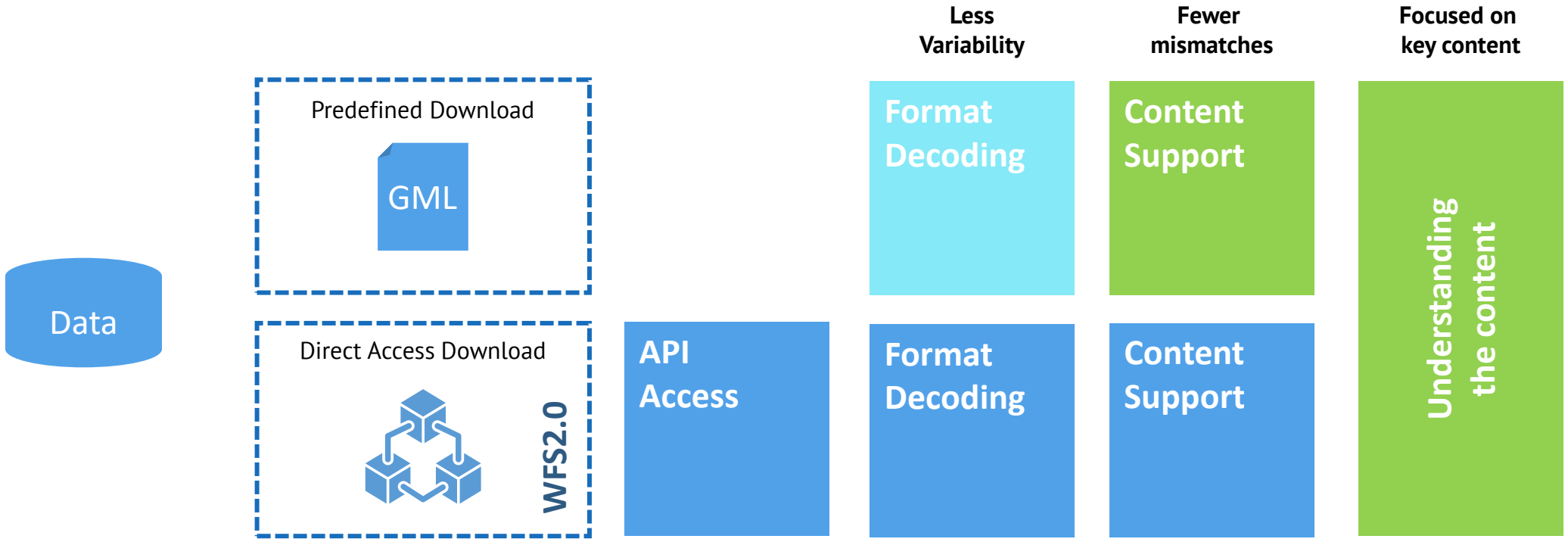
Starting to use the data

What hurdles are there before I can use the data?



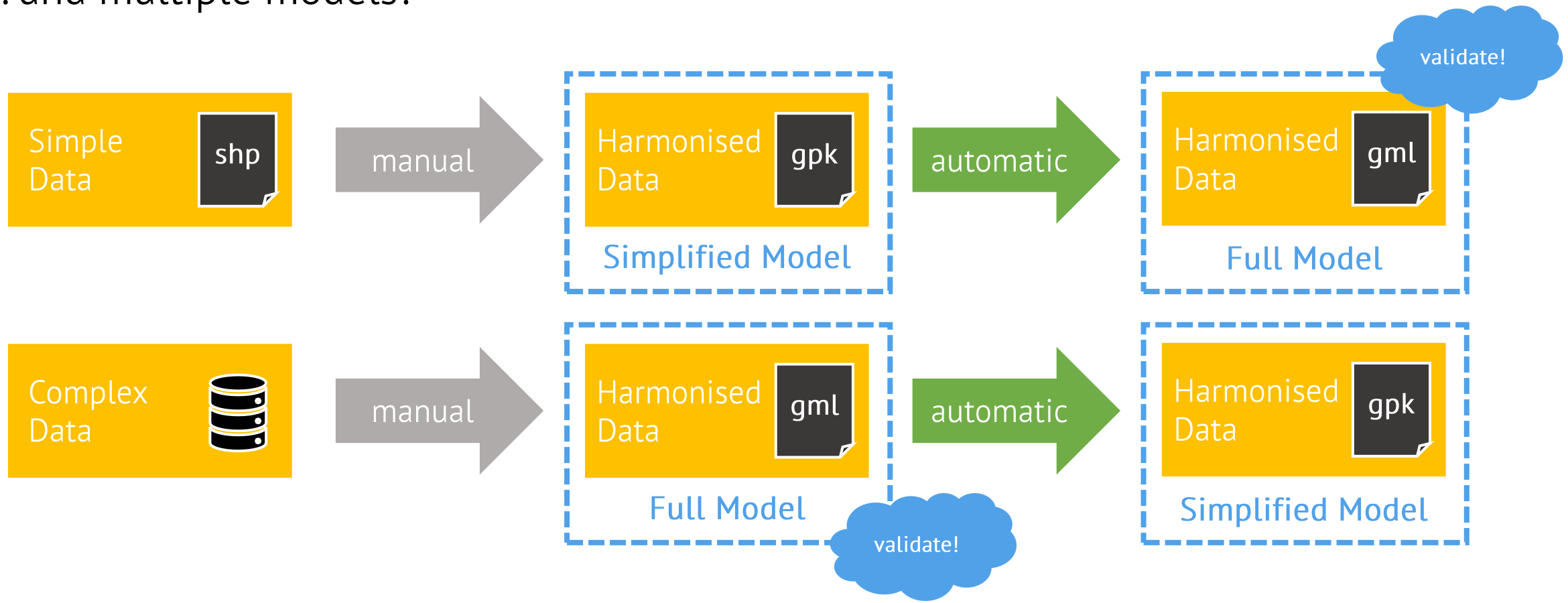
The Impact of GeoPackage

What does it change, compared to GML?



Using multiple encodings

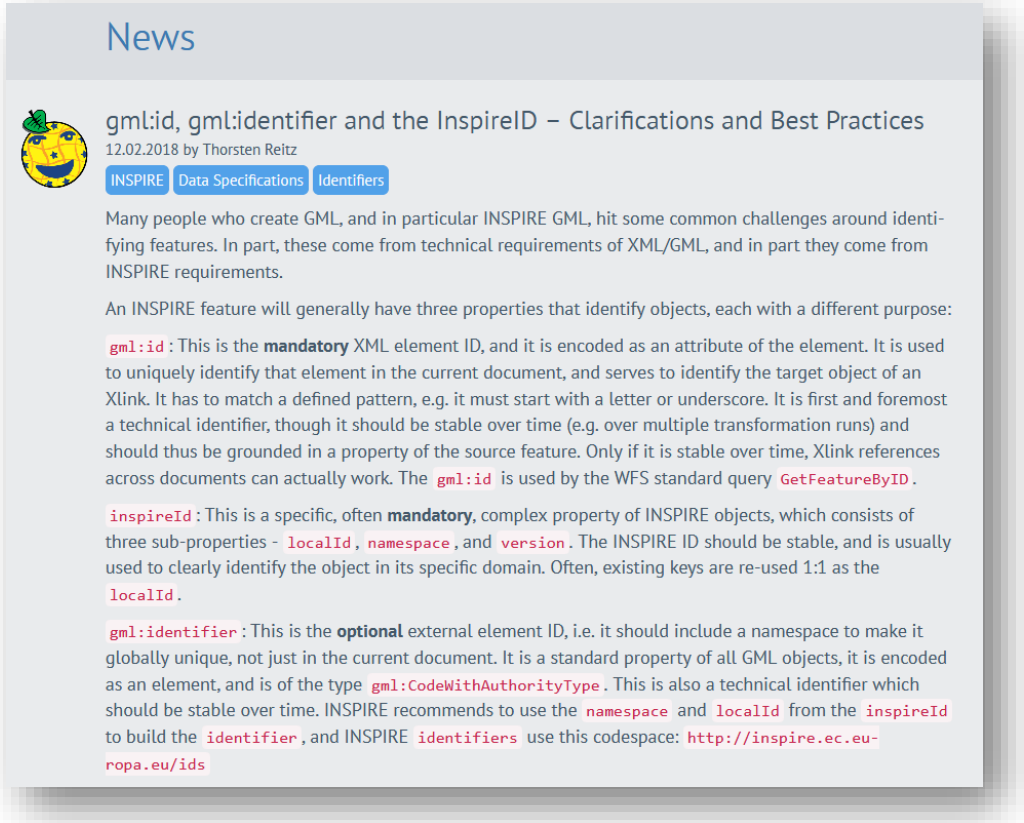
… and multiple models?



Apply Best Practices

Common approaches help improve interoperability

- Apply common rules for IDs (and Identifiers and inspireIds)
- Use or Agree on shared code lists, even if they are not perfect, and document mismatches
- For alternative encodings (GeoJSON, GeoPackage): Use existing guidelines where they exist already



News

gml:id, gml:identifier and the InspireID – Clarifications and Best Practices

12.02.2018 by Thorsten Reitz

INSPIRE Data Specifications Identifiers

Many people who create GML, and in particular INSPIRE GML, hit some common challenges around identifying features. In part, these come from technical requirements of XML/GML, and in part they come from INSPIRE requirements.

An INSPIRE feature will generally have three properties that identify objects, each with a different purpose:

`gml:id` : This is the **mandatory** XML element ID, and it is encoded as an attribute of the element. It is used to uniquely identify that element in the current document, and serves to identify the target object of an Xlink. It has to match a defined pattern, e.g. it must start with a letter or underscore. It is first and foremost a technical identifier, though it should be stable over time (e.g. over multiple transformation runs) and should thus be grounded in a property of the source feature. Only if it is stable over time, Xlink references across documents can actually work. The `gml:id` is used by the WFS standard query `GetFeatureByID`.

`inspireId` : This is a specific, often **mandatory**, complex property of INSPIRE objects, which consists of three sub-properties - `localId`, `namespace`, and `version`. The INSPIRE ID should be stable, and is usually used to clearly identify the object in its specific domain. Often, existing keys are re-used 1:1 as the `localId`.

`gml:identifier` : This is the **optional** external element ID, i.e. it should include a namespace to make it globally unique, not just in the current document. It is a standard property of all GML objects, it is encoded as an element, and is of the type `gml:CodeWithAuthorityType`. This is also a technical identifier which should be stable over time. INSPIRE recommends to use the `namespace` and `localId` from the `inspireId` to build the `identifier`, and INSPIRE `identifiers` use this codespace: `http://inspire.ec.europa.eu/ids`

Recommendation: Apply best practices instead of falling into the "Not-Invented-Here" Trap

Agenda

1. Introduction
2. Data usability and usefulness
3. **An overview of encodings: GML, GeoJSON, GeoPackage**
4. Model Transformations
5. The END use case (including demo)
6. INSPIRE and Geopackage: Next steps and Timeline, getting involved
7. Summary, Questions & Answers

Comparing Encodings

No silver bullets - each encoding has advantages and weaknesses

	...well suited for:	...not suited for:	Optimal Use Case
GML	- Production & consumption in back-end systems - Small to large payloads - All geometries - High precision geometries - Lossless encoding of complex models with many associations - Validation/Quality Assurance	- Performance critical web/mobile applications - Direct usage in Desktop GIS - Large data sets (> 10GB)	Systems data exchange (middleware), Compliance Testing
GeoPackage	- Consumption in desktop GIS - Delivery of whole data sets - Medium to large data sets(> 1 GB) - Most types of geometries - High-precision geodata - Data + (cartographic) Views	- Delivery of small sets of features (overhead) - Validation/Quality Assurance - Extremely large data sets (> 100GB)	Delivery of entire data sets with multiple types/ table to GIS users
GeoJSON	- Consumption in web & mobile applications - Smaller Payloads (< 10MB) - Simple geometries (limits on types of geometries) - Limited Precision geometries	- Performance critical web/mobile applications - Large payloads - Complex geometries, topologies* - High-precision geometries (< 1m) - Validation/Quality Assurance	Web application accessing individual sets of relatively simple spatial features

Performance, Compared

Benchmarks for common Geoformats, normalized to Shapefile

Format	Size	Size (zip)	Read Full*	Read Filter*	Write Full*
Shapefile	1.0	1.0	1.0	1.0	1.0
GeoPackage	0.7	1.1	0.9	0.2	0.6
GML	2.8	1.1	7.7	60	2.0
GeoJSON	2.6	1.2	15	100	2.7
FlatGeoBuf	0.9	1.0	0.5	0.1	0.4

All Formats use the same basic schema. Geometry is 50% of the data volume (100.000 polygon features).

* Source of the performance comparison:@bjornharrtell/@flatgeobuf

Recommendation: Use multiple encodings to best support different applications for your data

Can the Encoding be used?

Software has different support for different features of formats

gml_file_load

QGIS	ArcGIS Online	hale studio	OpenLayers	FME Desktop	Leaflet
2.18.24-Las Palmas					
3.4.4-Madeira	December 2018 update	3.4.1	5.3.0	2018.1	1.4.0

gml_file_display

QGIS	ArcGIS Online	hale studio	OpenLayers	FME Desktop	Leaflet
2.18.24-Las Palmas					
3.4.4-Madeira	December 2018 update	3.4.1	5.3.0	2018.1	1.4.0

gml_WFS2_load

QGIS	ArcGIS Online	hale studio	OpenLayers	FME Desktop	Leaflet
2.18.24-Las Palmas					
3.4.4-Madeira	December 2018 update	3.4.1	5.3.0	2018.1	1.4.0

gml_WFS2_display

QGIS	ArcGIS Online	hale studio	OpenLayers	FME Desktop	Leaflet
2.18.24-Las Palmas					
3.4.4-Madeira	December 2018 update	3.4.1	5.3.0	2018.1	1.4.0

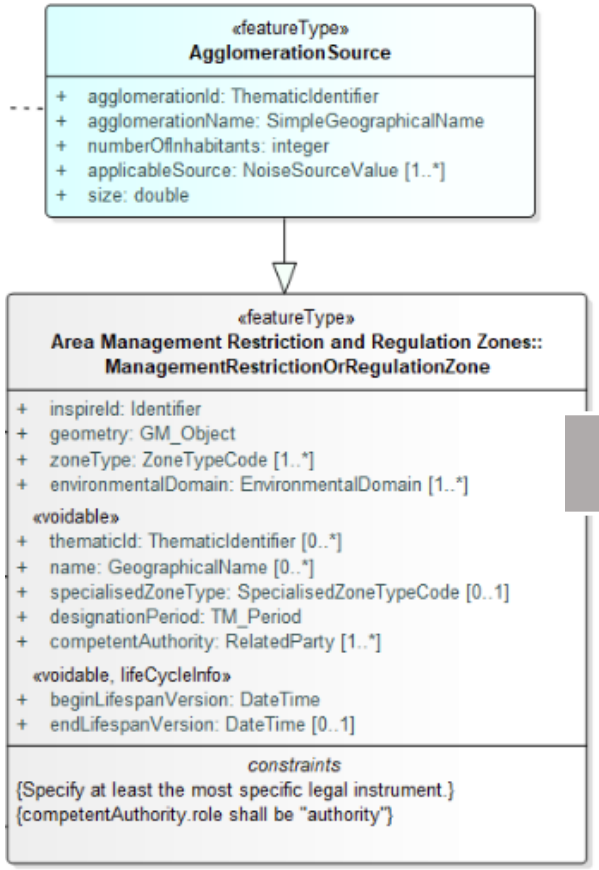
https://github.com/INSPIRE-MIF/caniuse/, https://inspire-mif.github.io/caniuse/generator/out.html

Agenda

1. Introduction
2. Data usability and usefulness
3. An overview of encodings: GML, GeoJSON, GeoPackage
4. **Model Transformations**
5. The END use case (including demo)
6. INSPIRE and Geopackage: Next steps and Timeline, getting involved
7. Summary, Questions & Answers

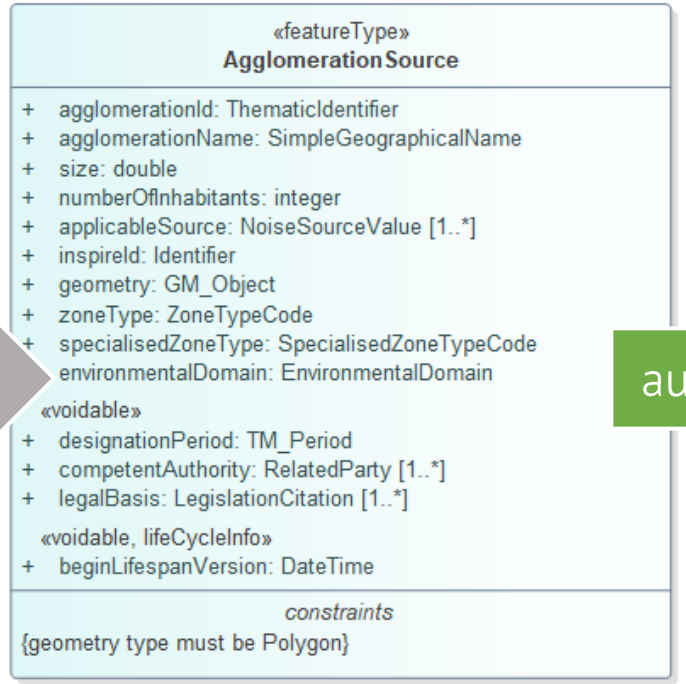
An example simplified Model

For European Noise Directive Reporting



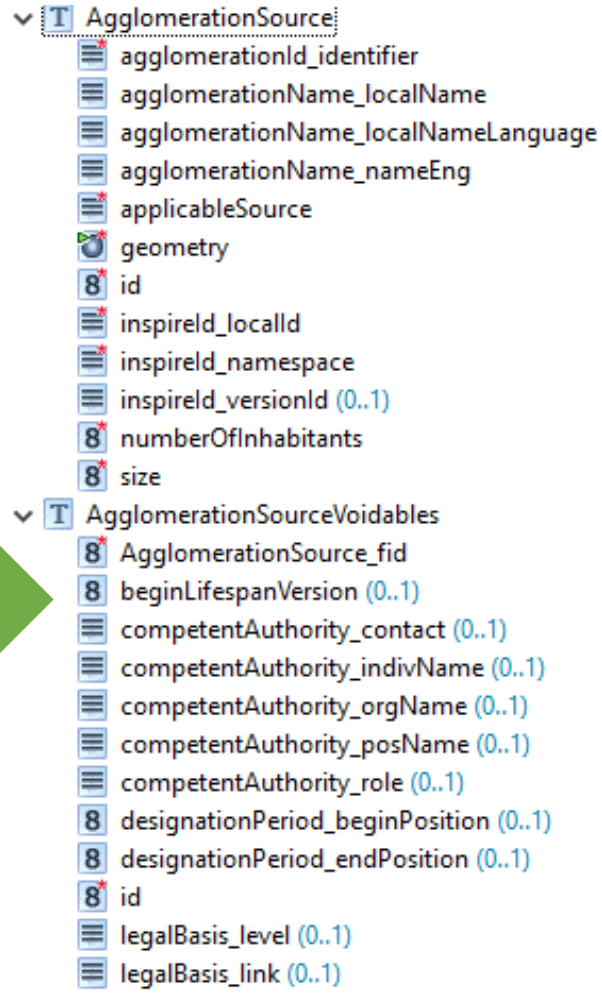
1. Conceptual Model

manual



2. Streamlined Conceptual Model

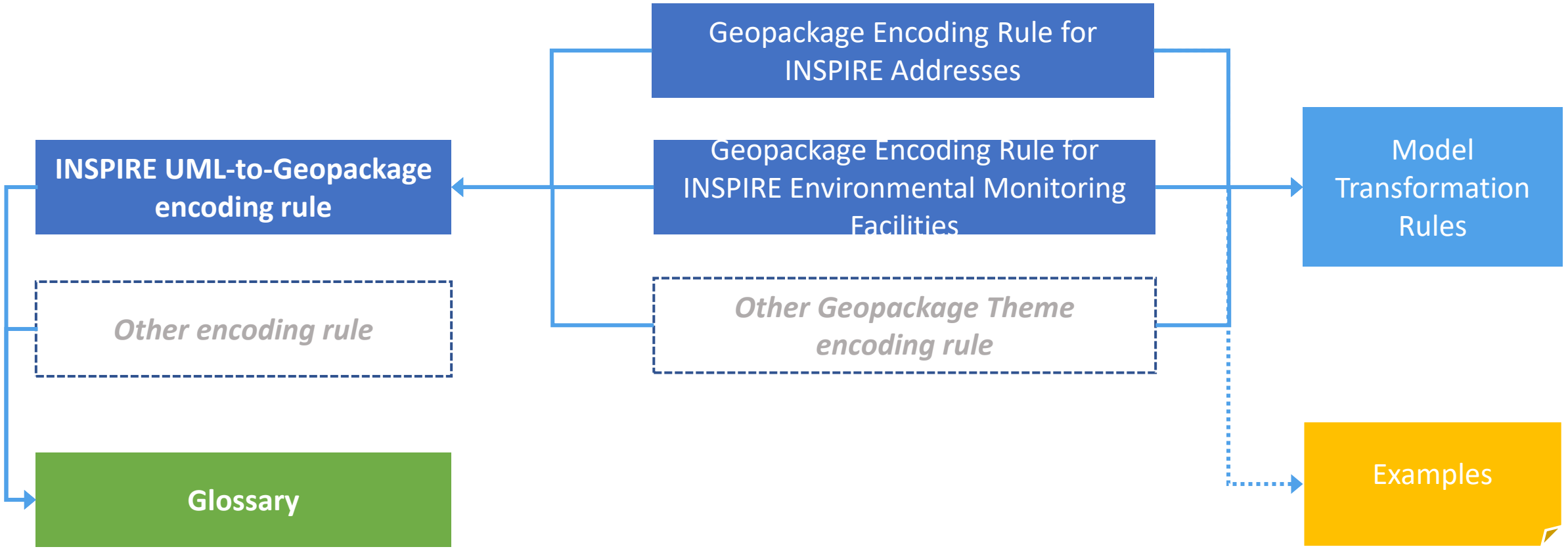
automatic



3. Simplified Logical Model

Encoding Rules and Model Transformation Rules

For European Noise Directive Reporting



Examples of Model Transformation Rules

From GeoJSON & GeoPackage UML-to… Encoding Rules

- Flattening of nested structures
- Extract Primitive Arrays
- Association/Aggregation to Composition with Hard Typing
- Property Composition to Association
- Nillable Properties to Voidables Table
- Common Properties to Dataset Defaults Table

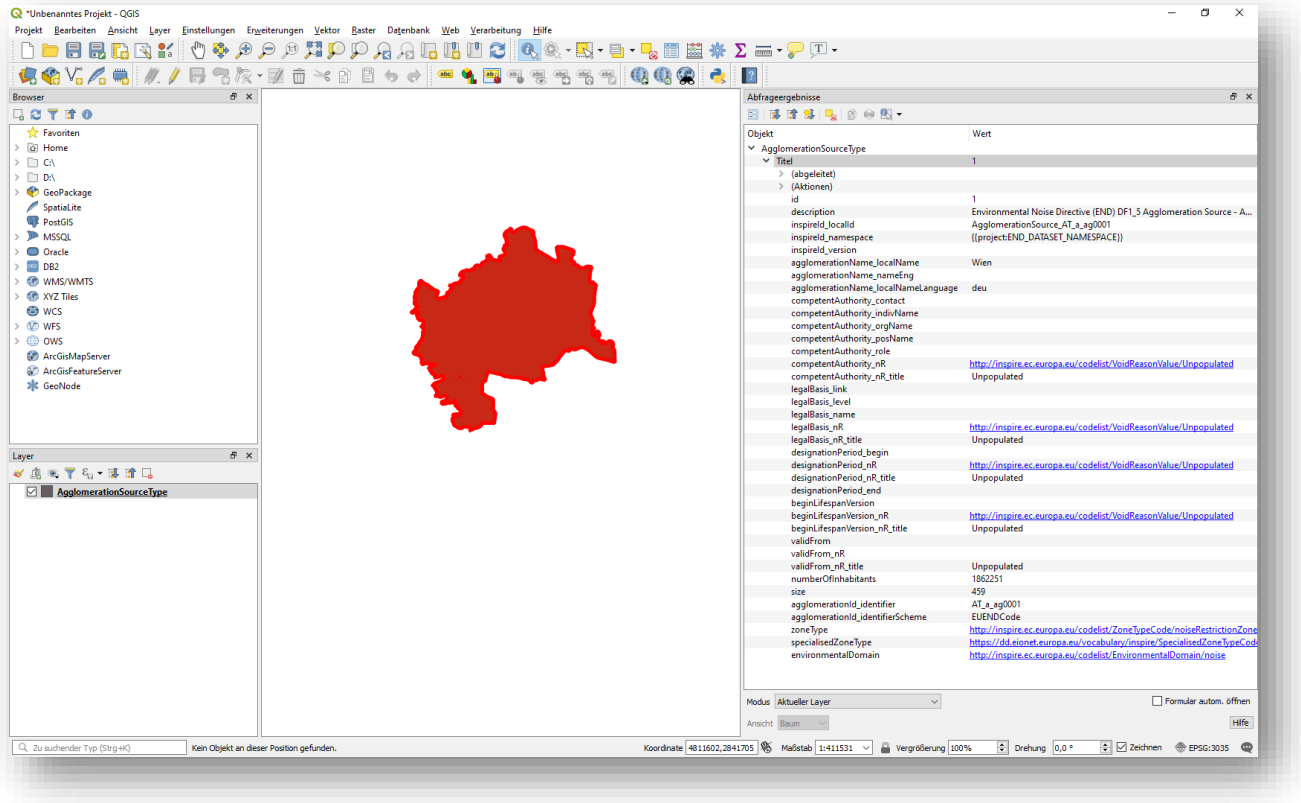
- Simple Geographic Name (Substitution)
- Simple Citation (Substitution)
- Simple Codelist Reference (Substitution)
- Simple Period (Substitution)

Agenda

1. Introduction
2. Data usability and usefulness
3. An overview of encodings: GML, GeoJSON, GeoPackage and others
4. Model Transformations
5. **The END use case (including demo)**
6. INSPIRE and Geopackage: Next steps and Timeline, getting involved
7. Summary, Questions & Answers

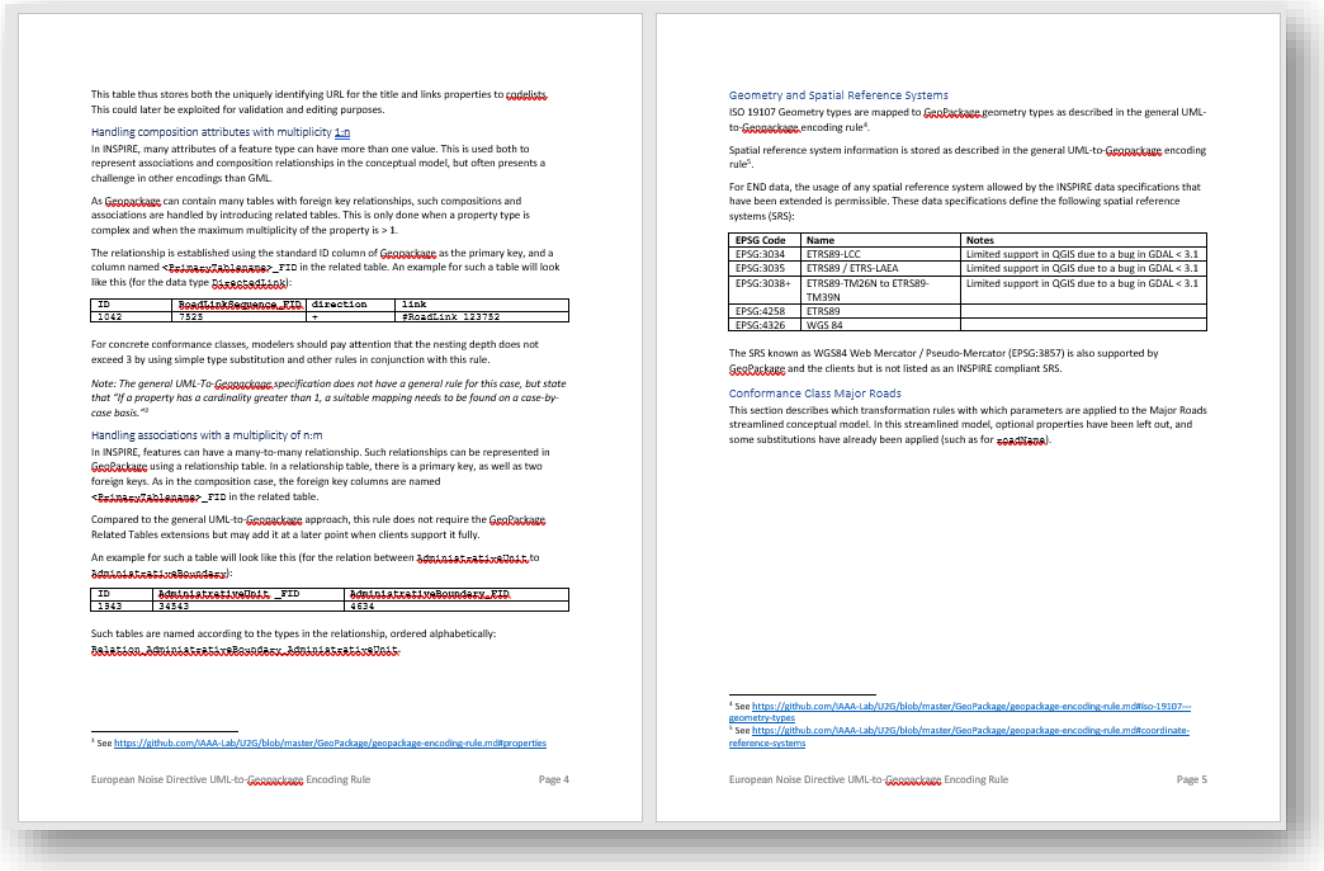
The objective

- Enable Delivery of Noise Directive Data using various encodings
  - INSPIRE-compatible GML
  - **GeoPackage**
- Fulfil INSPIRE IR requirements with both options
- Create GeoPackage templates and samples
- Evaluate Results in Clients



Creating the GeoPackage Schemas

- General Rules for all END topics
- Topic-specific rules
- Alignment to UNIZAR UML-to-GeoPackage encoding Rule for all INSPIRE themes



This table thus stores both the uniquely identifying URL for the title and links properties to codelists. This could later be exploited for validation and editing purposes.

Handling composition attributes with multiplicity 1..n

In INSPIRE, many attributes of a feature type can have more than one value. This is used both to represent associations and composition relationships in the conceptual model, but often presents a challenge in other encodings than GML.

As Geopackage can contain many tables with foreign key relationships, such compositions and associations are handled by introducing related tables. This is only done when a property type is complex and when the maximum multiplicity of the property is > 1.

The relationship is established using the standard ID column of Geopackage as the primary key, and a column named <PrimaryTablename>_FID in the related table. An example for such a table will look like this (for the data type DirectedLink):

ID	RoadLinkSequence_FID	direction	link
1042	7525	+	#RoadLink 123752

For concrete conformance classes, modelers should pay attention that the nesting depth does not exceed 3 by using simple type substitution and other rules in conjunction with this rule.

Note: The general UML-To-Geopackage specification does not have a general rule for this case, but state that "If a property has a cardinality greater than 1, a suitable mapping needs to be found on a case-by-case basis."[3]

Handling associations with a multiplicity of n:m

In INSPIRE, features can have a many-to-many relationship. Such relationships can be represented in GeoPackage using a relationship table. In a relationship table, there is a primary key, as well as two foreign keys. As in the composition case, the foreign key columns are named <PrimaryTablename>_FID in the related table.

Compared to the general UML-to-Geopackage approach, this rule does not require the GeoPackage Related Tables extensions but may add it at a later point when clients support it fully.

An example for such a table will look like this (for the relation between AdministrativeUnit to AdministrativeBoundary):

ID	AdministrativeUnit _FID	AdministrativeBoundary_FID
1343	34543	4834

Such tables are named according to the types in the relationship, ordered alphabetically:

Relation_AdministrativeBoundary_AdministrativeUnit

[3] See https://github.com/IAAA-Lab/U2G/blob/master/GeoPackage/geopackage-encoding-rule.md#properties

European Noise Directive UML-to-Geopackage Encoding Rule Page 4

Geometry and Spatial Reference Systems

ISO 19107 Geometry types are mapped to GeoPackage geometry types as described in the general UML-to-Geopackage encoding rule[4].

Spatial reference system information is stored as described in the general UML-to-Geopackage encoding rule[5].

For END data, the usage of any spatial reference system allowed by the INSPIRE data specifications that have been extended is permissible. These data specifications define the following spatial reference systems (SRS):

EPSG Code	Name	Notes
EPSG:3034	ETRS89-LCC	Limited support in QGIS due to a bug in GDAL < 3.1
EPSG:3035	ETRS89 / ETRS-LAEA	Limited support in QGIS due to a bug in GDAL < 3.1
EPSG:3038+	ETRS89-TM26N to ETRS89-TM39N	Limited support in QGIS due to a bug in GDAL < 3.1
EPSG:4258	ETRS89	
EPSG:4326	WGS 84	

The SRS known as WGS84 Web Mercator / Pseudo-Mercator (EPSG:3857) is also supported by GeoPackage and the clients but is not listed as an INSPIRE compliant SRS.

Conformance Class Major Roads

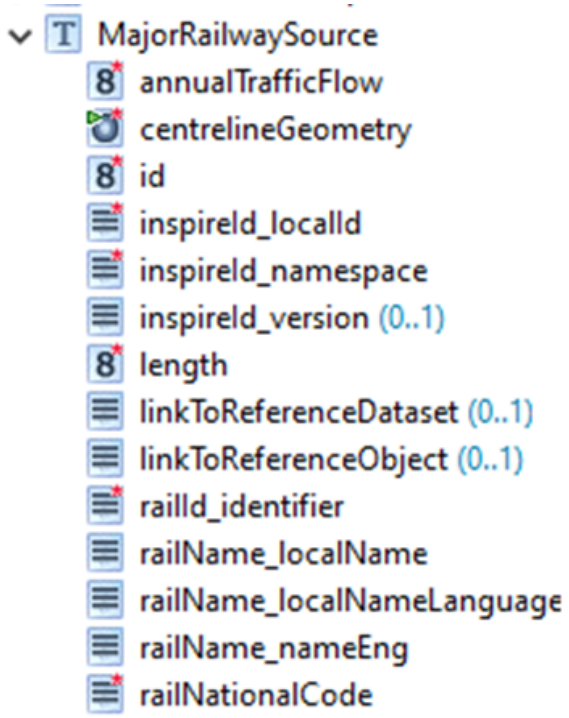
This section describes which transformation rules with which parameters are applied to the Major Roads streamlined conceptual model. In this streamlined model, optional properties have been left out, and some substitutions have already been applied (such as for roadName).

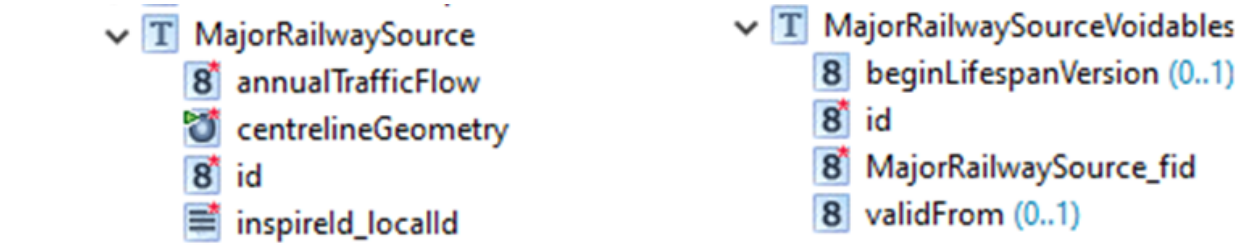
[4] See https://github.com/IAAA-Lab/U2G/blob/master/GeoPackage/geopackage-encoding-rule.md#iso-19107---geometry-types
[5] See https://github.com/IAAA-Lab/U2G/blob/master/GeoPackage/geopackage-encoding-rule.md#coordinate-reference-systems

European Noise Directive UML-to-Geopackage Encoding Rule Page 5

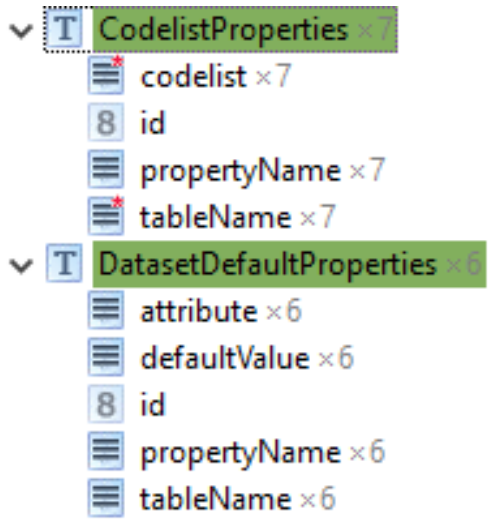
Principle: Keep core tables/types lean

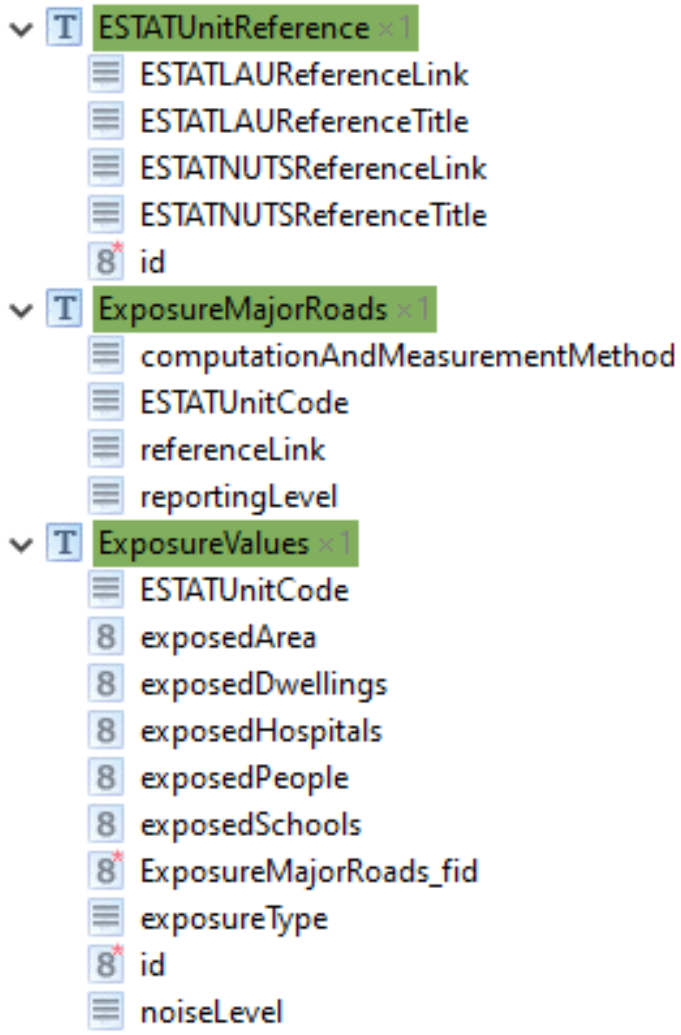
- Only include key properties that have per-feature values in core table
- Move values that are the same for all features to “DatasetDefaultProperties”
- Move properties that are nillable to “Voidables” table

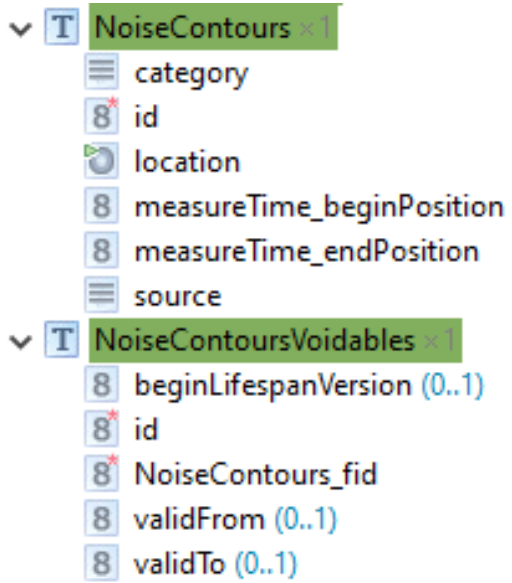




Example: DF4_8 Contours & Exposure Data

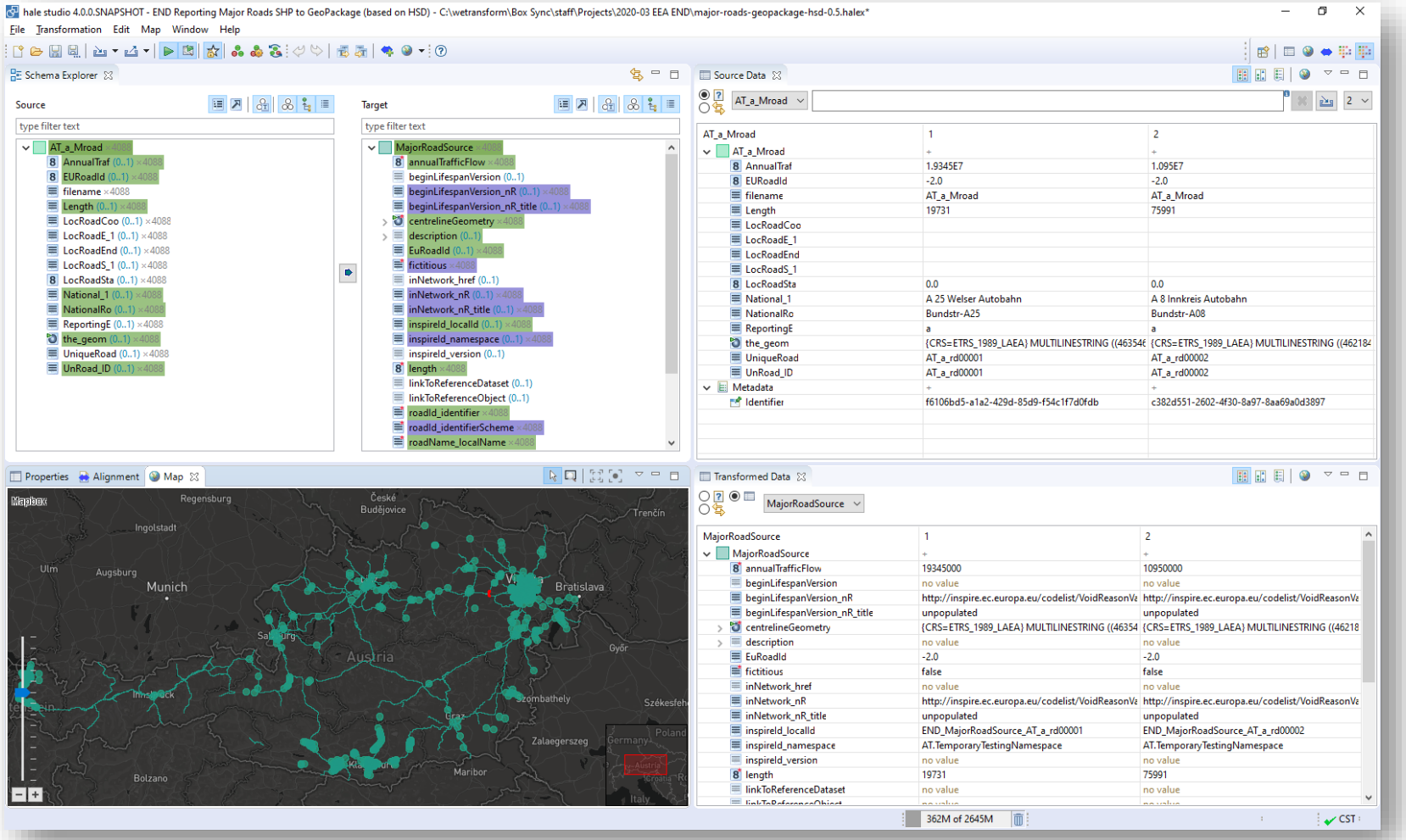






How to create END GeoPackages?

- hale»studio 4.0+
- Sources:
  - Shapefile
  - FileGeodatabase
  - GML
- Target: GeoPackage
- Live Demo



Agenda

1. Introduction
2. Data usability and usefulness
3. An overview of encodings: GML, GeoJSON, GeoPackage and others
4. Model Transformations
5. The END use case (including demo)
6. **INSPIRE and Geopackage: Next steps and Timeline, getting involved**
7. Summary, Questions & Answers

Next steps, getting involved

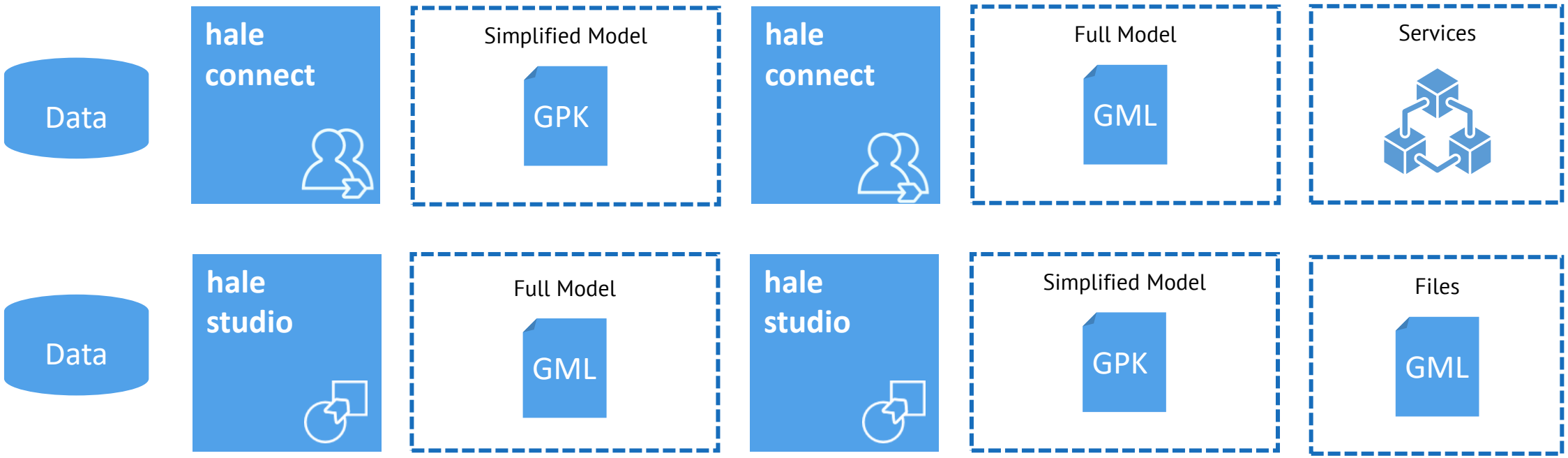
- For each INSPIRE theme, a simplified conceptual model needs to be designed and to go through the good practice process!
- Simplified conceptual models have been drafted for 7 of 29 relevant themes
  - UNIZAR: Addresses, Administrative Units
  - END: Area Management, Transport Networks, Human Health and Safety
- Acceptance of these will not happen before Q3/2021
- Collaboration on other theme models required
- Technical development required

Planned/On-going Work on Tooling

- Workflow 1:
  - Add schema presets and post-processing transformation projects for the INSPIRE schemas to hale studio
  - Advantage: Easy to implement
  - Disadvantage: Lots of effort to configure
- Workflow 2:
  - Add support for configuring and executing individual transformation rules to hale studio
  - Advantages: Applies to all models, no manual work on schemas or transformation required
  - Disadvantage: Requires much more effort to implement

Get multiple encodings without extra work ☺

hale tooling in development



Recommendation: Reduce effort for harmonization by first mapping to a simplified model or local standard (e.g. ALKIS in Germany)

Agenda

1. Introduction
2. Data usability and usefulness
3. An overview of encodings: GML, GeoJSON, GeoPackage and others
4. Model Transformations
5. The END use case (including demo)
6. INSPIRE and Geopackage: Next steps and Timeline, getting involved
7. Summary, Questions & Answers

Summary

- **Usability** and **Usefulness** can be improved through…
  - Improving compatibility and quality of source data
  - Applying best practices
  - Re-using solutions (models, transformations, products)
  - Providing multiple encodings in suitable models, including GeoPackge for GIS users
- To formally use GeoPackage as an INSPIRE implementation path, more technical and procedural work is required.

Webinars and 1:1 Sessions

Format	Description	Date	Price
Webinar	*INSPIRE Solutions for municipal service companies*	April 20	free
Webinar	*GeoPackage: An alternative encoding for INSPIRE*	May 18	free
Webinar	*Kommunale Lösungen für INSPIRE und XPlanung*	June 1	free
1:1 Session	*INSPIRE Monitoring 2021: Identify Compliance Gaps*	April 1-30	free
1:1 Session	*INSPIRE Monitoring 2021: Fix Compliance Gaps*	May 2-31	free

INSPIRE Online Trainings

Format		Description	Audience	Date	Price
15 hours		*Datentransformation nach INSPIRE mit hale»studio*	Beginners	June 14-18	800€
15 hours		*Transforming Data to INSPIRE with hale»studio*	Beginners	July 6–10	800€
8 hours		*Transformation for Environmental Monitoring Facilities & Observations and Measurements*	Advanced	May 25-27	400€
8 hours		*INSPIRE Data Transformation for Geology and Mineral Resources*	Advanced	June 21-23	400€
8 hours		*Mastering complex INSPIRE transformations with Scripts*	Advanced	June 8–10	400€

Any questions?
Reach out to us!

+49 6151 6290 890

info@wetransform.to

www.wetransform.to

www.linkedin.com/company/wetransform-gmbh

https://twitter.com/tr_xsdi