



Transferencia de Datos Estadísticos de Alemania a la Red Europea INSPIRE

Benjamin Quest

Dipl.-Geogr.



Visión General del proyecto

■ Asociados al proyecto

- > IT.NRW (representante de las instituciones estadísticas de los estados federales)
- > con terra GmbH

■ El desarrollo de un prototipo fue el objetivo del proyecto

- > Transferencia de los datos estadísticos de los estados federales de la base de datos GENESIS al modelo de datos INSPIRE del anexo III (SU, PD)

■ Documentación del procedimiento

- > Ajustamiento del procedimiento entre las instituciones estadísticas de los estados federales



Requisitos en los datos

■ Estadística

- > homogeneidad
- > Informaciones sobre la cantidad de los datos
- > Informaciones sobre la calidad de los datos

■ Cartografía y geoinformática

- > Mapa/ diseño
- > Plausibilidad geométrica
- > topología
- > Informaciones atributivas



¿Cómo se pueden utilizar los datos juntamente?

- **Spatial-ETL es la llave para el uso de datos espaciales**



Tecnología ETL

W Extract, transform and load x

es.wikipedia.org/wiki/Extract_transform_and_load

Crear una cuenta Ingresar

Artículo Discusión Leer Editar Ver historial Buscar

Extract, transform and load

Extract, Transform and Load (*Extraer, transformar y cargar* en castellano, frecuentemente abreviado a **ETL**) es el proceso que permite a las organizaciones mover datos desde múltiples fuentes, reformatearlos y limpiarlos, y cargarlos en otra base de datos, data mart, o data warehouse para analizar, o en otro sistema operacional para apoyar un proceso de negocio.

Los procesos ETL también se pueden utilizar para la integración con sistemas heredados.

Co	● ETL describe [...] un proceso para unir datos de múltiples fuentes de datos con estructuras diferentes en un sistema de destino.
1	Extraer
2	Transformar
3	Cargar
4	Procesos
5	Desafíos
6	Algunas Herramientas ETL
6.1	Libres
6.2	Licencia
7	Véase también
8	Enlaces externos

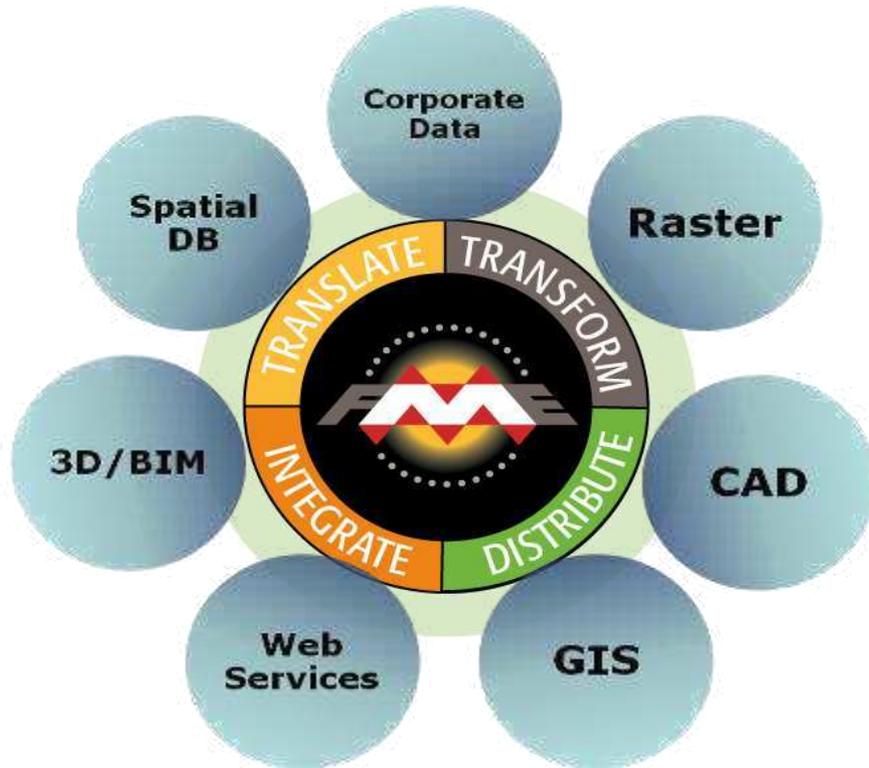
Extraer [editar]

La primera parte del proceso ETL consiste en extraer los datos desde los sistemas de origen. La mayoría de los proyectos de almacenamiento de datos fusionan datos provenientes de diferentes sistemas de origen. Cada sistema separado puede usar una organización diferente de los datos o formatos distintos. Los formatos de las fuentes normalmente se encuentran en bases de datos relacionales o ficheros planos, pero pueden incluir bases de datos no relacionales u otras estructuras diferentes. La extracción convierte los datos a un formato preparado para iniciar el proceso de transformación.

Una parte intrínseca del proceso de extracción es la de analizar los datos extraídos, de lo que resulta un chequeo que verifica si los

Spatial ETL – **E**xtract **T**ransform **L**oad

„Get the right data to the right system, in the right schema, at the right time“



Translate

Conversión de datos: > 275 formateos y fuentes de datos

Transform

Manipulación/ reestructuración/ modelación de Datos

Integrate

Unión y almacenamiento de datos y informaciones en sistemas de aplicación

Distribute

Distribución de datos en formatos y modelos específicos



La unión de informaciones con Spatial ETL

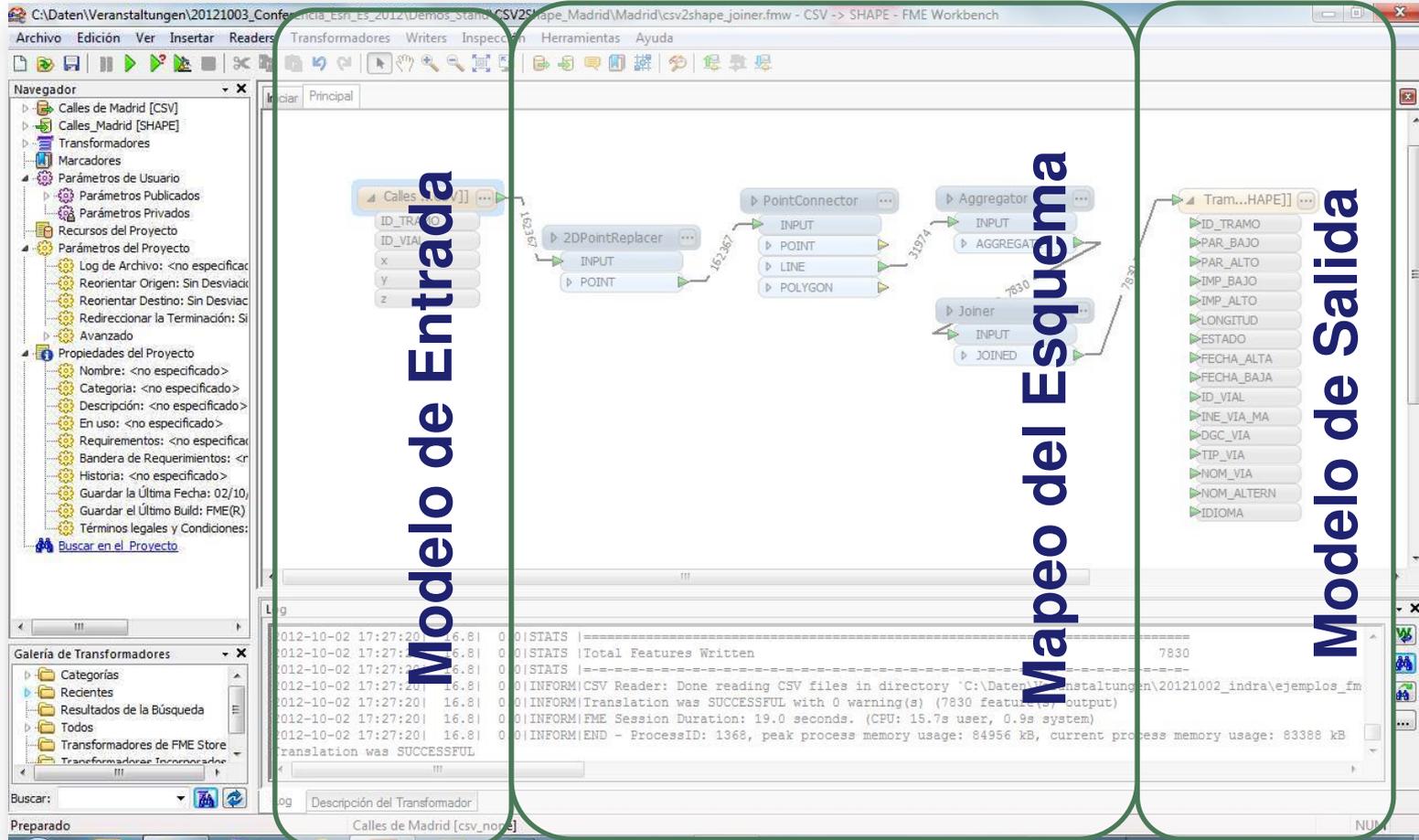
■ Extracción de datos estadísticos desde datos espaciales

- > Información atributiva y referencia espacial
- > Transferencia en sistemas típicos de almacenamiento de datos
 - > Excel, Access, Oracle, MS SharePoint, Informatica ...

■ Transferencia de datos estadísticos en IDEs

- > INSPIRE (Statistical Units, Population Distribution)
- > Transferencia en sistemas típicos de almacenamiento de datos para SIG
 - > ArcSDE, Geodatabase, ficheros Shape, ...

Definición de Procesos con FME Workbench



Modelo de Entrada

Mapeo del Esquema

Modelo de Salida



INSPIRE – Objetivos

Infraestructure for **S**Patial **I**nfo**R**mation in **E**urope

- **Interoperabilidad transfronteriza para acceso a datos espaciales de instituciones publicas**
 - > Interoperabilidad técnico y de contenido
- **Simplificación de reportes obligatorios para la Comisión UE**
 - > Directiva Marco del Agua (2000/60/CE)
- **Activación del potencial de valores añadidos de los datos espaciales existentes**



INSPIRE – Medidas

- Incorporación en IDEs existentes en los países miembros
- Gestión descentralizada
- Aplicación de normas existentes (OGC & ISO)
- Armonización de
 - > Metadatos, datos y servicios espaciales
 - > Condiciones de licencia (y modelos de costos)
- **Directiva europea obligatoria**



INSPIRE – orientación temática

Annex I (since 2012)

- Coordinate reference systems
- Geographical grid systems
- Geographical names
- Administrative units
- Addresses
- Cadastral parcels
- Transport networks
- Hydrography
- Protected sites

Annex II (2014)

- Elevation
- Land cover
- Ortho-imagery
- Geology

Annex III (2014)

- Statistical units
- Buildings
- Soil
- Land use
- Human health and safety
- Utility and governmental services
- Environmental monitoring facilities
- Production and industrial facilities
- Agricultural and aquaculture facilities
- Population distribution – demography
- Area management/restriction/regulation zones & reporting units
- Natural risk zones
- Atmospheric conditions
- Meteorological geographical features
- Oceanographic geographical features
- Sea regions
- Bio-geographical regions
- Habitats and biotopes
- Species distribution
- Energy Resources
- Mineral resources



Desafíos

■ Modelo complejo de datos de origen

- > Conocimientos expertos en el IT.NRW
- > Conocimientos sobre estructuras de datos y particularidades del modelo por con terra GmbH

■ Modelo complejo de datos de destino(SU, PD)

- > Estatus beta de los productos (ISP, Afl)
- > Especificaciones no finalizadas (Versión Draft, SU, PD)



Datos de origen

● **Gemeinsames Neues Statistisches Informations-System (GENESIS)**

> Nuevo Sistema común de información estadística

● **Formatos/ protocolos disponibles**

> Servicio SOAP

> Servicio web XML

> Ficheros CSV (preprocesados)

Date	Code	Content	new
18.05.2012	51900	Außenhandel (Revision: August 2011 - Januar 2012)	Mär 2012
16.05.2012	92211	Verfallfuge Tarifindex	Feb 2012
16.05.2012	61900	VGR des Bundes (Erwerbstätigkeit)	1.12.2012
15.05.2012	45213	Monatsstatistik im Gastgewerbe	Mär 2012
14.05.2012	31111	Statistik der Baugenehmigungen	Feb 2012
14.05.2012	41301	Stichtags- und Stichtagsgewerbesteuer	Mär 2012
14.05.2012	40111	Monatsbericht im verarbeitenden Gewerbe	Mär 2012
14.05.2012	45211	Monatsstatistik Großhandel und Handelsvermittlung	Mär 2012
14.05.2012	45212	Monatsstatistik im Einzelhandel und im Mehrhandels	Mär 2012
14.05.2012	51261	Index der Großhandelsverkaufspreise	Apr 2012
more...			

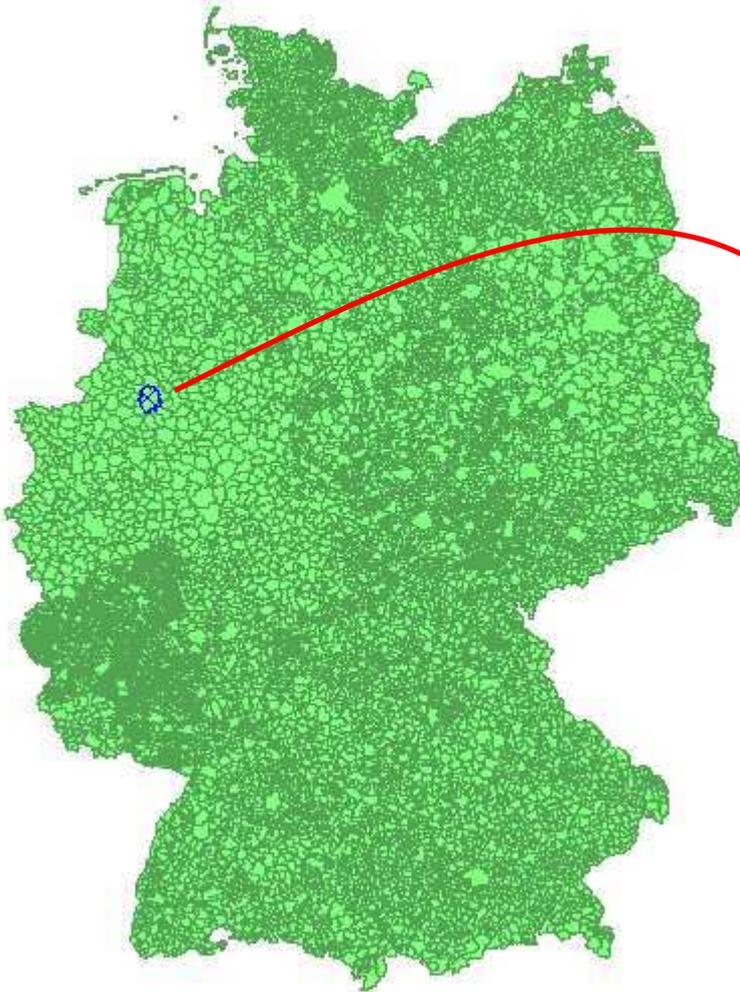
Code	Content
61111-0004	Verbraucherpreisindex
61121-0004	Harmonisierter Verbraucherpreisindex
61111-0002	Verbraucherpreisindex
51000-0018	Aus- und Einfuhr
61241-0002	Erzeugerpreisindex gewerblicher Produkte

Datos de origen – exporte de la BD GENESIS

PopulationbyAge-alles.csv

	A	B	C	D	E	F	G
1	GENESIS-Tabelle: Inspire-03						
2	populationAtResidencePlace by age						
3	demography						
4	populationAtResidencePlace (Anzahl)						
5				age	age	age	age
6				unter 3 Jahre	3 bis unter 6 Jahre	6 bis unter 10 Jahre	10 bis unter 15 Jahre
7	31.12.2008	1001000	Flensburg, krsfr. Stadt	2269	2140	2948	
8	31.12.2008	1002000	Kiel, krsfr. Stadt	6143	5557	7372	
9	31.12.2008	1003000	Lübeck, krsfr. Stadt	5375	5157	7089	
10	31.12.2008	1004000	Neumünster, krsfr. Stadt	2096	2015	3015	
11	31.12.2008	1051001	Albersdorf	88	100	167	
12	31.12.2008	1051002	Arkebek	3	6	10	
13	31.12.2008	1051003	Averlak	6	10	30	
14	31.12.2008	1051004	Bargenstedt	14	24	29	
15	31.12.2008	1051005	Barkenholm	4	3	7	
16	31.12.2008	1051006	Barlt	18	20	43	
17	31.12.2008	1051008	Bergewöhrden	0	1	4	
18	31.12.2008	1051010	Brickeln	10	12	14	
19	31.12.2008	1051011	Brunsbüttel	295	335	470	
20	31.12.2008	1051012	Buchholz	36	49	58	

Datos de origen IT.NRW – datos municipios

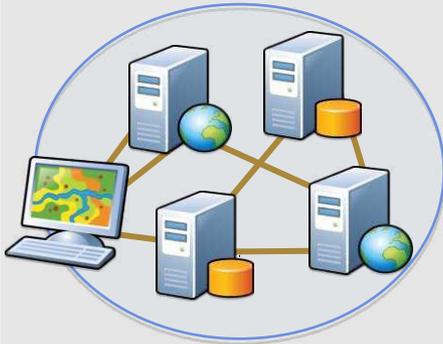


- Geometrías Shape
- ca. 12.000 objetos
- Llave municipal (ID)

Attribute Name	Attribute Value
fme_geometry	fme_polygon
fme_type	fme_area
Gemein	05515000
GEN	Münster
Shape_Area	303280921.156
SHAPE_GEOMETRY	shape_polygon

INSPIRE – Tareas nucleas y flujo de trabajo

INSPIRE Solution Pack for FME



Sistemas internos de información y almacenamiento de datos



Modelos de datos y servicios INSPIRE



ArcGIS for INSPIRE

Consumidores INSPIRE



Situación inicial de la transformación de esquema INSPIRE

- Conocimientos profundos de INSPIRE y de dominio requerido
- Alta complejidad de los datos de fuente y el modelo de destino INSPIRE
- Particularidades locales (calidad, historia, contenido) deben ser consideradas



Proceso ETL soportado por el INSPIRE Solution Pack for FME

■ Plantillas predefinidas de INSPIRE para todos los temas del Anexo I

The screenshot displays the FME Workbench interface for a workspace titled "Administrative Units - FME Workbench". The interface includes a menu bar (File, Edit, View, Insert, Readers, Transformers, Writers, Inspection, Tools, Help), a toolbar, and several panels:

- Navigator:** Shows the workspace structure, including "Transformers", "Parameters", and "Workspace Properties".
- Transformer Gallery:** Lists various INSPIRE templates, such as "INSPIRE Administrative Units" (with sub-templates like "auAdmBoundary_country", "auAdmBoundary_legalStatus", etc.) and "INSPIRE Cadastral Parcels".
- Main Canvas:** Displays a complex ETL workflow with multiple "INSPIRE_AdmUnits" transformers connected in a grid-like structure. The workflow is divided into three main sections: "Source Data", "Work Tables", and "Destination Data".
- Transformer Description:** A panel at the bottom provides details for the "INSPIRE Administrative Units Category", including a description: "Contains transformers for the INSPIRE Annex I Theme 'Administrative Units'." and the INSPIRE logo.

Proceso ETL soportado por el INSPIRE Solution Pack for FME

■ Conexión a datos individuales de origen con importancia INSPIRE

The screenshot displays the FME Workbench interface for a workspace named 'Administrative Units'. The main workspace is divided into two sections:

- Mapeo Individual (Red):** A complex network of transformers connected by lines, representing a custom ETL process for individual data mapping.
- Mapeo predefinido (Green):** A structured flow of transformers, representing a predefined mapping process.

The **Transformer Gallery** on the left lists various INSPIRE Administrative Units transformers, including:

- auAdmBoundary_country
- auAdmBoundary_legalStatus
- auAdmBoundary_nationalLevel_nat
- auAdmBoundary_technicalStatus
- auAdmUnit_country
- auAdmUnit_nationalLevel

The **Transformer Description** window at the bottom shows the **INSPIRE Administrative Units Category** with the following description:

INSPIRE Administrative Units Category
Description
Contains transformers for the INSPIRE Annex I Theme "Administrative Units".

The **INSPIRE Infrastructure for Spatial Information in Europe** logo is visible in the bottom right corner of the Transformer Description window.

Population Distribution

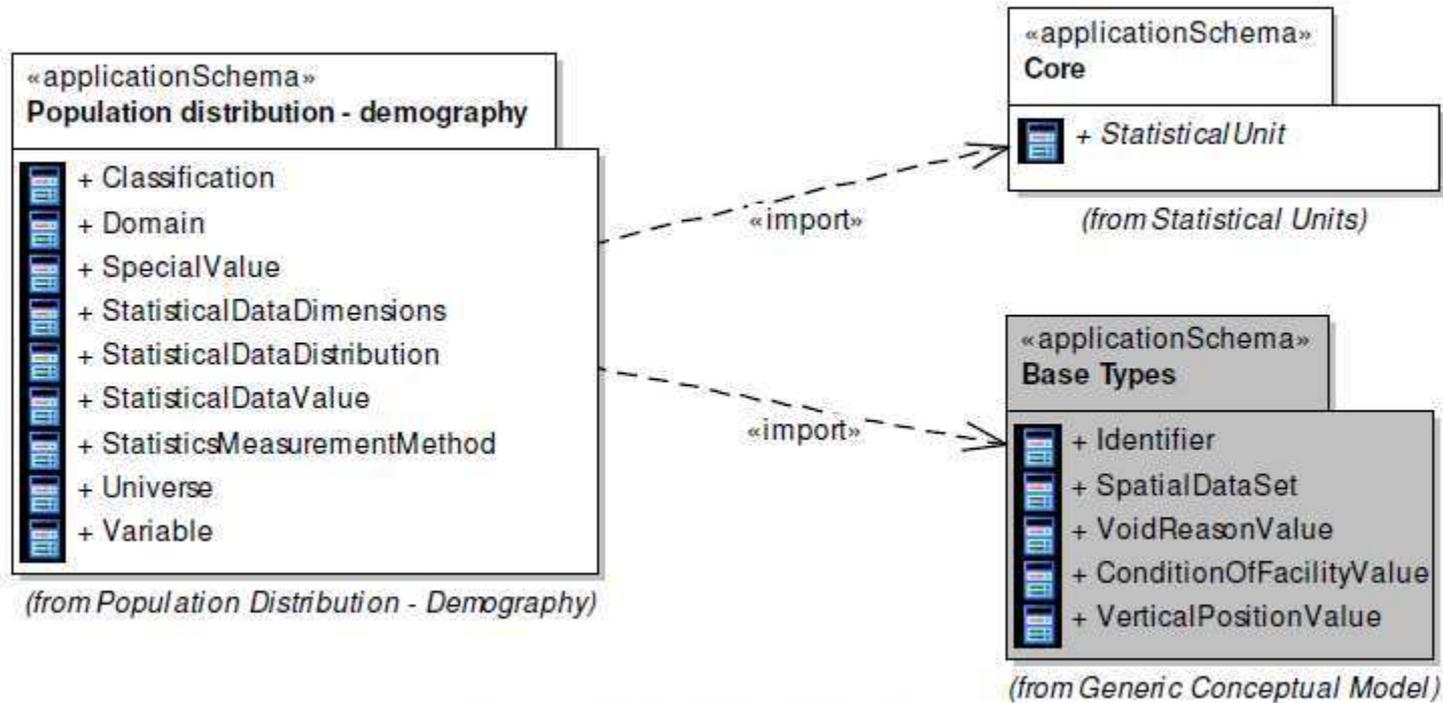
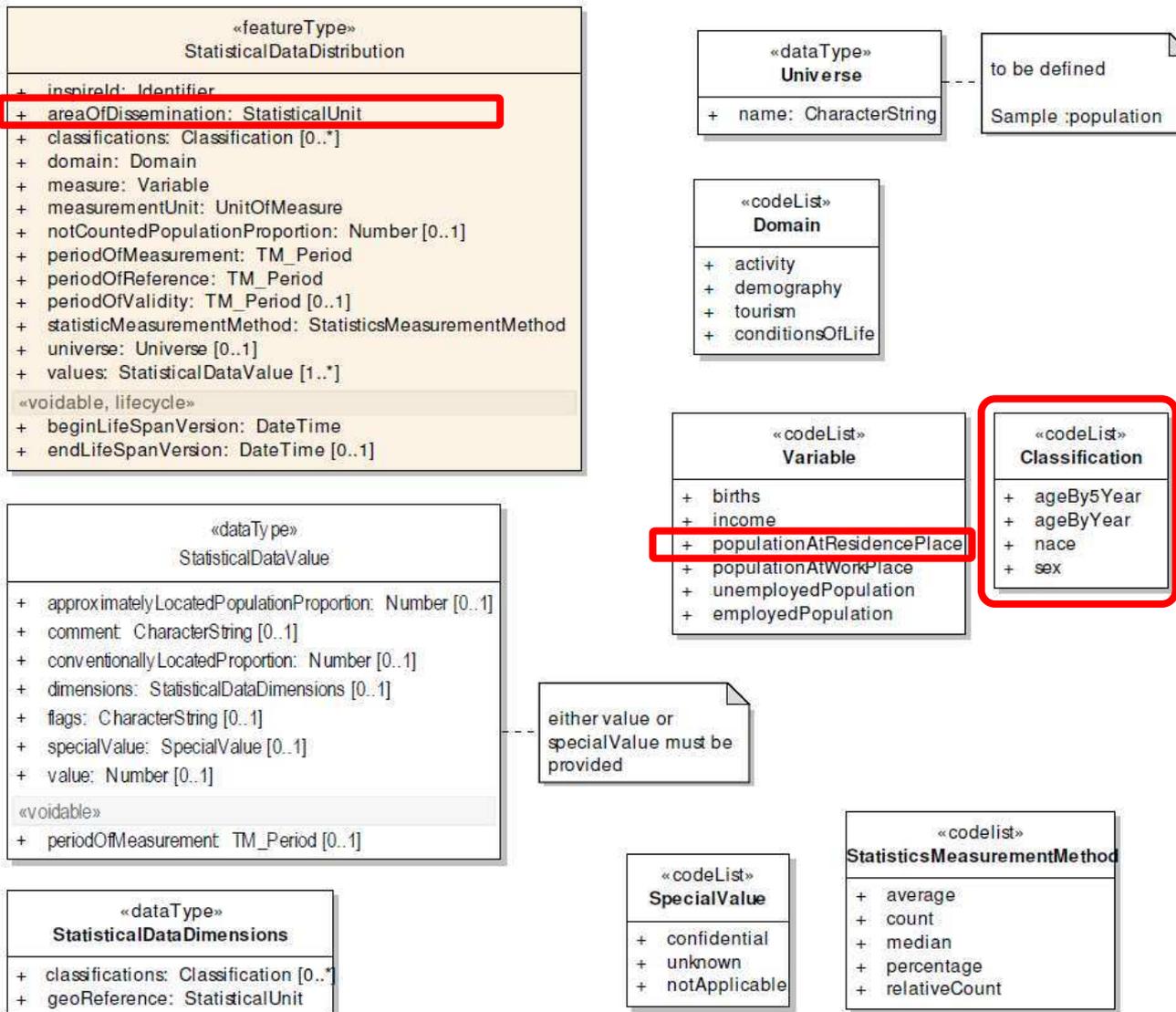


Figure 1: Package Overview

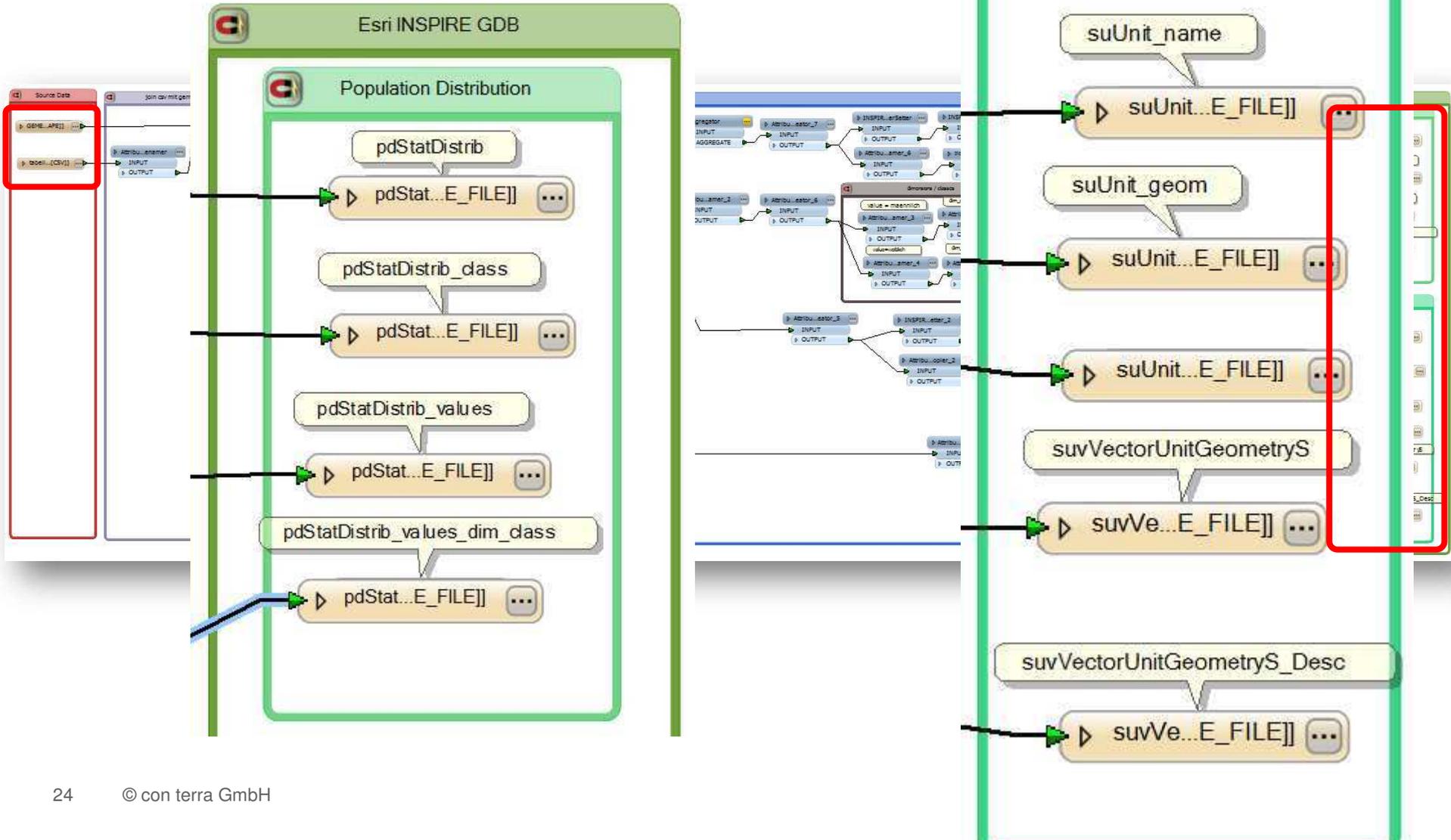
INSPIRE Population Distribution (class diagramm)



Descripción semántica del mapeo

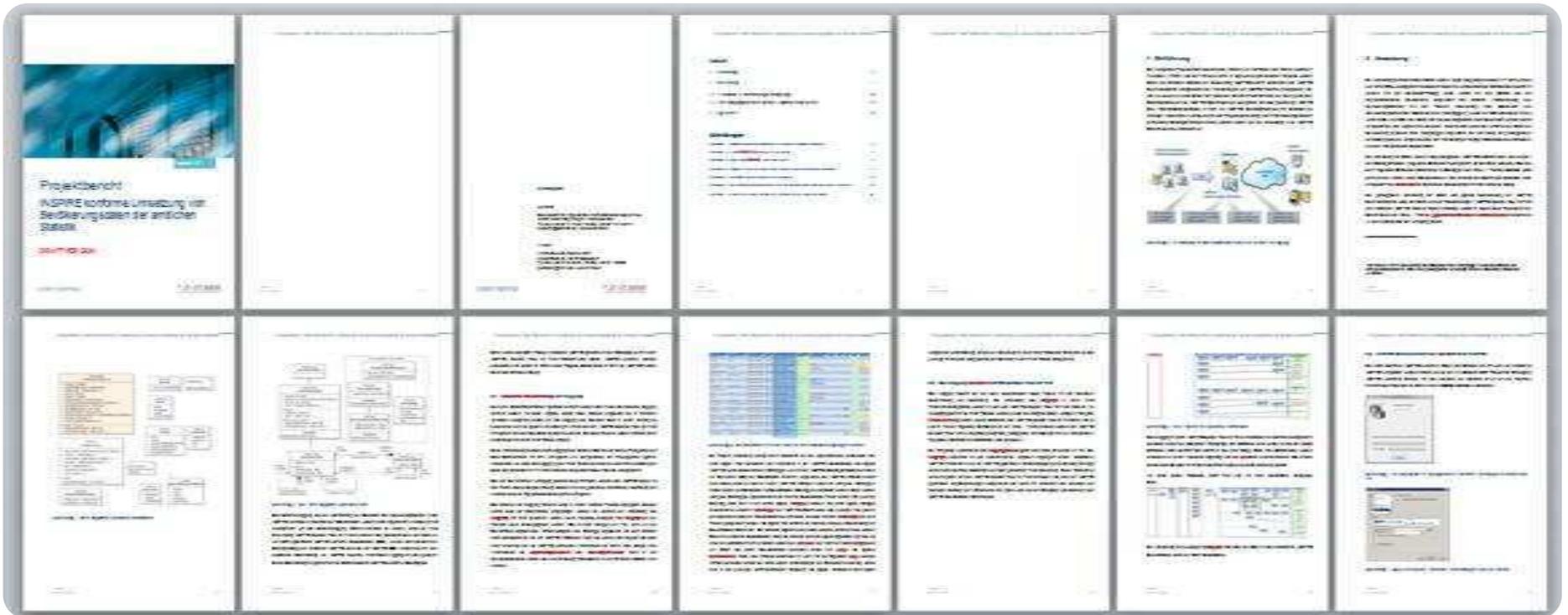
Field	INSPIREName	Subtype	Datatype	Text	Operation	Value	SourceData	Table	Attribute	Workflow	Kommentar ct
id_namespace	namespace	pd:StatisticalDataDistribution	String(255)	Namespace uniquely identifying the data source of the spatial object.	set value	DE.NRW.ITNRW.PD.<Jahr>					
id_versionId	versionId	pd:StatisticalDataDistribution	String(255)	The identifier of the particular version of the spatial object, with a maximum length	empty						
id_versionId_void	-	pd:StatisticalDataDistribution	SmallInteger(2)	This field indicates whether the property in field id_versionId is NIL. If the field	set value	unpopulated					Der Datensatz ist nicht im
areaOfInterest	areaOfInterest	pd:StatisticalDataDistribution	Integer(4)	The part of the 2D world the StatisticalDataDistribution describes.	create	umrisspolygon des untersuchungsraumes evtl.					Zu klären ob es sich um
beginLifeSpanVersion_void	-	pd:StatisticalDataDistribution	SmallInteger(2)	This field indicates whether the property in field beginLifeSpanVersion is NIL. If the	empty						
beginLifeSpanVersion	beginLifeSpanVersion	pd:StatisticalDataDistribution	Date(8)	Date and time at which this version of the spatial object was inserted or changed in	empty						
domain	domain	pd:StatisticalDataDistribution	String(255)	The part of statistical knowledge the data refers to.	map to code list value	demography (from domain codelist)					
domain_cl	-	pd:StatisticalDataDistribution	String(255)	This field contains the absolute URI of the code list dictionary that describes that	set value	http://services.interactive-instruments.de/download/cl-					
endLifeSpanVersion	endLifeSpanVersion	pd:StatisticalDataDistribution	Date(8)	Date and time at which this version of the spatial object was superseded or retired	empty						
endLifeSpanVersion_void	-	pd:StatisticalDataDistribution	SmallInteger(2)	This field indicates whether the property in field endLifeSpanVersion is NIL. If the	empty						
measure	measure	pd:StatisticalDataDistribution	String(255)	What the distribution measures. - EXAMPLE: Population by age and sex.	set value	populationAtResidencePlace					
measure_cl	-	pd:StatisticalDataDistribution	String(255)	This field contains the absolute URI of the code list dictionary that describes that	set value	http://services.interactive-instruments.de/download/cl-					
measurementUnit	measurementUnit	pd:StatisticalDataDistribution	String(255)	The unit of the measurement. - The value of this field must be an unit	set value	person					
notCountedPopProp	notCountedPopulationProportion	pd:StatisticalDataDistribution	Double(8)	The proportion of population of the area of interest that is not counted in any of its	empty						Minimale Multiplizität: 0
periodOfRef	periodOfReference	pd:StatisticalDataDistribution	String(4096)	The time period to which the data is supposed to give a picture of the area of	copy from source				Stichtag / create		create TM_Period
periodOfValidity	periodOfValidity	pd:StatisticalDataDistribution	String(4096)	The time period during which the data is usable. Outside this period the data	empty						
id_localId	localId	pd:StatisticalDataDistribution	String(255)	A local identifier, assigned by the data provider. The local identifier is unique	set value	Datensatz name	main	gemeindec	hlüssel		gemeindec
statisticMeasurementMethod	statisticMeasurementMethod	pd:StatisticalDataDistribution	String(255)	The description of the statistic measurement method.	map to code list value	count					el macht hier
statisticMeasurementMethod_cl	-	pd:StatisticalDataDistribution	String(255)	This field contains the absolute URI of the code list dictionary that describes that	set value	http://services.interactive-instruments.de/download/cl-					
universe_name	name	pd:StatisticalDataDistribution	String(255)	The name of the universe. - In the GML encoding, this field is	empty						universe entspricht name ?
IFCID	(IFC only)	(all)	Integer(2)	IFC identifier, used in references to the object/feature.	INSPIRE Solution Pack						
periodOfMeasurement	periodOfMeasurement	pd:StatisticalDataDistribution	String(4096)	The time period when the data was collected.	copy from source				Stichtag / create		create TM_Period

Proceso de mapeo (GENESIS to INSPIRE)



Resultados del proyecto

- Documentación del proyecto
- Servicio prototípico INSPIRE Download





Conclusiones

- **Conocimientos profundos de los datos de origen y de destino son necesarios**
 - > Conocimientos técnicos como del tema
 - > El modelo de destino es complejo!!!
- **Comunicación es necesario**
 - > Ajustamiento estrecho entra departamentos de SIG y estadístico
- **Modelos flexibles de datos de origen dificultan el mapeo**
 - > Spatial ETL para la armonización de los datos de origen
- **La realización del mapeo del esquema es posible entre 5 y 10 días**



Perspectivas

■ **Servicios y datos INSPIRE estarán disponibles para toda la UE**

- > Servicios View y Download

■ **Ampliación de la base de datos**

- > Dimensión espacial

- > Cantidad absoluta de valores de medición y observaciones

■ **Aprovechar de las informaciones**

- > Unión con tecnología ETL (INSPIRE reader)

- > Visualización a través de clientes web (clientes web, p.e. ArcGIS Online)



!Muchas Gracias!

Benjamin Quest

Venta técnica

Spatial ETL / FME

con terra GmbH

Teléfono +49 251 74745 2238

b.quest@conterra.de

<http://www.conterra.de/en/products/fme-es/>





con terra –
Gesellschaft für Angewandte
Informationstechnologie

Martin-Luther-King-Weg 24
48155 Münster
Telefon +49 251 747 45 0
Telefax +49 251 747 45 2111

