



# Comunicación

# 070

## **EL USO UNIVERSAL DEL FORMATO PDF PARA ALMACENAMIENTO ELECTRÓNICO DE DOCUMENTOS: EL GENERADOR DE PDFS DEL SISTEMA UNIFICADO DE RECURSOS (SUR) DE LA JUNTA DE ANDALUCÍA.**

### **Francisco Silveira García**

Jefe de Proyecto  
Dirección General de Sistemas de Información Económica-Financiera  
Consejería de Economía y Hacienda  
Junta de Andalucía

### **Rosario González Gómez**

Jefa del Servicio de Ingresos  
Dirección General de Sistemas de Información Económica-Financiera  
Consejería de Economía y Hacienda  
Junta de Andalucía

### **Fernando Rendón Quiles**

Analista  
SADIEL

## **Palabras clave**

*Generador de documentos, formato PDF, casilla-valor.*

## **Resumen de su Comunicación**

*El formato PDF (Portable Document Format) ha sido recientemente refrendado por la organización internacional de normalización ISO como el adecuado para almacenamiento electrónico de información.*

*La comunicación presenta el camino seguido por la documentación tratada por el Sistema de Gestión de Ingresos de la Junta de Andalucía (SUR), describiendo la herramienta empleada actualmente: un generador de PDFs basado en XML, que permite la creación, , almacenamiento, visualización e intercambio de todo tipo de documentos en dicho formato.*

---

# EL USO UNIVERSAL DEL FORMATO PDF PARA ALMACENAMIENTO ELECTRÓNICO DE DOCUMENTOS: EL GENERADOR DE PDFS DEL SISTEMA UNIFICADO DE RECURSOS (SUR) DE LA JUNTA DE ANDALUCÍA.

## 1. Introducción

El formato PDF (Portable Document Format) se ha convertido durante estos últimos años en un standard de facto para intercambio y archivo de documentos electrónicos. La reciente aprobación el año pasado de la norma ISO 19005-1:2005 'Document management- Electronic document file format for long-term preservation- Part1, Use of PDF 1.4 (PDF/A-1)' refrenda esta situación, y permite dirigir sin temor los esfuerzos de las organizaciones a unificar en esa dirección los formatos de documentación utilizados.

El Sistema Unificado de Recursos (SUR) es el sistema integrado de gestión de ingresos de la Junta de Andalucía. Está desarrollado y mantenido por la Dirección General de Sistemas de Información Económica-Financiera (DGSIEF) de la Consejería de Economía y Hacienda de la Junta de Andalucía, y está funcionando desde el año 1998.

SUR gestiona documentos procedentes de la aplicación cliente-servidor desarrollada con FORMS de Oracle y de la aplicación WEB ('Oficina Virtual') desarrollada con JAVA.

La aplicación tiene un volumen de unos 1.000 usuarios concurrentes, que han construido unos 600.000 (598.952) documentos el pasado año 2005, lo que supone una media (grosera) de 2.000 documentos al día.

En la actualidad, se manejan documentos PDF en la Oficina Virtual, estando en proceso de cambio la parte cliente-servidor, para convertir los documentos REPORT en PDFS.

## 2. Antecedentes

Toda la documentación de SUR responde desde el inicio a un esquema 'casilla-valor', es decir, la definición de unos formatos ('modelos') estructurados en 'casillas' que se convierten en 'documentos' cuando se les asigna 'valor'.

### 2.1. Documentos WORD

Cuando se implanta SUR en el año 1998, años en que triunfaba el paradigma cliente-servidor con ofimática radicada en la estación de trabajo, el modelo de creación de documentos estaba basado en el empleo de una serie de plantillas de Microsoft WORD, con creación de los correspondientes documentos en WORD. La deconstrucción de los citados documentos en componentes de la plantilla se almacenaba en un sistema de tablas de Oracle.

Esto planteaba, entre otros problemas, el continuo cambio del lenguaje de macros de las sucesivas versiones de WORD, lo que obligaba a una continua redefinición de las plantillas.

---

## 2.2. Documentos REPORT

Con tal motivo y otras muchas consideraciones que aconsejaban desligar la ofimática local de la aplicación, en el año 2002 se abordó la creación de un sistema de generación de documentos basado en plantillas propias basadas en una arquitectura de tablas de Oracle especialmente diseñada al efecto. El resultado fue a partir de entonces un documento REPORT, almacenando su estructura en el citado sistema de tablas.

## 2.3. Documentos PDF

Con el desarrollo de SUR para Internet ('Oficina Virtual'), los condicionantes del entorno empujaban a utilizar el formato PDF, y se estudiaron diferentes alternativas, de entre las cuales se eligió reutilizar el sistema de creación de documentos, produciendo esta vez como resultado un documento PDF que se almacena como BLOB en la base de datos.

A la vista del éxito del sistema, está a punto de implantarse la misma solución para la parte de SUR que funciona en Cliente/Servidor, que como comentamos previamente, todavía funciona con REPORTs.

## 3. Estructura del generador de documentos PDF

El sistema de generación de documentos en formato PDF se utiliza para la generación de documentos de "entrada", (autoliquidaciones, declaraciones, etc), realizada por los usuarios de la Oficina Virtual Tributaria y de Recaudación de la Consejería de Economía y Hacienda, así como para la generación de documentos de "salida" ( liquidaciones, certificados, etc.) generados en la Oficina Virtual.

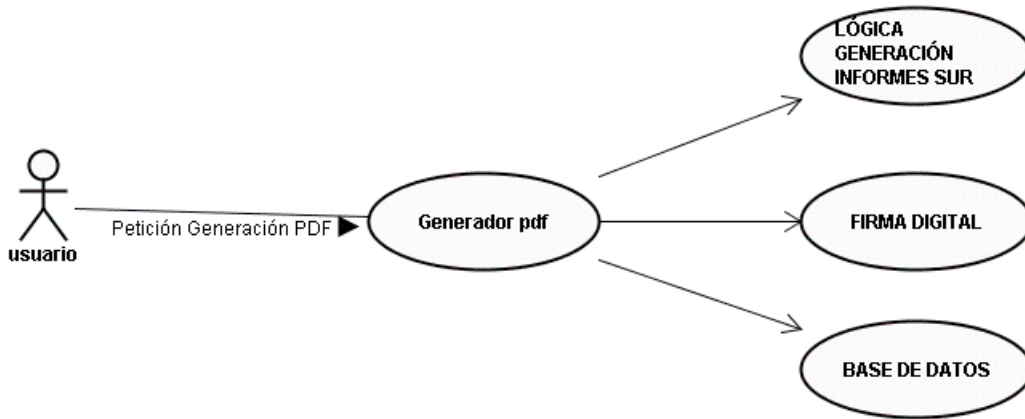
Actualmente existen dos tipos de generadores de documentos PDFs dependiendo del tipo de documento que se esté generando.

### 3.1. Documentos de salida

La tecnología empleada para la generación de este tipo de documento, es Java y XSL Formatting Objects (XSL-FO).

Este desarrollo reutiliza toda la infraestructura de la generación de documentos desarrollado en SUR cliente/servidor, utilizando las características de diseño aplicadas en el mismo.

Aparte de crear documentos PDFs individuales, el sistema es capaz de procesar lotes de documentos concatenándolos.



Caso de uso del subsistema

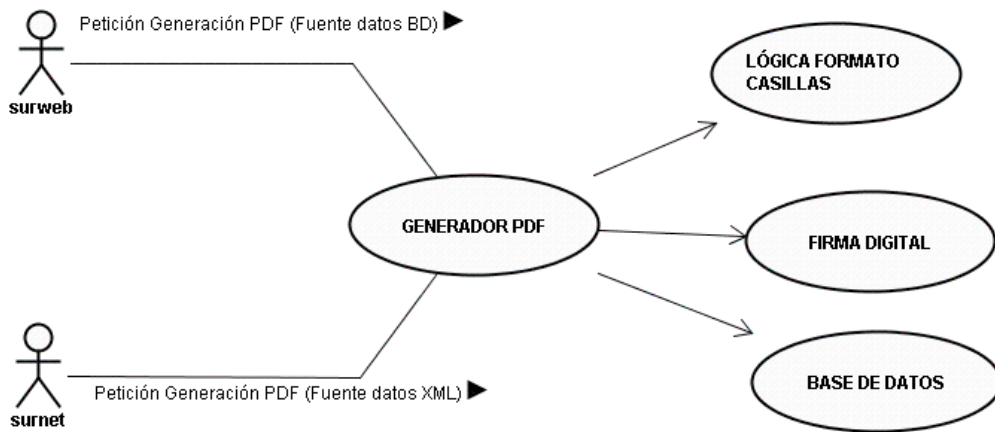
- Una vez generado el documento por el sistema con la lógica de negocio apropiada en cada caso, se inicia la generación de documentos.
- Llamada al punto de entrada del generador de PDFs de documentos de salida con la terna que identifica el documento (modelo, versión, número) y las distintas opciones de generación (Marca de agua, 'No se genere el elemento de firma', 'No se genere un ejemplar', etc. ) si se desea alterar las de defecto.
- Una vez determinado qué documento se genera se invoca a la lógica de generación de documentos que existe en SUR (paquete Oracle KGENREP), para la alimentación de las tablas temporales.
- Lectura de las tablas temporales que previamente han sido cargadas, para la confección de un documento siguiendo la especificación "XSL:FO".
- Transformación del documento a formato PDF.
- Firma electrónica del documento por la Consejería de Economía y Hacienda.
- Actualización de los indicadores de impresión del documento en la Base de Datos.
- Presentación del documento en formato PDF al usuario.

### 3.2. Documentos de entrada

Se emplea tecnología Java utilizando el framework de fuente abierta iText, conjuntamente con documentos en formato XML que definen el diseño del documento en formato PDF.

El diseño de los documentos se basa en el diseño aprobado por los usuarios expertos de SUR.

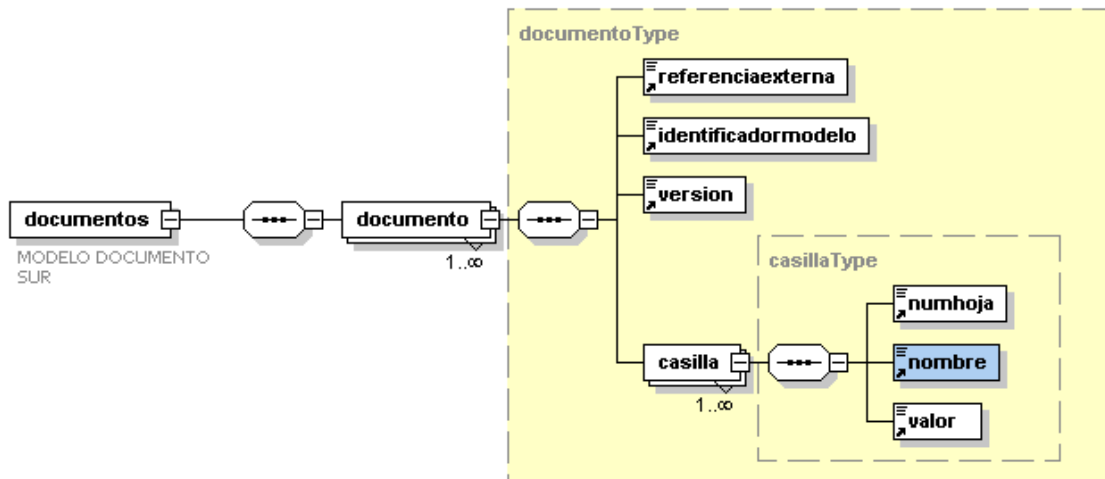
Asimismo, el sistema es capaz de generar lotes de documentos, con la única limitación que todos sean del mismo modelo y versión.



Caso de uso del subsistema

- Petición de generación, con los parámetros adecuados, obtenidos de base de datos, o bien fichero XML (ver anexo 1).  
Una vez que el sistema ha recuperado la información respecto al modelo, ya sea consultando una tabla de base de datos o leyendo un fichero XML, se procede con el siguiente paso.
- Clasificación de las casillas del modelo en una estructura jerárquica (ver Anexo 2) para su posterior manipulación.
- Llamada al procedimiento de bases de datos para formatear el contenido de las casillas.
- Lectura del fichero XML que describe el modelo, tanto en el ámbito del diseño como situación de casillas (ver Anexo 3).
- Generación del documento en formato PDF utilizando, para ello el framework iText, siguiendo las directivas que se encuentran en el documento XML.
- Firma del documento generado, si procede

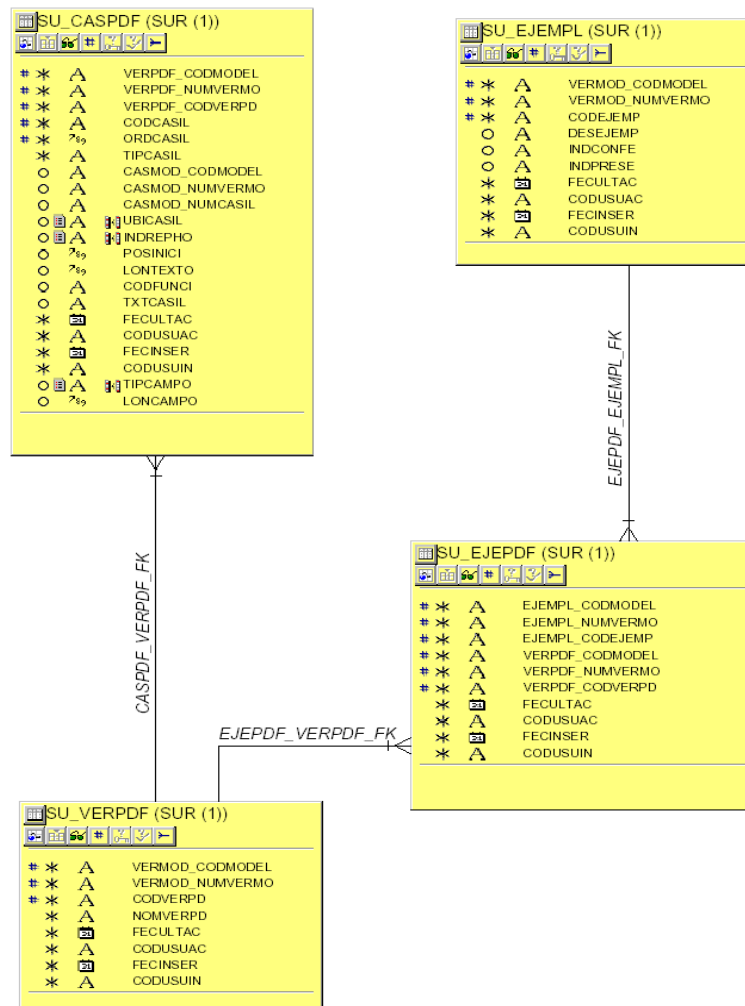
## Anexo 1. Fichero XML que recibe el generador de Pdfs (fuente de datos)



Este fichero puede contener un conjunto de documentos con la única limitación que deben de ser todos los documentos de la misma versión y modelo, para optimizar el tamaño de los PDFs.

## Anexo 2. Estructura de datos del documento PDF.

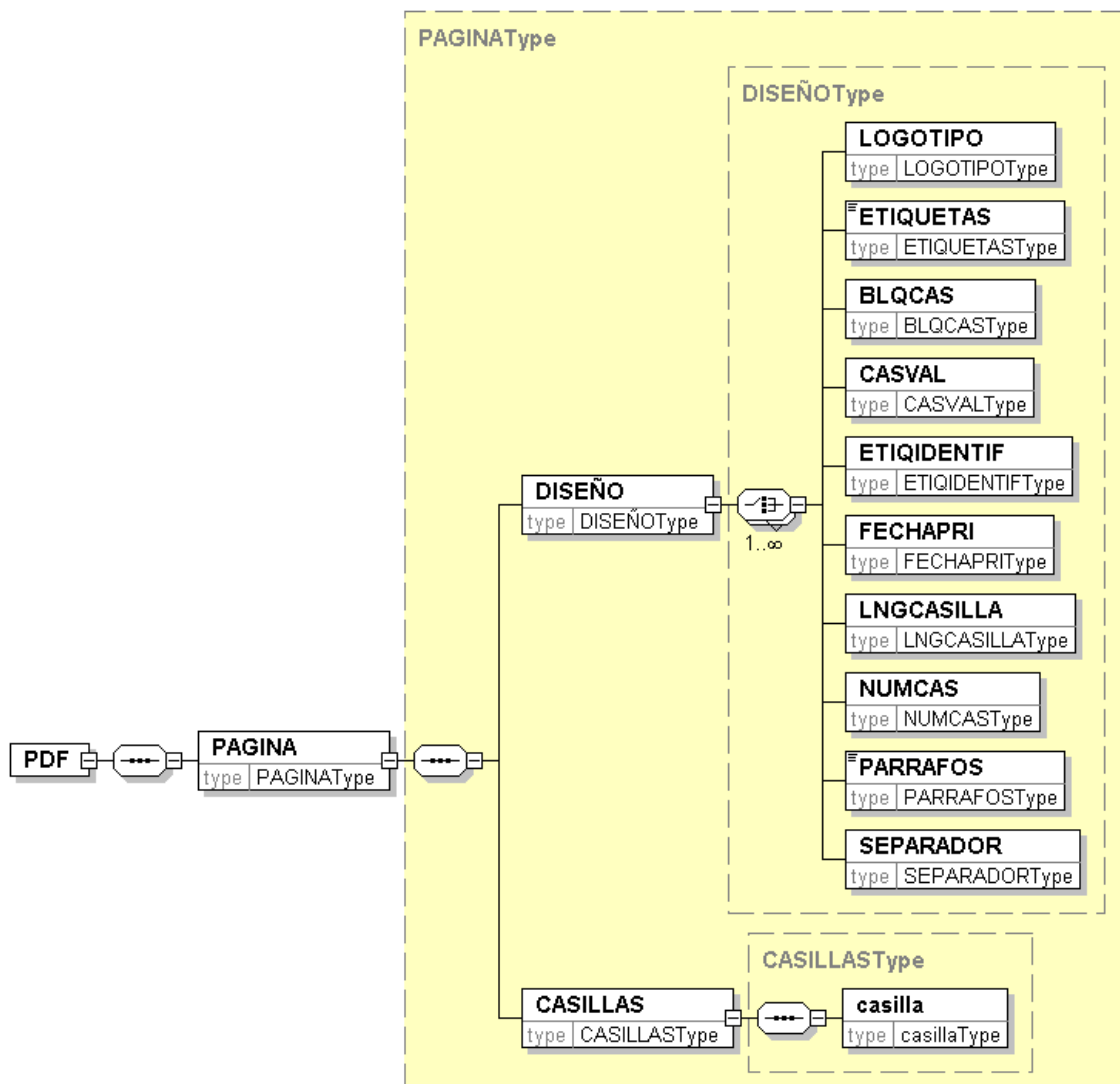
La estructura de base de datos en la cual se sustenta el generador de PDFs:



- SU\_EJEMPL: Número de ejemplares que contiene un modelo.
- SU\_EJEPDF: Relaciones de paginas físicas que contiene un ejemplar.
- SU\_VERPDF: Relaciona paginas físicas con hojas.
- SU\_CASPDF: Relaciona el conjunto de casillas de las hojas con el conjunto de casillas que define un modelo, como información de formato.



### Anexo 3. Fichero XML para la generación de Pdfs (parte gráfica)



Cada página distinta de un documento está identificada por un secuencial.

A su vez cada página consta de dos secciones:

- Diseño, que contiene la parte gráfica de la generación del documento, y
- Casillas, que contiene la identificación de la casilla y situación dentro de la página.