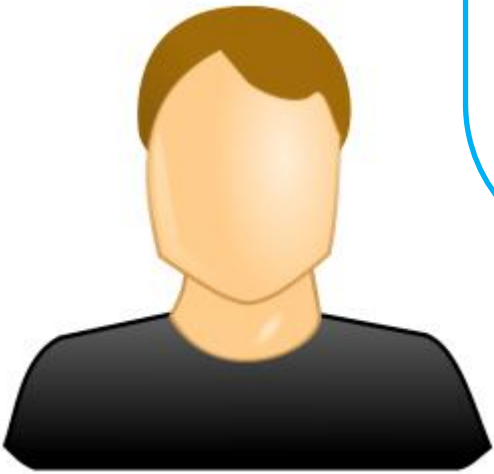


Descripción de recursos multimedia georreferenciados

Arturo Beltran, Carlos Granell, Joaquín Huerta



Motivación









BBOX





Hidrografía



World Countries



Límites Administrativos

Descripción



Hidrografía



World Countries



Límites Administrativos



BBOX:



Descripción: [Hidrografía](#)

Recurso: [Servicio](#)

Tipo: [WMS](#)

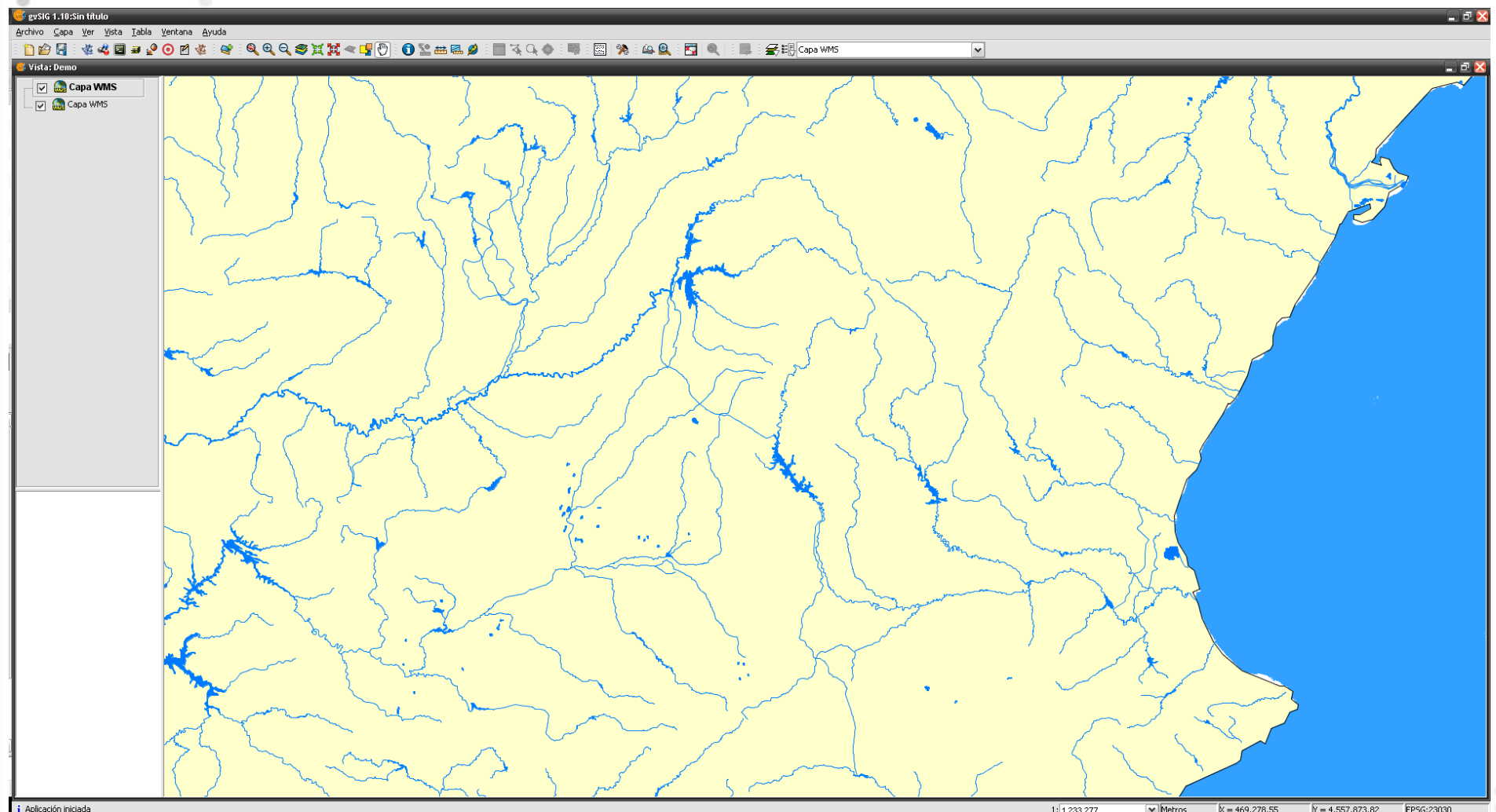
Versión: [1.3.0](#)

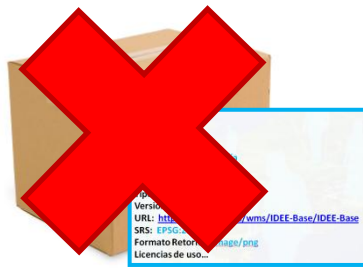
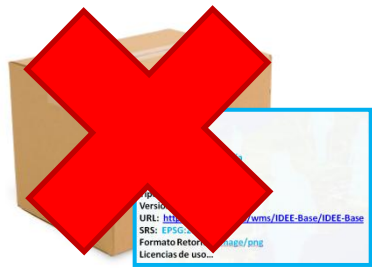
URL: <http://www.idee.es/wms/IDEE-Base/IDEE-Base>

SRS: [EPSG:23030](#)

Formato Retorno: [image/png](#)

Licencias de uso...





Objetivos

- ➔ **Facilitar descubrimiento y explotación de recursos**
 - ➔ **Organizar, publicitar y facilitar el acceso**
 - ➔ **Descripción de los recursos**
 - ➔ **Metadatos**

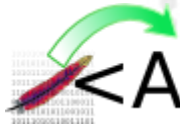
Metadatos

“Datos estructurados acerca de los datos”

“Datos que describen los atributos de un recurso”

“Información acerca de los datos”

Apache Tika



<Apache Tika/>™ (<http://lucene.apache.org/tika>)

+ Conjunto de herramientas para detectar y extraer metadatos de varios tipos de documentos

+ Texto, Imágenes, Video, Audio

+ Proyecto Apache



+ Apache Software License, v2.0

+ Java



```
c:\ Command Prompt
C:\tika-0.5\tika-app\target>
C:\tika-0.5\tika-app\target>java -jar tika-app-0.5.jar -m cartaArturoPDI_Bancaja.doc
Application-Name: Microsoft Office Word
Author: dlsi
Character Count: 710
Comments:
Company: uji
Content-Type: application/msword
Creation-Date: Wed Apr 22 12:19:00 CEST 2009
Edit-Time: 600000000
Keywords:
Last-Author: Arturo Beltran
Last-Printed: Mon Feb 23 13:50:00 CET 2009
Last-Save-Date: Mon Jan 25 10:20:00 CET 2010
Page-Count: 1
Revision-Number: 5
Template: Normal.dotm
Word-Count: 128
resourceName: cartaArturoPDI_Bancaja.doc
subject:
title: Joaquín Huerta, Profesor Titular del departamento de Lenguaje y Sistemas Informáticas y Co-
l Master en Geospacial Tecnologías de la Universitat Jaume I en Castellón, España
C:\tika-0.5\tika-app\target>
```

OSGeo FDO

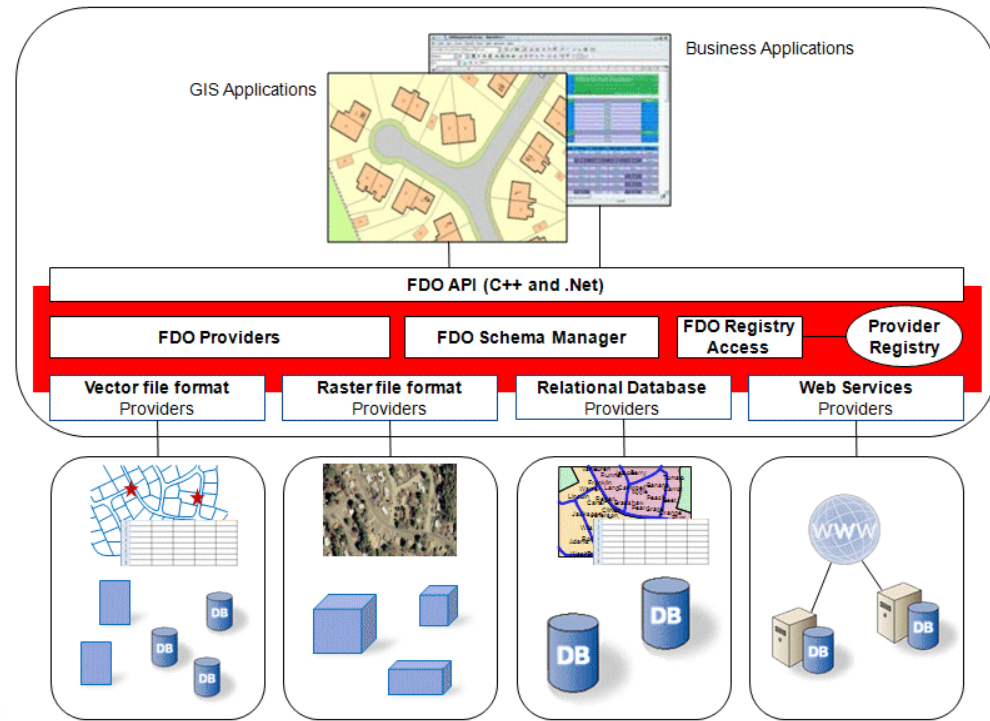
FDO Data Access Technology (<http://fdo.osgeo.org>)

- + API para manipular, definir y analizar información geográfica independientemente de donde esté almacenada
- + Soporta la mayoría de formatos de IG

+ Proyecto 

+ LGPL

+ C++ y .NET 



<Apache Tika/>



HTMLParser



EPubParser



Audio Parsers



XMLParser



Image Parsers



OpenDocumentParser



OfficeParser



Video Parsers



PDFParser



JavaParser



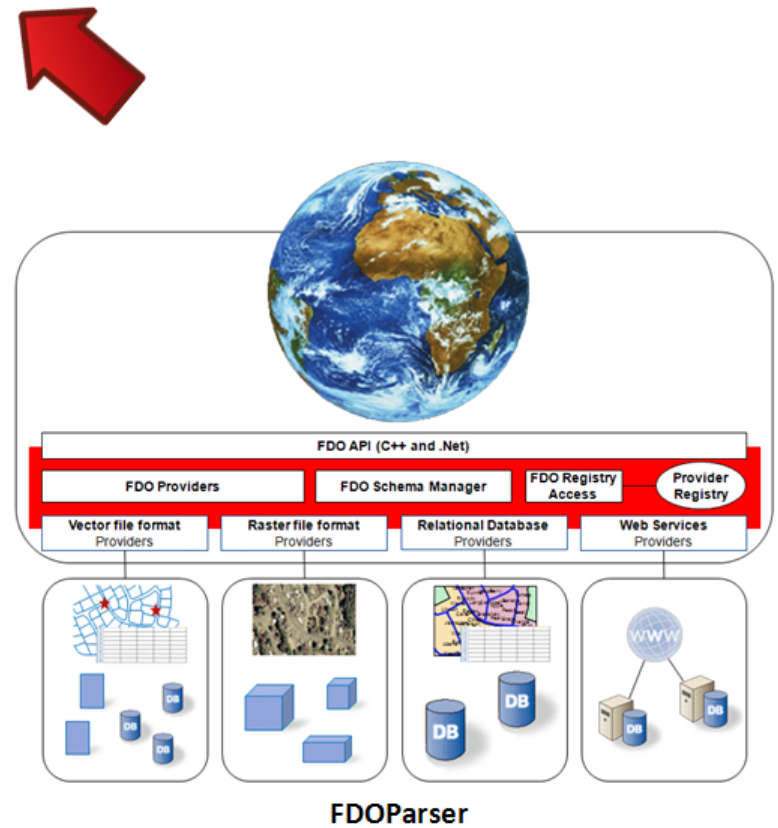
MBoxParser



Text Parsers



PackageParser



Integración Apache Tika/ OSGeo FDO

➔ Plataforma de extracción de metadatos para todo tipo de recursos

+ Arquitectura escalable de Apache Tika: Proveedores

+ Incorporar nuevo proveedor basado en OSGeo FDO

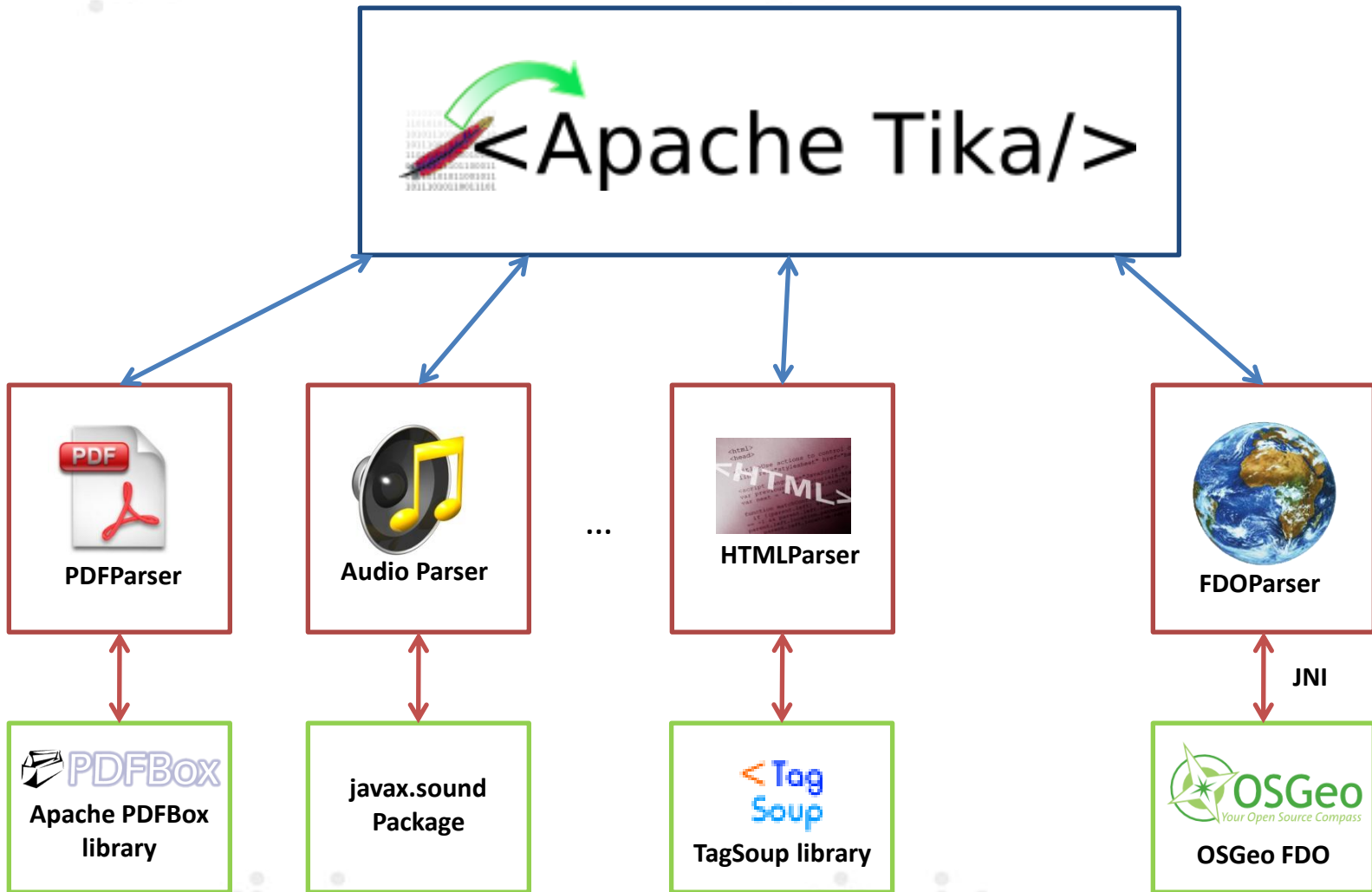
+ Apache Tika: **Java**

+ API de OSGeo FDO: **C++ y .NET**

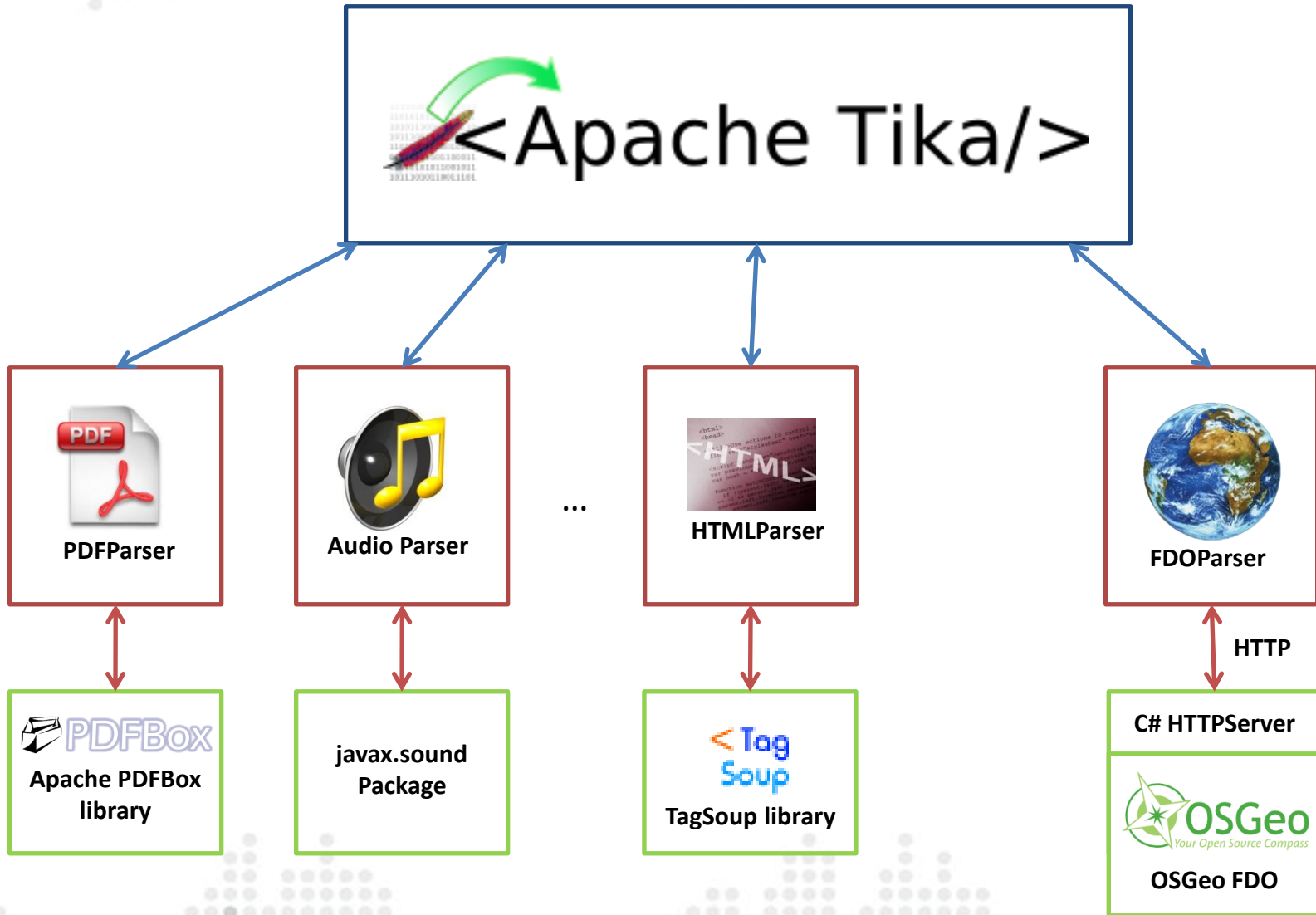


➔ Java Native Interface (**JNI**)
Simplified Wrapper and Interface Generator (**SWIG**)

Integración Apache Tika/ OSGeo FDO




Integración Apache Tika/ OSGeo FDO




Resultados

➔ Plataforma común de extracción de metadatos

- + Soporte para más de 200 formatos
 - Texto, Imágenes, Audio, Video (50)
 - Formatos geoespaciales (150)
- + Información contextual preconfigurada
- + Cantidad de información dependiente del tipo de recurso



Author: Test Artist
Content-Length: 39416
Content-Type: audio/mpeg
channels: 2
resourceName: testMP3id3v1.mp3
samplerate: 44100
title: Test Title
version: MPEG 3 Layer III Version 1
xmpDM:album: Test Album
xmpDM:artist: Test Artist
xmpDM:audioSampleRate: 44100
xmpDM:genre: Rock
xmpDM:logComment: Test Comment
xmpDM:releaseDate: 2008
xmpDM:trackNumber: 1



Application-Name: Microsoft Office Word
Application-Version: 12.0000
Author: Name Surname
Character Count: 57
Character-Count-With-Spaces: 66
Content-Length: 10189
Content-Type: application/vnd.openxmlformats-officedocument.wordprocessingml.document
Last-Modified: 2010-06-25T11:12:00Z
Last-Printed: 2010-06-25T11:11:58Z
Line-Count: 3
Page-Count: 1
Paragraph-Count: 1
Revision-Number: 2
Template: Normal.dotm
Total-Time: 1
Word-Count: 10
creator: Name Surname
date: 2010-06-25T11:11:58Z
publisher: Name Surname
resourceName: testWORD.docx
title: Sample Word Document

Resultados

Etiqueta	Significado
Resource	Inicia la descripción de un recurso
FormatName	Formato en el que se encuentra el recurso analizado
ResourceType	Tipo de recurso
Provider	Proveedor de FDO utilizado para analizar el recurso
Source	Origen de los datos
ResourceName	Nombre del recurso
ConnectionString	Cadena de conexión al recurso
dateStamp	Fecha de la creación de los metadatos
keywords	Palabras clave en la descripción del recurso
SpatialContexts	Descripción de los contextos espaciales del recurso. Incluyendo nombre, sistema de coordenadas, extensión...
Schemas	Descripción de los esquemas de datos del recurso. Incluyendo nombre, atributos, <i>features</i> ...
SchemaAttributes	Descripción de los atributos del esquema. Incluyendo nombre y valor de los mismos.
FeatureClasses	Descripción de los <i>features</i> del recurso. Incluyendo nombre, restricciones, sus propiedades...
BaseIdentityProperties y Properties	Descripción de las propiedades de cada <i>feature</i> . Incluyendo nombre, tipo y diferente información dependiente del tipo, como tamaño, valor por defecto, precisión...

Conclusiones

➔ Integración Apache Tika/OSGeo FDO

- + Plataforma común de extracción de metadatos
- + Soporta gran cantidad de tipos de recursos
- + Proporciona gran cantidad de información
- + Inicio para la descripción automática de recursos

➔ Difícil conseguir un sistema totalmente autónomo

- + Usuarios: Validar y completar metadatos
- + Campos poco intuitivos
- + Facilitar la labor de los creadores ➔ evitar errores
- + Esencial impulsar la investigación



¿Preguntas?

Muchas gracias

