

CAPTURA DE INFORMACIÓN POR INTERNET

Idoia Torrejón, Patxi Pizarro y Alejandro Ortúñez



EUSKAL ESTADISTIKA ERAKUNDEA
INSTITUTO VASCO DE ESTADISTICA

Donostia-San Sebastián, 1
01010 VITORIA-GASTEIZ
Tel.: 945 01 75 00
Fax.: 945 01 75 01
E-mail: eustat@eustat.es
www.eustat.es

CAPTURA DE INFORMACIÓN POR INTERNET

Idoia Torrejón, Patxi Pizarro y Alejandro Ortuñez

Indice

INDICE	3
INTRODUCCIÓN.....	4
EVOLUCIÓN.....	5
CONSECUENCIAS DEL USO DE FORMULARIOS INTERNET	7
APARTADOS DE UN CUESTIONARIO.....	8
PÁGINA DE CABECERA	8
IDENTIFICACIÓN DEL ENCUESTADO	9
PÁGINAS DE ENCUESTA.....	9
PÁGINA FINAL.....	10
ARQUITECTURA	11
ENTORNO DE TRABAJO	12

Introducción

El objetivo de la ponencia es contar la experiencia de Eustat en la recogida de información a través de cuestionarios electrónicos diseñados con tecnología para Internet.

Cuando Eustat inició el uso de estos cuestionarios se plantearon problemas, como por ejemplo, la duplicidad de formularios, la seguridad de la información, la especificidad de los accesos, ... que cambiaron el enfoque de la recogida de datos y que propiciaron soluciones que han ido evolucionando según el Instituto adquiría experiencia en esta materia.

Por otro lado, el diseño inicial de los formularios internet se ha ido mejorando hasta llegar a los formularios actuales, sin que los modelos se hayan dado por cerrados. A pesar de esto, se han identificado las características, diseño y funciones de todos los apartados.

Evolución

Este tipo de recogida se inició por primera vez en 2002 con la Encuesta sobre la Sociedad de la Información (Empresas) al calor del auge de Internet. Más adelante, en 2003, la Dirección de Eustat, dio una directriz según la cual se consideraba que la recogida de información a través de Internet era un asunto estratégico y de imagen para el Instituto y todas las nuevas operaciones estadísticas que tenían recogida de datos debían tener un formulario Internet. Para las operaciones ya rodadas, la recogida tradicional debía ser complementada con la recogida vía Internet.

Hoy día disponemos de formulario internet para 10 operaciones

- Encuesta de Condiciones de Vida (ECV)
- Encuesta de Innovación Tecnológica (EIT)
- Estadística sobre actividades de investigación científica y desarrollo tecnológico (I+D)
- Encuesta Industrial (EIA)
- Encuesta de Hostelería y otros servicios (EHS)
- Encuesta de Servicios (SER)
- Estadística de la construcción.
- Encuesta sobre la sociedad de la información (ESI)
- Índice de comercio en grandes superficies (ICGS)
- Programa Anual Estadístico 2006

y probablemente en el año 2006 todas las operaciones de Eustat con recogida de datos cuente con su correspondiente formulario internet.

Los porcentajes de respuesta a través de este medio no son muy altos por ahora, pero previsiblemente se incrementarían a medida que se extiende el uso de la tecnología y los informantes se acostumbren a utilizar este medio. Por otra parte las ventajas de recoger la información directamente del encuestado y de forma on line son claras.

A continuación se muestra un cuadro resumen con estadísticas de recogida de cuestionarios por internet para diferentes operaciones. Es interesante señalar la diferencia entre operaciones de carácter económico y de población.

	Año (primero y último)	Total Cuestionarios	Internet	% Internet
Panel de Servicios	2003	652	167	25,6
	2004	322	95	29,5
Condiciones de Vida Familias	2004	4985	216	4,3
Condiciones de Vida Individuos	2004	4908	164	3,3
ESI Empresas	2002	7094	448	6,3
	2004	7439	1103	14,8
EIT				
	2004	3042	310	10,2
I+D	2003	756	180	23,8
	2004	888	220	24,8
EHOS (Servicios)	2004	2136	481	22,5

Consecuencias del uso de formularios Internet

La introducción de formularios de este tipo ha tenido consecuencias inmediatas sobre la forma de enfocar la recogida de información en Eustat

La primera fue que la misma encuesta tuviera **dos entradas** muy diferentes por que algunas encuestas contaban ya con un formulario Blaise. La creación de un formulario específico para internet duplicó los cuestionarios.

Por un lado, la entrada Blaise permitía grabar datos a imagen y semejanza de los cuestionarios papel y la programación incluía la **validación con errores a diferentes niveles y control de grafo**. También permitía la **revisión** por parte de los inspectores y encuestadores de los cuestionarios una vez grabados.

La grabación del cuestionario necesita la muestra para, por lo menos hacer la autenticación del encuestado y eso origina problemas de seguridad al exponer datos de carácter personal en zonas no totalmente seguras.

Por **seguridad**, fue necesario planificar la extracción de registros/cuestionarios de la base de datos del cuestionario internet a los ficheros Blaise y mantener **la integridad entre ficheros**.

Para dar solución a estos problemas de integración, de control, de revisión de la información y de mantenimiento del cuestionario el modelo de recogida de datos actual solo tiene un **cuestionario de entrada** accesible desde internet y desde la LAN de Eustat al mismo tiempo. El cuestionario apunta a **una sola base de datos** activa solo durante el periodo de encuestación, accesible directamente para los técnicos del Area de Campo y que se vuelca a producción y se borra al finalizar el plazo. Tiene diferentes grados de validación: El primero y más simple, en el propio cuestionario de forma **ON-LINE** y para cualquier nivel de grabación (interno o externo), y el segundo mediante (normalmente) procedimientos PL/SQL en **Batch**, solamente para los técnicos del instituto.

Mención aparte merece un formulario especial para gestionar la Hoja de Ruta que regula la tipología de los accesos y las funciones a ejecutar para el personal de Eustat.

Apartados de un cuestionario

Los formularios que diseña y usa actualmente Eustat se dividen en tres apartados claramente diferenciados: La cabecera que es una pantalla de presentación, la encuesta que se divide en páginas internet ajustándose a los diferentes apartados del diseño de la encuesta en papel y la pantalla final que ofrece al usuario información sobre la propia encuesta.

Página de cabecera

La cabecera se divide a su vez en:

1. **Zona de títulos y logos.** Contiene de forma centrada el título de la encuesta en euskera y castellano y los logotipos del Gobierno Vasco y de Eustat en las esquinas superiores. Es importante que estas señales de identificación estén bien visibles y destacados para que el usuario tenga claro en que "site" está.
2. **Símbolos de seguridad.** También aparece en la pantalla claramente destacados el logotipo que recuerda que la información proporcionada esta sujeta a secreto estadístico. La imagen es idéntica a la tradicional de Eustat empleada en los cuestionarios papel.

Se incluye el link a VeriSign para que el usuario pueda confirmar que:

- Este sitio posee un Certificado Digital de Servidor emitido por VeriSign.
 - ACE ha verificado la identidad de la Empresa y EUSKAL ESTATISTIKA ERAKUNDEA ha aportado documentación que demuestra su derecho a utilizar ese nombre.
 - El sitio opera legítimamente en nombre de EUSKAL ESTATISTIKA ERAKUNDEA.
 - Toda la información enviada a este sitio bajo una conexión SSL se encuentra encriptada y protegida del acceso por terceras partes.
3. **Información.** Se proporciona información bilingüe sobre temas técnicos (resolución, versiones, exploradores,..) , notas sobre el funcionamiento de la encuesta (decimales, guarda de páginas completas,..), teléfono de consulta y disposiciones legales referidas al secreto estadístico, a la Ley de Estadística y al uso de los datos. Un ejemplo puede ser el utilizado para la EIA 2004

<ul style="list-style-type: none"> • Al abandonar la sesión se guardarán las páginas completadas
<ul style="list-style-type: none"> • Datos económicos en euros sin decimales. Datos referidos a 2004
<ul style="list-style-type: none"> • Teléfono de consulta: 900 840 500
<ul style="list-style-type: none"> • Todos los datos obtenidos mediante la cumplimentación de este cuestionario van a ser utilizados para la realización de la operación estadística "Estadística Industrial", regulada con el número 112311 en el Decreto 321/2003, por el que se aprueba el Programa Estadístico Anual de 2004, y a tal fin, van a ser incorporados para su tratamiento a un fichero automatizado.
<ul style="list-style-type: none"> • Conforme a los artículos 9 a 16 de la Ley de Estadística de la Comunidad Autónoma de Euskadi, en relación con la mencionada Ley del Plan Vasco de Estadística, Ud. está obligado a suministrar la información estadística que se le requiere.
<ul style="list-style-type: none"> • Si lo desea, tiene la posibilidad de ejercitar los derechos de acceso, rectificación y cancelación ante la Dirección General de Eustat, C/ Donostia-San Sebastián 1, 01010 Vitoria-Gasteiz.
<ul style="list-style-type: none"> • Todos los datos suministrados quedan amparados por el secreto estadístico regulado en los artículos 19 a 23 de la Ley de Estadística de la Comunidad Autónoma de Euskadi.

Identificación del encuestado

Consiste en dos campos de entrada que piden la identificación de usuario y la contraseña que le autentifica. Habitualmente, en la carta que se envía a la empresa/familia para comunicarle que forma parte de la muestra de una determinada encuesta, aparecen los dos códigos que se han generado previamente de forma aleatoria y forman parte del fichero muestral. También se generan códigos para los usuarios internos del Instituto (técnicos, inspectores e encuestadores), ya que según el código con el que se entra, el formulario ofrece diferentes niveles de funciones.

Páginas de encuesta

Se mantienen los logos de Eustat y del Gobierno Vasco y el título de la encuesta de forma semejante a la página inicial.

La página tiene un cuerpo de preguntas que normalmente hacen referencia uno o varios apartados (bloques) naturales del cuestionario en papel. No hay preguntas sueltas y cada página viene determinada por el tamaño (aproximadamente 100 ks) .

El control de grafo, si es necesario, se realiza solo en el ámbito de misma página, es decir, no se puede ir de una página a otra mediante el salto entre preguntas unidas por grafo. Se trata de evitar el desplazamiento horizontal (scroll horizontal) y, en algunas encuestas, los saltos de grafo se hacen automáticamente situando el cursor en la pregunta destino.

Las validaciones simples, como de tipo, de rango,.. se realizan en el cliente al salir del campo. Los campos calculados (sumas,...) también se actualizan cuando se sale de uno de los campos participes.

A pie de página se muestra un contador de páginas sobre el total de páginas de cuestionario para que el encuestado tenga conocimiento de en qué parte del formulario se encuentra. En la parte central se encuentran dos botones para moverse adelante y atrás entre las páginas del cuestionario.

Estos botones son muy importantes porque realizan las siguientes funciones antes de mover entre páginas:

Lanza todas las validaciones de servidor de la página en curso marcando todos los campos erróneos con un símbolo específico (❗), muestra un mensaje genérico anunciando que existen errores y no permite cambiar de página hasta que los errores estén solucionados.

Pasando el curso por encima del símbolo ❗ aparece la descripción específica del error.

En el caso de no existir errores, graba los datos de la página en la base de datos y carga la siguiente en secuencia hacia delante y atrás.

En cualquier otra circunstancia la información no se graba y se pierden los datos de la página actual

Si el texto no es excesivamente largo, la descripción de las preguntas suele ser bilingüe. En caso contrario se crean dos cuestionarios distintos y el idioma se selecciona en la primera pantalla.

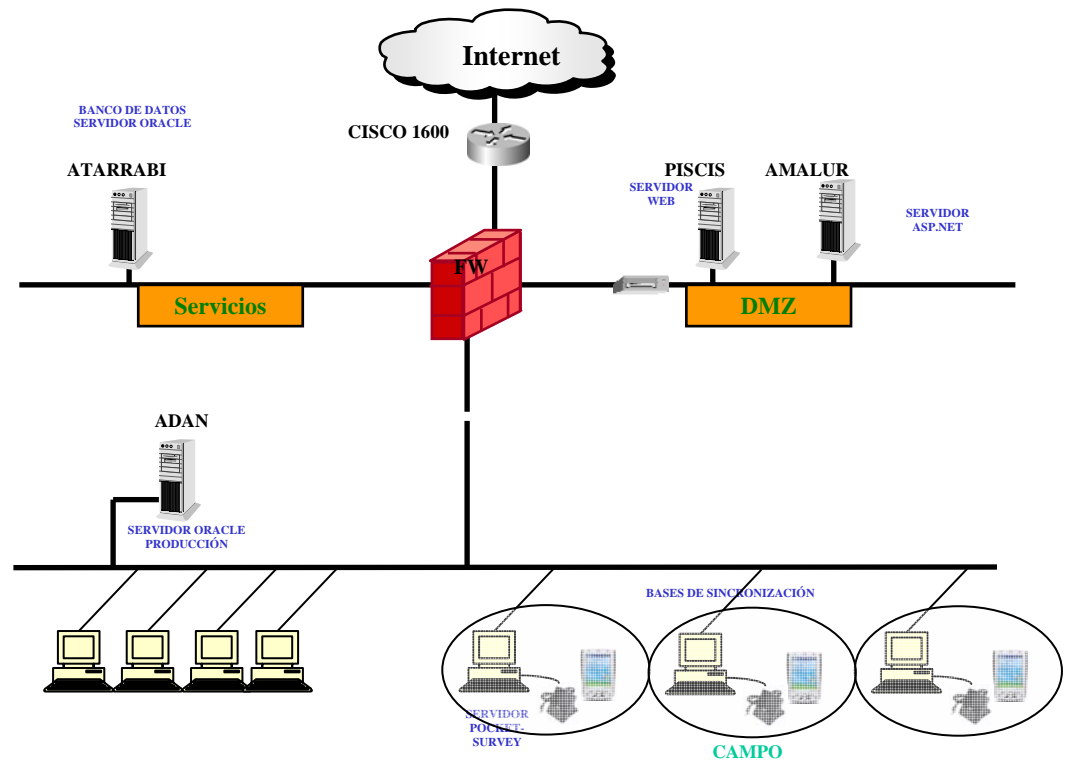
Página final

Esta página, que también mantiene el formato de títulos y logos, suele proporcionar información de contacto con Eustat, un campo de observaciones y alguna información referida a qué periodo corresponden los datos.

El objeto principal es el botón de terminar que una vez pulsado permite al usuario dar por finalizada la encuesta y graba en la base de datos una marca especial (fecha de cierre) para tener constancia de esta acción. Las consecuencias de terminar la encuesta son:

- La encuesta queda en modo de consulta, es decir, el usuario no puede modificar ningún dato. En algunas ocasiones se deja un día o dos de latencia antes de cerrarla definitivamente por si son necesarios cambios de última hora.
- El usuario puede imprimir un resumen de los datos con el formato de la encuesta.
- La encuesta queda inaccesible para el usuario una vez pasados un periodo preestablecido que puede ser de tres o cinco días.

Arquitectura



Los datos están en la DMZ2 (ATARRABI), el formulario en la DMZ1 (AMALUR) y la entrada se realiza a través del servidor web (PISCIS).

Periódicamente, hay un acceso desde el servidor ORACLE de producción (ADAN) hasta el servidor de datos de la DMZ2 (ATARRABI) para comprobar qué encuestas están ya cerradas con el plazo de visibilidad caducado para copiarlas de la BD de encuestación y crearlo en la BD de producción.

Entorno de trabajo

El entorno de trabajo para la programación de los formularios de Internet es el mismo que APRA el resto de trabajos informáticos del Instituto.

En principio se utilizó como SGBDE MS-Access por problemas de arquitectura, en la actualidad se emplea ORACLE 9.i

En el cliente de desarrollo se instala como sistema operativo Microsoft Windows XP Professional y como herramienta de desarrollo y producción, Microsoft Visual Studio.NET 2003 (Microsoft Development Environment 2003 y Microsoft.NET Framework 1.1) y dentro de este entorno, se programa con el lenguaje Microsoft Visual Basic.NET.

En el servidor se instala como sistema operativo Microsoft Windows 2000 Server con IIS.

Todas las páginas se diseñan para Microsoft Internet Explorer (versión 6.0) ya que cuando se inició la encuestación por internet se vio que era muy costoso producir versiones APRA otros exploradores y quedaba fuera de nuestro alcance.