Metodología de análisis de sistemas de información y diseño de bases de datos documentales: aspectos lógicos y funcionales [1]

Lluís Codina
Universitat Pompeu Fabra (Barcelona)
lluis.codina@cpis.upf.es

RESUMEN

Presentación de una metodología de análisis de sistemas de información y de diseño de bases de datos documentales basada en la teoría de sistemas. Se exponen los instrumentos de análisis, las bases conceptuales y los procedimientos para interpretar problemas de información y diseñar bases de datos documentales.

RESUM

Presentació d'una metodologia d'anàlisi de sistemes d'informació i de disseny de bases de dades documentals, basada en la teoria de sistemes. S'exposen els instruments d'anàlisi, les bases conceptuals i els procediments necessaris per a interpretar problemes d'informació i dissenyar bases de dades documentals.

Introducción

En el contexto de los sistemas de información, el término *metodologías* suele generar equívocos a menudo. Es frecuente que se espere de ellas cosas que, en realidad, no pueden dar. En concreto, se suele esperar de ellas lo mismo que proporcionan, por ejemplo, los algoritmos en matemáticas, es decir, una solución segura a un problema bien planteado.

Por desgracia, en el desarrollo de sistemas de información no existe nada parecido a los algoritmos (ni a las recetas de cocina). ¿Para qué sirve entonces una metodología en este contexto? La experiencia indica que una metodología sirve, exactamente, para que el resultado final de un proyecto documental se deba *en lo más posible a la planificación consciente y, en lo menos posible, al azar o al método de ensayo y error*. Nada más, pero nada menos.

No parece necesario insistir mucho en que, mediante la planificación consciente, un profesional tiene derecho a esperar un grado de éxito mucho mayor que si toma las decisiones al azar o por el método del ensayo y error. Por contra, por muy correcta que sea una metodología, un lego no hará nada bueno con ella. Por tanto, la diferencia entre utilizar una metodología o no utilizarla está en qué proporción la parte final del producto puede atribuirse: *a)* al azar; *b)* al ensayo y error; *c)* a la planificación consciente.

De ello se desprende que siempre se introduce algo de azar en el diseño de sistemas de información, así como siempre existe la necesidad de recurrir al ensayo y error para refinar el resultado final. La cuestión clave radica en que *la parte de planificación consciente debe ser la que tenga mayor influencia en el resultado final*, tanto por razones de eficiencia como por razones de economía.

Lo contrario, que el azar y el ensayo y error tengan un gran peso, sólo puede producir sistemas desastrosos, principalmente porque los sistemas mal diseñados e ineficientes son mucho más probables, porque hay un nú-

1. Esta es la tercera versión pública que se presenta de esta metodología, desarrollada por el autor en su tesis doctoral (1994). La primera vez fue presentada en un Congreso sobre Documentación organizado por el Departamento de Documentación de la Universidad de Zaragoza (1996). La segunda, a través de la revista *Information World en Español* (1997). A su vez, los fundamentos científicos de esta metodología fueron publicados en la *Revista Española de Documentación Científica* (1994b). La versión que se presenta aquí ha sido revisada y actualizada en febrero de 1998, aunque su gestación, como ya se ha indicado, es anterior a 1994. Desde entonces, viene siendo testada y puesta a prueba en proyectos reales de diseño y creación de bases de datos documentales.

mero virtualmente infinito de hacer mal cualquier cosa, que los bien diseñados y eficientes y siempre que dejamos algo al mero azar sucede lo más probable. Esto no es más que una forma un poco más fisicalista de enunciar la conocida *Ley de Murphy*.

1. El peligro del sentido común

Es también habitual que las metodologías suenen como un mero puñado de consejos de sentido común, lo cual induce a algunos a un peligroso menosprecio hacia ellas.

El problema ante esta postura radica en que, si bien muchas recomendaciones acertadas parecen de sentido común, sus contrarias también lo parecen. Es decir, aunque una recomendación dada suene a sentido común, es peligroso no observar que, si nos fuera dada la recomendación contraria, también nos parecería de sentido común.

Así pues, con una metodología, por lo menos sabemos cuáles de las muchas cosas que *parecen* razonables *son*, de hecho, razonables. Pongamos un ejemplo, supongamos que alguien afirma que el mejor procedimiento para diseñar una base de datos es escoger un buen equipo informático, después elegir un programa que sea compatible con el mismo y, a continuación, diseñar la base de datos.

No sé que le parecerá al lector, pero se sabe de muchos equipos de diseñadores a los cuales el consejo le pareció tan adecuado que lo llevaron a la práctica con resultados, por supuesto, bastante lamentables. No les hubiera sucedido así si hubieran conocido uno de los aspectos más básicos del diseño de sistemas de información que aconseja comenzar siempre un proyecto estudiando primero los aspectos lógicos y no los físicos, o comenzar por la fase de análisis y no por la de implantación, etc. Sin embargo, cuando se explican esa clase de principios en un aula (o se leen en un artículo), invariablemente, se tiene la sensación de estar ante un mensaje de sentido común.

2. Qué es una metodología

Por otro lado, unas meras reflexiones o unos consejos no son, a pesar de todo, una auténtica metodología. ¿Qué cosas forman parte, por tanto, de una auténtica metodología? Entendemos que, en sistemas de información documentales, una metodología debería contemplar, como mínimo, tres grupos de elementos o aparatos conceptuales:

a) Aparato conceptual
b) Aparato instrumental
c) Aparato procedimental

El primer aparato, o grupo de elementos conceptuales, tiene la misión de proporcionar a los responsables de desarollo de sistemas de información unas bases conceptuales mínimas que faciliten su entendimiento de todo el proyecto y que faciliten, así mismo, la comunicación entre los diferentes actores involucrados en el proceso. Por tanto, en el *aparato conceptual* se definen las entidades básicas que intervienen en el proyecto y se proporcionan puntos de vista estratégicos.

El *aparato instrumental* es el responsable de proveer los instrumentos de análisis y de diseño, es decir, es aquella parte de la metodología que, precisamente, a veces se ha confundido, incorrectamente, con un algoritmo.

Finalmente, el *aparato procedimental* establece las fases y los procedimientos básicos, señalando sus objetivos, así como identifica y describe los productos que deben obtenerse de cada fase de análisis, incluido el producto final.

Así pues, y de acuerdo con lo expuesto, se describirá aquí una metodología de desarrollo de bases de datos documentales que no es un algoritmo, es decir, que no libera, mágicamente, de la obligación de tener una buena formación para poder aplicarla con éxito, pero que ayuda a reducir al mínimo posible los riesgos debidos a la improvisación.

Por otro lado, importa señalar que la metodología que se expone aquí se ha obtenido, básicamente, por la utilización de tres tradiciones científicas y académicas distintas, que este autor ha intentado fusionar en una metodología unificada y, hasta cierto punto, consistente. Se trata de las siguientes tradiciones académicas y/o tecnológicas:

a). La tradición del análisis de sistemas, proviniente de las ciencias informáticas. Unos de los autores más representativos y cualificados sería Yourdon (1993).

b). La tradición de la teoría de sistemas y de la metodología general de resolución de problemas. Concretamente, se ha utilizado teoría general de sistemas adaptada a problemas de información (Baiget, 1986; Currás, 1988) y aportaciones de la SSM *(Soft System Metodology),* una metodología elaborada principalmente, pero no únicamente, por Checkland (Checkland, 1981; Checkland y Scholes, 1990; Lewis, 1994; Underwood, 1996).

c). La tradición, naturalmente, de los métodos y procedimientos de trabajo de las ciencias de la documentación.

Una vez expuestas estas consideraciones de tipo meta-metodológicas, se exponen en las secciones siguientes los elementos de una metodología que, a su vez, tiene sus fundamentos teóricos en un modelo conceptual sobre sistemas de información documental expuesto con más detalle en otro lugar (Codina, 1994a y Codina 1994b).

3. Aparato conceptual

Un primer punto de partida muy útil en el diseño de todo sistema de información y, por tanto, también en el diseño de una base de datos documental, consiste en definir un sistema de información como un sistema, S1, denominado sistema de información, que mantiene registros sobre otro sistema del mundo real, S2, denominado sistema objeto.

De este modo, el proceso de análisis y diseño puede concebirse como el intento de obtener un modelo de aquella parte de la realidad, o sistema objeto (S2) que resulta de interés para el sistema de información (S1).

Tenemos entonces el par conceptual <sistema de información, sistema objeto>, o <S1, S2>, y la relación que les une es que el primero (S1) es un modelo del segundo (S2), exactamente en el mismo sentido en que un mapa será un buen sistema de información justo en la medida en que sea un buen modelo del territorio sobre el que informa.

El segundo punto de partida consiste en considerar que, desde el punto de vista de los intereses de la Documentación, todo sistema objeto (S2) se compone de dos subsistemas, que denominamos:

i) Sistema de actividades humanas (SAH)

ii) Sistema de conocimiento (SCO)

El SAH es el sistema social –es decir un sistema formado por personas y cosas– que justifica la existencia del sistema de información, porque en él desarrollan sus actividades los futuros usuarios que necesitarán que exista un sistema de información.

Para citar un ejemplo bien conocido en nuestro entorno, si pensamos en una biblioteca universitaria como en un sistema de información, entonces el sistema de actividades humanas al cual intenta modelar consiste en la creación y difusión del conocimiento, actividad que, a su vez, necesita a la biblioteca y a otros recursos documentales.

¿En qué sentido la biblioteca modela al mencionado sistema de actividades humanas? En el sentido, por ejemplo, de los temas y disciplinas científicas que cubre la biblioteca, la clase de documentos que procesa, el modo como los procesa, los procedimientos de trabajo, la clase de servicios que presta, el tipo de usuarios que admite, etc. Todas las características señaladas son un reflejo de las características de una actividad concreta centrada sobre la creación y la difusión del conocimiento.

Otro ejemplo. Si consideramos el proyecto de una base de datos para automatizar total o parcialmente el centro de documentación de un organismo, por ejemplo, el centro de documentación de un instituto o de un departamento de investigación, entonces ese centro de documentación será considerado el sistema de actividades humanas (SAH) que debe ser estudiado para diseñar a la base de datos, y el organismo del que depende el centro de documentación actúa de entorno del sistema. Como el entorno de un sistema siempre influye en él de alguna forma, los diseñadores de la base de datos, aunque deberán concentrarse en las características del centro de documentación, también deberán conocer las características de su entorno, esto es, de la empresa.

Los ejemplos podrían multiplicarse fácilmente. Por ejemplo, si se trata de diseñar la base de datos de un museo, el SAH vendrá determinado por el museo en cuestión y sus actividades; si se trata de crear una base de datos de cine para una filmoteca, el SAH vendrá determinado por las características de esta filmoteca, etc.

Por su parte, el sistema de conocimiento (SCO) está formado por la clase de documentos o por los tipos de entidad sobre los cuales el sistema de información debe mantener alguna clase de registros.

En el caso de la base de datos de un museo, por seguir con uno de los ejemplos mencionados, el SCO podría consistir en los objetos expuestos y conservados en el museo. Una variante podría ser que el SCO consistiera en los documentos de su centro de documentación, o en ambos. En todo caso, este ejemplo nos demuestra que, a veces, determinar qué forma parte del SCO puede ser una decisión técnica e incluso intuitiva, fruto de un análisis superficial, pero otras veces será el resultado de una decisión política más o menos elaborada.

Con los dos principios fundamentales anteriores se dispone ya de un mínimo aparato conceptual que permite iniciar la discusión de los otros elementos de la metodología. Se observará que algunas herramientas del aparato instrumental, tal como el modelo entidad-relación (que se explica más adelante) incluyen también aspectos conceptuales. En realidad, es en buena parte arbitrario decidir qué elementos pertenecen al aparato conceptual y qué elementos pertenecen al procedural o al instrumental. Aquí se he hecho una elección concreta, pero probablemente son posibles otras interpretaciones.

4. Aparato instrumental

El aparato instrumental de una metodología proporciona los instrumentos de análisis que puede utilizar el analista. En concreto, tres son los instrumentos principales que se pueden emplear: el modelo entidad-relación, desarrollado originalmente por Chen (1976), el diccionario de datos y la norma ISBD.

4.1. Modelo Entidad-Relación

El modelo entidad-relación (o modelo E-R) ayuda a detectar sin ambigüedades las entidades que formarán parte de la base de datos, es decir, los objetos que forman parte del sistema de conocimiento. Estas entidades son las que habrán de ser descritas en la base de datos e importa, por tanto, identificarlas con la mayor precisión posible. Además, el modelo E-R proporciona una terminología adecuada para las primeras fases de diseño y un método para discriminar entre entidad y atributo de entidad, cosa que a veces puede resultar trivial pero que en otras ocasiones no lo es en absoluto. El modelo E-R utiliza los siguientes conceptos:

- Entidad
- Atributo
- Relación

Según este modelo, si las bases de datos representan a cosas u objetos del mundo real, tales cosas deben ser identificables y deben tener algunas propiedades. A las cosas sobre las cuales almacena información una base de datos se las denomina entidades, y pueden ser cosas materiales (libros, personas, etc.) o conceptuales (ideas, teorías científicas, etc.). La única restricción aplicable es que las entidades que han de estar representadas en una base de datos deben ser identificables y, por tanto, debe ser posible señalar a una cualquiera de ellas sin ambigüedad.

Los atributos, por su parte, son las propiedades relevantes que caracterizan a una entidad. En este sentido, el término relevantes significa lo siguiente: relevantes para el problema de información que se está considerando. Teniendo en cuenta que, en principio, los atributos de una entidad son virtualmente ilimitados, será labor del documentalista seleccionar en cada caso cuáles son los que se consideran más relevantes.

El modelo distingue entre tipo de entidad y ocurrencia de entidad. Un tipo de entidad define un conjunto de entidades constituidas por datos del mismo tipo, mientras que una ocurrencia de entidad es una entidad determinada y concreta. Cuando se diseña una base de datos el objetivo del documentalista debe consistir en definir un tipo de entidad, que obtiene estudiando ocurrencias concretas de entidades.

Un registro es una representación de una entidad en la base de datos y, por lo tanto, cada registro describe a una entidad. Por ejemplo, en una base de datos bibliográfica, cada documento se describe en un registro. Por tanto, si los registros describen entidades del mundo real, los campos corresponden a los atributos de la entidad. De este modo, si un tipo de entidad posee los atributos A, B, C, el modelo de registro debe poseer los campos *A, B, C.*

En este punto, necesitamos diferenciar entre los siguientes conceptos:

1. *Etiqueta* del campo
2. *Valor* del campo
3. *Dominio* del campo

La etiqueta es el nombre del campo, es decir, una constante que identifica una zona del registro. El valor se refiere al contenido concreto de un campo concreto y puede ser distinto para cada campo de cada registro. El dominio, por su parte, es el conjunto del cual puede tomar sus valores un campo. Por ejemplo, el dominio del campo *Año de publicación*, es el conjunto formado por los años de publicación de documentos.

Título	Multimediaand hypertext: the Internet and beyond
Autor	Jakob Nielsen
Fuente	Boston: Academic Press, 1995
Año	1995
Páginas	480
ISBN	0-12-518408-5
Descriptores	Hipertextos, Multimedia, Sistemas de información, Publicaciones digitales, Documentación, Bases de datos, Internet, World Wide Web

Figura 1. Un registro en representación de un libro

Veámoslo con otro ejemplo. De acuerdo con el registro de la figura 1, el segundo campo o zona de información se puede analizar o descomponer así:

– *Nombre del campo:* Autor
– *Valor del campo:* Jakob Nielsen
– *Dominio del campo:* El conjunto de los nombre de responsables intelectuales de los documentos.

4.2. Generalizaciones y abstracciones

Al igual que distinguimos ente tipo y ocurrencia de entidad, debemos diferenciar también entre modelo de registro y ocurrencia de registro. Un tipo de entidad se forma por abstracción y/o generalización. Abstracción o generalización significa que se ignoran ciertos aspectos distintos de diversas ocurrencias de entidad y se forma con todas ellas un tipo unitario, o que se generalizan a todas las entidades ciertos rasgos que presentan regularmente ciertas entidades.

Por ejemplo, supongamos que aplicando el modelo E-R a un problema de información (por ejemplo, una base de datos para automatizar el archivo de un medio de comunicación), nos muestra como primer resultado los siguientes tipos de entidades:

1. Artículos de revistas
2. Artículos de prensa
3. Capítulos de libros
4. Libros
5. Informes
6. Fotografías de personajes
7. Fotografías de sucesos
8. Fotografías de estudio
9. Infografías

Una simple generalización reduce los nueve tipos de entidades a dos, puesto que las entidades 1 a 5 pueden reducirse, por abstracción, a una sola: *documentos escritos*, y los tipos de entidades 5 a 9 al tipo de entidad: *documentos icónicos.*

Para efectuar esa generalización el tipo de entidad *documentos escritos* deberá poseer un atributo denominado *Tipo de documento*, que permitirá describir qué clase de documento es: artículo, libro, etc. Por su parte, la entidad *documentos gráficos*, deberá tener también un atributo denominado *Tipo de documento*, que permitirá indicar si es una fotografía de personas, fotografía de paisajes, o si es una infografía, etc.

4.3. Relaciones

Las entidades del mundo real pueden tener relaciones entre ellas y, mientras las entidades suelen nombrarse mediante sustantivos, las relaciones se nombran mediante verbos.

Por ejemplo, consideremos el encargo hipotético de diseñar una base de datos sobre teatro. Un análisis superficial sería suficiente para revelar la existencia de dos entidades relevantes para el sistema: *[obras de teatro]* y *[autores teatrales]*, y veríamos que entre ambas entidades existe la relación <escriben>, que significa más explícitamente que *[autores teatrales]* <escriben> *[obras de teatro]*.

Un aspecto importante de la relación es su *grado*, el cual indica el número de elementos que pueden participar en cada uno de los extremos de la relación, en este caso *[autores]* y *[obras de teatro]*. Este grado puede ser de uno a uno (1:1), de uno a muchos (1:N) y de muchos a muchos (N:M). Una manera típica de representar estas relaciones y su grado es utilizando diagramas y expresiones textuales. En estos diagramas, las entidades se representan como rectángulos y las relaciones como rombos. A su vez, las entidades se identifican con sustantivos y las relaciones con verbos. En la figura 2 podemos ver un ejemplo de tales diagramas:

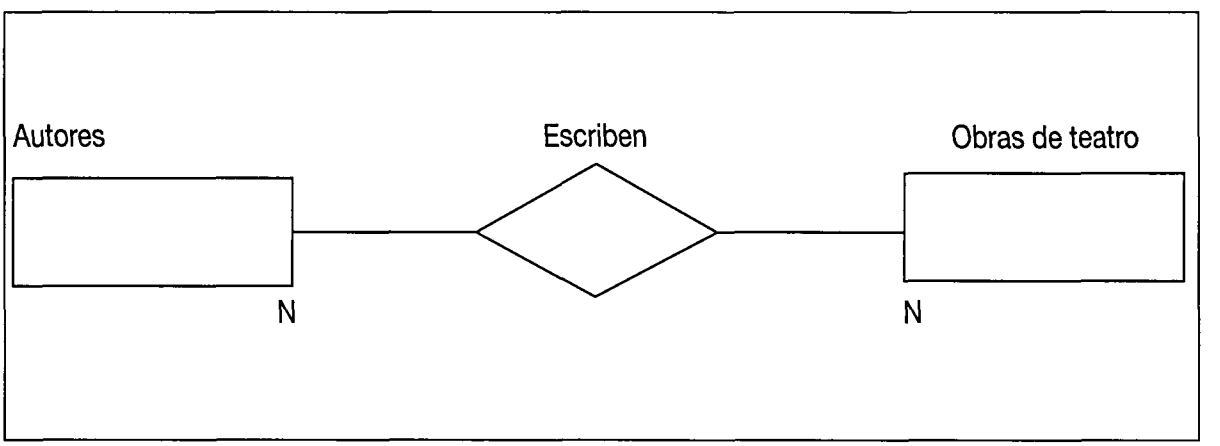

Figura 2. Diagrama entidad-relación de las entidades <autores de teatro> y <obras de teatro>

Así, por ejemplo, la relación que existe entre el número de ISBN y un libro es una relación de 1:1 (se lee «relación de uno a uno») porque un número de ISBN se asigna a un solo libro, y cada libro tiene un solo número de ISBN.

En cambio, la relación entre catedráticos de universidad y universidades es de 1:N (de uno a muchos), porque cada catedrático pertenece a una sola universidad, y una universidad tiene diversos catedráticos.

Finalmente, una relación de N:M (de muchos a muchos) seria la que existe entre autores de teatro y obras de teatro, porque un autor puede escribir diversas obras de teatro, y una obra de teatro puede estar escrita por varios autores y justamente ese es el significado de las letras N y M que hemos puesto en el diagrama anterior.

Además, la participación de la entidad puede o no ser obligatoria, lo cual significa que una entidad obligatoria interviene siempre en la relación. Por ejemplo, en la relación entre ISBN y libros, la participación de la entidad *[libros]* es obligatoria, porque siempre que hay un número de ISBN hay un libro, en cambio lo contrario no es cierto, porque hay libros que no tienen número de ISBN.

Esta última parte del análisis entidad-relación (grado y participación) es muy importante en el diseño de bases de datos de gestión que suelen utilizar tecnología relacional, porque ayuda a modelar los datos de la empresa y a representarlos en tablas normalizadas.

En cambio, en sistemas documentales no es tan importante porque estos no suelen utilizar tecnología relacional, ni necesitan modelar relaciones complejas entre entidades, como las que se dan en los sistemas de gestión administrativos.

4.3.1. Toma de decisiones

En conclusión, el modelo E-R aporta una importante claridad conceptual y proporciona una terminología común a todos los miembros que participan en el diseño. Sin embargo, el propósito de las herramientas de diseño no es tanto proporcionar soluciones para situaciones que son bien conocidas, sino para las situaciones no conocidas o menos típicas y, en este sentido, el modelo E-R puede resultar de ayuda también para determinar otros elementos del diseño.

Por ejemplo, y volviendo al caso anterior, donde se nos pide diseñar una base de datos sobre teatro. Supongamos que tenemos dudas sobre el siguiente aspecto: no sabemos si considerar que el autor (y todos sus datos biográficos) son atributos de la obra de teatro, o bien si considerar que autor y obras de teatro son entidades distintas, como hemos dado por supuesto en el diagrama.

Si adoptáramos el primer punto de vista, tendríamos que diseñar un único modelo de registro, donde los atributos del autor serían otros tantos campos, junto con los atributos de la obra de teatro. En cambio, si adoptamos el segundo punto de vista, necesitaremos diseñar dos modelos de registro, uno para obras de teatro y otro para autores. Puede ser que la simple intuición no indique cuál es el camino correcto en este o en otros casos parecidos, pero si queremos estar seguros de no equivocarnos en nuestra decisión, siempre podemos aplicar el siguiente procedimiento:

1. En caso de duda, tratar las cosas como entidades distintas.

2. Determinar la relación entre entidades.

3. Determinar su grado.

4. Si la relación es de grado 1:1, entonces se trata de una sola entidad, y un solo modelo de registro es suficiente para representarla. Por ejemplo, el número de ISBN es, de hecho, un atributo de la entidad libro, y para representarla es suficiente un solo registro, con un atributo que incluya el número de ISBN.

5. Si la relación es de grado N:1, o N:M, se trata de dos entidades y, por lo tanto, necesitamos dos modelos de registro, uno para cada entidad. Los dos modelos de registro deben contar con un campo compartido, lo cual proporciona dos campos con un dominio común. Esto último permitirá el cruce de datos. Por ejemplo, en el supuesto que estamos discutiendo, deberían utilizarse estos dos modelos de registro:

   - Autores
   - Obras

   Tanto el modelo de registro *Autores* como el modelo de registro *Obras* debería tener un campo cuyo dominio fuera el nombre de los autores, aunque en cada campo la etiqueta fuera distinta.

¿Qué sucedería si no procedíeramos como indica esta norma? En tal caso, la carga de datos sería poco eficiente, porque para autores muy prolíficos tendríamos que entrar los mismos datos tantas veces como obras de teatro hubieran escrito.

En general, si un autor ha escrito *n* obras de teatro, tendríamos que repetir sus datos *n* veces. Además, la redundancia, como es sabido, genera inmediatamente inconsistencias, y tendríamos enseguida, por ejemplo, diversas fechas de nacimiento para un mismo autor. Es evidente que si no detectamos ese error de diseño a tiempo, no tardará en hacerse evidente en algún momento de la fase de carga de datos, pero no debería ser menos evidente que si podemos evitar el error en la fase de diseño estaremos trabajando con mucha mejor calidad (ahora que está tan de moda este tema) que si necesitamos llegar a la implantación para detectar los errores, tal vez después de meses de trabajo que, de golpe, se revelarán inútiles.

Una advertencia final, muy importante, sobre la aplicación del modelo E-R. Primero, cuando se utiliza para diseñar bases de datos relacionales, las reglas para tomar decisiones son más complejas, porque la descomposición de datos a la que obliga el modelo relacional implica la necesidad de representar no sólo las entidades, sino también las relaciones entre entidades mediante una tabla más. Los interesados en esos aspectos de diseño pueden consultar Jackson (1990).

En general, la tecnología relacional debería ser necesaria cuando se trata sobre todo de modelar actividades (relaciones) y los datos relativos a cada entidad son relativamente simples o están muy estructurados. La mayoría de las actividades de gestión administrativa de una empresa son de esa clase y por eso utilizan sistemas relacionales. En cambio, deberíamos utilizar sistemas documentales en la situación simétricamente opuesta a la anterior, es decir, cuando se trata de modelar depósitos de conocimiento más que actividades, y los datos no son en realidad datos, sino información no estructurada o extremadamente compleja. La mayoría de las actividades de la Documentación responden a ese perfil y por eso utilizan sistemas documentales.

4.4. El diccionario de datos

El diccionario de datos es una herramienta que ayuda al diseñador de una base de datos a garantizar la calidad, la fiabilidad, la consistencia y la coherencia de la información introducida en la base de datos, de tal manera que el diccionario de datos marcará decisivamente el rendimiento y la calidad global del sistema de información.

Consiste en la lista detallada de cada uno de los campos que forman los distintos modelos de registro de la base de datos. A cada campo de cada modelo de registro se le aplica una parrilla de análisis que contempla, como mínimo, los siguientes aspectos:

1. Etiqueta
2. Dominio
3. Tipo de datos
4. Indexación
5. Tratamiento documental
6. Lengua
7. Otros controles de validación u observaciones
8. Ejemplos válidos
9. Otros controles o especificaciones según el tipo de campo

Por ejemplo, supongamos, a efectos de esta explicación, una base de datos documental imaginaria sobre noticias de actualidad con sólo tres campos: <Título>, <Descriptores> y <Fecha de publicación>. El diccionario de datos tendría entonces esta forma:

Etiqueta: Título

Dominio:
Título del documento. El título se transcribe de la siguiente forma: *Título: antetítulo: subtítulo*.

Tipo:
Alfanumérico

Indexación:
Sí

Tratamiento documental:
Lenguaje libre

Lengua:
Lengua del documento

Controles de validación:
No puede quedar vacío. Si por alguna razón, el documento careciera de título, el documentalista asignará un título descriptivo.

Etiqueta: Descriptores

Dominio:
Palabras clave normalizadas que expresan los conceptos principales contenidos en el documento, según el siguiente principio general: si el artículo contiene *n* conceptos relevantes se asignan *n* descriptores, procurando no asignar más de 20 descriptores por documento.

Tipo:
Alfanumérico

Indexación:
Sí

Tratamiento documental:
Lenguaje controlado

Lengua:
Del centro de documentación

Controles de validación:
No puede quedar vacío y sólo admite valores extraídos de una lista de términos autorizados.

Etiqueta: Fecha de publicación

Dominio:
La fecha de publicación de la noticia, indicada con el siguiente formato: DD/MM/AAAA.

Tipo:
Fecha

Indexación:
Sí

Tratamiento documental:
No procede

Lengua:
No procede

Controles de validación:
No admite valores fuera de rango.

Estudiando el ejemplo de diccionario de datos anterior, formado únicamente por tres campos, podemos observar cuatro aspectos importantes para el diseño de bases de datos:

1. Que el *Dominio,* en el contexto del diccionario de datos, se refiere al conjunto del que un campo puede obtener sus valores.

2. Que el *Tipo* se refiere, en cambio, al tipo de dato que admite el campo. Los tipos de datos suelen ser: numérico, alfanumérico, fechas y lógico.

Recordemos que un tipo de dato (*data type*) define un conjunto de operaciones válidas y un rango de valores aceptable. Por ejemplo, el tipo de datos alfanumérico define operaciones de comparación de cadenas de caracteres, entre otras, así como cualquier letra de la *a* a la *z* y cualquier número del *0* al *9,* así como cualquier combinación de esos caracteres en palabras, frases, párrafos, etc. En cambio, no admite operaciones aritméticas, aunque admita números. Por el contrario, un tipo de dato numérico admite sólo números así como cualquier operación aritmética, etc.

Por su parte, un campo de fechas sólo admite fechas en un formato establecido y permite búsquedas por rangos de fechas o por valores superiores o inferiores a una fecha dada. Un campo lógico sólo admite uno de dos valores: Sí o No; Verdadero o Falso.

3. Que el *Tratamiento documental* establece si se debe utilizar algún lenguaje documental para entrar los valores del campo, como así sucede en el campo Descriptores, donde el diccionario de datos establece que ese campo sólo admite palabras clave autorizadas extraídas de un thesauro o de una lista de autoridades.

4. Que la *Lengua* puede ser, o bien la lengua del documento, o bien la del centro de documentación. Eso significa, en el caso de un documento escrito en inglés, que el título estaría en inglés, pero los descriptores en castellano, siempre de acuerdo con el diccionario de datos precedente.

La descripción funcional, por su parte, debe incluir los siguientes elementos:

1. Qué clase de información se tratará y cómo entrará la información en el sistema.
2. Qué procesos documentales se llevarán a cabo.
3. Qué servicios y productos generará el sistema, y/o a qué aplicaciones podrá dar soporte.

El primer punto debe describir en qué consisten las entradas del sistema. El punto dos debe proporcionar una idea sobre qué procesos de tratamiento documental automatiza la base de datos, y el punto siguiente debe explicar en qué consisten las salidas del sistema.

4.5. La norma ISBD y los modelos canónicos

No deberíamos olvidar que, en Documentación, la experiencia previa ha dejado bien sentados cuáles son los atributos de algunas entidades e incluso cuál es la forma más conveniente de representarlos. Podemos hablar entonces de situaciones canónicas que han generado un modelo. La mejor herramienta de análisis y de diseño, en tal caso, consiste precisamente en aplicar ese modelo bien conocido y testeado.

Por ejemplo, los atributos estructurales de cualquier clase de documento pueden ser adecuadamente modelados siguiendo la norma internacional ISBD.

Recordemos que esa norma internacional representa un gran esfuerzo de abstracción para proporcionar un marco general de descripción, válido para cualquier clase de documento, desde una partitura musical, hasta una filmación audio-visual, pasando por un archivo de ordenador, un fonograma o un artículo de revista, de manera que las ISBD constituyen una herramienta de diseño de primera magnitud para cualquier problema documental donde debamos representar documentos.

Sobre el uso de las ISBD, cabe señalar que algunos centros de documentación se han sentido intimidados ante la aparente complejidad de la norma y la supuesta obligación de adoptarla como un todo, incluyendo la puntuación que prescribe y, en tal sentido, se ha argumentado que utilizar la norma ISBD puede tener sentido en algunos contextos de lectura pública, pero no necesariamente para el diseño de bases de datos documentales.

Entiendo que tal postura es un error: primero, porque siempre podemos utilizar la estructura de las ISBD como una orientación en el análisis de los documentos convencionales así como una fuente de información para situaciones más exóticas, independientemente de que incorporemos o no la norma en toda su complejidad.

5. Aparato procedimental

El principio general de diseño de sistemas de información indica que todo proyecto comienza siempre por un diseño lógico y que, una vez aprobado éste, se procede al diseño físico o implantación, en un proceso que es tan circular como lineal, ya que la fase de diseño, por ejemplo, puede obligar a repensar aspectos de la fase de análisis.

El aspecto importante aquí es que la metodología nos dice claramente que el proceso de creación de una base de datos debe ir siempre desde los aspectos lógicos hacia los aspectos físicos, y no al revés, como, sin embargo, suele suceder, ya que, en la práctica, existen muchas formas de violar ese principio general a causa de malos hábitos de trabajo.

Otra manera de enfocar incorrectamente este proceso consiste en querer abordar directamente el diseño del sistema de información e, incluso en querer visualizarlo por completo en nuestra mente, sin saber antes nada del sistema objeto.

El resultado, claro está, será una visión caótica. Todas las interrogantes se agolparán en nuestra mente y seremos incapaces de despejar una sola de ellas. Lo correcto en ambos casos es comenzar a diseñar los aspectos lógicos (nivel conceptual) ignorando de momento los aspectos físicos; de la misma manera, hay que comenzar por analizar el sistema objeto y, sólo después de conocerlo bien, podremos iniciar el diseño del sistema de información.

Así pues, el proceso de diseño de un sistema de información debe ajustarse siempre al siguiente ciclo de vida que, por otro lado, es universal para todo sistema de información:

1. Análisis
2. Diseño
3. Implantación

Otra forma de enfocar el ciclo de vida de un proyecto de desarrollo es indicar que la dirección del diseño debe proceder de lo conocido a lo desconocido, y no al revés, como sucede cuando se desea visualizar el sistema de información antes de conocer el sistema de actividades humanas y el sistema de conocimiento.

Finalmente, y por la misma razón, la dirección del diseño debe ir de lo general a lo específico y de los aspectos lógicos a los aspectos físicos, y nunca al revés, es decir, nunca se debe empezar a discutir o a considerar cuestiones concretas (¿cómo se imprimirá la información?) o físicas (¿qué tamaño tendrán las estanterías de los documentos?) antes de plantear las cuestiones generales (¿cuál es el propósito de la base de datos?) o lógicas (¿qué entidades formarán parte de la base de datos?). El siguiente cuadro sinóptico sintetiza estas ideas:

- De lo conocido a lo desconocido.
- De los aspectos lógicos a los aspectos físicos.
- De lo general a lo concreto.

Figura 3. Cuadro sinóptico de la dirección del diseño en el ciclo de vida de un sistema de información

En cuanto al ciclo de vida, cada una de las tres fases enunciadas antes (Análisis, Diseño, Implantación) puede dividirse en cuantas subfases sean necesarias según el proyecto concreto y la clase de sistema que se está diseñando.

En el caso de una base de datos documental, las dos primeras fases se pueden subdividir en otras dos subfases (a y b). Las fases de implantación pueden subdividirse en cuatro subfases (a, b, c, d, e). Nuevamente debe indicarse que tales divisiones tienen siempre algo de arbitrario. Aquí se hace una propuesta concreta, pero pueden ser válidas otras formas de dividir el ciclo de vida. En concreto, en esta metodología se propone la división de fases del cuadro sinóptico de la figura 4:

1. Análisis

1a. Análisis del sistema de actividades humanas
1b. Análisis del sistema de conocimiento

2. Diseño

2a. Diseño del modelo conceptual
2b. Determinación de los procedimientos de tratamiento documental (descripción, análisis e indexación documental, etc.) si es el caso.

3. Implantación

3a. Elaboración del presupuesto y del calendario de implantación, en su caso.
3b. Selección del soporte informático *(software* y *hardware)* de acuerdo con los requerimientos expresados en el modelo conceptual de la base de datos producido en la fase 2a y de acuerdo con los requerimientos expresados en 2b.
3c. Instalación, pruebas de rendimiento y re-elaboración, en su caso, de los puntos previos de este ciclo de vida.
3d. Elaboración del libro de estilo de la base de datos.
3e. Carga de datos, formación de usuarios y promoción del producto.

Figura 4. Cuadro sinóptico del ciclo de vida de una base de datos documental

Aunque expresado en fases y enumeradas secuencialmente el proceso parece estrictamente lineal, en realidad, el proceso de diseño también tiene mucho de circular, porque aunque siempre se empieza por la fase de análisis y se sigue con la de diseño, llegados a la fase 2b, por ejemplo, es posible que el diseñador desee considerar de nuevo algunos aspectos de 2a, o que necesite aclarar mejor algunas cuestiones de 1b, etc., idea que recoge la figura 5:

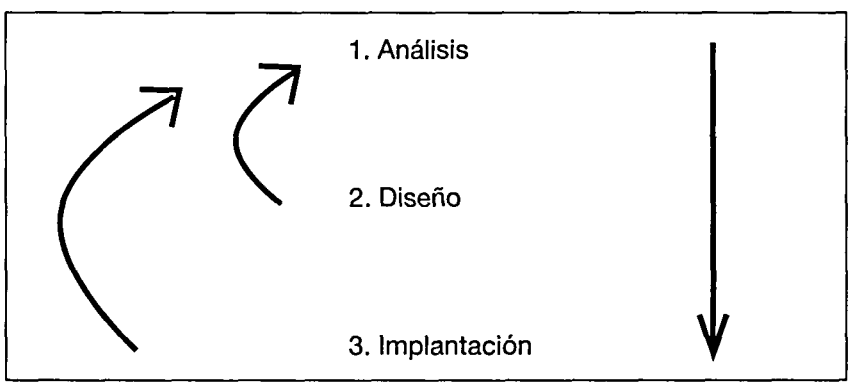

Figura 5. El ciclo de vida de un sistema de información como un proceso circular

En este sentido, debe hacerse notar que la metodología no excluye totalmente el procedimiento del ensayo y error, como ya se advirtió, sino que lo integra como un modo natural de refinar el producto. En particular, es prácticamente imposible producir un modelo conceptual correcto en el primer intento, y la experiencia indica que lo más probable es que el modelo elaborado en los puntos 2a y 2b haya que rehacerlo más de una vez, por lo menos en alguno de sus aspectos, principalmente a la vista de las primeras pruebas de rendimiento (3c).

Naturalmente, tiene que llegar un momento en el cual el diseñador dé por finalizado el proceso, pero la cuestión de cuántas veces conviene iterarlo antes de darlo por bueno, no puede establecerse *a priori,* sino que, antes bien, es una cuestión sensible al contexto y que debe decidir el diseñador en cada caso.

En todo caso, es importante que se llegue a la fase de implantación con un modelo lo más sólido posible porque a partir de tal fase ya no resulta tan fácil reconsiderar el proyecto, por lo menos no sin pagar algún precio, de manera que el punto 3c debería considerarse el punto de despegue, de alguna manera, el punto de no retorno del proyecto.

La fase de implantación puede llevarla a cabo un equipo distinto del que hizo el diseño. De hecho, en algunas empresas, sobre todo en empresas medianas y grandes, puede ocurrir que la fase de implantación corra a cargo del departamento de informática, aunque el análisis y el diseño lo haya hecho el de documentación. En empresas pequeñas, lo más habitual es que todo el proceso lo ejecute un mismo equipo o una misma persona.

Cada una de las fases precedentes (Análisis, Diseño, Implantación) tiene unos objetivos, debe producir unos resultados concretos y utilizar unas herramientas determinadas.

5.1. La fase de análisis

El objetivo de esta fase es conocer bien aquella parte del mundo real, llamada sistema objeto, que justifica y requiere la creación del sistema de información, de una base de datos en este caso.

Como ya vimos anteriormente, a efectos de análisis, el sistema objeto se considera dividido en:

1. Un sistema de actividades humanas (SAH)

2. Un sistema de conocimientos (SCO)

Por lo tanto, y dado que las características del sistema de actividades humanas (SAH) determinarán las características de la base de datos, deberá conocerse lo mejor posible antes de iniciar cualquier actividad de diseño.

El resultado que debe producir esta fase de análisis es una descripción textual que puede incluir gráficos de ser necesario, sobre el SAH, que suele denominarse *Modelo Esencial*, y que debe incluir, como mínimo, los siguientes aspectos:

1. Propósito y objetivos del SAH
2. Actores principales del SAH
3. Descripción de las actividades más relevantes del SAH
4. Datos principales sobre el entorno del SAH

La herramienta principal aquí es la realización de entrevistas con representantes del SAH y el análisis de cualquier documentación, del y sobre el SAH, que pueda aportar una comprensión global del sistema. Entre tales documentos podemos citar organigramas, documentos fundacionales, memorias, etc.

Aunque el *Modelo Esencial* consiste básicamente en una descripción textual, puede incluir, si el documentalista lo considera necesario, diagramas o gráficos que facillten su comprensión. El *Modelo Esencial* no debe ser muy extenso, sino, que tal como indica su nombre, debe consistir únicamente en una descripción que recoja los aspectos esenciales de la naturaleza y de las actividades del SAH. Además, como una base de datos documental no persigue el modelado de esas actividades, probablemente cinco o seis párrafos deberían ser suficientes para aportar el conocimiento necesario para los objetivos perseguidos.

Este modelo podrá formar parte del producto final, pero no es necesario que sea así, ya que, principalmente su misión es asegurarse de que el responsable del proyecto y otros actores que intervengan en él tienen una adecuada concepción de la naturaleza del SAH.

Por su parte, el propósito de la fase del análisis del sistema de conocimiento consiste en conocer el componente clave en este caso del sistema objeto, a saber, los documentos o las cosas sobre las cuales la base de datos deberá recoger información.

El resultado de esta fase debe consistir en la identificación clara y sin ambigüedades de los documentos o las cosas (entidades) sobre las cuales la base de datos deberá mantener información, así como debe poner de manifiesto las propiedades más relevantes de esas entidades.

La herramienta más adecuada para esta fase es el modelo entidad-relación (modelo E-R), un modelo bastante intuitivo que, sin embargo, resulta de gran utilidad para enfocar este tipo de análisis. Este modelo se explicará en el apartado dedicado a las herramientas.

5.2. La fase de diseño

El propósito de la fase de diseño es obtener un *Modelo Conceptual* de la base de datos y una *Propuesta de tratamiento documental*. El primero contiene los elementos necesarios para orientar el proceso de implantación. El segundo establece criterios y orientaciones sobre el proceso de descripción y de representación del contenido semántico de los documentos o entidades de los que tratará la base de datos.

Los dos modelos mencionados son el resultado de la fase de diseño y deben ser aprobados por quien encargó el proyecto, antes de que puedan servir como guías de implantación. Por tanto, el modelo conceptual no sólo debe ser acertado, sino que además debe parecerlo.

El *Modelo Conceptual* debe contener, por lo menos, los siguientes elementos:

1. El Modelo Esencial, mencionando el propósito de la base de datos e identificando el tipo de usuarios del sistema.
2. Una definición del tema o dominio de la base de datos, aunque puede estar recogido en el punto anterior.
3. El diccionario de datos completo.
4. Una descripción funcional del sistema, si no ha quedado bien establecido en los apartados anteriores.

El *dominio* de la base de datos es el conjunto de los temas o entidades sobre los que mantiene información la base de datos. Como todo dominio, puede definirse por extensión o por comprensión. Por tanto, puede ser tan breve como el nombre de una o más disciplinas científicas, por ejemplo, el dominio de la base de datos LISA consiste en en que trata sobre *Ciencias de la Documentación*. O puede consistir en una frase, por ejemplo, el dominio de la base de datos *Teseo* se enuncia diciendo que está formado por *las tesis doctorales publicadas por universidades españolas*.

Las herramientas para producir el documento anterior son, entre otras, las siguientes:

1. El Modelo esencial.
2. El modelo entidad-relación.
3. El diccionario de datos.

5.3. La fase de implantación

La implantación de una base de datos, cuando forma parte de un proyecto completo, requiere toda una metodología propia en donde se contemplen las diversas fases de proyecto y se pueda realizar la estimación de costos y el calendario de realizaciones.

Aquí hemos presentado una metodología de análisis y de diseño de bases de datos. Para la fase de implantación, únicamente podemos ofrecer aquí algunas orientaciones muy generales.

Estas orientaciones pueden ser válidas en tanto son muy generales, pero debe tenerse en cuenta que en determinado proyectos, por ejemplo en proyectos de tipo GED (Gestión Electrónica de Documentos) que incluyen a veces digitalizaciones masivas de documentos, cada una de las fases que se indicarán a continuación debe ser objeto de estudio y evaluación específica.

Por tanto, una vez aprobado el modelo conceptual de la base de datos, puede procederse a su implantación, la cual suele seguir el siguiente proceso:

1. Se realiza un análisis de costos y un primer calendario de realizaciones. Ambas cosas deben hacerse en función de las características del proceso de análisis de la información; del valor añadido que se quiera dar a los documentos mediante tratamiento documental; del volumen de datos a tratar y de las características y número de usuarios del futuro sistema.

2. Se selecciona el sistema informático (*software* + *hardware*) que pueda satisfacer mejor los requerimientos del modelo conceptual y del modelo de normativa de indización que contempla aquel. Naturalmente, en algunos entornos la arquitectura de información corporativa impondrá restricciones en el rango de soluciones informáticas a contemplar. En estas situaciones, la selección del sistema informático suele ser una responsabilidad exclusiva del Departamento de Información. En tales casos, el equipo de documentalistas únicamente debe ocuparse de presentar las especificaciones funcionales que debe satisfacer el sistema al Departamento de Informática.

   De ser necesario, se examinarán varios programas candidatos hasta que exista la certeza de que el programa elegido se ajustará lo mejor posible a los requerimientos funcionales del diseño. Se realiza la primera instalación y se nombra a un administrador de la base de datos que, a partir de ahora, será el máximo responsable de ella.

3. Se realizan pruebas con una colección-test de documentos o de entidades a ser representadas para comprobar la consistencia de los modelos, esquemas de registros, vocabularios controlados, etc.

4. Se realizan los cambios o ajustes necesarios, para refinar el modelo final.

5. Se determina la política final de mantenimiento y explotación.

6. Se edita la versión 1 del *Libro de estilo de la base de datos*, que incluye:

   a. La versión definitiva del modelo conceptual.
   b. La normativa de tratamiento documental, en su caso.

7. Se procede a la formación del personal técnico y de los usuarios finales.

8. Acciones de promoción, etc., en su caso.

Conclusiones

El valor de esta metodología radica, como ya se dijo al principio, en que ayuda a que el producto final sea más resultado del diseño consciente que de las fuerzas ciegas del azar y/o del ensayo y error, pero, particularmente entendemos que su utilidad aumenta conforme se aplica a situaciones poco canónicas o a situaciones atípicas, como las que el entorno cambiante de nuestra profesión introduce en cada momento y, al parecer, tal como el nuevo horizonte de las autopistas de la información y de un futuro mundo digital parece prometer.

Esperamos que, entonces, la aplicación de esta clase de metodologías sirva para que los profesionales de nuestro campo puedan demostrar los beneficios de una adecuada formación académica, del trabajo bien realizado y de la planificación, porque en nuestro campo de actividades también es rigurosamente cierto que el éxito se debe invariablemente a la famosa proporción de «un diez por ciento de inspiración y un noventa por ciento de transpiración».

Bibliografía

AENOR. (1990). *Norma UNE 50-106-90. Documentación. Directrices para el establecimiento y desarrollo de tesauros monolingües.* Madrid: AENOR.

BAIGET, T. (1986). *Análisis de sistemas de información*. Barcelona: Institut Català de Tecnologia.

CHECKLAND, P. B. (1981). *Systems thinking, systems practice* Chichester: Wiley.

CHECKLAND, P. B.; SCHOLES, J. (1990). *Soft systems methodology in action*. Chichester: Wiley.

CHEN, P.P-S. (1976). «The entity-relationship model: towards a unified view of data». *ACM transactions on databases systems.* Vol. 1, Nº 1, 1976, p. 9-36.

CODINA, Lluís. (1994a). *Sistemes d'informació documental: concepció, anàlisi i disseny de sistemes de gestió documental amb microordinadors*. Barcelona: Pòrtic.

CODINA, Lluís (1994b). «Modelo conceptual de un sistema de información documental». *Revista española de documentación científica.* Vol. 17, Nº 4, (Octubre/Diciembre), p. 440-449.

CODINA, Lluís (1996). «Análisis de sistemas y metodología de diseño de bases de datos documentales (Conferencia). *I Encuentro sobre sistemas de información y documentación IBERSID96.* Facultad de Filosofía y Letras. Universidad de Zaragoza, 12, 13 y 14 de marzo de 1996.

CODINA, Lluís (1997). Una propuesta de metodología para el diseño de bases de datos documentales. *Information world en español*. Parte I., Vol. 6, Nº 11 (noviembre) p. 23-29. Parte II. Vol. 6, Nº 12 (diciembre), p. 20-26.

CURRAS, E. (1988). *La información en sus nuevos aspectos.* Madrid: Paraninfo.

IFLA. (1983). ISBD(G): *Descripció bibliogràfica normalitzada internacional general*. Barcelona: Generalitat de Catalunya. Institut Català de Bibliografia.

JACKSON, G. A. (1990). *Introducción al diseño de bases de datos relacionales*. Madrid: Anaya.

LEWIS, P. (1994). *Information systems development*. London: Pitman.

UNDERWOOD, P. G. (1996). *Soft Systems Analysis and the management of libraries, information services and resource centres*. London: Library Association.

VAN SLYPE, Georges. (1991). *Los lenguajes de indización: concepción, construcción y utilización en los sistemas documentales*. Madrid: Fundación Germán Sánchez Ruipérez.

WALKER, D.W. (1991). *Sistemas de información basados en ordenador.* Barcelona: Marcombo.

YOURDON, E. (1993). *Análisis estructurado moderno*. México: Prentice-Hall Hispanoamericana.