

IX

El área de archivo en el proyecto de digitalización de ETB

Jesús Andérez

Jefe del Departamento de Documentación y Archivo de ETB
ETBko Dokumentazio eta Artxibo Saileko Burua

Laburpena: Labur esanda, teknologia digitalek dokumentazioan eta telebista saioen gordeketan dakartzaten aldaketen inguruko ideia orokorra emango du aurkezpenak. Digitalizazio proiektuaren aldien ikuspegi orokorra emango da, zein den egungo egoera azalduz. Prozesu digital integratu bateko lan-fluxuetan dauden aldaketak eta aurre egin beharreko zailtasunak osoko katalogazio eta kudeaketarako tresna berrien adibide batzuk ere emango dira.

Resumen: La presentación tratará de ofrecer una idea general de los cambios que las tecnologías digitales suponen para la documentación y la conservación de programas de televisión. Se dará una visión general de las etapas del proyecto de digitalización, situación actual, cambios en los flujos de trabajo en un proceso digital integrado y dificultades a encarar. Se mostrarán también ejemplos de nuevas herramientas para la catalogación y gestión integral de los contenidos audiovisuales.

1. INTRODUCCIÓN

La situación actual del archivo de ETB es de transición, desde un archivo basado en la cinta de vídeo a otro basado en la infraestructura informática, prácticamente la que hemos vivido desde 2004 a 2008. Continuamos manejando los soportes tradicionales (cintas de vídeo) y también los ficheros digitales que se crean a diario en el nuevo entorno de producción: por ejemplo, las piezas que se hacen para los programas informativos, los programas propios o incluso los programas ajenos. En muchos casos disponemos de un archivo «duplicado» que incluye tanto cintas de vídeo como ficheros digitales.

La labor del archivo continúa siendo la misma: conservamos los programas que se emiten en nuestras cadenas y ofrecemos soporte documental para elaborar nuevos programas, a veces con viejas imágenes. Lo que intentamos ahora es atender y poner en valor ese contenido desde tres vertientes: su valor de producción (el valor que pueda tener un contenido para ser nuevamente utilizado), su valor comercial y su valor como bien cultural, en cuanto dirigido al público en general.

La cronología del proceso de cambio se resume de la siguiente manera. En julio de 2004 comenzamos el archivo digital de las noticias de informativos. Continuando con el proceso normal de emisión, se añade un sistema de conservación en discos que permite conservar las piezas de los informativos. Además, hemos digitalizado en parte el archivo histórico de programas informativos: las cintas de vídeo de los programas inmediatamente anteriores a 2004 se ingestan, se digitalizan y se trocean pieza a pieza. Cada una de estas piezas se liga a la información histórica de que dispone el archivo, creando un nuevo contenido digital que incluye esencia (vídeo + audio) e información relacionada (metadatos). Adicionalmente, hemos desarrollado herramientas para la catalogación en entorno digital. Más adelante mostraré cómo están estas herramientas en la actualidad.

El punto de inflexión del cambio hacia el archivo digital se ha producido en septiembre de 2007, coincidiendo con el traslado del centro de emisión de Iurreta a Bilbao. En la actualidad, estamos emitiendo en digital en cinco cadenas y disponemos de un sistema de gestión de contenidos (MAM) corporativo y centralizado. Estamos ya produciendo y archivando programas informativos, conservando programas propios y conservando programas ajenos. El área de documentación y archivo se sitúa así en consonancia con la línea natural de producción: se comienza con la digitalización de los informativos diarios, para continuar con el resto de los programas propios y, finalmente, con los programas ajenos.

2. FONDOS DEL ARCHIVO

¿Qué almacenamos en el archivo? Los depósitos y almacenes de Bilbao contienen 120.000 cintas de vídeo aproximadamente; en Miramón tenemos otras 100.000 cintas de vídeo. Disponemos también de fondos digitalizados en el nuevo sistema digital que cubren aproximadamente 11.000 horas de piezas y brutos de informativos. Estos fondos son controlados mediante distintas herramientas que, aunque están separadas, están integradas en parte: los catálogos documentales del sistema tradicional (AIRS), el catálogo que tenemos en la base de datos de préstamos (ORACLE archivo, que controla el préstamo y el movimiento de cintas) y el sistema de gestión de contenidos digitales (MAM) que nos permite documentar las noticias o los programas que van llegando diariamente.

Respecto al archivo histórico, lo que hacemos es ingestar las cintas de vídeo, metiéndolas en un reproductor, para que un proceso interno cree ficheros o copias digitales de esos contenidos. Se conservan en un sistema centralizado y se asocia ese vídeo, que ahora llamaremos «esencia», a una ficha que es la ficha correspondiente del archivo histórico.

Este proceso hacia el archivo digitalizado, en desarrollo desde hace varios años, tiene dos retos importantes. En primer lugar, cómo conseguir el control

intelectual de los contenidos, esto es, cómo organizar los metadatos asociados que describen o contextualizan el contenido de esas «esencias». En nuestro caso, hemos migrado todas las bases históricas al nuevo sistema, pasando los datos del archivo documental de AIRS al nuevo sistema digital. Los datos están por tanto y de alguna manera duplicados. Hemos añadido también la información de catálogo de ORACLE y las bases históricas con información sobre emisión y otros datos técnicos. Estas informaciones sobre los programas se han ido generando durante años. Tenemos, por ejemplo, información fiable sobre las fechas de emisión de los programas, en qué soporte se conservan, qué segmentos tienen, etc. Es decir, hemos integrado las viejas informaciones en el nuevo sistema. En muchos casos, cuando se trate de programas no informativos o programas ajenos, esos contenidos se encuentran en cinta de vídeo y los metadatos pueden consultarse, aunque no pueda verse el vídeo directamente. Para hacer esto, hay que recuperar la cinta y visionarla mediante magnetoscopios tradicionales.

Sobre el control material de los contenidos, existe un plan de recuperación de soportes históricos. Este plan de digitalización se basa en digitalizar los programas informativos desde mayo de 2004. Existe también la digitalización a demanda: los materiales que piden los periodistas, si están en cinta de vídeo, son digitalizados y pueden conservarse en el nuevo sistema como nuevas versiones digitales. En definitiva, se trata de asociar los nuevos ficheros que se generan en entorno digital con los registros históricos que hay en la base de datos de archivo.

En relación con el control material de los contenidos, conviene mencionar el modo en que los materiales diarios se mueven desde los discos duros a los robots de cintas de datos, desde el *online* al *nearline*. Aquí surge un interrogante de gran interés para el trabajo y la definición del archivo: ¿Qué es el Archivo? ¿Es el Archivo el departamento? ¿Pertenecen al Archivo todos los contenidos que son conservados? ¿Es el Archivo únicamente aquello que se documenta? ¿A qué nivel de detalle debemos documentar? En fin, nuevas y también viejas preguntas. Sobre el control material debe subrayarse que el movimiento de soportes, antes muy intenso, se ha reducido mucho.

3. EL SISTEMA MAM

Voy a describir brevemente el sistema MAM de gestión de contenidos, aunque luego Koldo Lizarralde hablará de los aspectos más técnicos. El sistema MAM es la plataforma corporativa, del conjunto de la organización, que permite gestionar los contenidos audiovisuales. Los contenidos se definen como el conjunto formado por las «esencias» (que pueden ser audio, vídeo o, generalmente para contenidos televisivos, vídeo con audio) y los metadatos correspondientes. MAM, acrónimo de «Media Asset Management», es el sistema de gestión de contenidos media, que pivota sobre una base de datos y permite gestionar distin-

tas colecciones de contenidos, organizar la información relacionada, recuperarla, controlar quién la usa y (para los documentalistas algo muy importante) también visionar y catalogar desde los propios equipos informáticos.



En esta diapositiva se puede ver el esquema del sistema de gestión y cómo se relaciona con otras aplicaciones informáticas o técnicas. Por un lado se encuentra la gestión corporativa (las empresas tienen generalmente sus aplicaciones de ORACLE, sus aplicaciones de ofimática para controlar pedidos, administración, publicidad, producción, etc.) Por otro lado, se encuentra el entorno de automatización y técnico, es decir, las cámaras, las redes, etc. El MAM se sitúa fuera de estos dos entornos de gestión, ya que lo que maneja son los contenidos audiovisuales. Por medio de él podemos ingestar nuevos contenidos (tanto en ficheros como en cinta), limpiar y editar materiales, recoger la información que se está generando en las escaletas de los programas, obtener datos de emisión, o realizar labores de administración del sistema, por ejemplo, conocer quién está usando los contenidos, cómo se usan o documentar los programas o las piezas que los componen.

La conservación de contenidos digitales se realiza en los tres distintos niveles en que se almacenan. El almacén de cintas de vídeo tradicional, en baldas, se encuentra extendido, comenzando por el núcleo de contenidos que están *online* (contenidos recientes, en discos duros, disponemos en la actualidad de aproximadamente mil horas de contenidos en línea). Estos pueden ser borrados o pasar al sistema *nearline*, con contenidos recuperables en minutos y basado en cintas de datos. En la actualidad disponemos de unas 10.000 ho-

ras en este nivel de almacenamiento. Por último, tenemos el nivel *offline*, las estanterías tradicionales donde se conservan las cintas de vídeo o las cintas de datos que puedan ser sacadas de los robots. Tenemos actualmente en los almacenes 220.000 cintas de vídeo, aproximadamente 100.000 horas de programas.

4. LOS METADATOS

¿Cómo llevamos a cabo el control intelectual de los contenidos? Los sistemas digitalizados permiten incorporar metadatos desde el comienzo mismo de la producción. Hasta entonces, no disponíamos de información prácticamente hasta que no llegaba la cinta al archivo. Ahora podemos tener datos de producción, incluso si no hay propiamente vídeo, o disponer de datos de catalogación técnica antes de comenzar propiamente la catalogación documental (por ejemplo, conocer la duración o los segmentos).

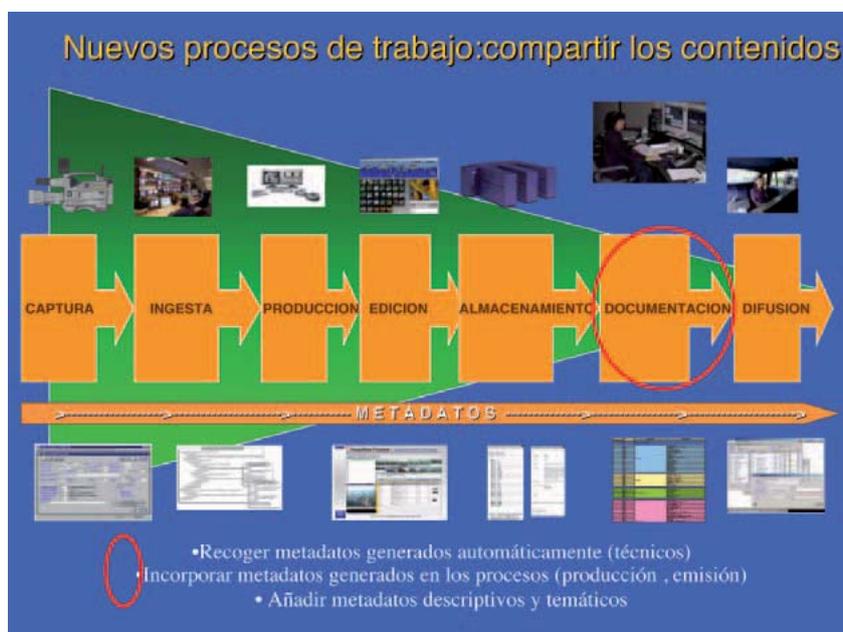
Para el área de archivo es muy interesante poder disponer de estos metadatos sin intervención directa de los documentalistas, aunque, naturalmente, haya que preparar los sistemas para que esos datos se incorporen. Vemos en la siguiente diapositiva un ejemplo de cuáles son estos datos de producción, datos técnicos o de emisión; por ejemplo, las fechas de emisión de un programa, en qué canal, datos de derechos...

Hacia el archivo (totalmente) digitalizado

Control intelectual de los contenidos →
Incorporar metadatos desde el comienzo de la producción:
pre-catalogación y catalogación técnica

PRODUCCION	• FECHA PRODUCCION / DIRECTOR /...
TECNICOS	• DV25
ESTRUCTURALES	• Segmento 1 / Segmento 2 / ...
EMISION	• ETB2 2004/01/01 / ETB SAT 2004/01/05 / ...
DERECHOS	• COPRODUCCION / RESTRINGIDO ...
...	

La siguiente diapositiva muestra un esquema de los nuevos procesos de trabajo y el modo en que se comparten los contenidos:



Con un sistema único, el personal de producción, los periodistas que necesitan materiales para editar, los documentalistas del archivo, todos pueden acceder a los contenidos desde el instante en que el contenido entra al sistema. Lo que vemos aquí como un esquema lineal, debe más bien verse como un círculo: los contenidos pueden reutilizarse en nuevas producciones. Lo importante es subrayar que los metadatos se generan en cada parte del proceso, desde la captura de cámara, datos ingesta, de producción, edición, etc., para que puedan ser utilizados por el resto.

5. EL TRABAJO DE DOCUMENTACIÓN

¿Qué añade el archivo? Lo primero, recoger todos los metadatos que se generan automáticamente, organizarlos y presentarlos de manera coherente. Después, verifica los datos que se generan en la producción y la emisión, muy importantes para conocer el contexto en el que un programa se ha hecho. También, y como hemos hecho siempre, añadimos metadatos descriptivos y temáticos.

En esta diapositiva vemos un ejemplo del uso que hacemos de las clasificaciones y las tipologías que hemos creado para controlar los distintos contenidos. Estas clasificaciones se utilizan antes de que lleguen al proceso de documentación.

Control de los contenidos ANTES de su archivo: tipología estructural / funcional

Ejemplo de clasificación de FORMATOS

Code	Format 1	Format 2
010	ELEMENTO DE CONTENIDO	BRUTO ORIGINAL
011	ELEMENTO DE CONTENIDO	BRUTO LIMPIO
020	ELEMENTO DE CONTENIDO	COMPAÑADO ARCHIVO
030	ELEMENTO DE CONTENIDO	EDITAJE C/DITE
031	ELEMENTO DE CONTENIDO	EDITAJE OFF
032	ELEMENTO DE CONTENIDO	EDITAJE VIDEO / REPORTAJE
040	ELEMENTO DE CONTENIDO	CONEXIÓN DUPLEX
041	ELEMENTO DE CONTENIDO	CONEXIÓN DIRECTO
042	ELEMENTO DE CONTENIDO	CONEXIÓN TELEFÓNICA
050	ELEMENTO DE CONTENIDO	PRESENTACIÓN
051	ELEMENTO DE CONTENIDO	ENTREVISTA
100	ELEMENTO DE REALIZACIÓN	CARETA ENTRADA
101	ELEMENTO DE REALIZACIÓN	CARETA SALIDA
102	ELEMENTO DE REALIZACIÓN	COPYRIGHT
103	ELEMENTO DE REALIZACIÓN	GRAFICO ANIMADO
120	ELEMENTO DE REALIZACIÓN	IMAGEN BRUTO
121	ELEMENTO DE REALIZACIÓN	IMAGEN GRAFISMO
130	ELEMENTO DE REALIZACIÓN	LOGOTIPO ANIMADO
131	ELEMENTO DE REALIZACIÓN	LOGOTIPO ESTÁTICO
132	ELEMENTO DE REALIZACIÓN	MÁSCARA
133	ELEMENTO DE REALIZACIÓN	GRAFICA / BREAK BUMPER
134	ELEMENTO DE REALIZACIÓN	SEPARADOR
135	ELEMENTO DE REALIZACIÓN	SIN FIN
136	ELEMENTO DE REALIZACIÓN	SOBREVANCE
137	ELEMENTO DE REALIZACIÓN	SOBRESIMPRESION
138	ELEMENTO DE REALIZACIÓN	TEMPLATE GRAFICO /
139	ELEMENTO DE REALIZACIÓN	VENTANILLA
140	ELEMENTO DE REALIZACIÓN	VUELTA PROGRAMA
200	PROGRAMA TV	INFORMATIVO DIARIO
201	PROGRAMA TV	INFORMATIVO NO DIARIO
202	PROGRAMA TV	INFORMATIVO ESPECIAL
210	PROGRAMA TV	ENTREVISTA
211	PROGRAMA TV	DEBATE
212	PROGRAMA TV	MAGAZINE
220	PROGRAMA TV	METEOLOGIA
221	PROGRAMA TV	BOLSA / NEGOCIOS
230	PROGRAMA TV	RETRANSMISIÓN

Estos ejemplos recogen cómo se utilizan en televisión materiales diversos: programas completos, elementos de programas, elementos de realización (por ejemplo caretas o sintonías), brutos de grabación, piezas editadas, elementos de radio, programas de radio, publicidad... Son todos, en fin, contenidos muy diferentes que, en cuanto son ingestados, son ya clasificados según su formato. Se puede ver también que manejamos brutos originales, brutos limpios, editajes, off, vídeo... La tipología de géneros es otra clasificación muy útil, en el siguiente ejemplo puede verse una clasificación de contenidos de ficción... como película, con qué género se relaciona (puede ser de aventuras, fantástica, religiosa, de terror...), si es un programa informativo, en qué sección se ha originado, etc.

Tipologías de géneros : temática

Ejemplo clasificación de GENEROS

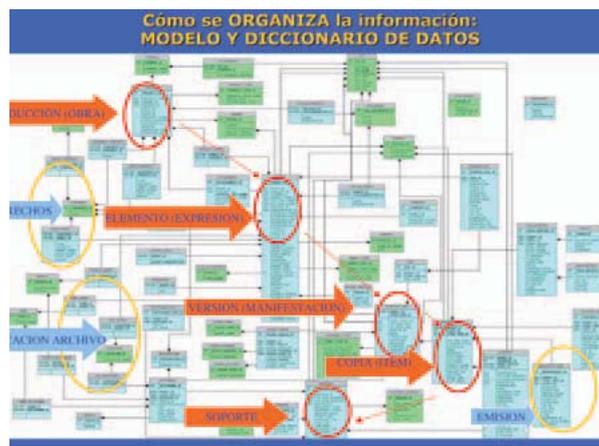







Code	Genero 1	Genero 2
1100	FICCION	GENÉRICO
1101	FICCION	AVENTURAS / ACCIÓN
1102	FICCION	BELICO
1103	FICCION	CARTELABO
1104	FICCION	CATATROFES
1105	FICCION	CIENCIA FICCION
1106	FICCION	CINE MUDO
1107	FICCION	CINE X
1108	FICCION	COMEDIA
1109	FICCION	COMEDIA ROMANTICA
1110	FICCION	DRAMA
1111	FICCION	EROTICO
1112	FICCION	FANTASTICO
1113	FICCION	HISTORICO
1114	FICCION	INFANTIL
1115	FICCION	JURIDICA
1116	FICCION	LIBRE
1117	FICCION	POLITICO
1118	FICCION	POLITICO
1119	FICCION	RELIGIOSO
1120	FICCION	TERROR
1200	MUSICA	GENÉRICO
1201	MUSICA	DANZA / BALLE
1202	MUSICA	JAZZ
1203	MUSICA	MUSICA CLASICA
1204	MUSICA	MUSICA CORAL
1205	MUSICA	MUSICA POPULAR
1206	MUSICA	MUSICA VASCA
1207	MUSICA	OPERA
1208	MUSICA	POP / ROCK
1209	MUSICA	ZARZUELA
1300	INFORMATIVO	GENÉRICO
1310	INFORMATIVO	PSYLP
1320	INFORMATIVO	CULTURA
1330	INFORMATIVO	ECONOMIA
1340	INFORMATIVO	INTERNACIONAL
1350	INFORMATIVO	TECNOLOGIA
1360	INFORMATIVO	SOCIEDAD
1370	INFORMATIVO	POLITICA
1400	DEPORTIVO INFORMATIVO	GENÉRICO

El modo de organizar la información en el sistema se concreta en el modelo de datos, el conjunto de todos los campos de información agrupados en entidades. Estos campos pueden también verse como un diccionario de datos o conjunto de campos disponibles, que se presentan alfabéticamente con una definición de ese campo, qué tipo de información va dentro, su longitud y modo de utilización. Aquí vemos un esquema del modelo de datos utilizado por nosotros:



Este modelo nos permite trabajar con los contenidos a distintos niveles: por ejemplo, a nivel de producción, si se trata de una teleserie que tiene 30 capítulos, tendremos los datos compartidos de producción; también, a un nivel inferior, el nivel de expresión o de elemento. Otro nivel inferior es el de manifestación. Por ejemplo, un elemento se puede manifestar como película en 16:9, estar o no subtulado, etc. siendo todas versiones del mismo elemento. También el nivel de copia y el nivel de soporte físico. El modelo soporta informaciones de distintas áreas, desde datos de emisión a índices temáticos generados en el archivo, datos sobre derechos de emisión o de explotación, etc.

ENTIDADES PRINCIPALES de información para programas de tv

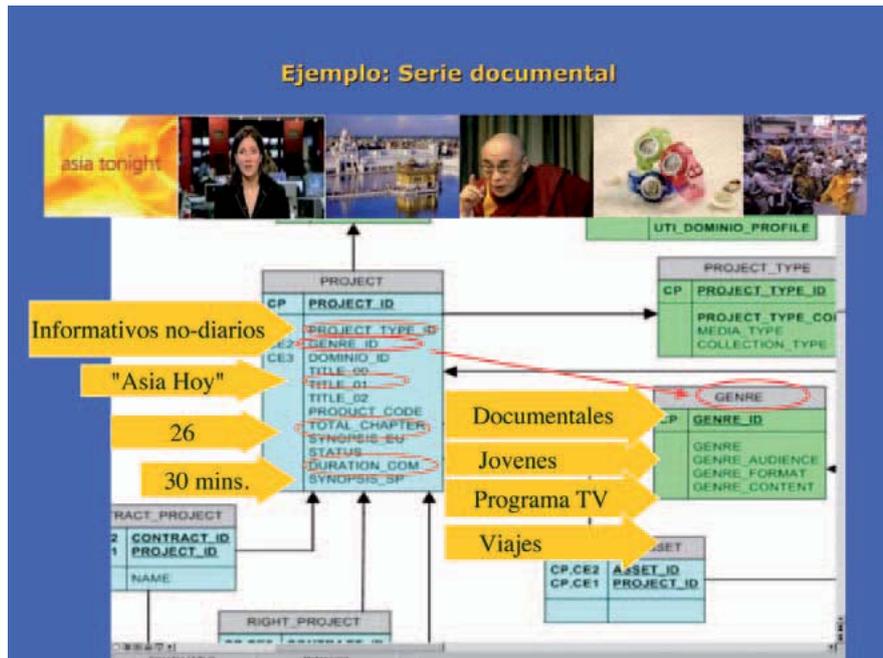
Entidad	Concepto	Ej. 1	Ej. 2
• PRODUCCION	• OBRA	• Documentales World today	• Mihiluze
• ELEMENTO	• PARTE DE OBRA	• "Irak en guerra"	• Cap. 19
• VERSION	• EXPRESION	• MPEG2	• Subtitulada
• COPIA	• MANIFESTACION FISICA	• Fichero XXX	• Betacam SP
• SOPORTE	• ELEMENTO FISICO	• Disco 1 / Disco 2 / ...	• Cinta video 1 / Cinta 2/ ...
• Derechos	• Comunicación, reproducción, ...	• ...	• ...
• Emisión	• Datos de difusión	• PROD. AJENA	• PROD. PROPIA
• Clasificación archivo	• Indización	• CANAL 2	• CANAL 2
• ...		• GUERRA / ATENTADO / ...	• CONCURSO / INFANTIL / ...




Un ejemplo: la producción *World Today* es una serie documental, con un número de producción concreto, unos datos del contrato, etc. Un elemento de esa producción podría ser un capítulo que se llamaría «Irak en guerra»; una versión estaría en un determinado fichero, conservada en un disco determinado, se ha emitido en un canal...

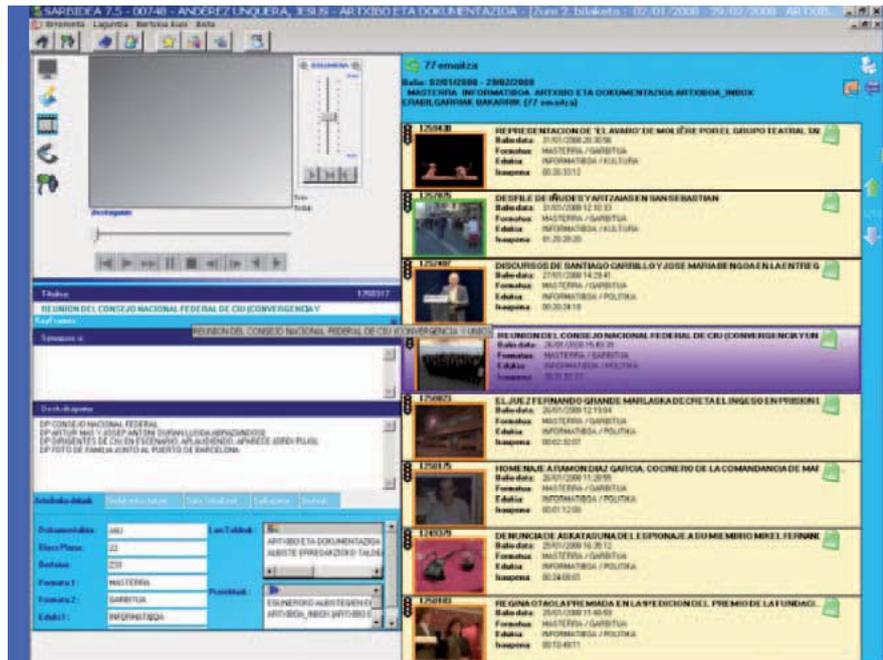
Otro ejemplo: un programa de producción propia, un concurso que se llama *Mihiluze*, capítulo 19, del que tenemos la versión subtulada, que se encuentra grabada en dos cintas de vídeo, etc.

Aquí vemos otro ejemplo con una serie documental:

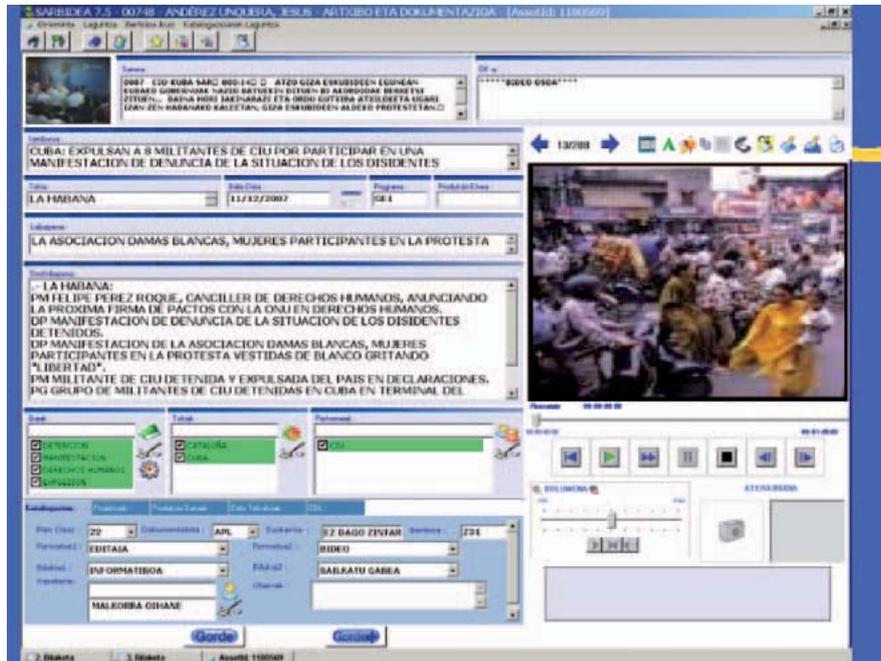


Más ejemplos: una noticia del programa informativo *Teleberri*. Son elementos que documentamos con mucha precisión, porque son de producción propia y tienen un gran potencial para poder ser reutilizados. Puede verse cómo nos llegan los metadatos desde la escaleta, antes del proceso de documentación. Nos llega un título, un vídeo titulado «Irán» tiene un número de escaleta determinado, por ejemplo el 60, lo genera la sección de internacional, en determinada fecha, con esta duración, lo ha elaborado tal reportero... Estos datos se recogen desde las herramientas de producción o de edición. Archivo visiona esos contenidos mejorando el título (por ejemplo, en lugar de «Irán», titula «Terremoto de escala seis en Irán»), haciendo un pequeño resumen, describiéndolo a nivel de secuencia o de planos, añadiendo los índices temáticos correspondientes, por ejemplo «terremoto», «campana de ayuda», lugares, personas, etc.

Vemos en la siguiente diapositiva una pantalla del catalogador:



Esta sería la primera pantalla, la pantalla de consulta; si hacemos una búsqueda, nos devuelve una lista de resultados, en la que podemos seleccionar un *asset* o elemento concreto y catalogarlo. La ficha se abre, tenemos los datos de precatalogación y podemos empezar a ver el vídeo.



Aquí vemos un ejemplo de la información que nos llega procedente de escaletas, en este caso la locución del presentador durante el programa; el documentalista añade el título, el lugar, la fecha y una descripción de planos, dependiendo del material a un nivel de mayor o menor detalle... Vemos también cómo se utilizan los campos de indización, los temas, lugares, personas... Se incluye una posibilidad muy interesante que permite automatizar el texto de algunos campos, trocarlo y compararlo con las listas del tesoro. Obtenemos así una serie de términos candidatos, muy interesante a la hora de catalogar y documentar. En la lista tenemos otro tipo de datos como formatos, contenidos, etc.

significativos de color, movimiento, etc., y asociarlos con un fotograma para crear una especie de guión gráfico, un *storyboard* del contenido. Los fotogramas significativos y los segmentos pueden también ser generados de modo manual.

En la actualidad, la información que se genera en el nuevo sistema se exporta al sistema documental tradicional, pero es previsible que en algún momento habrá que abandonar el sistema tradicional y utilizar únicamente el nuevo.

Estamos abordando el problema de cómo describir y catalogar teniendo en cuenta las nuevas condiciones, ya que los usuarios pueden ahora buscar y ver directamente los contenidos. Las descripciones que siempre hemos hecho se dirigían principalmente a transcribir a texto el contenido visual. Ahora, cuando los usuarios pueden ver el vídeo, las descripciones deberían cambiar.

Esta diapositiva muestra cómo se realiza el proceso de segmentación:



El software detecta los cambios, por ejemplo de movimiento o de color, y crea los segmentos correspondientes. En este ejemplo se aprecia cómo la segmentación puede trabajar en ayuda de un catalogador humano. Se crean secuencias automáticamente y eso nos permite documentar a nivel de *asset* completo o de cada segmento particular.

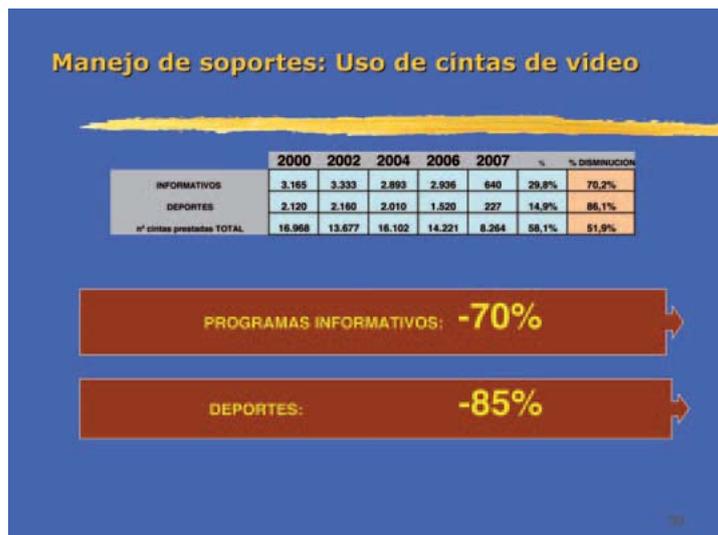
Automatización de la catalogación

Descripción (general / secuencias / planos)

Asignación de descriptores a partes de un elemento

En este último caso, se asocian descriptores sólo a una parte de un contenido; por ejemplo, imaginemos un programa, una retransmisión parlamentaria, que dura seis horas. Si no se trocea ese contenido, en su momento habrá que recuperarlo completo (seis horas), lo que resultaría muy pesado para los sistemas.

Otro efecto claro de la introducción del entorno digital es la disminución en el uso de cintas de vídeo:



Una estimación sobre datos de cuatro meses indica que los programas informativos están utilizando un 70% menos de cintas de vídeo que hace un año, y en los deportes incluso más. Las cintas de vídeo se están dejando de lado. Los programas, especialmente los que producen productoras externas, piden todavía muchas cintas de vídeo.

7. CAMBIOS EN EL ÁREA DE DOCUMENTACIÓN

Dedicamos más tiempo a establecer normas y estructuras de información, cómo debe moverse esa información de un sitio a otro, cómo presentarla de manera eficiente, control de calidad... Desde luego, disponemos de mayores posibilidades de acceso y visionado de los contenidos, y la selección de materiales para archivo es mejor y más ajustada a las necesidades.

Internet ha supuesto otro cambio muy importante, tanto para los usuarios que buscan allí información de referencia como para el propio documentalista. Es ya una herramienta de uso diario, por ejemplo, cuando llegan vídeos con poca información, ya que podemos buscar en la red y obtener de manera fácil información abundante sobre el contexto de una noticia.

Respecto a la precatalogación, disponemos de muchos datos que antes nos costaba mucho esfuerzo conseguir. Hemos trabajado especialmente en la automatización y la ayuda al análisis documental, pero tenemos que hacer todavía más, incluyendo el uso de técnicas lingüísticas para hacer más eficaz nuestro trabajo. Sobre la interpretación y valoración de los contenidos, no se han producido grandes cambios: continúa siendo necesario que una persona vea los contenidos y cubra los aspectos semánticos.