

Ejemplos Prácticos de Manejo e Integración de Datos en el Proyecto TRUST implementado en la Cuenca del Río Lurín en Perú

Vanessa Rojas
Directora Proyectos
Disy Informationssysteme
Alemania

- Perfil empresa
- Integración de datos
- Aplicaciones
- Implementación Proyecto TRUST
- Otros proyectos
- Resumen
- Links



- Proveedor de soluciones de software para análisis de datos, reportes y SIG en Alemania
- Sede central: Karlsruhe, Alemania. 100 empleados.
- Enfoque empresarial: gestión de datos y análisis con enfoque en datos espaciales
- Producto Cadenza: SIG & plataforma de análisis Desktop, Web & Mobile
- GeoSpatial Integration for Talend (Integración de datos)

- Principal segmento de clientes: grandes autoridades públicas en Alemania y Austria
- Clientes internacionales: Proyectos Investigación/Plugin

Áreas:

- Seguridad
- Medio Ambiente
- Protección al Consumidor
- Infraestructura
- Tráfico





Medio ambiente: ¿Qué empresas usan el agua río arriba? ¿Qué propiedades están ubicadas en áreas de protección contra inundaciones? ¿Qué tan limpia es el agua en una sección del río?



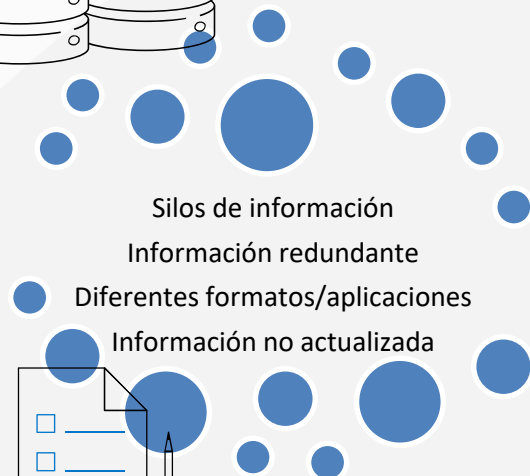
Infraestructura & Ruido: ¿Cuántas personas viven en un distrito afectado por el tráfico aéreo / vial o ruido del ferrocarril? ¿Cuáles son los efectos de las medidas de control de ruido? ¿Dónde se pueden crear áreas de compensación para nuevos planes y proyectos?



Policía & Seguridad regional: ¿Dónde se reportaron robos en las últimas dos semanas? ¿Cómo preparo a mi equipo para un evento importante? ¿Ha aumentado o disminuido el número de casos por región en los últimos años?



Agricultura: Evaluación y control de fondos de la UE. ¿Qué granja recibirá fondos para qué tipo de campos? ¿Es consistente? ¿Cuál es el rendimiento de cultivo esperado en una determinada región? ¿Qué rendimientos se esperan en caso de tormenta severa o inundación?

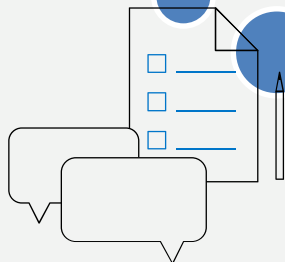


Silos de información

Información redundante

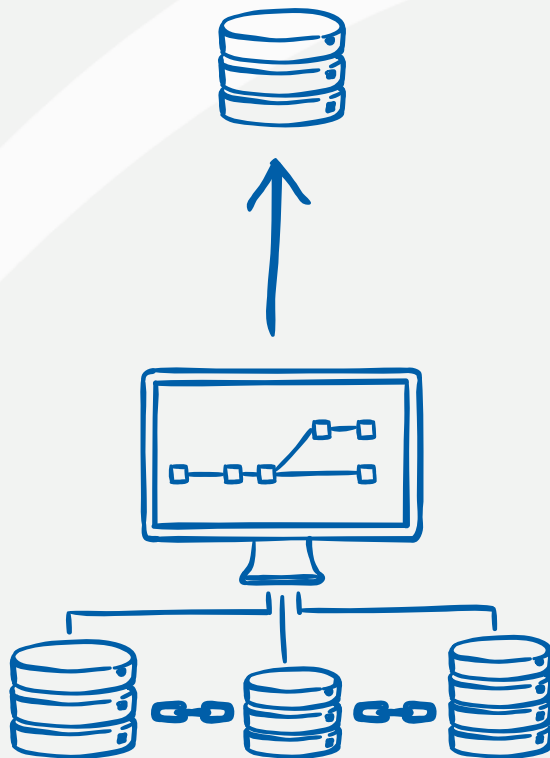
Diferentes formatos/aplicaciones

Información no actualizada



Información
Actual
Fiable
Apoyo toma
decisiones





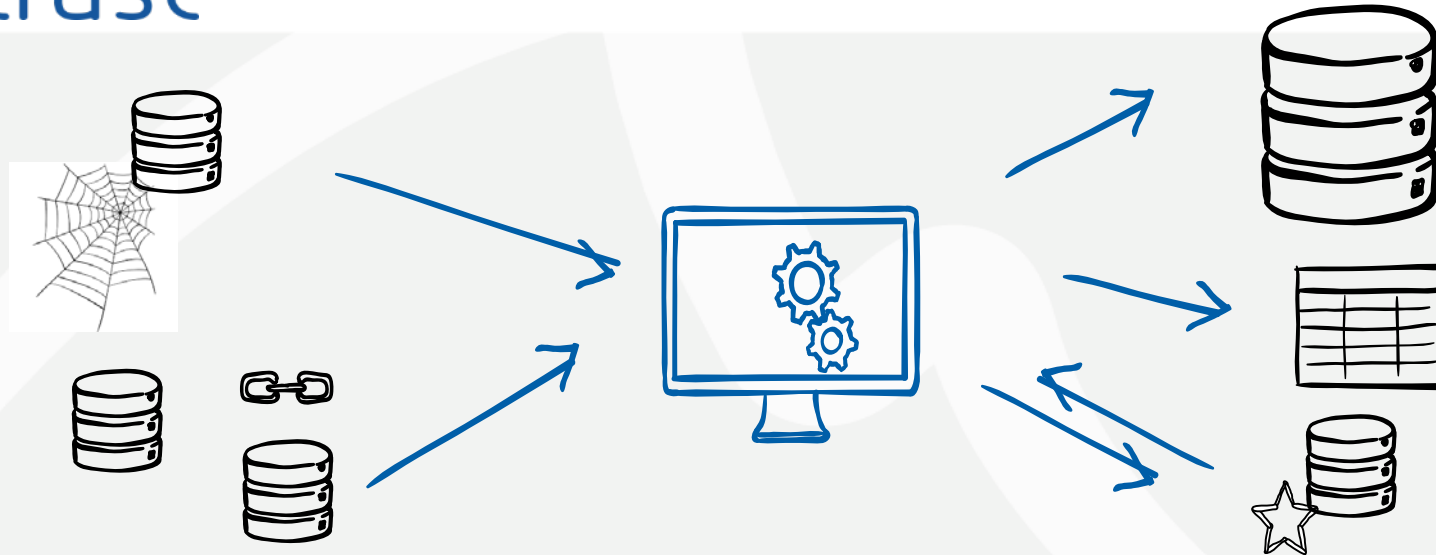
"L" Load / Cargar



"T" Transform / Transformar



"E" Extract / Extraer



• ... Extraer:

- Acceso diferentes formatos
- Definición de datos
- Interfaces

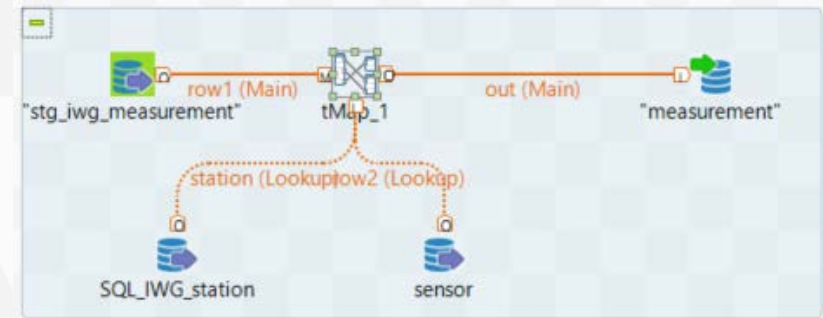
• ... Transformar:

- Conversión formatos
- Cálculos
- Estandarización
- Calidad

• ... Load/Cargar:

- Modo de carga (Insertar, Actualizar, Eliminar)
- Estadísticas/Errores
- Formatos de salida

- Aplicación gráfica
- > 800 componentes
- Organización de procesos
- Soporte múltiples formatos de entrada y salida de datos
- Generación de código automático (Java)
- Talend Open Studio (sin costo)



Conectar datos



Crear procesos



Generar datos



Lista componentes

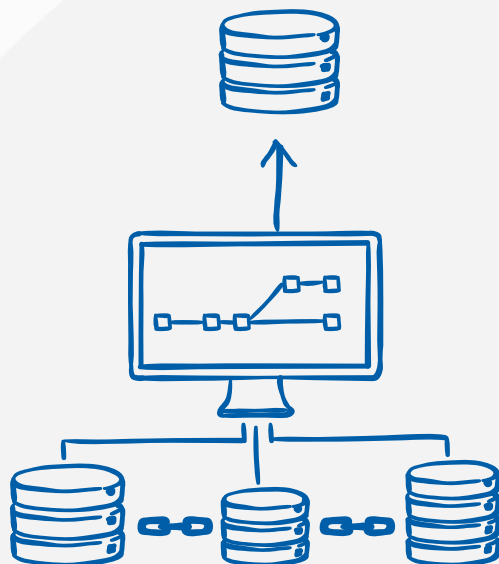
Proceso (Job)

Componente

Repositorio

- Integración combinada de atributos y datos espaciales
- Componentes geoespaciales
- Integración con Talend Open Studio, Talend Data Fabric y Plataformas
- Soporte de diferentes formatos (Oracle, PostGIS, SpatiaLite, Shapefile, WKT, WKB, GeoJSON, LRS)
- Numerosos componentes de transformación.

Disy GeoSpatial
Calculators
dAreaCalculator
dCoordinateFetcher
dLengthCalculator
Data quality
Databases
File
GeoJSON
dGeoJSONObjectToGeometry
dGeometryToGeoJSONObject
Shapefile
dShapeFileInput
dShapeFileOutput
Geom Operations
dAggregateGeometryRow
dBufferCalculator
dCentroidCalculator
dConvexHullCalculator
dEnvelopeCalculator
dIntersection
dPointFromCoordinates
dProject
Input



Análisis, Visualización

"L" Load / Cargar

+ Datos espaciales
(Formatos, Bases de Datos, Interfaces)

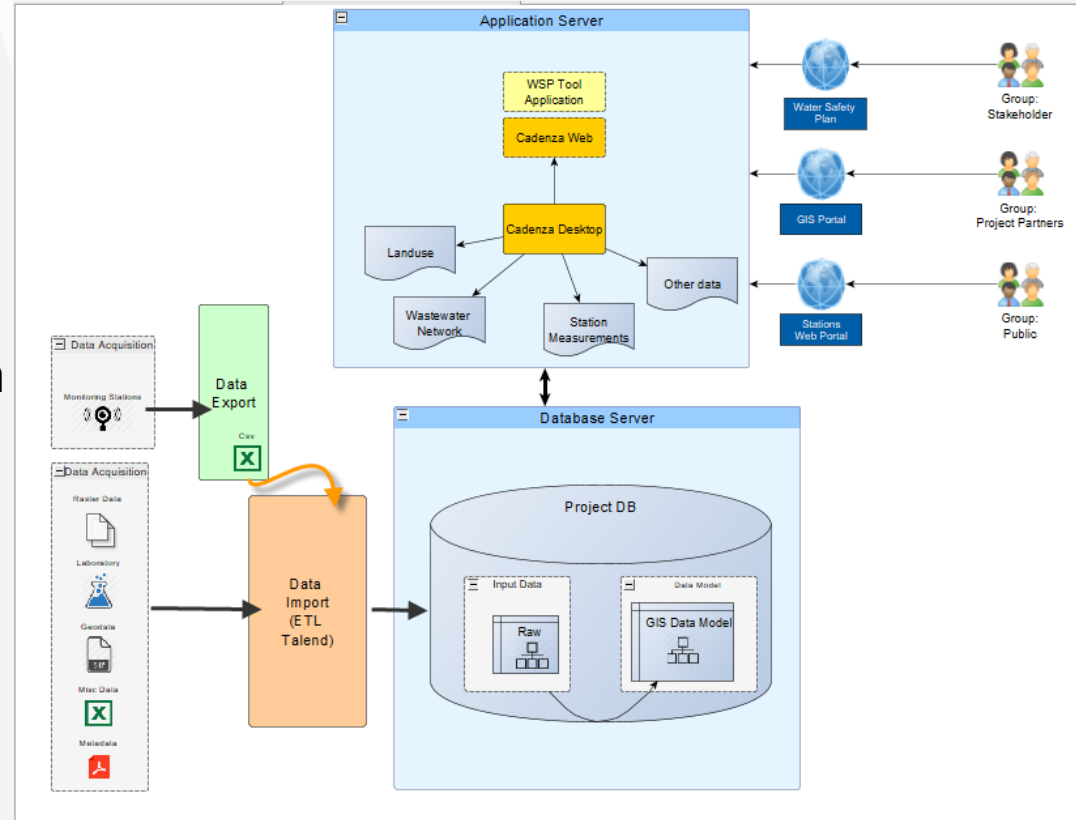
"T" Transform / Transformar

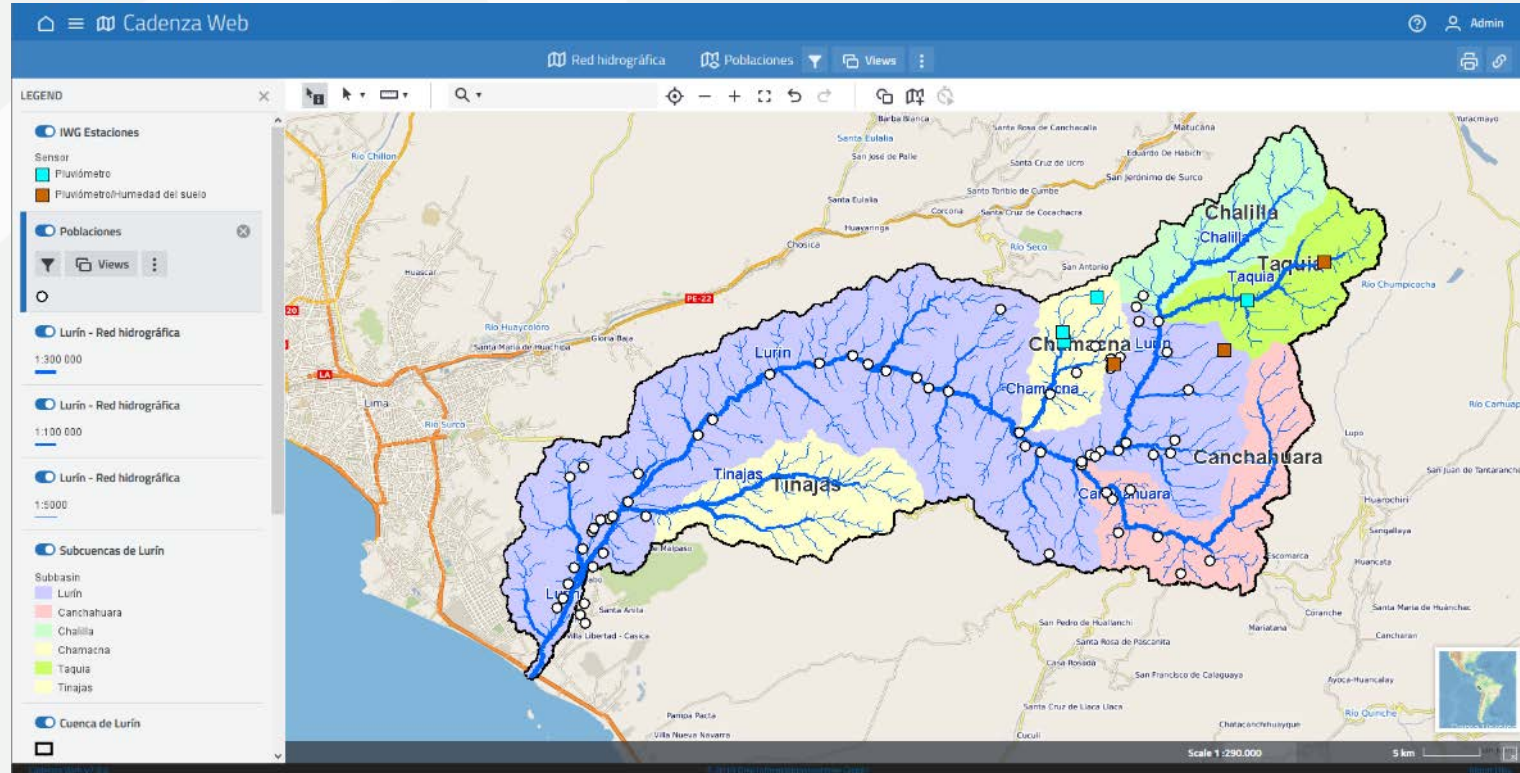
+ Operaciones geométricas
+ Validación y corrección (espacial)
+ Combinación atributos y datos espaciales

"E" Extract / Extraer

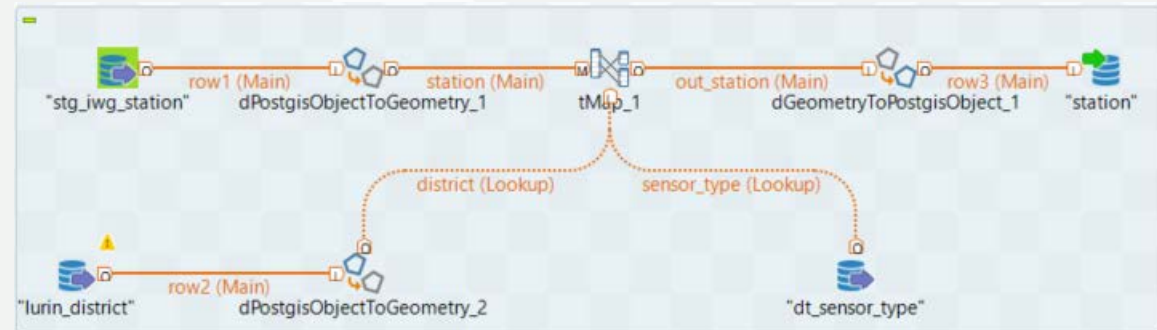
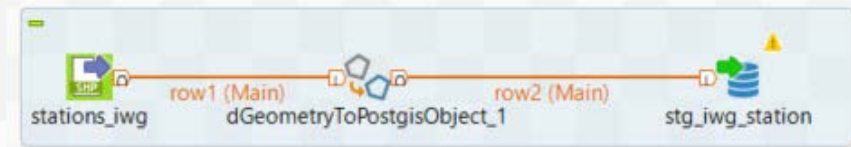
+ Datos espaciales
(Formatos, Bases de Datos, Interfaces)

- Diferentes formatos de entrada
- Base de Datos PostgreSQL + PostGIS
- Procesos ETL para la integración de datos



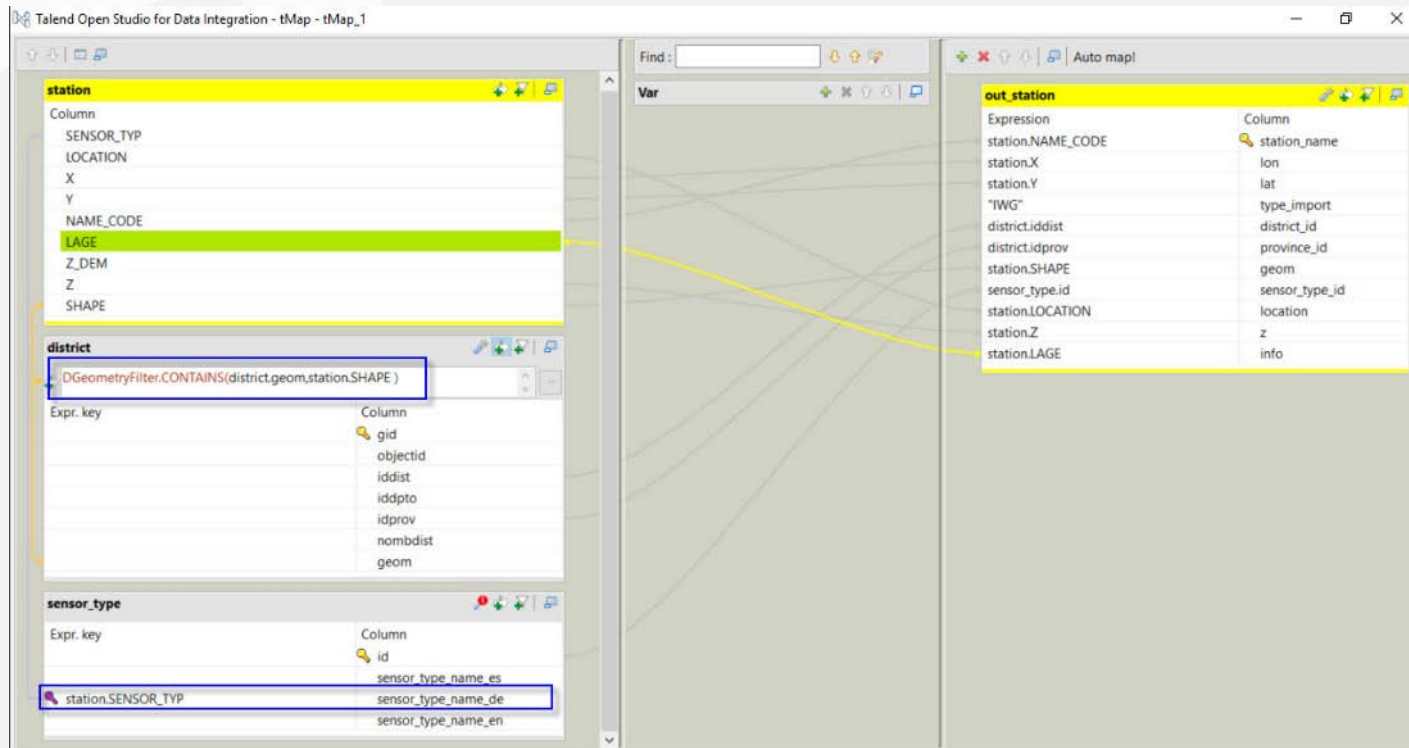


- Estaciones IWG
 - Pluviómetro
 - Humedad del suelo
- Reconocimiento de geometrías
 - Conversión Objeto <-> DGeometry
- Funciones espaciales
- Lectura Shapefile
- Lectura/escritura PostgreSQL
- Mapeo de atributos (tMap)



Componente tMap



- Filtro espacial
– CONTAINS
- Tablas paramétricas
- Eliminación datos redundantes








The screenshot shows the Talend Open Studio interface for a tMap component. The 'station' table is selected, and the 'LAGE' column is highlighted. The 'district' table is also selected, and the 'DGeometryFilter.CONTAINS(district.geom,station.SHAPE)' expression is entered in the 'Expr. key' field. The 'sensor_type' table is selected, and the 'station.SENSOR_TYP' expression is entered in the 'Expr. key' field. The 'out_station' table is shown on the right, with columns for 'station_name', 'lon', 'lat', 'type_import', 'district_id', 'province_id', 'geom', 'sensor_type_id', 'location', 'z', and 'info'.

Portal SIG


- Vista Usuario:
Estaciones




Cadenza Web

 IWG Estaciones

 Views

SELECTION




Filter
7 Results

Estación


Sensor


Distrito

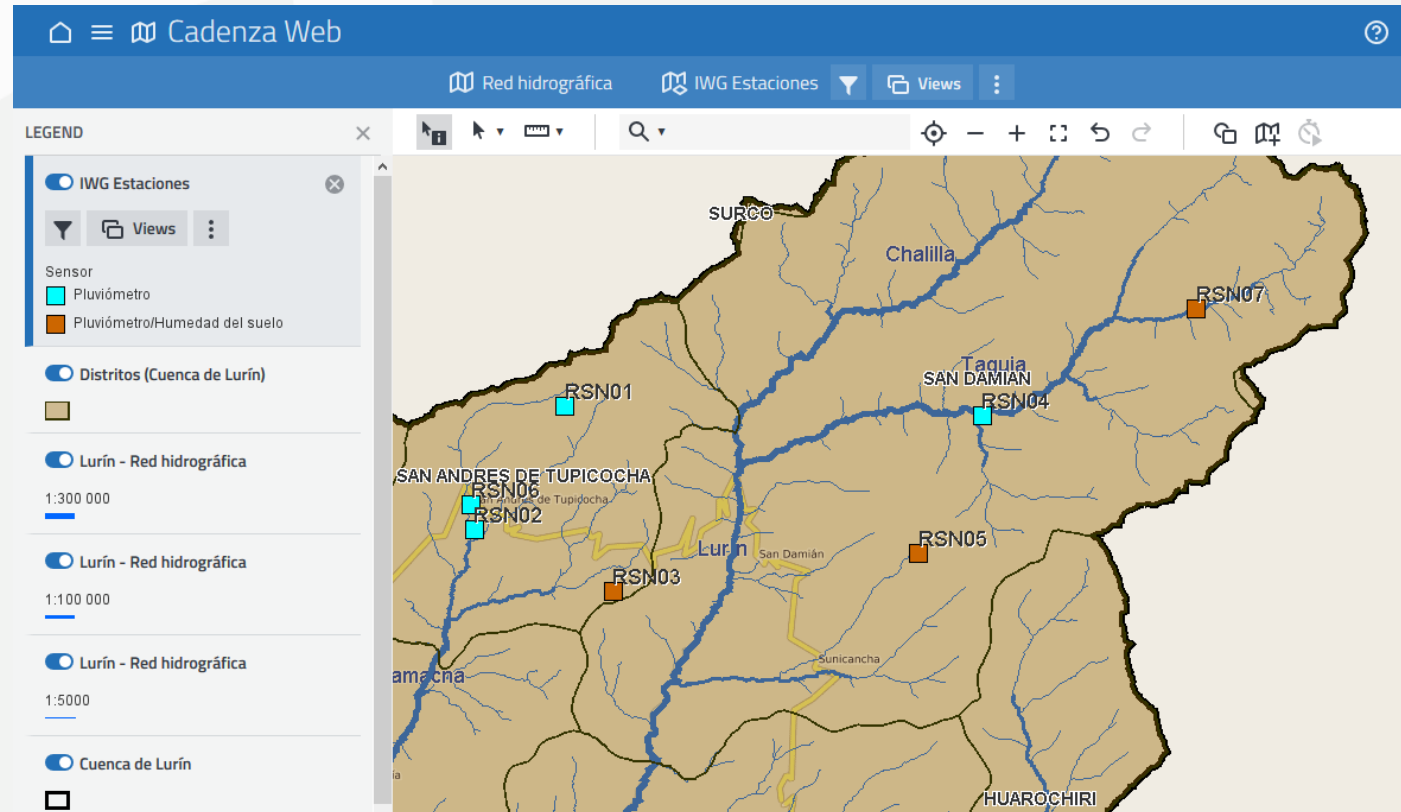
SAN ANDRES DE TUPICOCHA
SAN DAMIAN

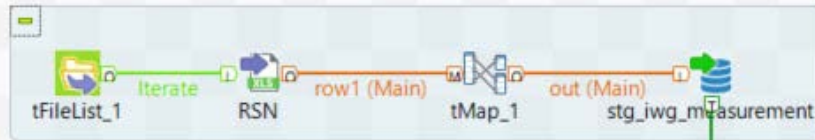


	Estación	Lon	Lat	Z	Sensor	Distrito	Ubicación
1	RSN01	-76.4482	-11.9735	4358.33	Pluviómetro	SAN ANDRES DE TUPICOCHA	Chullaca
2	RSN02	-76.4744	-12.0083	0.00	Pluviómetro	SAN ANDRES DE TUPICOCHA	Tupicocha
3	RSN03	-76.4343	-12.0257	3603.53	Pluviómetro/Humedad del suelo	SAN ANDRES DE TUPICOCHA	Culpe
4	RSN04	-76.3275	-11.976	3680.37	Pluviómetro	SAN DAMIAN	Piscigranja
5	RSN05	-76.346	-12.015	4454.51	Pluviómetro/Humedad del suelo	SAN DAMIAN	Yanascocha
6	RSN06	-76.4754	-12.0011	3322.60	Pluviómetro	SAN ANDRES DE TUPICOCHA	Tupicocha
7	RSN07	-76.2656	-11.9459	4335.74	Pluviómetro/Humedad del suelo	SAN DAMIAN	Taqua alto

Portal SIG

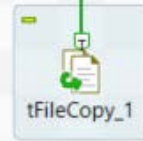
- Mapa Estaciones





Name	Date modified	Type
RSN01.xlsx	9/16/2019 11:38 AM	Microsoft Excel-Arbeits
RSN02.xlsx	9/16/2019 11:38 AM	Microsoft Excel-Arbeits
RSN03.xlsx	9/16/2019 11:38 AM	Microsoft Excel-Arbeits
RSN05.xlsx	9/16/2019 11:38 AM	Microsoft Excel-Arbeits

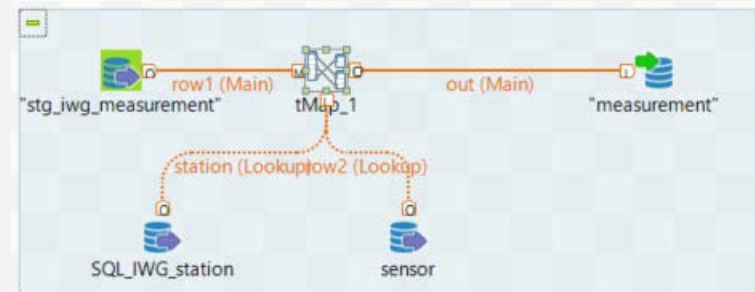
OnComponentOk



RSN01.xlsx - Microsoft E

time			
A	B	C	D
time	P_mm	temp_C	
2018-02-23 00:00:00	0	7.803858	
2018-02-24 00:00:00	0.193925	4.672615	
2018-02-25 00:00:00	8.14486	9.910642	
2018-02-26 00:00:00	8.338785	4.513861	

- Mediciones (Excel)
- Formato definido para la carga
- Carga de varios archivos



- Portal SIG
- Vista usuario:
Mediciones
- Filtros
 - Parámetro
 - Estación
 - Fecha

Cadenza Web
Mediciones (IWG) ▼ Views 📄

SELECTION

Filter

103 Results

Parámetro ✕

Precipitación (diaria),PTD ✕

Estación ✕

RSN01 ✕

Sensor ✕

Fecha ✕

01.01.2019 - 31.12.2019

OK

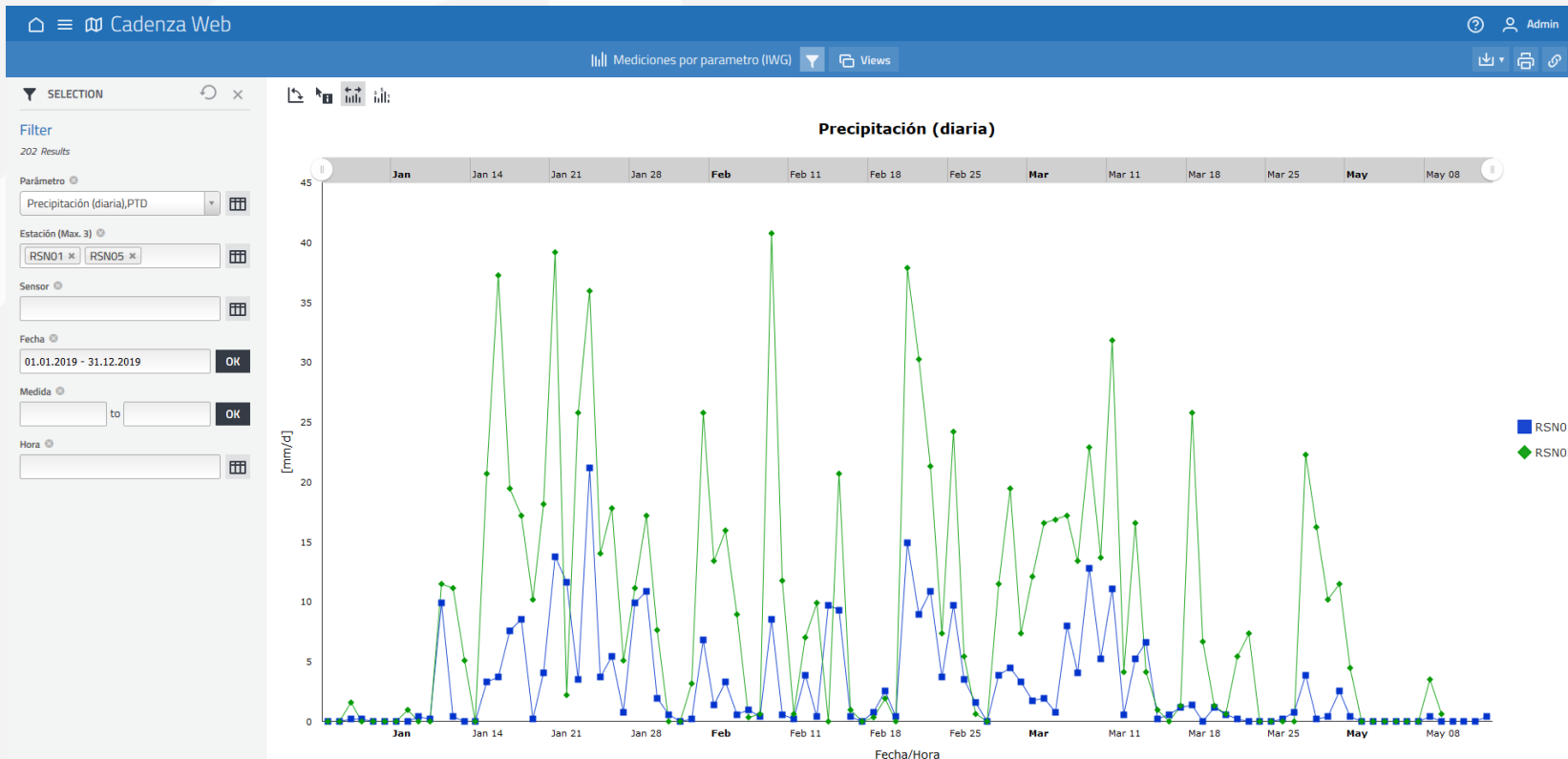
Medida ✕

to

OK

Hora ✕

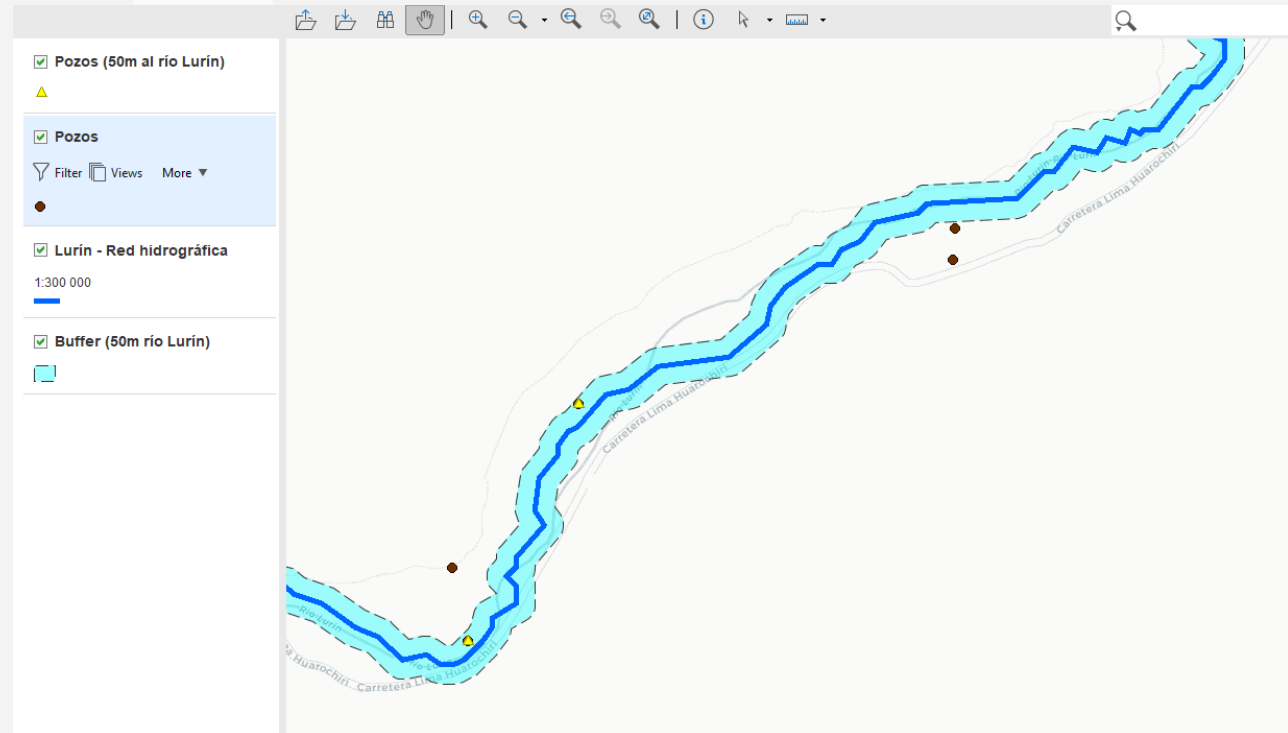
	Station	Sensor Name	Fecha/Hora	Medida	Unidad	Param. Abbr.	Parámetro
13	RSN01	RSN01	01.04.2019 00:00	0.3879	[mm/d]	PTD	Precipitación (diaria)
14	RSN01	RSN01	31.03.2019 00:00	2.521	[mm/d]	PTD	Precipitación (diaria)
15	RSN01	RSN01	30.03.2019 00:00	0.3879	[mm/d]	PTD	Precipitación (diaria)
16	RSN01	RSN01	29.03.2019 00:00	0.1939	[mm/d]	PTD	Precipitación (diaria)
17	RSN01	RSN01	28.03.2019 00:00	3.8785	[mm/d]	PTD	Precipitación (diaria)
18	RSN01	RSN01	27.03.2019 00:00	0.7757	[mm/d]	PTD	Precipitación (diaria)
19	RSN01	RSN01	26.03.2019 00:00	0.1939	[mm/d]	PTD	Precipitación (diaria)
20	RSN01	RSN01	25.03.2019 00:00	0	[mm/d]	PTD	Precipitación (diaria)
21	RSN01	RSN01	24.03.2019 00:00	0	[mm/d]	PTD	Precipitación (diaria)
22	RSN01	RSN01	23.03.2019 00:00	0	[mm/d]	PTD	Precipitación (diaria)
23	RSN01	RSN01	22.03.2019 00:00	0.1939	[mm/d]	PTD	Precipitación (diaria)
24	RSN01	RSN01	21.03.2019 00:00	0.5818	[mm/d]	PTD	Precipitación (diaria)
25	RSN01	RSN01	20.03.2019 00:00	1.1636	[mm/d]	PTD	Precipitación (diaria)
26	RSN01	RSN01	19.03.2019 00:00	0	[mm/d]	PTD	Precipitación (diaria)
27	RSN01	RSN01	18.03.2019 00:00	1.2535	[mm/d]	PTD	Precipitación (diaria)



Control distancia pozos

Proceso:

- Cálculo buffer alrededor del río Lurín (50m)
- Determinar pozos cercanos al río

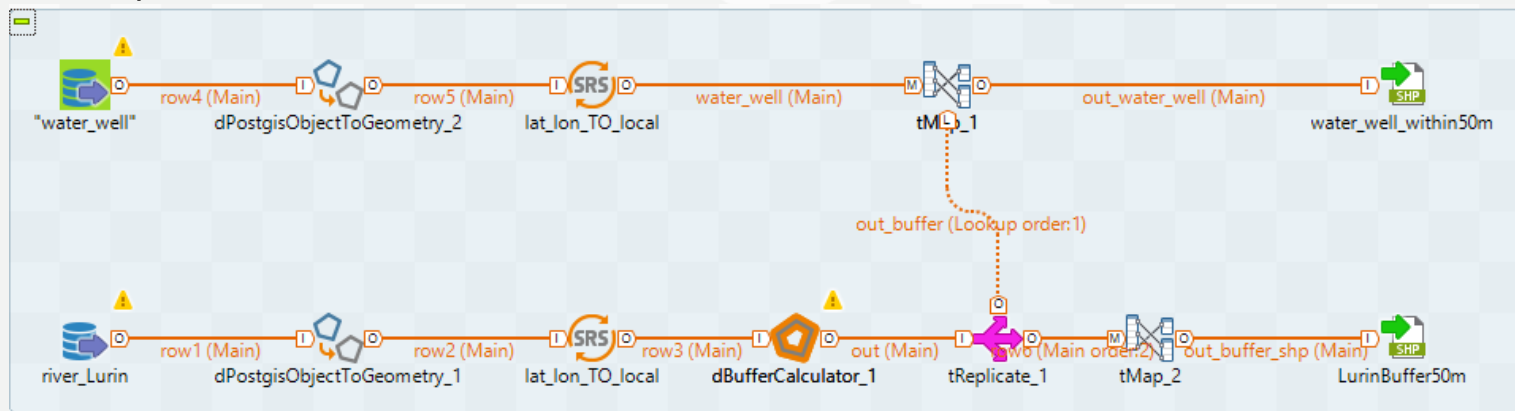


Proceso:

- Diferentes flujos de datos
- Proyección de sist. de coordenadas
- Cálculo buffer
- Filtro espacial

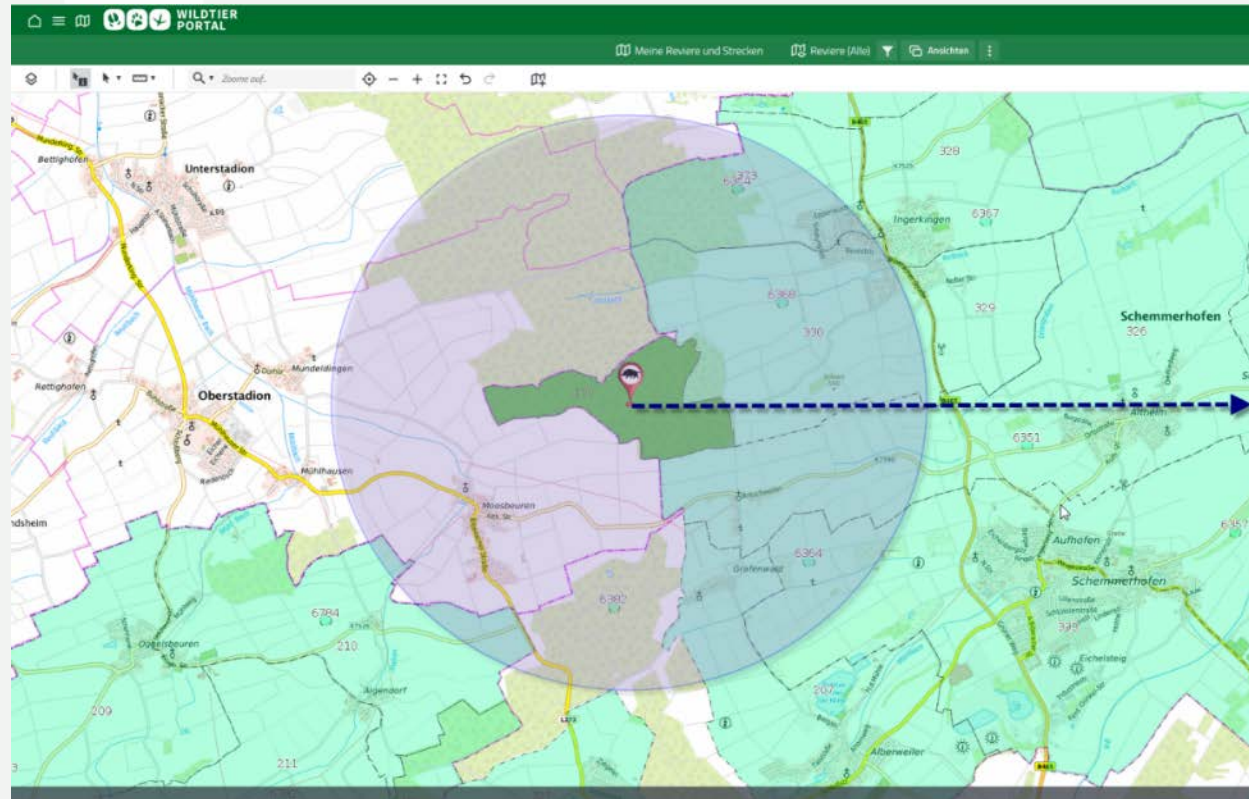
Resultado:

- Archivo con pozos e información relevante
- Archivo con buffer

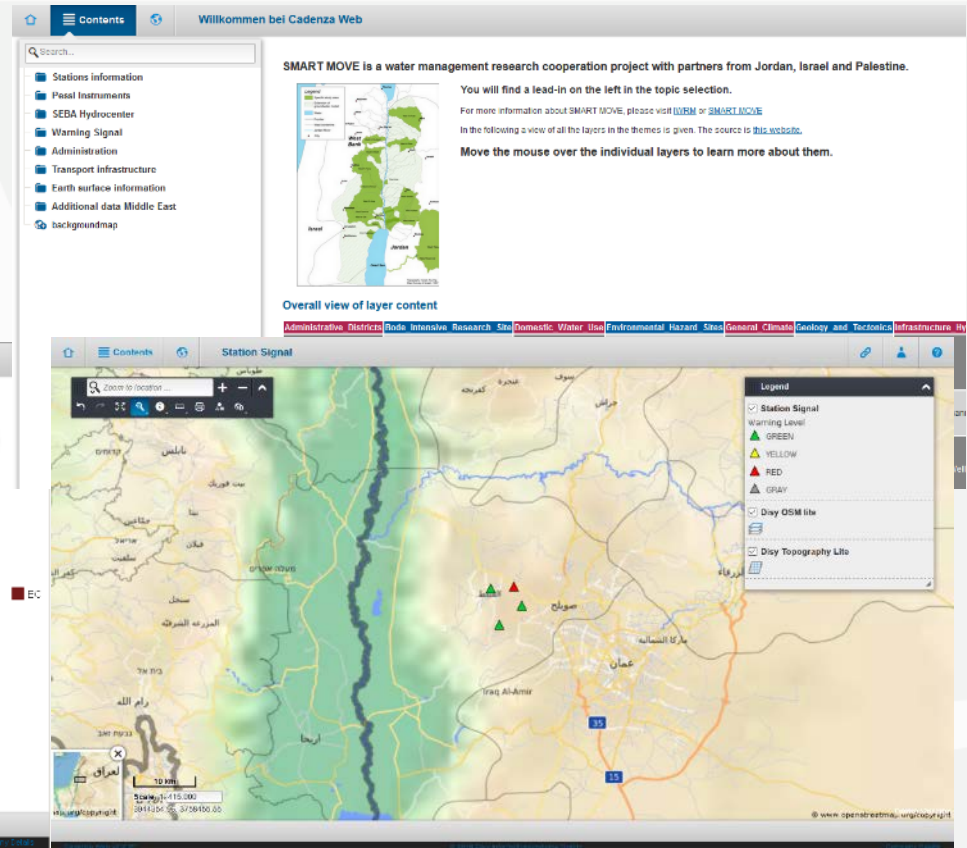
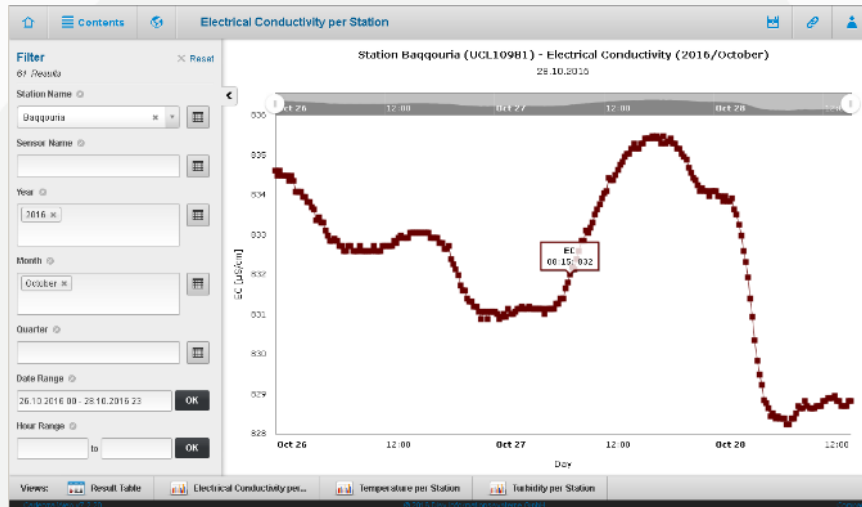


LGL: Oficina Estatal de Geoinformación y Desarrollo de Tierra en Baden-Württemberg (Wildtierportal)

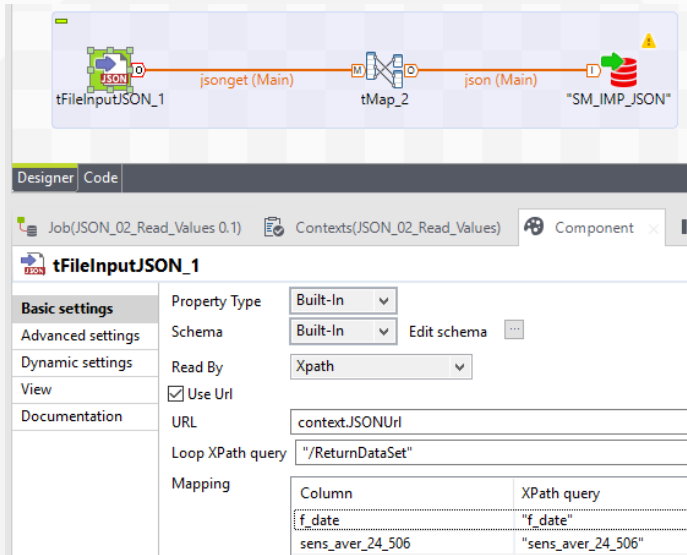
- Control de caza en la región
- Protección contra epidemias
 - Proceso automático
 - Cálculo de áreas de caza 15km radio
 - Generación de Shapefiles con áreas en peligro y posición
 - Notificación por Email a todos los responsables



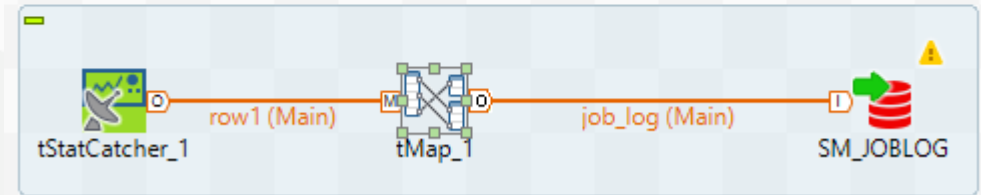
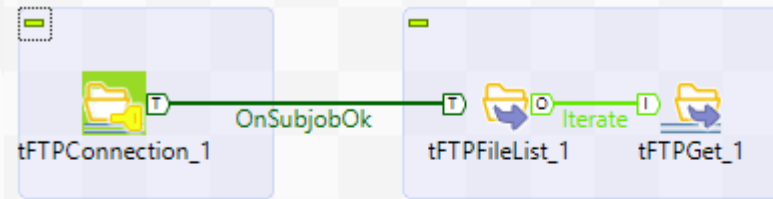
- Lectura automática mediciones (x hora)
 - Conductividad eléctrica, turbulencia, temperatura, precipitación
- Datos de entrada:JSON / FTP (csv)
- Generación señal alerta: Email



- Lectura de datos FTP (csv)
- Generación de estadísticas
- Lectura de datos interface JSON



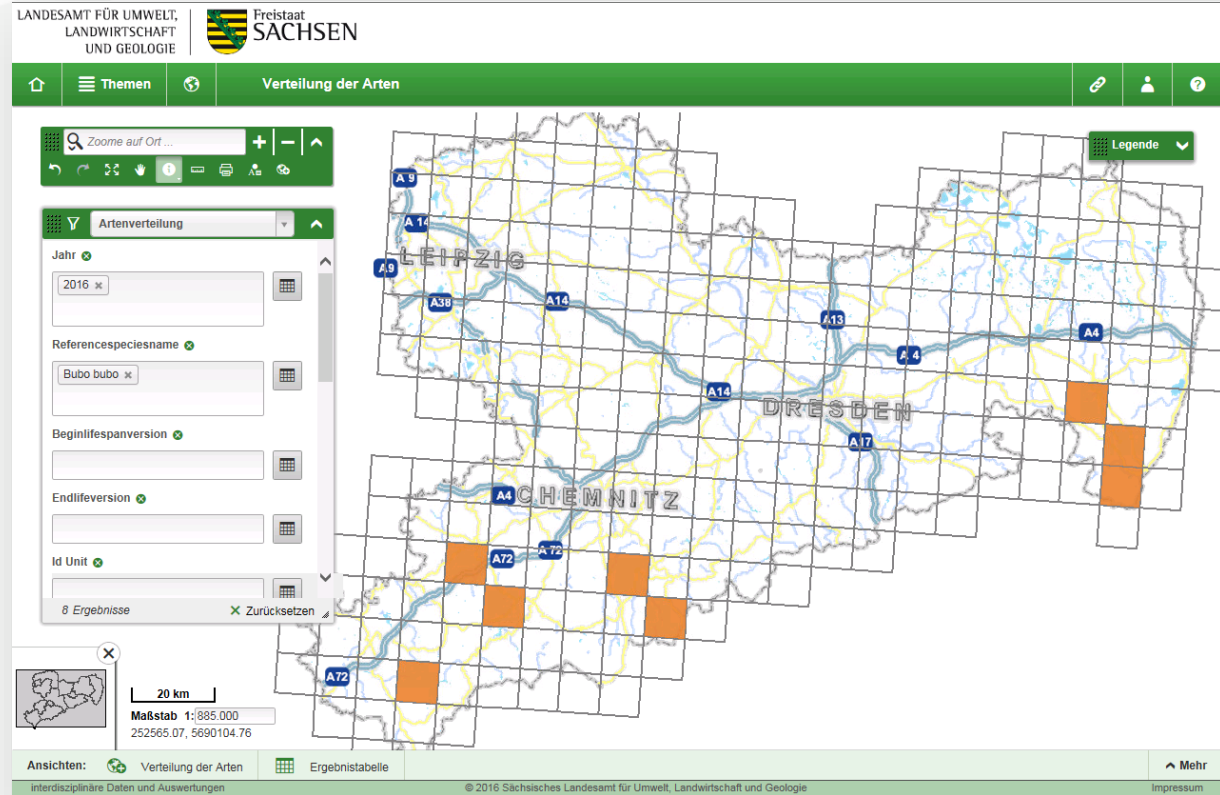
Column	XPath query
f_date	"f_date"
sens_aver_24_506	"sens_aver_24_506"



MOMENT	PID	FATHER_PID	ROOT_PID	SYSTEM_PID	PROJECT	...	JOB_REP...	JOB...	CONTEXT	...	MESS...	MESSAGE	DURATION
10-NOV-15 05.51.06	TS2ImQ	TS2ImQ	TS2ImQ	4172	SMARTMOVE	...	_fUDK4I...	0.1	Default	...	begin	(null)	(null)
10-NOV-15 05.51.08	TS2ImQ	TS2ImQ	TS2ImQ	4172	SMARTMOVE	...	_fUDK4I...	0.1	Default	...	end	success	1794
11-NOV-15 10.15.51	Fcmc03	Fcmc03	Fcmc03	10776	SMARTMOVE	...	_fUDK4I...	0.1	Default	...	end	success	4271
11-NOV-15 10.54.55	IteTaQ	IteTaQ	IteTaQ	6468	SMARTMOVE	...	_fUDK4I...	0.1	Default	...	begin	(null)	(null)
11-NOV-15 10.58.54	0iDeKy	0iDeKy	0iDeKy	6024	SMARTMOVE	...	_fUDK4I...	0.1	Default	...	begin	(null)	(null)
11-NOV-15 10.58.55	0iDeKy	0iDeKy	0iDeKy	6024	SMARTMOVE	...	_fUDK4I...	0.1	Default	...	end	success	1326

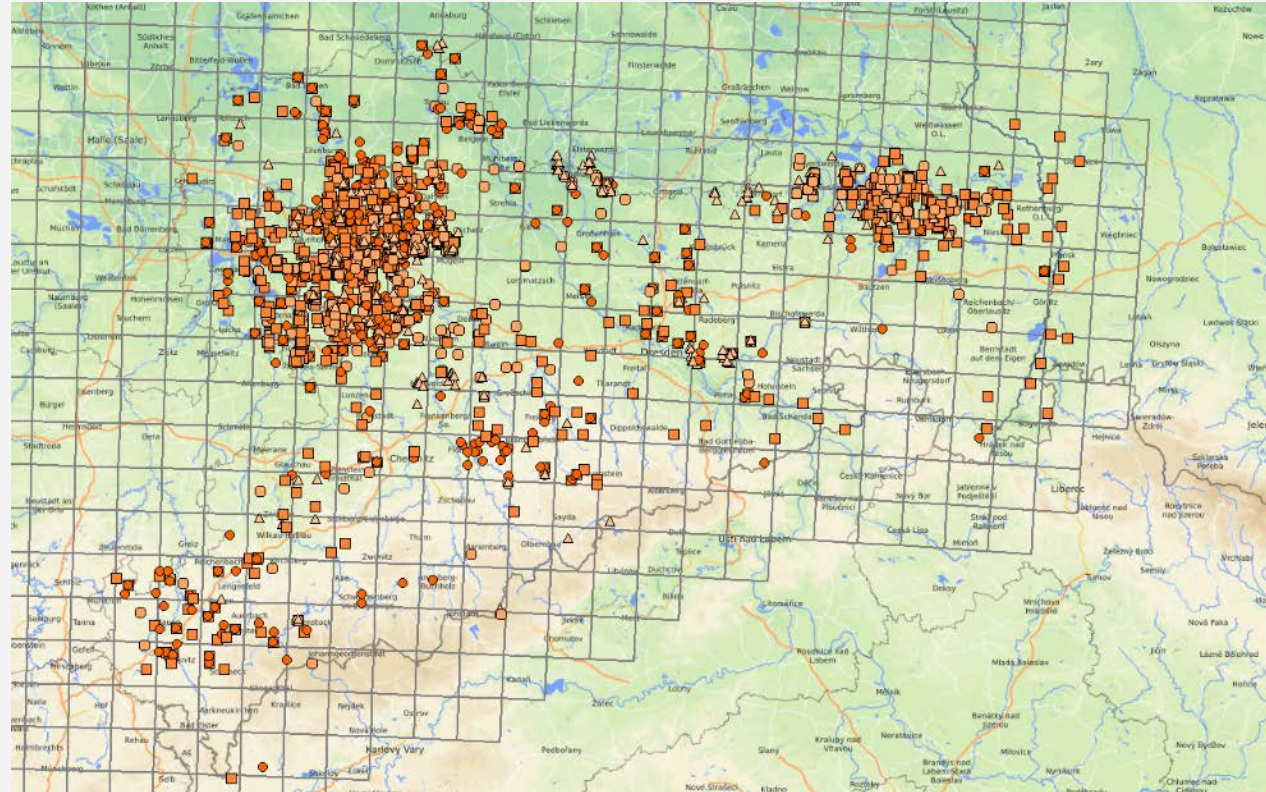
Oficina estatal del medio ambiente, agricultura y geología en Sachsen

- Proyecto de distribución de especies en la región



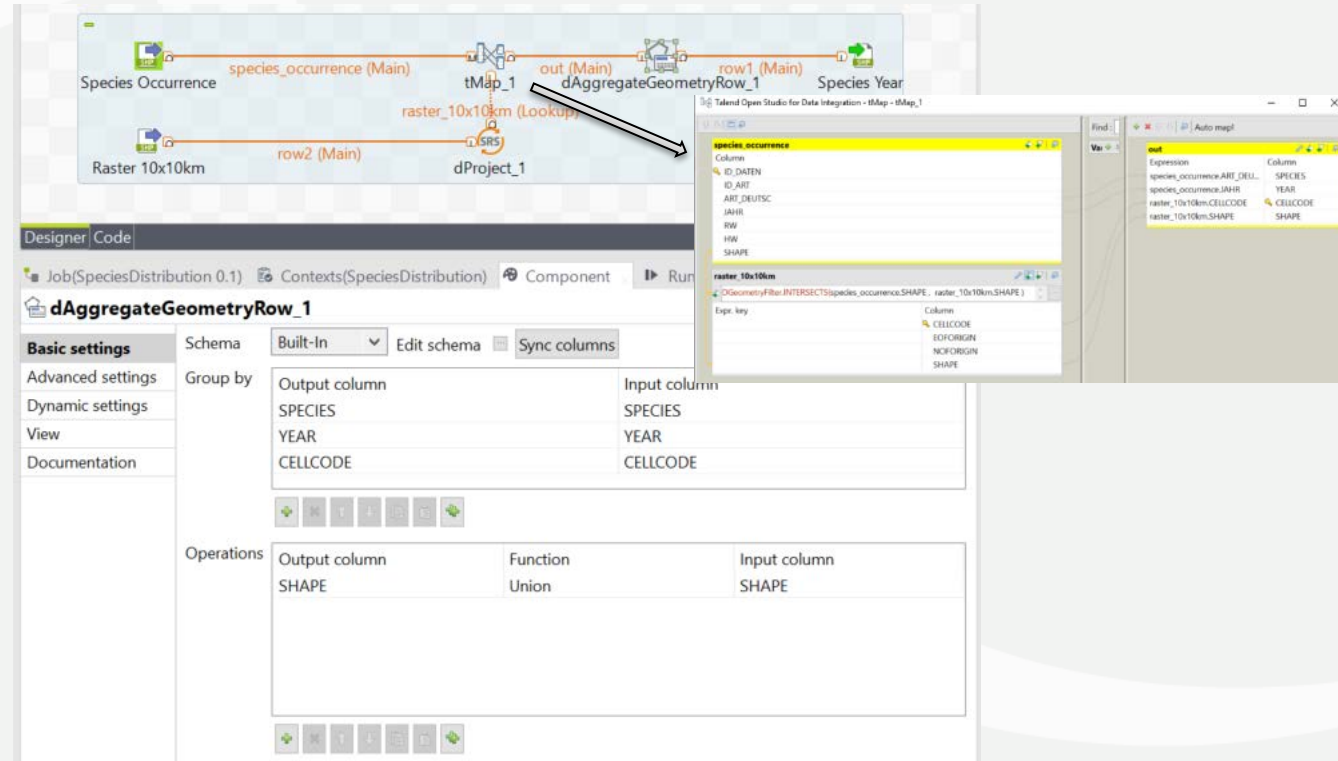
Datos de entrada:

- Raster 10x10 km de Alemania
- Ejemplo: Puntos con la distribución de especies de patos en Sachsen



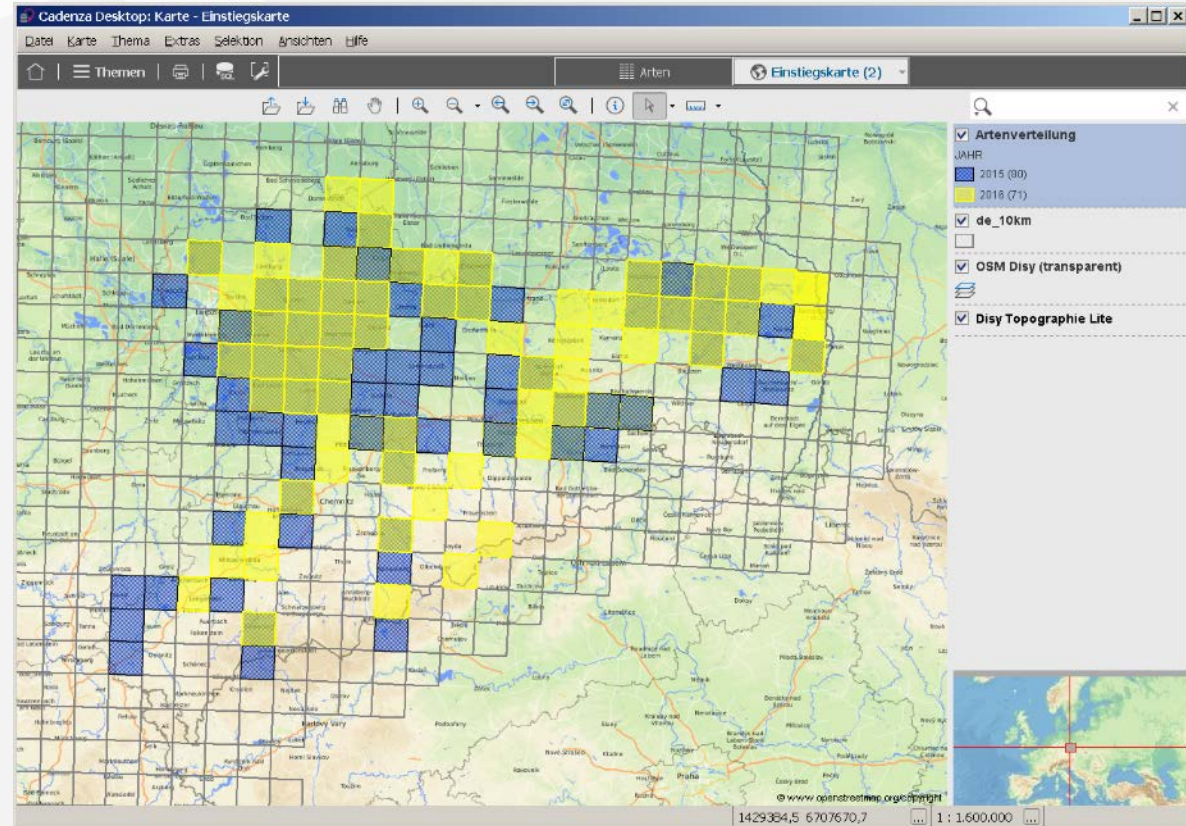
Proceso:

- Proyección sistema coordenadas
- Intersección puntos/raster
- Agrupación año/especie/raster



Resultado:

- Distribución anual de especies



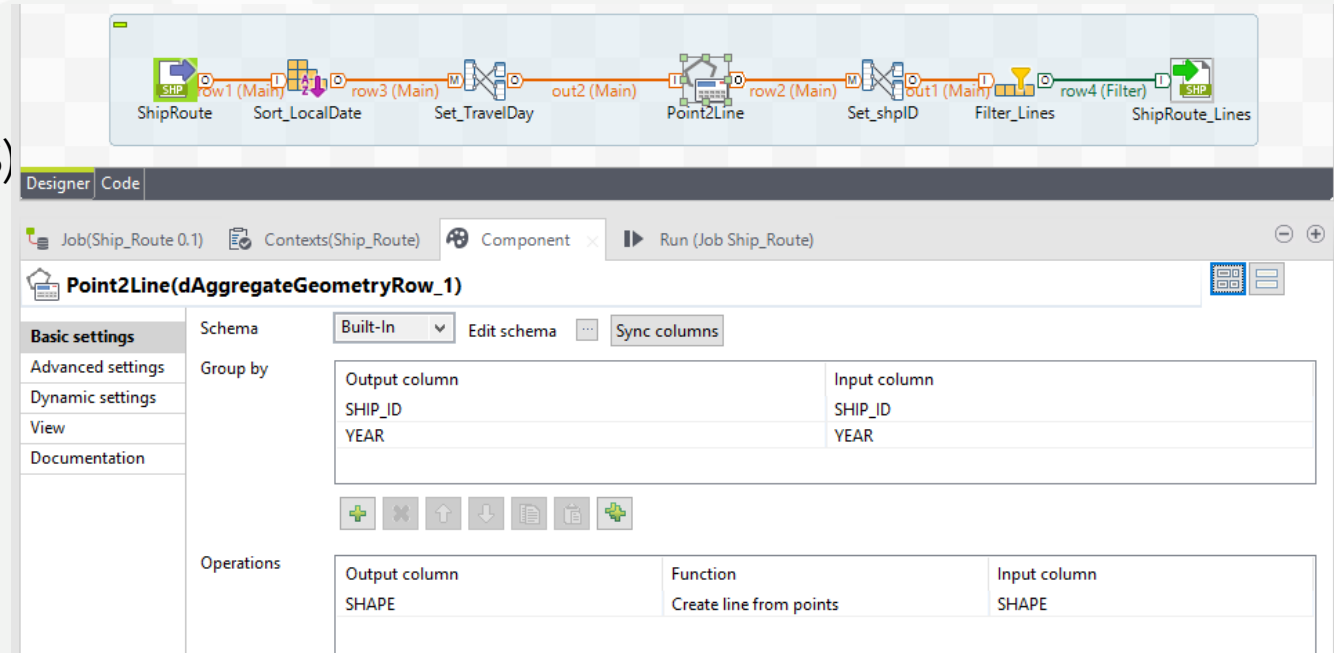


San Francisco Municipal Transportation Authority (SFMTA)

- Todos los vehículos envían posición cada minuto
 - Metro, bus, autos y bicicletas compartidas
- Talend Big Data Platform + GeoSpatial Plugin for Talend
 - Análisis de rutas
 - Visión: Servicios de datos en tiempo real para otros departamentos que brindan servicios públicos y corporativos.

Generación de rutas:

- Datos entrada:
ubicación (puntos GPS)
con hora y fecha
- Proceso: Generación
de rutas (líneas)



The screenshot shows a data processing workflow in a software interface. The workflow consists of several components connected in a sequence: ShipRoute, Sort_LocalDate, Set_TravelDay, Point2Line, Set_shipID, Filter_Lines, and ShipRoute_Lines. The 'Point2Line' component is currently selected, and its configuration is displayed in the 'Designer' tab.

Point2Line(dAggregateGeometryRow_1)

Basic settings

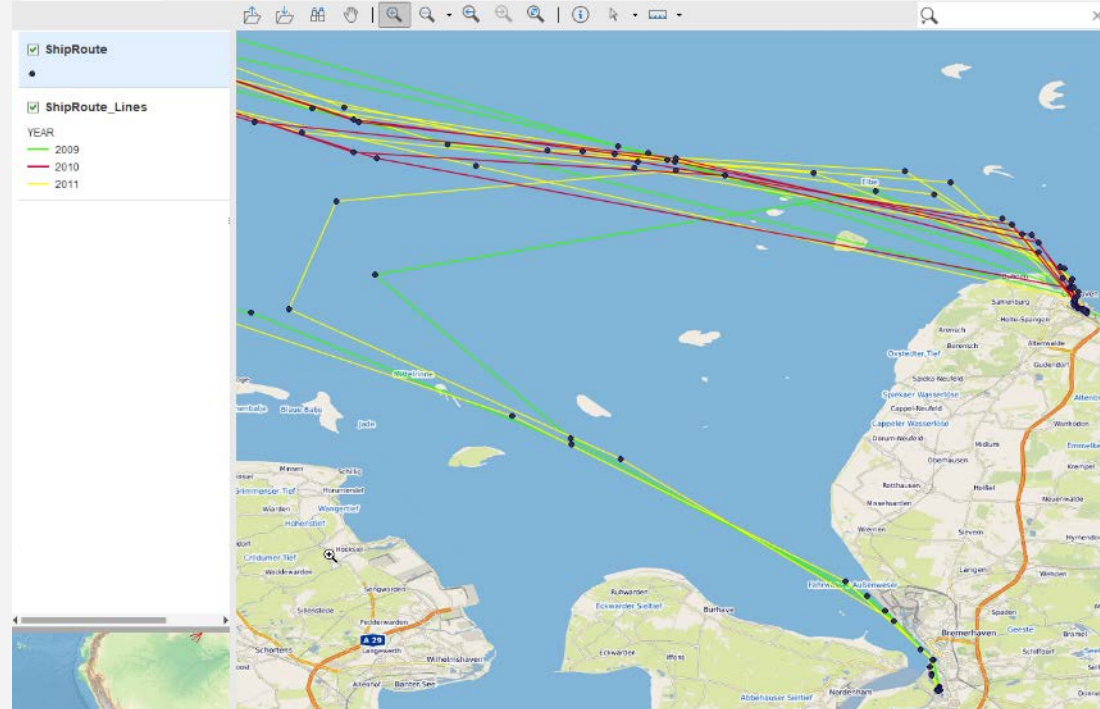
- Schema: Built-In
- Group by: SHIP_ID, YEAR
- Output column: SHIP_ID, YEAR
- Input column: SHIP_ID, YEAR

Operations

Output column	Function	Input column
SHAPE	Create line from points	SHAPE

Generación de rutas:

- Líneas clasificadas por año y barco



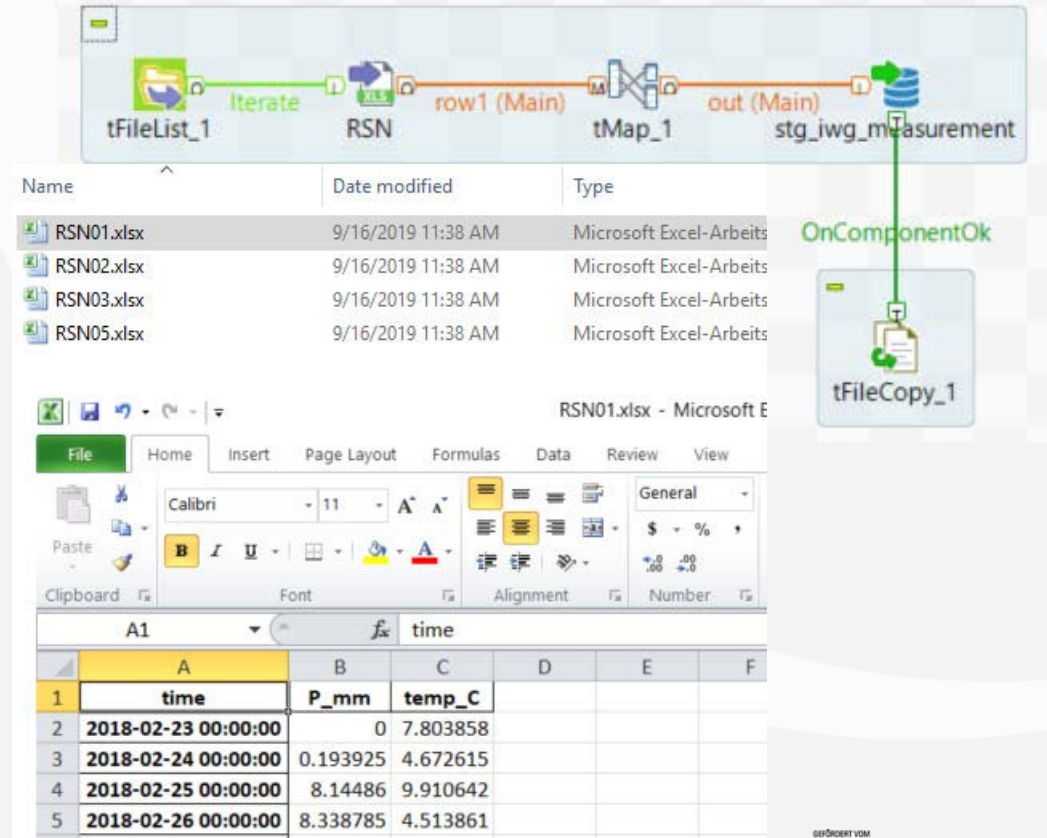
Integración de datos - Ventajas

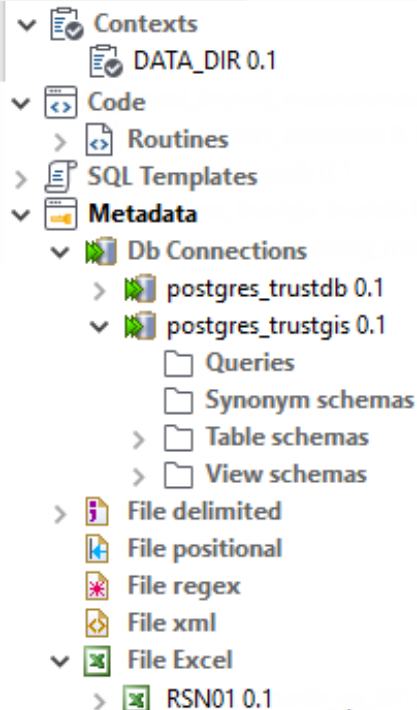
- Combinar diferentes formatos y fuentes de datos
- Calidad e integridad de los datos
- Integración de atributos y datos espaciales
- Generación de diferentes tipos de salidas
- Procesos manuales y automáticos
- Aplicaciones en tiempo real
- Disponibilidad de datos
- Soporte a la toma de decisiones

- Perfil Empresa – Soluciones
<https://www.disy.net/en/solutions/>
- Producto Geospatial Integration for Talend
<https://www.disy.net/en/products/geospatial-for-talend/>
- Proyecto SMART-Move
<http://www.iwrm-smart-move.de/index.php/project-outline>
- Proyecto Portal de Animales Salvajes (Wildtier Portal)
<https://www.disy.net/de/unternehmen/presse/presse-2018/startschuss-fuer-digitales-wildtiermonitoring/>
- Talend Open Studio
<https://www.talend.com/download/>

Proceso de carga de mediciones:

- Datos de entrada: archivos Excel con estructura pre-definida





Proceso de carga de mediciones

Paso Inicial:

- Definición contextos
- Meta-datos: Estructura archivo Excel, conexión base de datos

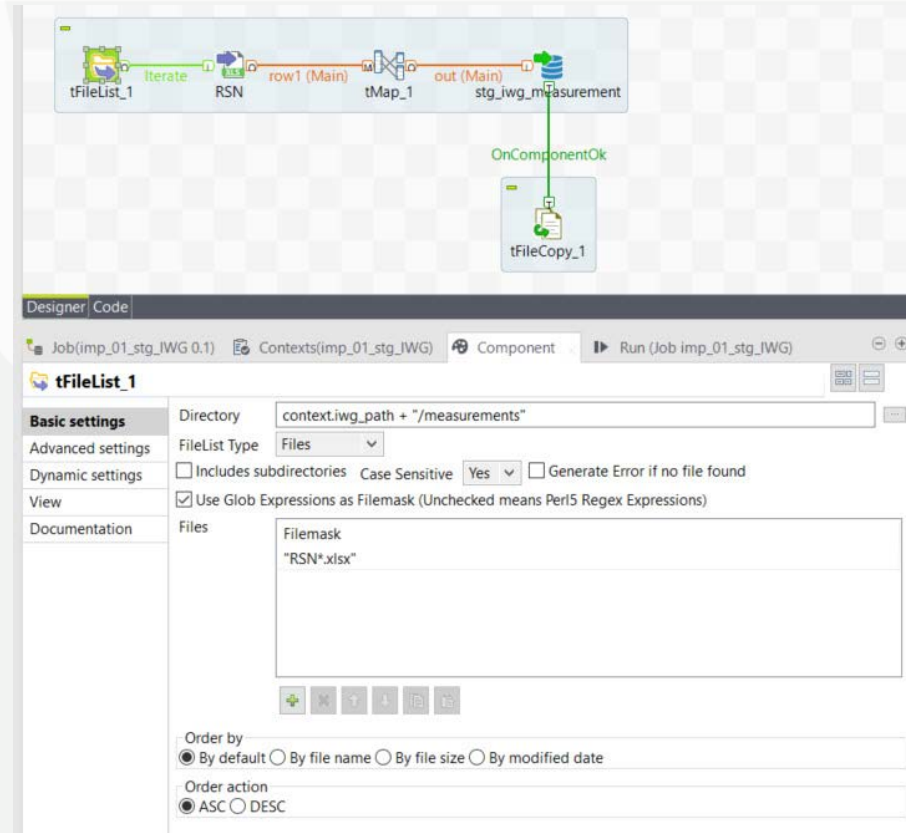
Job(imp_01_stg_IWG 0.1) Contexts(imp_01_stg_IWG) Component

	Name	Type	Comment	Value
1	DATA_DIR (from repository co			
2	iwg_path	String		"C:\Proyectos\TRUST_data\
3	postgres_trustgis_trustdb (fro			
4	postgres_trustgis_Port	String		5932
5	postgres_trustgis_Schema	String		trustgis
6	postgres_trustgis_Server	String		docker.disy.net
7	postgres_trustgis_Login	String		postgres
8	postgres_trustgis_Password	Password		*****
9	postgres_trustgis_Database	String		trustdb

Proceso de carga de mediciones

Paso 1 (tFileList):

- Indicar carpeta donde se encuentran los datos



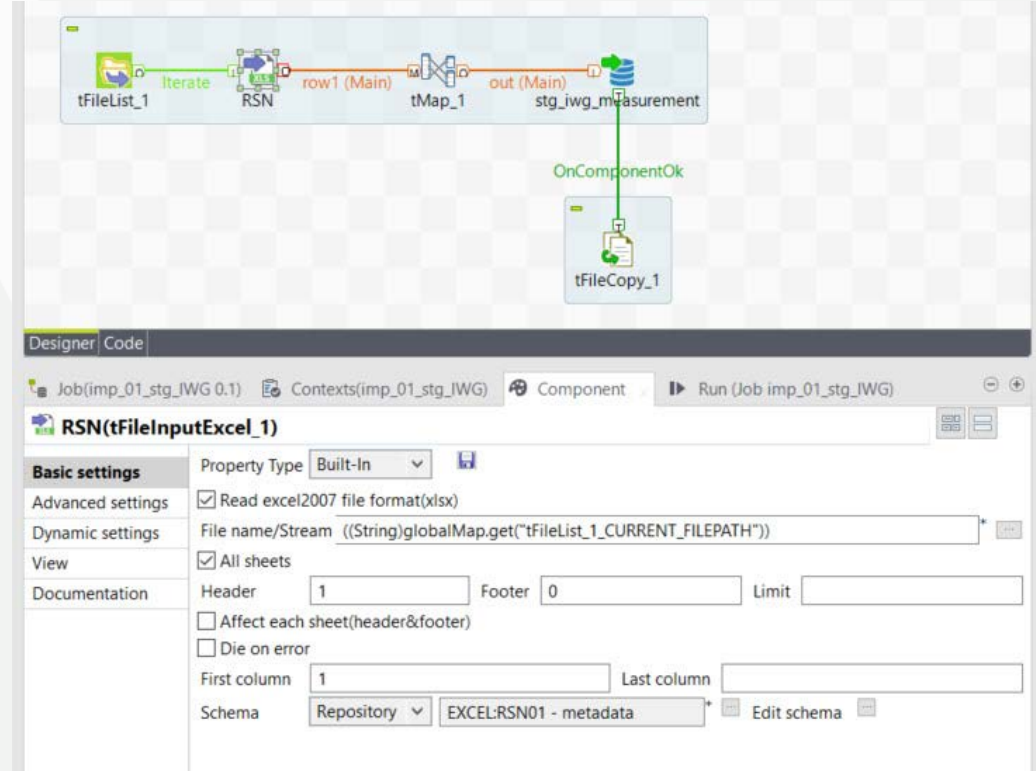
The screenshot displays the configuration for the 'tFileList_1' component. The 'Basic settings' tab is active, showing the following fields:

- Directory:** context.iwg_path + "/measurements"
- FileList Type:** Files
- Includes subdirectories:** ☐
- Case Sensitive:** Yes
- Generate Error if no file found:** ☐
- Use Glob Expressions as Filemask (Unchecked means Perl5 Regex Expressions):** ☒
- Filemask:** "RSN*.xlsx"
- Order by:** ☒ By default, ☐ By file name, ☐ By file size, ☐ By modified date
- Order action:** ☒ ASC, ☐ DESC

Proceso de carga de mediciones

Paso 2 (tFileInputExcel):

- Este componente permite la lectura de archivos excel
- La estructura del archivo Excel fue definida como parte de los metadatos del proceso



The screenshot displays the SAP DWB Designer interface. At the top, a process flow is visible with components: tFileList_1, Iterate, RSN, row1 (Main), tMap_1, out (Main), and stg_iwg_measurement. Below the flow, the configuration for the **RSN(tFileInputExcel_1)** component is shown.

Basic settings

- Property Type: Built-In
- ☒ Read excel2007 file format(xlsx)
- File name/Stream: ((String)globalMap.get("tFileList_1_CURRENT_FILEPATH"))
- ☒ All sheets
- Header: 1, Footer: 0, Limit:
- ☐ Affect each sheet(header&footer)
- ☐ Die on error
- First column: 1, Last column:
- Schema: Repository, EXCEL:RSN01 - metadata

Proceso de carga de mediciones - Paso 3 (tMap):

- Mapeo de atributos de entrada y salida

Talend Open Studio for Data Integration - tMap - tMap_1

Find Auto map!

Var

out

Expression

```

Numeric.sequence("s1",1,1)
((String)globalMap.get("tFileList_1_CURRENT_FILE"))
((String)globalMap.get("tFileList_1_CURRENT_FILE")).split(".xlsx")[0]
row1.time
row1.p_mm
row1.temp_c
    
```

Column

id
filename
station_name
meas_date
p_mm
temp_c

Schema editor Expression editor

row1

Column	K...	Type	✓ N.	Date Pattern (...)	Length	Precision	Defa...	Comme...
time	<input type="checkbox"/>	Date	<input checked="" type="checkbox"/>	"yyyy-MM-dd"		0		
p_mm	<input type="checkbox"/>	Double	<input checked="" type="checkbox"/>			17		
temp_c	<input type="checkbox"/>	Double	<input checked="" type="checkbox"/>			16		

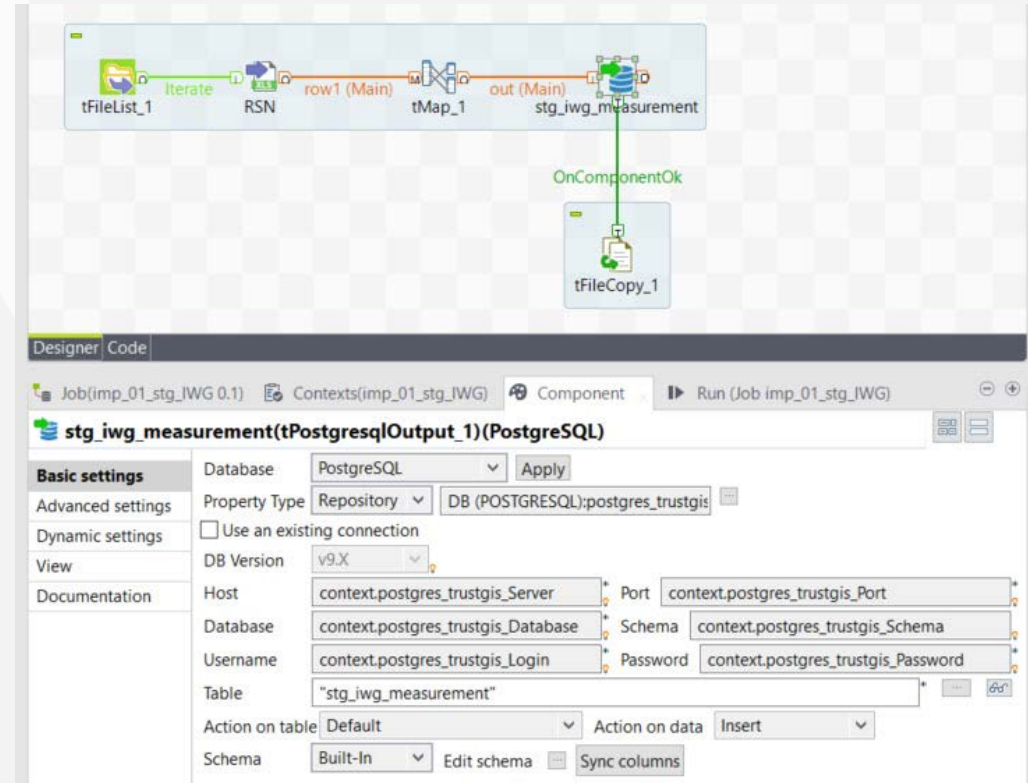
out

Column	K...	Type	✓ N.	Date Pattern (C...	Length	Precision	Defa...	Comme...
id	<input type="checkbox"/>	Integer	<input checked="" type="checkbox"/>					
filename	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>					
station_name	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>					
meas_date	<input type="checkbox"/>	Date	<input checked="" type="checkbox"/>	"yyyy-MM-dd"		0		
p_mm	<input type="checkbox"/>	Double	<input checked="" type="checkbox"/>		18	17		
temp_c	<input type="checkbox"/>	Double	<input checked="" type="checkbox"/>		17	16		

Proceso de carga de mediciones

Paso 4 (tPostgresqlOutput/tDBOutput):

- Salida a una base de datos PostgreSQL
- Se insertan los datos en la tabla destino
- La conexión de la base de datos fue definida en los meta-datos



Proceso de carga de mediciones

Paso 5 (tCopy):

- Copia los archivos cargados en otra carpeta

