

TRANSFORMACIÓN Y EXPLOTACIÓN DE **DATOS HIDROGRÁFICOS** DEL **IGN** CON HERRAMIENTAS **ETL**

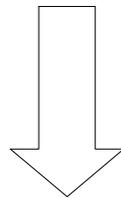
Rafael Sierra Requena

Instituto Geográfico Nacional (IGN)

1. INTRODUCCIÓN

1.1. Objetivo

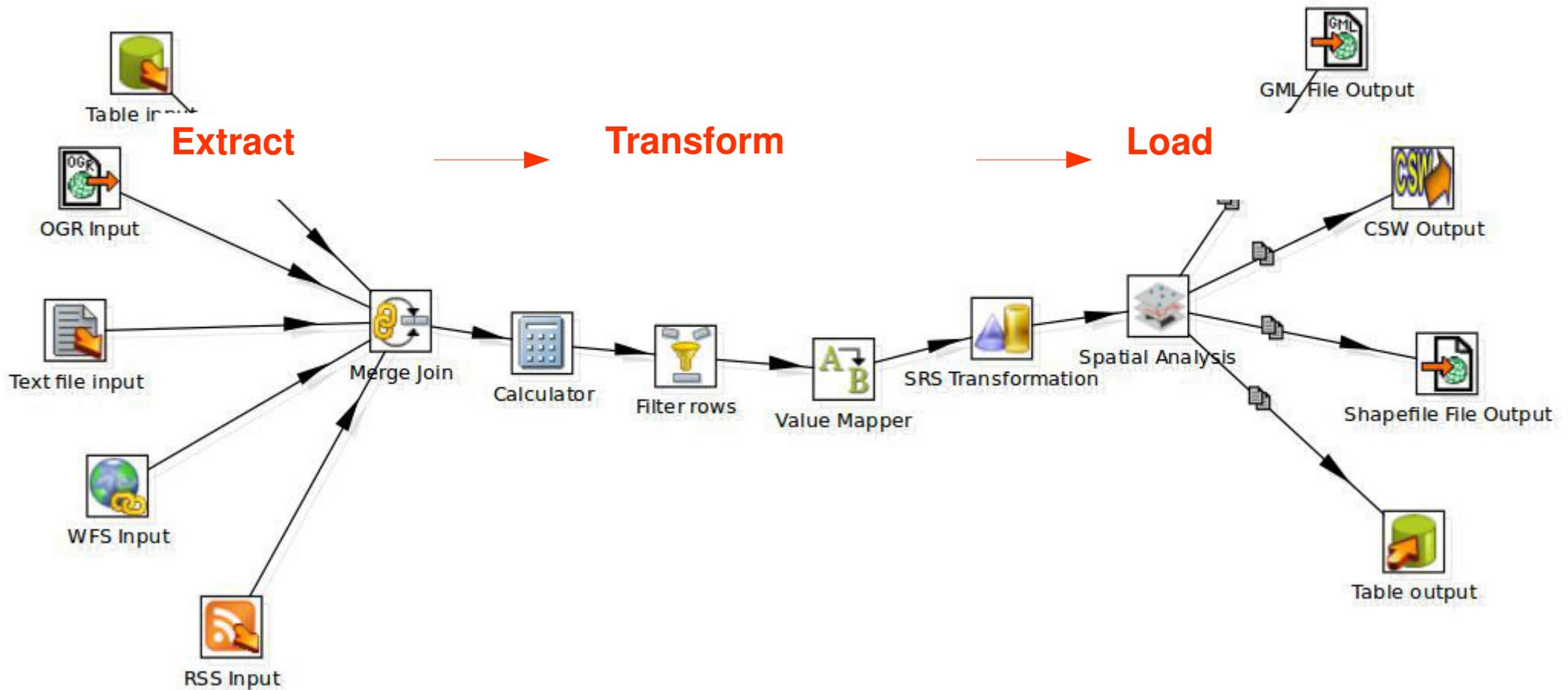
- **Adaptación información geográfica IGN** a nuevos escenarios en tecnológicos (INSPIRE)
- Uso **herramientas Extract-Transform-Load (ETL)** para automatizar procesos de transformación de datos y análisis espacial



Aplicación: **DATOS HIDROGRÁFICOS**

1. INTRODUCCIÓN

1.2. Uso de herramientas ETL opensource

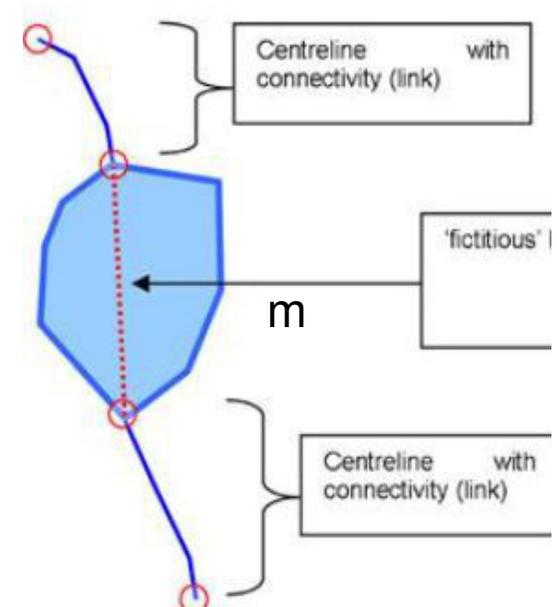
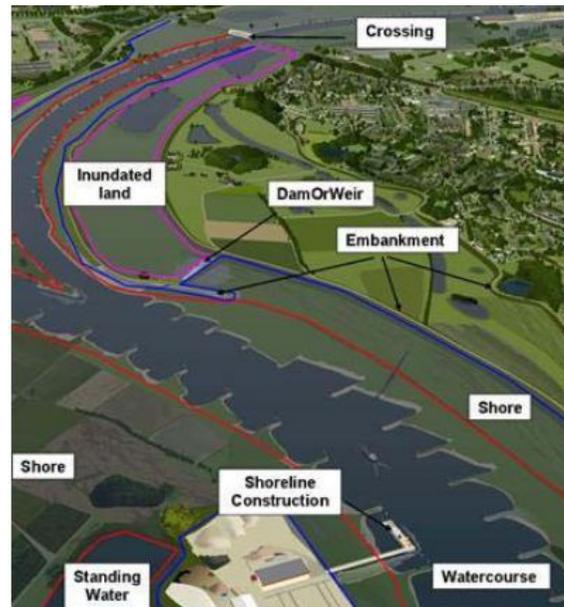


1. INTRODUCCIÓN

1.3. Ejemplo transformaciones de datos: Hidrografía

Necesidades cambio de modelo de datos hidrografía:

- Mejorar la **adaptación** de inf. geográfica existente **al esquema hidrografía INSPIRE**, con modelos de datos unificados: **representación y análisis de red.**



2. MODELO DE DATOS HIDROGRÁFICOS IGN

2. MODELO DE DATOS HIDROGRÁFICOS IGN

2.1. Base Topográfica Nacional (BTN25)



La **Base Topográfica Nacional a escala 1:25.000 (BTN25)** es un conjunto de datos vectoriales tridimensionales de carácter topográfico, que contiene múltiples datos hidrográficos.

Estos datos son **capturados** mediante *restitución fotogramétrica* o **digitalización sobre ortofotos**, junto con datos de *otras fuentes organismos oficiales* (estatales, CCAA, diputaciones, municipios, etc)

Usos:

- **Mapa Topográfico Nacional 1:25.000 (MTN25)**
- **Sistemas de información Geográfica (SIG)**

2. MODELO DE DATOS HIDROGRÁFICOS IGN

2.1. Base Topográfica Nacional (BTN25)



La **Base Topográfica Nacional a escala 1:25.000 (BTN25)** es un conjunto de datos vectoriales tridimensionales de carácter topográfico, que contiene múltiples datos hidrográficos.

Estos datos son **capturados** mediante *restitución fotogramétrica* o **digitalización sobre ortofotos**, junto con datos de *otras fuentes organismos oficiales* (estatales, CCAA, diputaciones, municipios, etc)

Usos:

- **Mapa Topográfico Nacional 1:25.000 (MTN25)**
- **Sistemas de información Geográfica (SIG)**

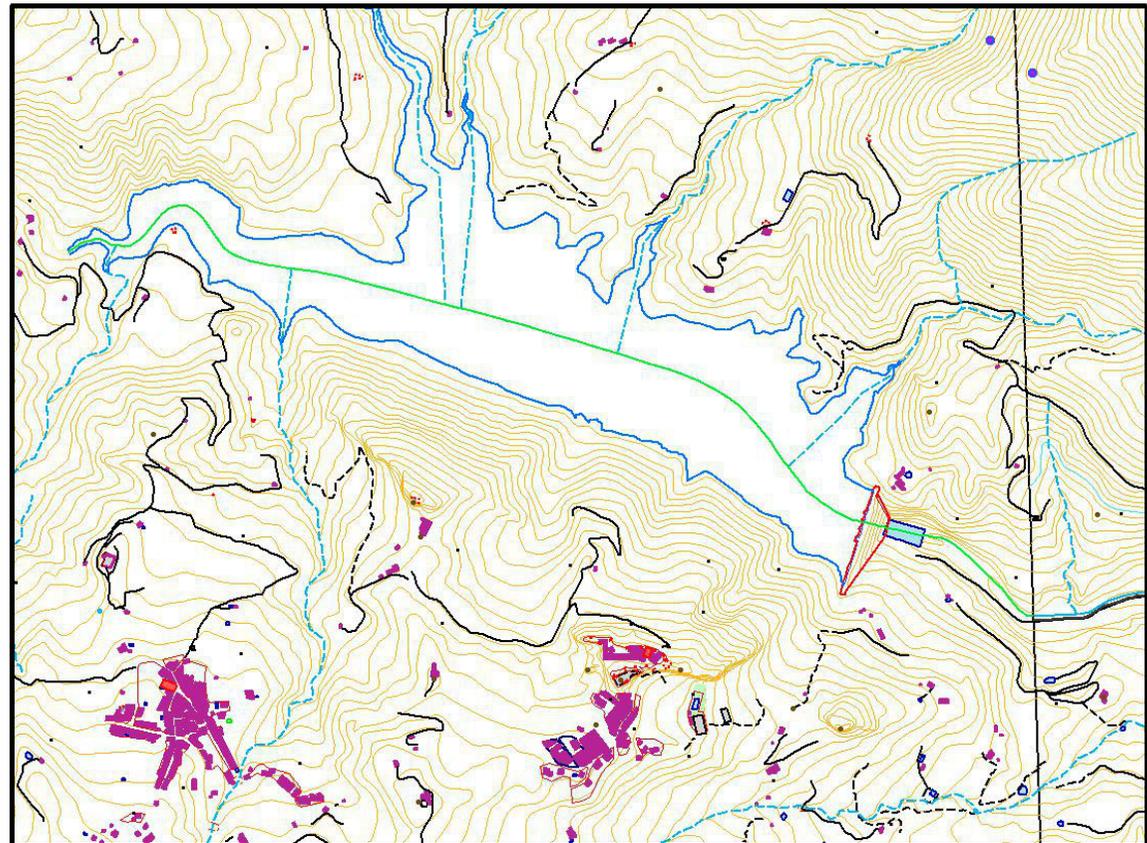
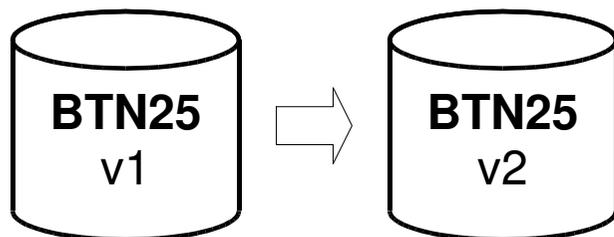
2. MODELOS DE DATOS HIDROGRÁFICOS IGN

2.2. Cambios en Base Topográfica Nacional (BTN25)

Se ha *modificado la estructura de BTN25 (v1)* para orientarlo a un modelo de explotación SIG, **BTN25 (v2)**, y al mismo tiempo siga utilizándose como base para actualizar el MTN25.

Requisitos:

- Continuidad geométrica
- Codigos identificadores de rio (DGA)



2. MODELOS DE DATOS HIDROGRÁFICOS IGN

2.3. Generación de información geográfica de referencia (IGR)

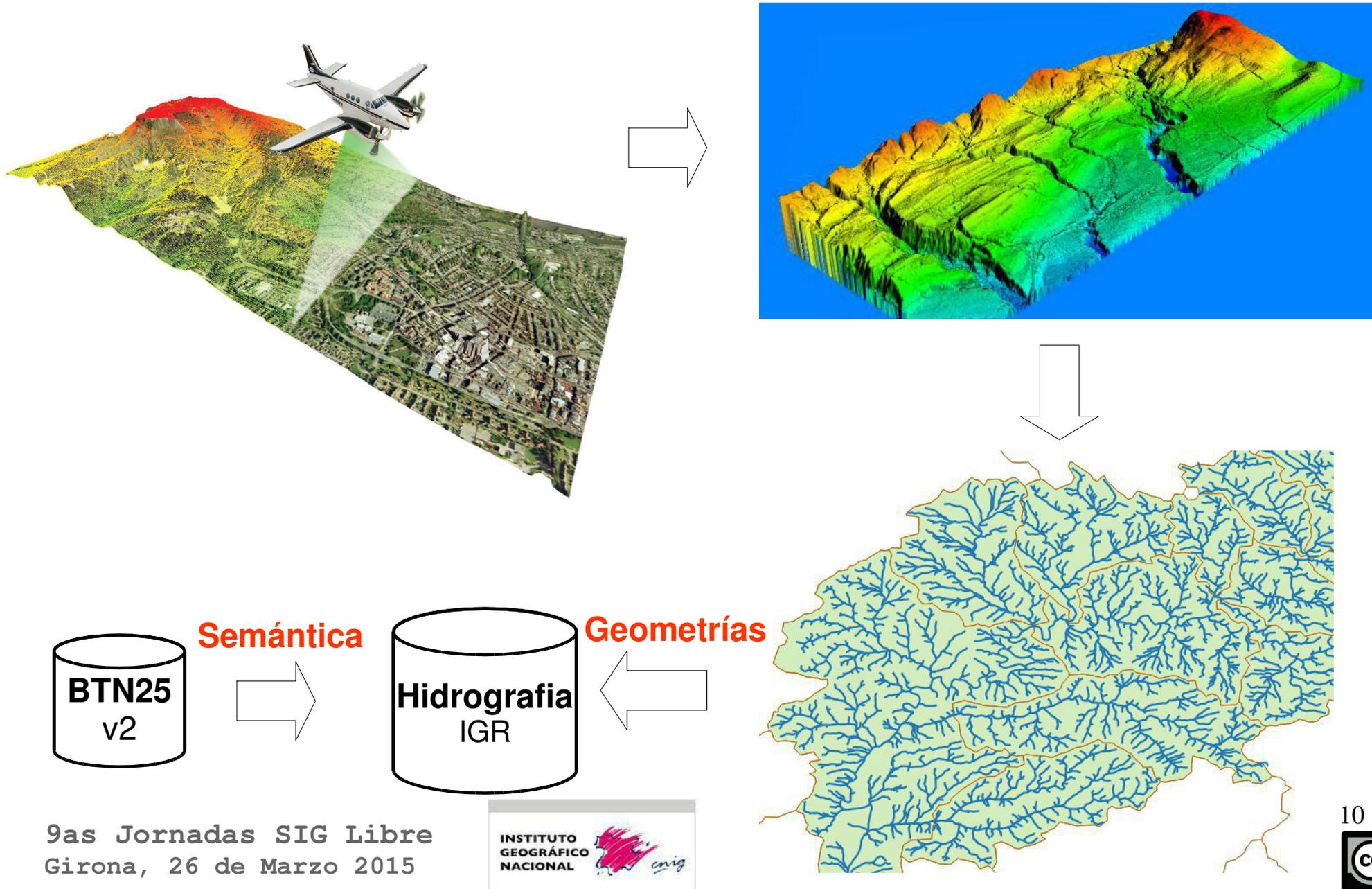
Nueva línea de producción de información geográfica de referencia (IGR)

Requisitos:

- Actualización semi-automática continua datos sensores aerotransportados (imágenes aéreas, satélite o LIDAR)
- Mayor resolución
- Mayor homogeneidad de datos
- Disminuir tiempo de actualización

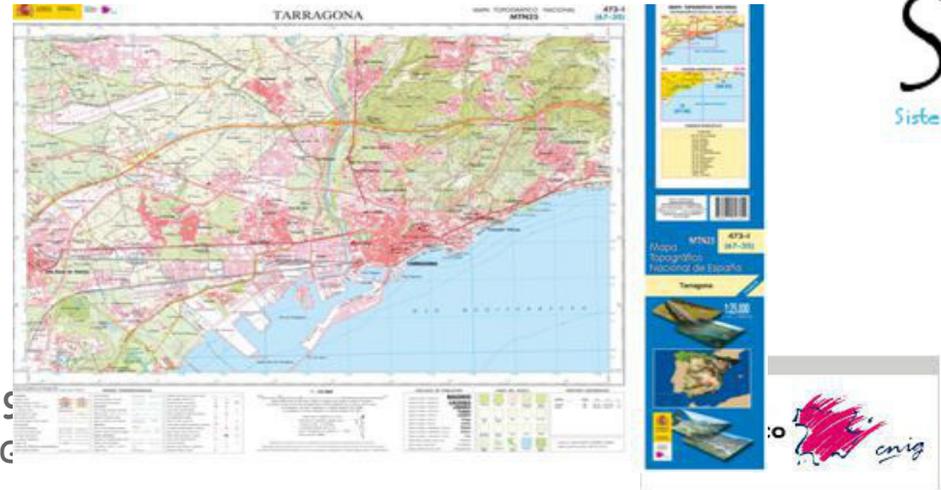
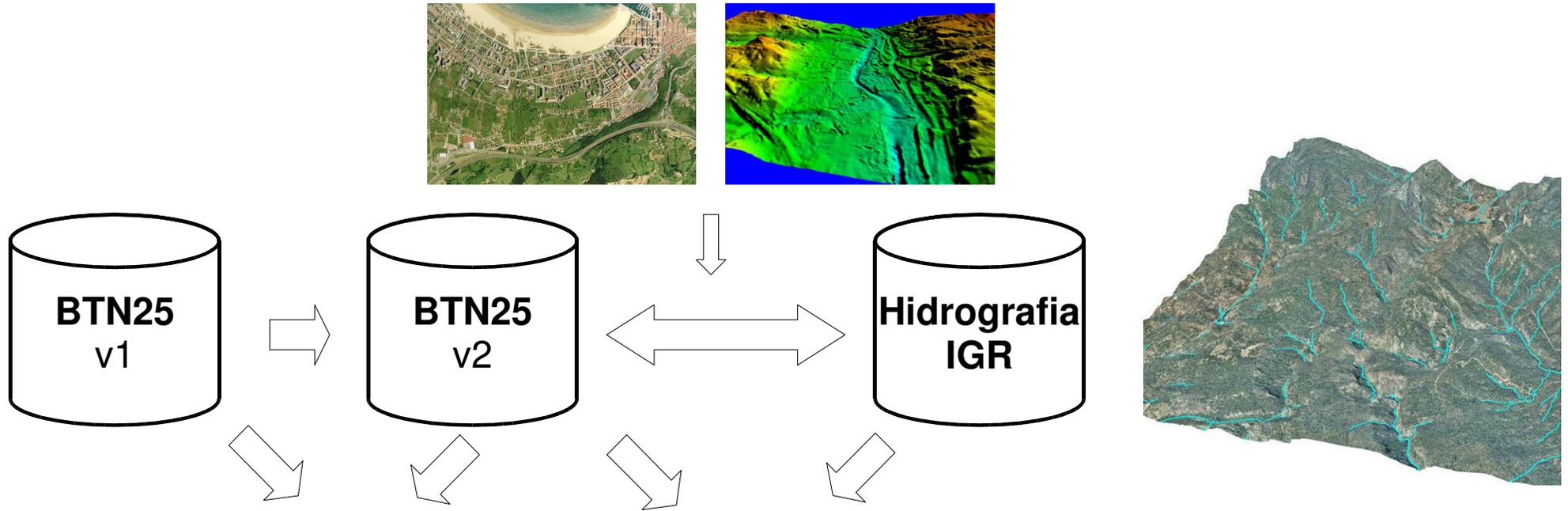
2. MODELOS DE DATOS HIDROGRÁFICOS IGN

2.3. Generación de información geográfica de referencia (IGR)



2. MODELOS DE DATOS HIDROGRÁFICOS IGN

2.4. Transformaciones y relación entre productos del IGN



3. TRANSFORMACIONES CON ETL

3. TRANSFORMACIONES CON ETL

3.1. Software.

Las herramientas **Extract - Transform - Load (ETL)** testeadas son **software libre (opensource)**

Geokettle 2.5



Kettle: <http://community.pentaho.com/projects/data-integration/>



Geokettle: <http://www.spatialytics.org/projects/geokettle/>

**Talend Spatial
Data Integrator 5.4**



Talend: <http://www.talend.com/>

Talend Spatial: <http://talend-spatial.github.io/>

**Humboldt Alignment
Editor
(Hale)**



Humboldt: <http://www.esdi-humboldt.eu/home.html>

HALE:
<http://www.dhpanel.eu/humboldt-framework/hale.html>

3. TRANSFORMACIONES CON ETL

3.2. Tipos de transformaciones

Las **transformaciones de estructura** (objetos, atributos y valores) que se realizan mediante un "mapeo" con herramientas ETL.

a) Mapeo entre modelos de datos

Las **transformaciones de geometrías** son procesos mas complicados y en algunos casos se han podido abordar de forma semi-automática mediante **análisis espacial** con herramientas ETL .

b) Control de calidad y análisis espacial



3. TRANSFORMACIONES CON ETL

3.3. Mapeo entre objetos, atributos y valores.

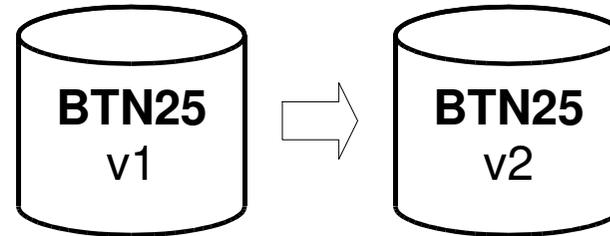


TABLA	ATRIBUTO	NOMBRE_ATR	VALOR	DESCRIPCION
BCN0301L_RIO (BTN25 v1)	0301L01	COMPO_0301	01	EJE
			02	MARGEN
	0301L02	REGIM_0301	01	PERMANENTE
			02	NO PERMANENTE
	0301L03	CATEG_0301	01	PRIMERA
			02	SEGUNDA
			07	TERCERA Y CUARTA
			08	QUINTA Y SEXTA
			98	NO DISPONIBLE
	0301L04	ETIQUETA		



TABLA	ATRIBUTO	NOMBRE_ATR	VALOR	DESCRIPCION
BCN0302L_RIO (BTN25 v2)	0302L01	COMPO_0302	01	EJE
			02	CANALIZADO
			03	CONEXIÓN
			04	TUBERÍA
			05	CORTA
	0302L02	PERSL_0302	01	PERMANENTE
			02	NO PERMANENTE
	0302L03	JERAR_0302	01	PRIMERA
			02	SEGUNDA
			03	TERCERA Y CUARTA
			04	QUINTA Y SEXTA
	0302L04	FUENTE		
	0302L05	ID_CURSO		
	0302L04	ETIQUETA		

3. TRANSFORMACIONES CON ETL

3.3. Mapeo entre objetos, atributos y valores

Ejemplo: cursos de agua (RIOS)

GeoKettle
Spatialytics.org ETL Tool

Value Mapper

Step name: Value Mapper

Fieldname to use: CATEG_0301

Target field name: JERAR_0302

Default upon no match:

Field values:

#	Source value	Target value
1	01	01
2	02	02
3	07	03
4	08	04

OK Cancel

Execution Results

Execution History | Logging | Step Metrics | Performance Graph

2015/03/14 01:28:14 - Shapefile File Output.0 - Finished processing (I=0, O=1, R=955, W=955, U=955, E=0)
2015/03/14 01:28:20 - GeoKettle's GUI - The transformation has finished!!
2015/03/14 01:28:20 - bcn0302l_rio.0 - Finished processing (I=0, O=1, R=33285, W=33285, U=33285, E=0)

3. TRANSFORMACIONES CON ETL

3.3. Mapeo entre objetos, atributos y valores.

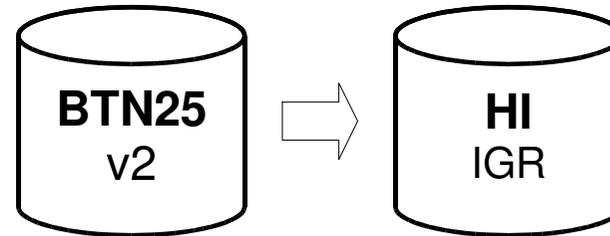


TABLA	ATRIBUTO	NOMBRE_ATR	VALOR	DESCRIPCION
BCN0302L_RIO (BTN25 v2)	0302L01	COMPO_0302	01	EJE
			02	CANALIZADO
			03	CONEXIÓN
			04	TUBERÍA
			05	CORTA
	0302L02	PERSI_0302	01	PERMANENTE
			02	NO PERMANENTE
	0302L03	JERAR_0302	01	PRIMERA
			02	SEGUNDA
			03	TERCERA Y CUARTA
			04	QUINTA Y SEXTA
	0302L04	FUENTE		
	0302L05	ID_CURSO		
	0302L04	ETIQUETA		



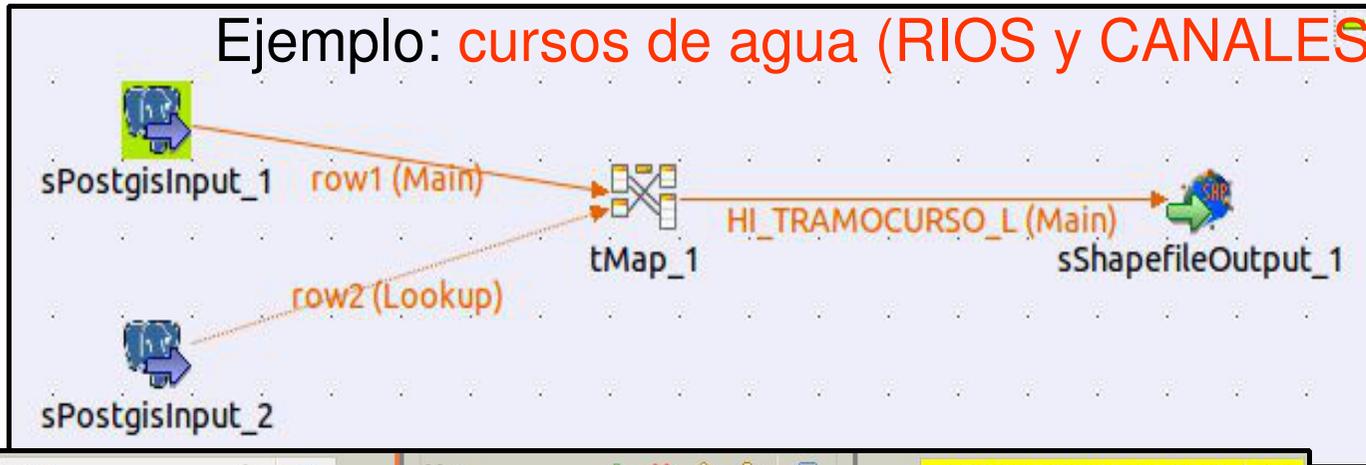
TABLA	atributo	valor	
HI_TRAMOCURSO_L (HI IGR)	id		
	idgn		
	id_curso		
	nombre		
	id_enlace		
	id_demarc		
	fecha_alta		
	fecha_baja		
	tipo_tramo	Eje	
		Conexión	
	tipo_curso	Curso natural	
		Canal	
		Acequia	
	persist	Permanente	
		Estacional	
		Esporádico	
	jerarquia	Primera	
		Segunda	
		Tercera	
Cuarta			
Quinta			
Sexta			
origen	natural		
	artificial		
situacion	En uso		
	En construcción		
	Proyectado		
	Abandonado		

3. TRANSFORMACIONES CON ETL

3.3. Mapeo entre objetos, atributos y valores



Ejemplo: cursos de agua (RIOS y CANALES)



row1

Column
gid
ID_CURSO
COMPO_0302
PERSI_0302
JERAR_0302
ETIQUETA
FUENTE

row2

Column
gid
ID_CURSO
COMPO_0305
SITUA_0305
ESTAD_0305
TIPO_0305

Var

Expression	Type	N	Variable
row1.COMPO_0302	String		var1
row1.COMPO_0305	String		var2
	int		idunique

HI_TRAMOCURSO_L

Expression	Column
Numeric.sequence("i	id
row1.ID_CURSO&&ro	id_curso
row1.ETIQUETA	nombre
	id_enlace
	id_demarc
	fecha_alta
	fecha_baja
Var.var1	tipo_tramo
row2.TIPO_0305	tipo_curso

```
3 // Copyright (c) 2006-2013, Talend Inc.
4 //
5 // This source code has been automatically generated by Talend Open Studio for
6 // Licensed under the Apache License, Version 2.0 (the "License");
7 // you may not use this file except in compliance with the License.
8 // You may obtain a copy of the License at
9 // http://www.apache.org/licenses/LICENSE-2.0
10 //
11 // Unless required by applicable law or agreed to in writing, software
12 // distributed under the License is distributed on an "AS IS" BASIS,
13 // WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
14 // See the license for the specific language governing permissions and
15 // limitations under the License.
16
17 package transf_ign.mapeo_hi_igr_0_1;
18
19 import routines.Mathematical;
20 import routines.DataOperation;
21 import routines.OpenStreetMap;
22 import routines.Relational;
23 import routines.TalendDate;
24 import routines.TalendDataGenerator;
```

Schema editor Expression editor

Column	Key	Type	NotNull	Date P.	Len	Pre	Del	Con	Column	Key	Type	NotNull
gid	<input checked="" type="checkbox"/>	int	<input checked="" type="checkbox"/>						persist	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>
ID_CURSO	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>						jerarquia	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>
COMPO_C	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>						origen	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>

3. TRANSFORMACIONES CON ETL

3.4. Análisis espacial y control de calidad

- **Orientación y validez de las geometrías lineales** de ríos en BTN25

GeoKettle
Spatialytics.org ETL Tool

Table input

Step name: StartPoint_0302L_RIO
Connection: PostGIS

SQL

```
SELECT  
"ID",  
"ETIQUETA", ST_Z(ST_StartPoint(the_geom) AS cota_inicio,  
ST_StartPoint(the_geom) AS the_geom  
FROM bcn03021_rio;
```

Script Values / Mod

Step name: Modified Java Script Value

Java script functions :script:

```
Script 1  
//Script here  
var geomreverse = the_geom.reverse()  
var valid = the_geom.isValid()  
var simple = the_geom.isSimple()
```

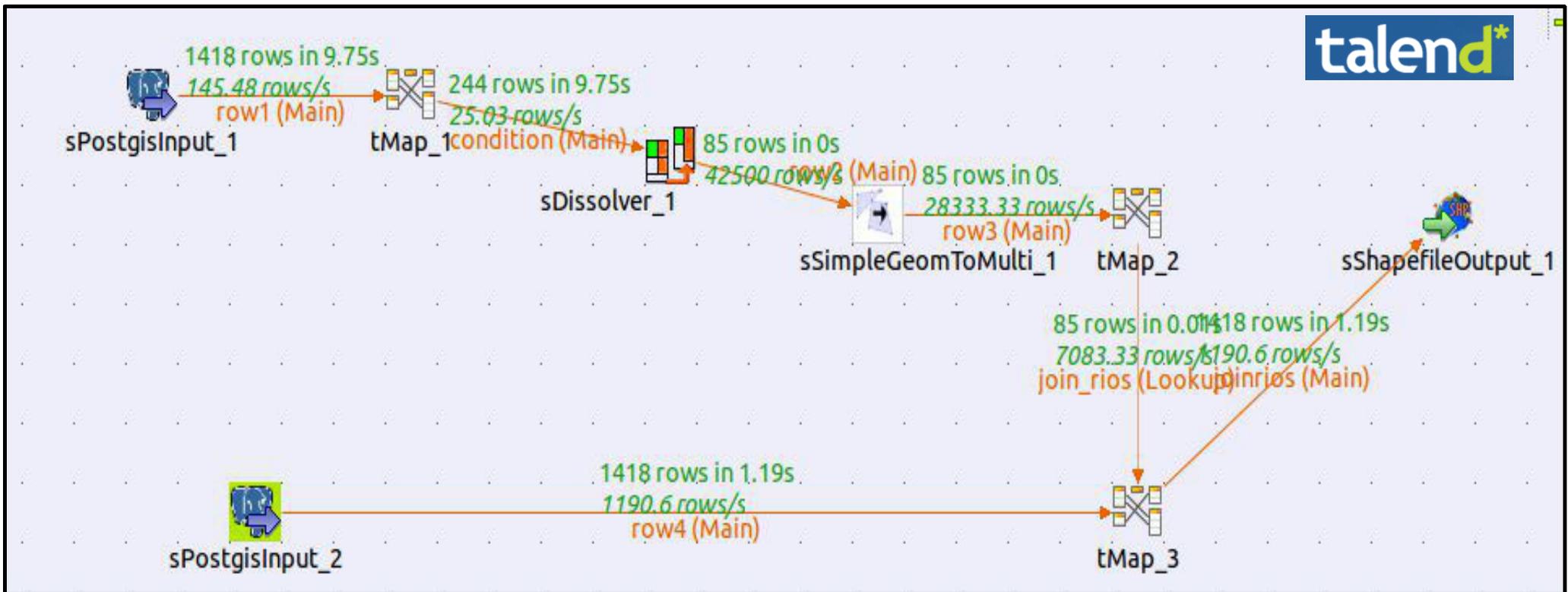
Positio
Compatibil

#	Fieldname	Rename to	Type	Length	Precision
1	geomreverse		Geometry		
2	valid		Boolean		
3	simple		Boolean		

3. TRANSFORMACIONES CON ETL

3.4. Análisis espacial y control de calidad

- **Asignación de códigos identificadores de ríos y nombres geográficos**



3. TRANSFORMACIONES CON ETL

3.5. Transformación a INSPIRE

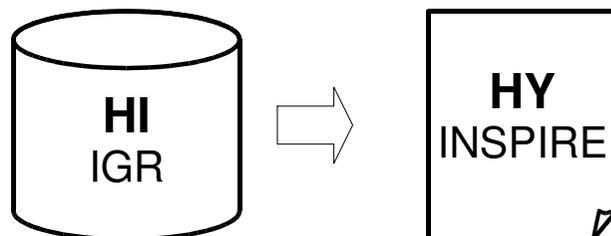
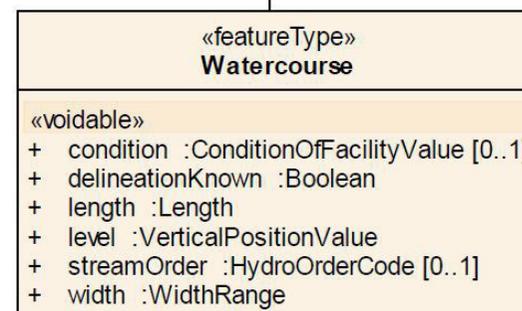
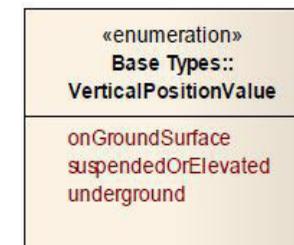
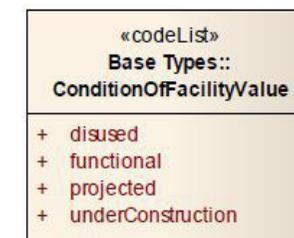
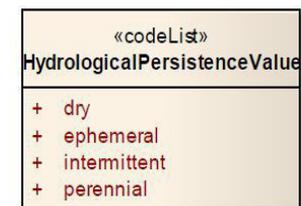
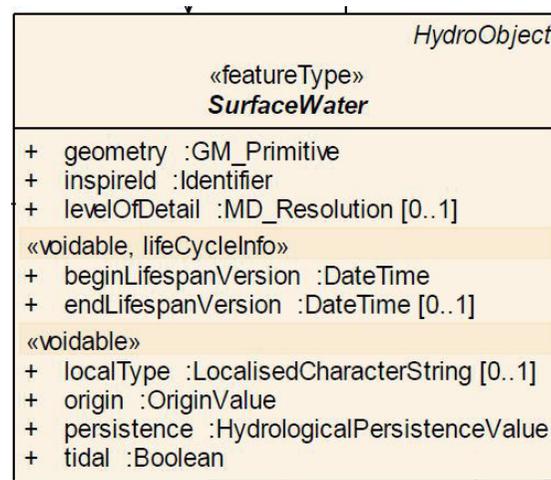


TABLA	atributo	valor
	id	
	idgn	
	id_curso	
	nombre	
	id_enlace	
	id_demarc	
	fecha_alta	
	fecha_baja	
	tipo_tramo	Eje Conexión
	tipo_curso	Curso natural Canal Acequia
	persist	Permanente Estacional Esporádico
		Primera Segunda
	jerarquia	Tercera Cuarta Quinta Sexta
	origen	natural artificial
	situacion	En uso En construcción Proyectado Abandonado

HI_TRAMOCURSO_L
(HI IGR)



3. TRANSFORMACIONES CON ETL

3.5. Transformación a INSPIRE



The screenshot displays the HUMBOLDT Alignment Editor 2.9.0 interface. The main window is titled "HUMBOLDT Alignment Editor 2.9.0 - hyign - /home/rafa/Tranformation/02_Transformations/03_HALE/hyign.hale". The interface is divided into several panes:

- Schema Explorer:** Shows the source schema "HI_TRAMOCURSO_L" and the target schema "Watercourse". The source schema includes fields like ANCHO_MAX, ANCHO_MIN, CANALIZA, CONDICION, FECHA_ALTA, FECHA_BAJA, FUENTE, gid, ID_CURSO, ID_DEMARC, and IDGN. The target schema includes fields like location, bank, beginLifespan, boundedBy, condition, delineationK, description, descriptionR, drainsBasin, endLifespan, geographical, and geometry.
- Alignment:** Shows the mapping process. Fields from the source schema are being mapped to fields in the target schema. For example, "ID_CURSO" is mapped to "...old.HydroIdentifier", "ID_DEMARC" is mapped to "drainsBasin", and "MAREA" is mapped to "tidal".
- Map:** Displays a map of the source data, showing a river network and a town labeled "GUALDA". The map is overlaid with the source data.
- Source Data:** Shows a table of source data for "HI_TRAMOCURSO_L".
- Transforme:** Shows a table of transformed data for "Watercourse".
- Error Log:** Shows a list of operations, including "HALE project export" (successful) and "HALE project archive export" (failed).

The "Source Data" table is as follows:

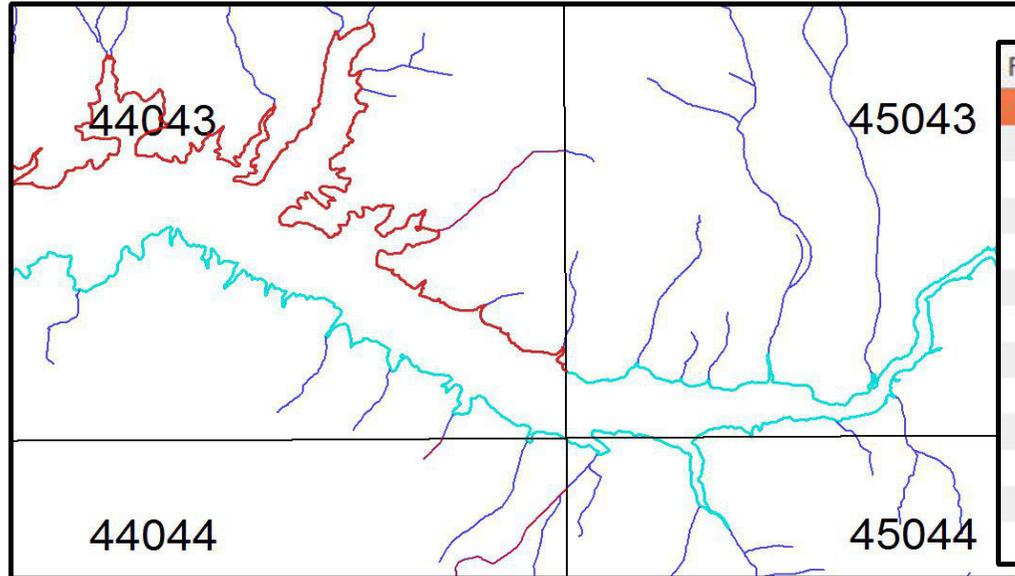
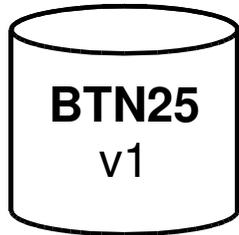
HI_TRAMOCURSO_L	1	2
HI_TRAMOCURSO	+	+
ANCHO_MAX	no value	no value
ANCHO_MIN	5	5
CANALIZA	NO	NO
CONDICION	NO APLICABLE	NO APLICABLE
FECHA_ALTA	20140916	20140916
FECHA_BAJA	no value	no value
FUENTE	IGN	IGN
gid	1	2
ID_CURSO	10069354864	1006935486

The "Transforme" table is as follows:

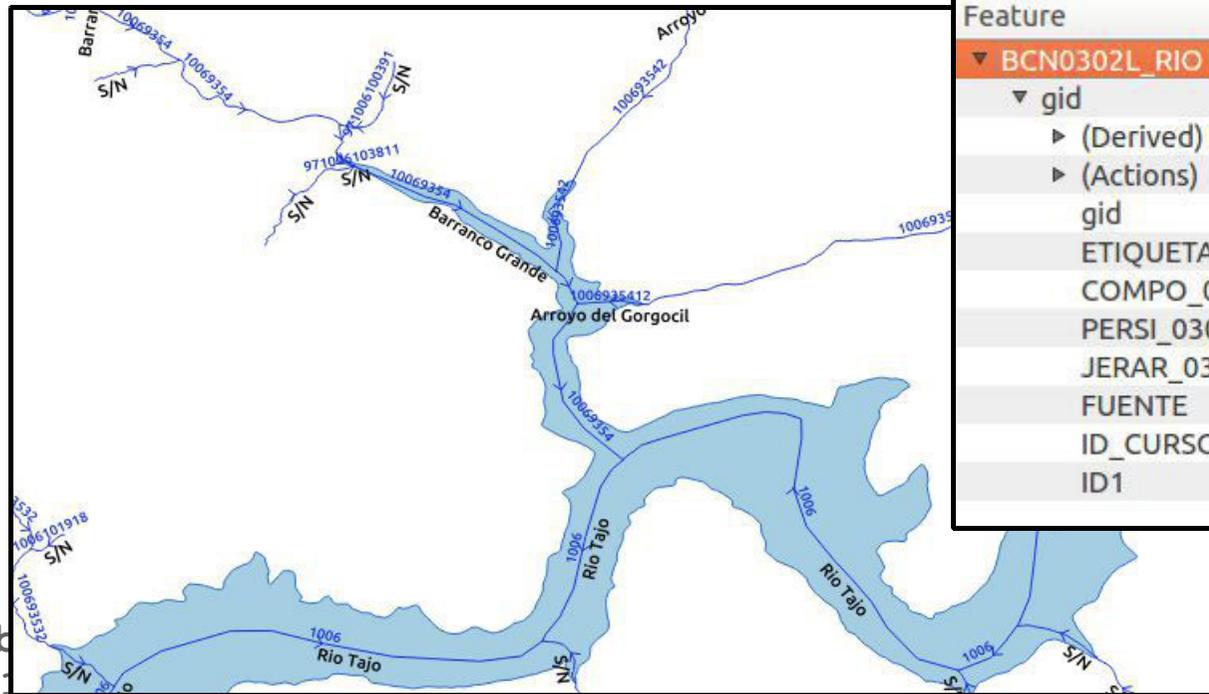
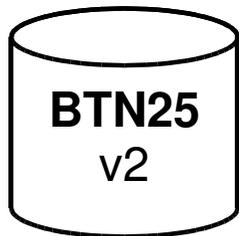
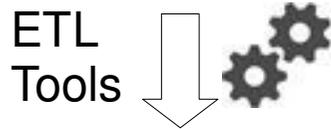
Watercourse	1	2
spelling	+	+
SpellingOfN	+	+
script	spa	spa
text	Arroyo de Navafria	S/N
translitera		
nilReason	no value	no value
geometry	+	+
hydroid	no value	no value
id	Watercourse_id_7b97e	Watercourse_id_c63c6

3. TRANSFORMACIONES CON ETL

3.6. Resultados



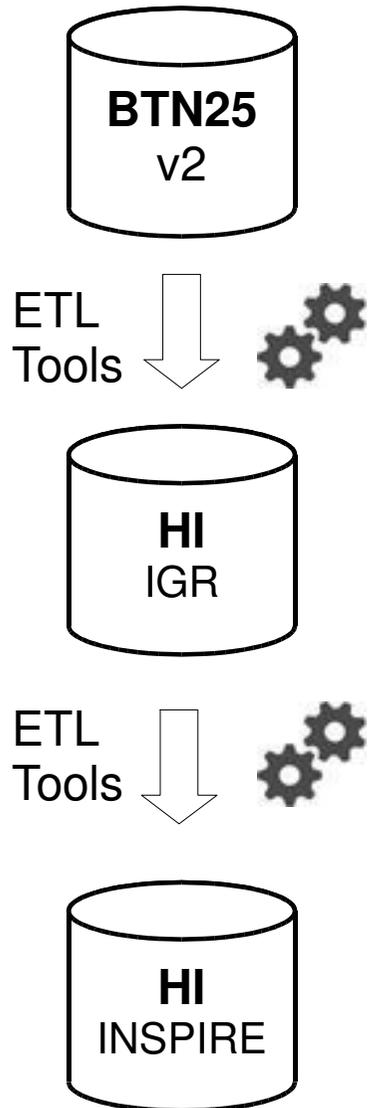
Feature	Value
BCN0301L_RIO	
gid	313
(Derived)	
(Actions)	
gid	313
ID	4127504
ID_HOJA	5124
F_ALTA	11/13/2008
F_BAJA	NULL
ETIQUETA	Río Cifuentes
COMPO_0301	01
REGIM_0301	01
CATEG_0301	98



Feature	Value
BCN0302L_RIO	
gid	5819
(Derived)	
(Actions)	
gid	5819
ETIQUETA	Río Tajo
COMPO_0302	03
PERSI_0302	01
JERAR_0302	01
FUENTE	NULL
ID_CURSO	1006
ID1	7261

3. TRANSFORMACIONES CON ETL

3.6. Resultados



2012-07

Feature	HI_TRAMOCURSO_L	Value
gid		7927
IDGN		44597
ID_CURSO		100694
NOMBRE		Río Ablanquejo
FUENTE		IGN
ID_DEMARC		ES030
FECHA_ALTA		20140916
FECHA_BAJA		NULL
TIPO_CURSO		CURSO
ORIGEN		NATURAL
PERSIST		PERMANENTE
MAREA		0
CONDICION		NO APLICABLE
NIVEL_VERT		SUPERFICIAL
ORDEN		NULL
ORDEN_ESQ		NULL
ORDEN_AMB		NULL
ANCHO_MAX		NULL
ANCHO_MIN		5

WMS PNOA

Layers

- Watercourse
- WMS PNOA
- Mosaico

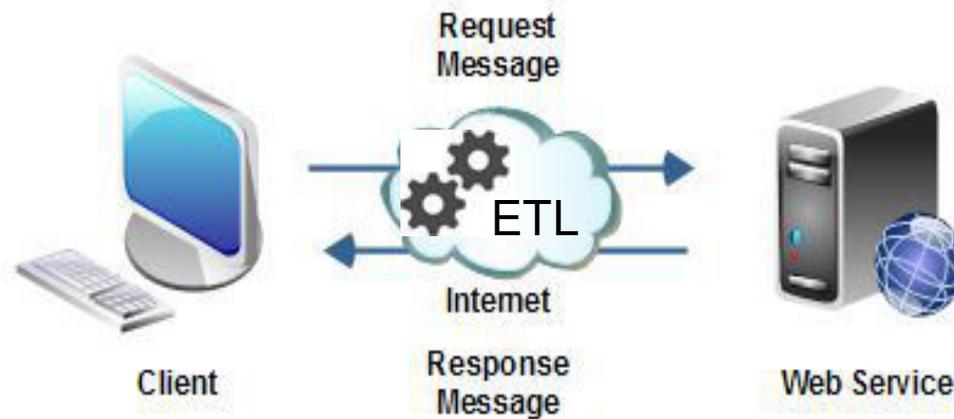
Identify Results

Feature	Value
▼ Watercourse	
▼ nameStatus	official
▶ (Derived)	
▶ (Actions)	
gml_id	Watercourse_id_6...
language	spa
nativeness	endonym
nameStatus	official
text	Arroyo de la Fuen...
script	spa

4. EXPLOTAR DATOS DESDE SERVICIOS WEB CON ETL

4. EXPLOTAR DATOS SERVICIOS WEB CON ETL

4.1. Explotar datos de un servicio web con ETL



- Funciones automatizadas por una **ETL**:

a) **Petición de datos** a un servidor que devuelve fichero (*.xml, *.gml)

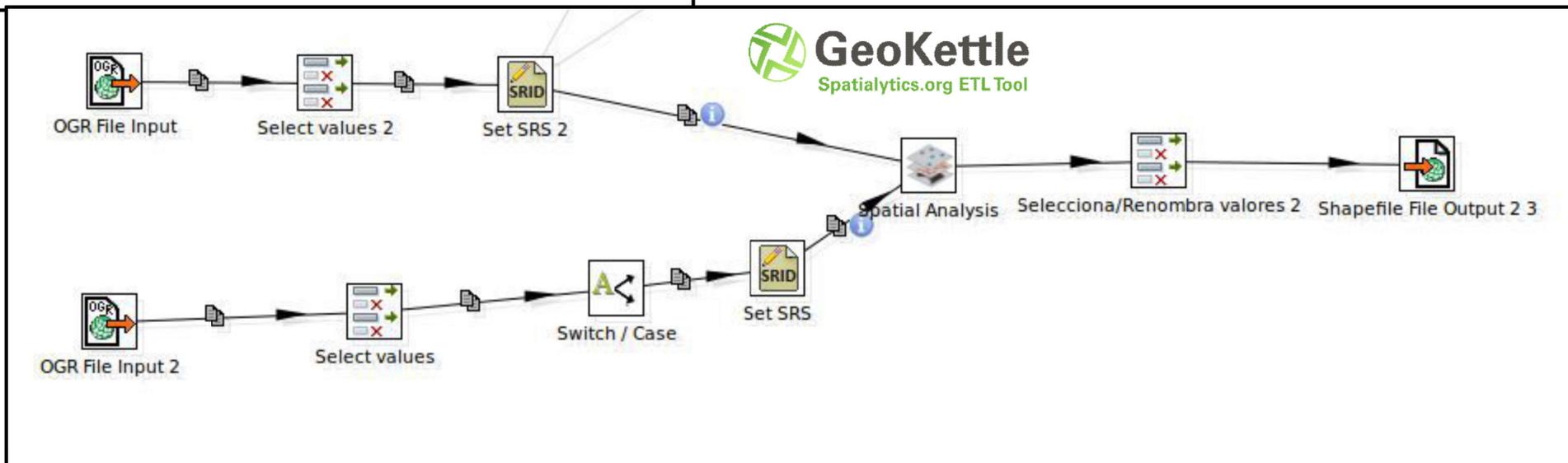
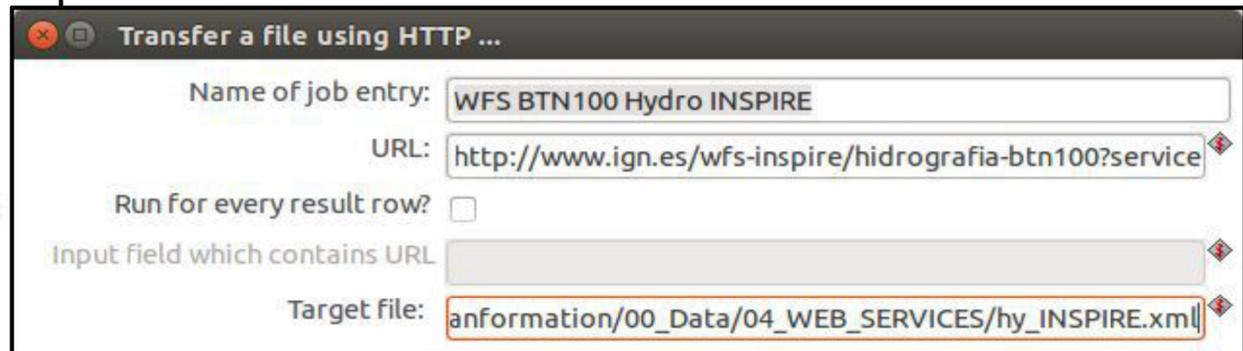
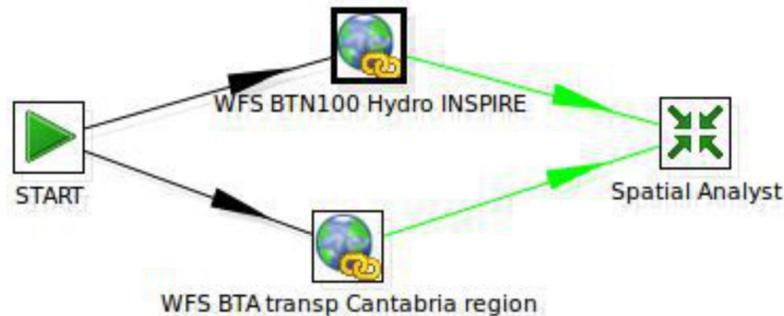
c) **Lee y Transformar** estos datos (mapeo o análisis espacial)

d) **Carga el resultado** en un almacén de datos

4. EXPLOTAR DATOS SERVICIOS WEB CON ETL

4.1. Web Feature Service (WFS)

- Nos **devuelve elementos geométricos** alojados en un fichero *.xml
(Cuidado: gran tamaño fichero acotar tamaño en la petición)



4. EXPLOTAR DATOS SERVICIOS WEB CON ETL

4.2. Metadata.

- **Catálogos de Metadatos (CSW)** permitiendo **realizar consultas y descubrir información geográfica** de una organización.
- Permite a los productores **transformar y cargar metadatos** dentro del propio catálogo con una estructura determinada (**NEM v1.2**)

The screenshot shows the 'CSW Input' configuration window. The 'Step Name' is 'CSW Input'. Under 'General', the 'csw Url' is 'http://www.ign.es/csw-inspire/srv/spa/csw?'. The 'Method' is 'GET', 'Version' is '2.0.2', and 'Constraint Language' is 'CQL_TEXT'. There is an option for 'Activate Login?'. Under 'login parameters', the 'Login Url' is 'http://localhost:8080/geonetwork/srv/en/xml.user.login'. There are fields for 'Username' and 'Password'. The 'Query Detail' section has a 'Keyword' of 'Hidrografía', 'Advanced search' checked, and 'Start position and MaxRecord' set to 'Start Position: 1' and 'Max Records: 100'. An 'Advanced Query' table is visible at the bottom.

#	Queryable Elements	Comparison Operators	Value to Search
1			



5. CONCLUSIONES

1. Análisis herramientas ETL opensource:

- **Resultados aceptables** (falta testeo en producción)
- Necesidad de **aumentar funciones análisis espacial**
- **Mejorar accesibilidad** a librerías espaciales (GDAL, jts topology suite)

2. Metodología datos hidrográficos:

- Los datos siguen en fase de producción
- **Mayor actualización e interoperabilidad**
- **Mejor explotación** con SIG y Servicios web espaciales

Gracias por su atención

Rafael Sierra Requena

rsierra@fomento.es