

**TÉRMINOS DE REFERENCIA PARA EL SERVICIO DE CAPTURA DE DATOS CON  
TECNOLOGÍAS OCR/OMR/ICR DE FICHAS ÓPTICAS Y DIGITALIZACIÓN Y CAPTURA DE  
DATOS EN UNA MUESTRA DE PRUEBAS DE LA EVALUACIÓN CENSAL DE  
ESTUDIANTES 2011 (ECE 2011)**

**1. CAPTURA DE DATOS OCR/OMR/ICR DE FICHAS ÓPTICAS Y CONSISTENCIA Y  
VALIDACIÓN DE DATA:**

La captura se realizará con **valor legal** a 510 150 páginas A4 aproximadamente. Deberá realizarse la captura del total de fichas ópticas que se reciban, aún cuando algunas de éstas estuvieran en blanco.

Previo a la captura de datos OCR/OMR/ICR de las fichas ópticas, éstas deberán clasificarse por operativo y *correlativo MED* por cada provincia. Las fichas ópticas tienen una grapa en la parte superior izquierda, la que deberá desengraparse para la captura e inmediatamente después de ésta, deberán volverse a engrapar para su posterior modulado y almacenaje.

**Se deberá realizar la captura OCR/OMR/ICR (en simultáneo) de acuerdo al siguiente detalle:**

- Provisión de hardware, incluyendo lectores ópticos, PC's, servidores y otros.
- Provisión de software de captura, validación, consistencias y generación de reportes.
- Preparación de formularios (proceso de conteo y loteo) agrupándolos por IE de cada UGEL.
- Captura de información de los cuadernillos de fichas ópticas empleando tecnología ICR para los campos abiertos escritos a mano, tecnología OMR para los campos de respuestas (burbujas) y tecnología OCR para los caracteres preimpresos como data variable. El estimado de caracteres **ICR** a capturar por cuadernillo es de **1 242 caracteres alfabéticos y 350 caracteres numéricos** aproximadamente. El estimado de campos **OMR** a capturar por cuadernillo es de **2 430** aproximadamente. Para los campos **OCR** (código modular, nombre de IE, aplicador, sección evaluada) se calcula **64 caracteres** aproximadamente.
- Cada ficha leída deberá tener un número de "secuencia de captura" que se imprimirá al momento mismo de la captura (en línea).
- Proceso de control de calidad vía módulos KFI (*key from image*).
- Consistencias y validaciones.
- Proceso de generación de exportación a base de datos.
- Los campos OMR (burbujas) del área de respuestas de las fichas ópticas serán capturados por medio de un equipo lector (*hardware*), empleando cabezales para captura óptica de marcas con capacidades de discriminación de cuando menos 16 niveles de intensidad de marca, para discriminar un borrón o mancha de una marca bien hecha. Los datos capturados serán transferidos directamente del equipo lector al computador sin posteriores verificaciones o manipulación de datos.
- Los campos ICR/OCR (caracteres escritos a mano y preimpresos en las fichas ópticas) serán capturados a partir de las imágenes transferidas del mismo equipo lector (*hardware*) al computador. El *software* de este equipo lector deberá enlazar de forma automática los campos OMR con los campos ICR/OCR de una misma ficha óptica.
- Este proceso de captura OMR/ICR/OCR deberá realizarse de manera simultánea, es decir el mismo equipo lector deberá realizar la captura de los datos OMR/ICR/OCR a la misma vez.
- El proceso de reconocimiento de los caracteres escritos a mano (ICR), deberá ser verificado por personal especializado para lograr 100% de calidad de data.
- Los campos leídos con la tecnología OCR (caracteres pre impresos, data variable) se deberá utilizar para la consistencia y verificación de la data.
- Para los casos que lo requiera, se deberá realizar un proceso de recuperación de cuadernillos deteriorados o con datos incompletos.



- El contratista asignará un área de al menos 70 m<sup>2</sup> para uso exclusivo del servicio de captura.
- Al finalizar la captura de datos, las fichas ópticas (previamente reengrapadas) deberán ser moduladas por operativo (ECE, MC y EQP) en paquetes por IE (en el orden del *correlativo MED*) y embaladas por provincia y región.

**Para la recepción de los formularios y cuadernillos de evaluación y almacenamiento de las cajas:**

- El contratista deberá asignar un área de 250 m<sup>2</sup> para la recepción de las 10,000 cajas aproximadamente, conteniendo las fichas ópticas y cuadernillos de evaluación, entre otros.
- Tanto las fichas ópticas como los cuadernillos de evaluación deberán ser contados en presencia del operador logístico de la ECE 2011 (responsable de las entregas de éstos). Este conteo se realizará a través de lectores laser de códigos de barras manuales (escáners manuales), los que a través del código de barras que tienen los cuadernillos de evaluación y las fichas ópticas identificarán la cantidad recibida. Adicionalmente a la lectura del código de barras de todos los instrumentos, se deberá verificar que todas las FOAR tengan la última página, cuyo número está especificado en la primera página. Luego de ello, elaborarán un reporte detallando la cantidad de fichas ópticas y cuadernillos recibidos, y el dato exacto de aquellos que faltaran (si fuera el caso). Este reporte deberá estar firmado por un representante del contratista y del operador logístico de la ECE 2011.
- La recepción se realizará por provincia completa, con excepción de los casos de la provincia de la Convención (Cusco), cuyos distritos Kimbiri y Pichari vendrán incluidos en la provincia de la Mar (Ayacucho); y la provincia de Manú (Madre de Dios), cuyos distritos Huetupe y Madre de Dios están incluidos en la provincia de Tahuamanu (Madre de Dios).
- La entrega de cada camión se realizará mediando un Acta, la que detallará la cantidad de cajas entregadas por operativo y será firmada por los representantes del operador logístico de la ECE 2011 y el contratista.
- Para ello, el contratista asignará la cantidad de personal necesario para que realice el conteo (uno por uno) tanto de las fichas ópticas (incluye FOAR y Cuestionarios) como de los cuadernillos de evaluación.
- Al concluir con el inventario del total de instrumentos recibidos, se verificará si el total entregado/recibido coincida con la cantidad programada de entrega. En el improbable caso que faltara alguna ficha óptica (llena o en blanco) o algún cuadernillo de evaluación (llenó o en blanco), el contratista deberá informarlo inmediatamente a la UMC.

**Para la consistencia de la data:**

- Consistencia entre las bases provistas de Códigos Modulares, Anexos, Instituciones Educativas, *correlativo MED*, etc. utilizados en la impresión de los documentos (FOAR – cuestionarios) y las bases con la información capturada con tecnologías OCR/OMR/ICR.
- Consistencia con los números de lithocódigos (de acuerdo a la distribución logística establecida).
- Verificación de que los campos necesarios para la identificación estén completamente llenos, deteniéndose el proceso si algún código se encuentra incompleto o si no existe en la base de datos.

**Para la entrega de la base de datos:**

- Emisión de reportes de lo entregado y lo procesado.
- Grabación y entrega de la data capturada y de los reportes finales, de acuerdo a la estructura solicitada.

- Grabación y entrega de las imágenes capturadas con sus respectivos índices, en un disco duro externo con visores – buscadores para fácil acceso.
- Acta firmada por el Fedatario que certifica el valor legal de las imágenes capturadas.

**Validación de la data entregada:**

- Asignación de un ambiente de trabajo para 12 personas designadas por la UMC (con conexión para 12 PC).
- Asignación de 6 PC durante 4 semanas (contabilizadas después de la entrega final de la base de datos), con las siguientes características: Core-i3, 4gbRAM, HD 500. Solo 1 de las PC deberá tener salida a internet.
- Asignación de una impresora (blanco y negro), para papel tamaño A3.

**Plazo para captura de datos OCR/OMR/ICR de fichas ópticas y consistencia y validación de data:**

**12 días** a partir de la entrega del total de las fichas ópticas MC y EQP aplicadas.

**15 días** a partir de la entrega del total (las entregas serán parciales) de las fichas ópticas ECE aplicadas.

**2. DIGITALIZACIÓN Y CAPTURA DE DATOS ICR DE UNA MUESTRA DE CUADERNILLOS DE PRUEBAS:**

La digitalización y captura de datos ICR se realizará con **valor legal** a 719 200 cuadernillos aproximadamente (5 753 550 imágenes aproximadamente).

Los cuadernillos cuyas imágenes deberá capturarse tienen un tamaño de 21.5 x 27.5 cerrado aproximadamente. Estos cuadernillos tienen 2 grapas en el lomo, las que deberán retirarse para la captura de datos e inmediatamente después de ésta, deberán volverse a engrapar para su posterior modulado y almacenaje.

Cada cuadernillo tiene aproximadamente 100 campos tipo marca (ICR), 50 caracteres del tipo OCR y 60 caracteres del tipo ICR.

La estructura de la base de datos, resultado de la digitalización, será establecida por la UMC.

**Los cuadernillos a digitalizar serán procesados por un software de reconocimiento de tecnología ICR para capturar la información contenida en ellos.** El proceso de digitalización y captura ICR deberá contemplar lo siguiente:

- Creación de formato *output* y definición de áreas de captura óptica.
- Creación de módulo de consistencias y validaciones.
- Creación de módulo de control de calidad.
- Creación de módulo de exportación de datos.
- Configuración de visor para búsqueda posterior de imágenes.
- Pruebas generales.

**Producción:**

- Provisión de *hardware*, incluyendo PC's, servidores y otros.
- Provisión de *software* legal (licencias autorizadas) de reconocimiento y verificación ICR.
- Recepción de documentación.
- Proceso de conteo y loteo por unidades de producción.
- Proceso de corte y preparación de cuadernillos.
- Proceso de digitalización de las fichas empleando escáners con capacidades de *dropout*.
- Proceso de reconocimiento.



- Proceso de verificación en estaciones de trabajo.
- Control de calidad de imágenes vs data verificada para una óptima calidad de datos.
- Verificación y consistencia de la información de la data capturada de los cuadernillos de pruebas respecto a la data obtenida de las fichas ópticas.
- Proceso de generación de exportación de base de datos.
- Proceso de grabación de imágenes en disco duro externo y generación de visor/buscador.
- Los cuadernillos deberán ser modulados en paquetes por IE (en el orden del *correlativo MED*) y embalados por provincia y región.

#### **Productos a entregar:**

- Emisión de reportes de lo entregado y lo procesado.
- DVD en original y copia con la grabación de los documentos digitalizados y entrega de la data capturada y de los reportes finales, de acuerdo a las estructuras solicitadas.
- Acta firmada por el Fedatario que certifica el valor legal de los documentos digitalizados y la data capturada.

#### **Plazo para la digitalización y captura de datos ICR:**

**15 días** a partir de la entrega del total (las entregas serán parciales) de los cuadernillos en el almacén del contratista en Lima.

#### **\*Medidas mínimas de seguridad para los servicios 1 y 2:**

- Control de acceso general vigilado durante las 24 horas del día.
- Control de acceso restringido a las diferentes áreas del local del contratista.
- Revisión del personal que labora en el local del contratista, tanto al ingreso como a la salida de la jornada laboral.
- Cámaras de video para monitoreo de las áreas de producción. El contratista deberá realizar lo necesario para que la UMC monitoree en tiempo real las áreas de producción a través de una página web autorizada por el contratista.

### **3. ALMACENAMIENTO DE LOS DOCUMENTOS:**

El periodo de almacenaje será de **6 meses**, contados a partir de la finalización de los servicios de captura de datos OMR/ICR/OCR de las fichas ópticas y digitalización y captura de datos ICR de los cuadernillos de pruebas.

La cantidad a almacenar será alrededor de 4 000 cajas aproximadamente y este almacenamiento deberá realizarse en un lugar que cuente con las medidas de seguridad para preservar la conservación y mantener segura la documentación, de acuerdo a los siguientes requisitos mínimos:

- La construcción del almacén debe estar realizada en material noble. El techo debe ser no inflamable y estar equipado con extractores de aire y canaletas de drenaje para prevenir daños por lluvias.
- El almacén debe tener sensores para la prevención temprana de incendios y sistemas de alarmas y monitoreo permanente y una adecuada distribución de extintores de PQS y agua.
- Estrictos controles de seguridad y control de acceso y vigilancia las 24 horas al día.
- Sistemas de Circuito Cerrado de TV (CCTV).

Al finalizar el periodo de almacenaje, todas las fichas ópticas y cuadernillos deberán ser devueltos a la UMC; en el lugar que éste lo indique.

**\* SUPERVISIÓN DEL SERVICIO POR PARTE DE LA UMC:**

Se permitirá el ingreso a las instalaciones del contratista a las personas señaladas por la UMC para la supervisión de los trabajos de manera permanente o en los momentos que éste lo requiera.

**\* CONFIDENCIALIDAD:**

Para asegurar la confidencialidad de los instrumentos y la data a capturar, este servicio deberá ser realizado por un contratista que deberá comprometerse a usar con reserva y confidencialidad toda la información proporcionada y a no entregar ningún documento o informe a ninguna persona o institución sin autorización expresa de la UMC.

**\* EXPERIENCIA SOLICITADA:**

Deberá sustentar experiencia en captura de datos con tecnología OMR/ICR.

**\* FORMA DE PAGO**

El pago total se realizará de la siguiente manera, luego de la conformidad de servicio emitida por el área usuaria:

80% a la entrega de la base de datos y los reportes de las capturas OMR/ICR/OCR y la digitalización y captura ICR.

20% a la entrega de las Actas de inicio y fin del almacenamiento y de la entrega física de los documentos.

**\* FINALIDAD PÚBLICA**

El servicio requerido tiene como finalidad pública el conocer el nivel de logro de aprendizaje de los estudiantes que están terminando segundo grado de primaria en comprensión lectora y matemática y de los estudiantes de cuarto grado de primaria, en las IE que están implementando el Programa de Educación Intercultural Bilingüe (EIB), en comprensión lectora. Esto permitirá informar los resultados obtenidos a los padres de familia de los estudiantes evaluados, IE, UGEL, DRE y Gobiernos Regionales, de tal manera que, cada uno en su ámbito, planifique y desarrolle estrategias necesarias para que los estudiantes evaluados logren superar las dificultades evidenciadas en los resultados y consigan los aprendizajes esperados.



11/11/11

11

11