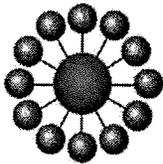


CCC
AI
1478



CONEICC

Guadalajara, Jal.
25 de junio de 2006

Ing. Javier Coronado
DIGIX, S. A.
Presente:

Estimado Javier:

Por medio de la presente te solicito, de la manera más atenta, sirvas elaborar una cotización del servicio de digitalización del acervo documental de tesis de licenciatura en Ciencias de la Comunicación, que se encuentran en el Centro de Documentación del Consejo Nacional para la Enseñanza y la Investigación de las Ciencias de la Comunicación. (Coneicc)

El Coneicc es una asociación civil, sin fines de lucro, que agrupa a las universidades a nivel nacional, que cuentan con programas de licenciatura en Ciencias de la Comunicación y que tiene el propósito de prestar servicios de apoyo en la formación de comunicadores. Consideramos que uno de estos servicios es facilitar la consulta del trabajo de investigación que se realiza, a nivel licenciatura, en nuestro país, y que queda plasmado en las tesis que elaboran los estudiantes que han cursado sus estudios en las universidades afiliadas al Consejo.

El acervo cuenta actualmente con cerca de 3,900 tesis y un aproximado de 600,000 hojas que requieren ser trasladadas a archivos digitales a fin de que puedan ser consultadas en línea, de manera remota.

Es importante señalar que el Centro de Documentación del Coneicc ya cuenta con un sistema y plataforma de clasificación, y una base de datos en la que todos los documentos tienen clave de catalogación y los elementos básicos para la búsqueda de los documentos.

El servicio que solicitamos a Digix, en caso de ser nuestra mejor opción, es, solamente, el traslado de los documentos a formato digital. Estamos muy interesados en que una empresa profesional, con una trayectoria significativa en este campo, como lo es Digix, sea la empresa a la que le confiemos esta tarea.

Sin más por el momento, quedo a tus órdenes para cualquier aclaración o comentario

Atentamente

Mtra. Magdalena Sofía Paláu C.
Coordinadora de documentación
Coneicc



Catálogo del Consejo Nacional para el Estudio y
la Investigación de las Ciencias de la
Comunicación, **CONEICC**

IT E S O



CONTENIDO

Introducción	3
1. Porque digitalizar el contenido	4
2. Objetivos de la propuesta	4
3. Parámetros de trabajo requeridos	5
3.1 Sitio	5
3.2 Operaciones de digitalización	6
4. Equipo	6
4.1 Dispositivos de escaneo	6
4.2 Reconocimiento Óptico de Caracteres (OCR)	7
4.3 Personal de DIGIX. S. A. DE C.V.	7
5. Proceso de digitalización	8
5.1 Barrido óptico.	10
5.2 Control de calidad para páginas incompletas	10
5.3 Recorte (manual y automático)	10
5.4 Enderezamiento	10
5.5 Sorteado / control de calidad (páginas faltantes)	11
5.6 Nuevo barrido	11
5.7 Índice con dos campos llave	12
6. Las ventajas de los servicios de Digix	14
7. Recomendaciones	15
8. Características del proyecto	15
9. Plazos de entrega	15
10. Servicios incluidos	16
11. Costos	16



Introducción

La tecnología es una herramienta básica utilizada casi para cualquier actividad de la vida contemporánea.

A nivel internacional la aceptación de nuevas tecnologías como herramienta de trabajo facilita la productividad y lo que es mejor, permite a las empresas o instituciones ser más competitivas; lo que da como resultado una estandarización en procesos o mejor conocido como *globalización*.

México no es la excepción a este fenómeno, aunque con reservas se ha integrado al mismo; lo que ocasiona que las empresas asuman un mayor esfuerzo con responsabilidad para estar al mismo nivel competitivo que los mercados internacionales exigen en el corto y largo plazo.

DIGIX S.A. DE C.V. Nace como una empresa comprometida con las organizaciones mexicanas y, sobretodo, con el avance tecnológico de México al ser experta en digitalización, manejo de documentos históricos y comercio electrónico.

En el presente documento se muestran como se desarrollan dichos procesos y las ventajas que con lleva el uso de tecnologías de este tipo para una organización como el **CONEICC**.



1. Porque digitalizar el contenido

Las razones por las que se propone digitalizar el contenido del archivo del **CONEICC** se mencionan a continuación:

El formato digital permite optimizar el manejo de información mediante una manera más ágil y sencilla, pues en cuestión de segundos la información puede ser presentada al usuario para el fin requerido; estos usos van de consultas hasta investigaciones formales.

El contenido digital incrementa considerablemente la accesibilidad pues su acceso puede ser remoto (Internet), las 24 horas del día, los 365 días del año, en cualquier parte del mundo, sin restricción alguna.

Los archivos digitales ofrecen un número interminable de consultas, sin que esto incremente el costo del contenido.

La conservación de documentos es uno de los principales objetivos ya que el microfilm y papel son formatos caducos, lo cual representa una cuestión importante pues está latente la pérdida del patrimonio en caso de siniestro o sustracción del contenido.

La información es una herramienta que además de ser muy utilizada es también valiosa, lo cual representa una oportunidad de negocio para la empresa u organización que tome la decisión de modificar su acervo a formato digital.

2. Objetivos fijados

- a) Demostrar que la tecnología de **DIGIX S.A. DE C.V.** ofrece una solución aceptable a los requisitos técnicos definidos por el **CONEICC**.
- b) Describir como los procesos y tecnologías empleadas por **DIGIX S.A. DE C.V.** complementan la experiencia y la capacidad instalada del archivo **CONEICC**.
- c) Demostrar la utilidad a largo plazo y la accesibilidad de los formatos digitales a ser usados; propuestos por **DIGIX S.A. DE C.V.**
- d) Describir los procesos de producción, los recursos humanos y de equipo necesario para el proyecto.



3. Parámetros de trabajo requeridos

Dichos parámetros son fijados una vez que la negociación se formaliza con la empresa o institución y todo está sustentado en el número de imágenes a realizar, el tiempo de entrega (que indiscutiblemente varía por el número de imágenes y procesos pactados). Más adelante se explicará cuántos y cuáles son los mejores procesos para el proyecto.

Cabe mencionar, que estos procesos son adecuados a las necesidades específicas del archivo del **CONEICC**. A continuación se enlistan los parámetros de trabajo:

- a) Demostrar que la tecnología de **DIGIX S.A. DE C.V.** ofrece una solución aceptable a los requisitos técnicos definidos por el **CONEICC**.
- b) Los procesos de digitalización son realizados en su totalidad fuera de las instalaciones del archivo **CONEICC**, para facilitar el trabajo por etapas sugerido.
- c) **DIGIX S.A. DE C.V.** entregará el resultado digital en un formato físico seleccionado por el **CONEICC**.
- d) Demostrar la utilidad a largo plazo y la accesibilidad que muestran los formatos digitales al ser usados.
- e) Describir los procesos de producción, los recursos humanos y de equipo necesarios para el proyecto.

3.1 Sitio

El sitio donde se han de realizar los procesos digitales esta ubicado en la planta de un edificio en la zona centro de Guadalajara debido a que posee los siguientes atributos:

1. Situado en una zona central
2. Servicio de seguridad las 24 horas/ 7días a la semana/ todo el año
3. Instalaciones eléctricas y de fibra óptica funcionales
4. Conexión a Internet operacional
5. Espacio adecuado disponible para expandir la operación a varios tamaños



3.2 Operaciones de digitalización

Las operaciones de digitalización convertirán el papel de las obras del **CONEICC**, en imágenes digitales bitonales (blanco y negro), a 300 DPI en formato **PDF**. Este proceso incluye los siguientes puntos:

1. Digitalización de las obras por medio de scanners especiales para libros
2. Control de calidad del proceso de escaneo
3. Perfeccionamiento automático y manual de imágenes para corregir bordes irregulares, manchas, enfoque, iluminación o algún otro aspecto de la imagen que se pueda mejorar con un esfuerzo razonable
4. La clasificación de las imágenes escaneadas según el número de obra
5. Almacenamiento de los archivos de texto é imágenes en la estructura de alojamiento

Cabe mencionar que cualquiera de los pasos mencionados anteriormente puede ser excluido a consideración del **CONEICC**.

4. Equipo

4.1 Dispositivos de Escaneo

La calidad de la impresión de las colecciones varía década a década. Es la intención de **DIGIX S.A. DE C.V.** digitalizar las obras del archivo del **CONEICC** con la mejor calidad posible que permitirá un proceso de actualización técnica más eficiente.

1. El establecimiento y el ajuste delicado de los parámetros técnicos para los dispositivos con el fin de obtener una cierta calidad de imágenes, combinado con el tiempo requerido para digitalizar cada obra individual determinará el número de dispositivos que se requerirán.
2. Las imágenes que se encuentran dentro de cada obra pueden variar en claridad y tono del papel, por lo que algunos materiales no llenan las expectativas de calidad de **DIGIX**. Así, en la medida en que la calidad de la imagen se mejore y se ingrese al sistema, mejor será el resultado de un futuro proceso de **Reconocimiento Óptico de Caracteres (OCR)**.



4.2 Reconocimiento Óptico de Caracteres (OCR)

Este es un software exclusivo y desarrollado por **DIGIX** empleado en el reconocimiento de texto dentro de las imágenes. Además, crea una base de datos la cual hace posible concretar una búsqueda específica dentro de un mar de información. Cabe mencionar que este software es único en su especie, por lo tanto ninguna empresa nacional o internacional puede ofrecer este servicio.

4.3 Personal de DIGIX S.A. DE C.V.

El personal de **DIGIX S.A. DE C.V.** estará involucrado en la totalidad de los procesos y procedimientos descritos en este documento. Todos los procesos llevados a cabo tienen amplia supervisión para garantizar su plena funcionalidad.

El personal está seleccionado bajo los siguientes parámetros indispensables para poder realizar el trabajo con responsabilidad y calidad:

- Un mínimo de uno o dos años de un programa micro-computacional de aplicaciones, estudios y tecnología de información
- Habilidad en la entrada de base de datos
- Capacidad de aprender nuevos sistemas rápidamente
- El mejoramiento y procesamiento de imágenes digitales con productos de software especializados
- Que pueda hablar inglés con facilidad
- Que trabaje dentro de los procedimientos especificados del flujo de trabajo
- Que trabaje con las expectativas de control de calidad y de aseguro de calidad
- Entienda la manipulación de documentos.
- Manejo de los dispositivos de barrido óptico para libros.
- Confiable, bien organizado y adaptable
- Requerimientos de horas extras y de trabajo de turno



5. Proceso de digitalización

DIGIX ofrece la solución de principio a fin en servicios de digitalización, de alojamiento y de una base de datos de Internet buscable.

El proceso de digitalización de DIGIX permite que los archivos históricos de periódicos, documentos o cualquier colección de microfilm puedan ser barridos, digitalizados y publicados por vía Internet.

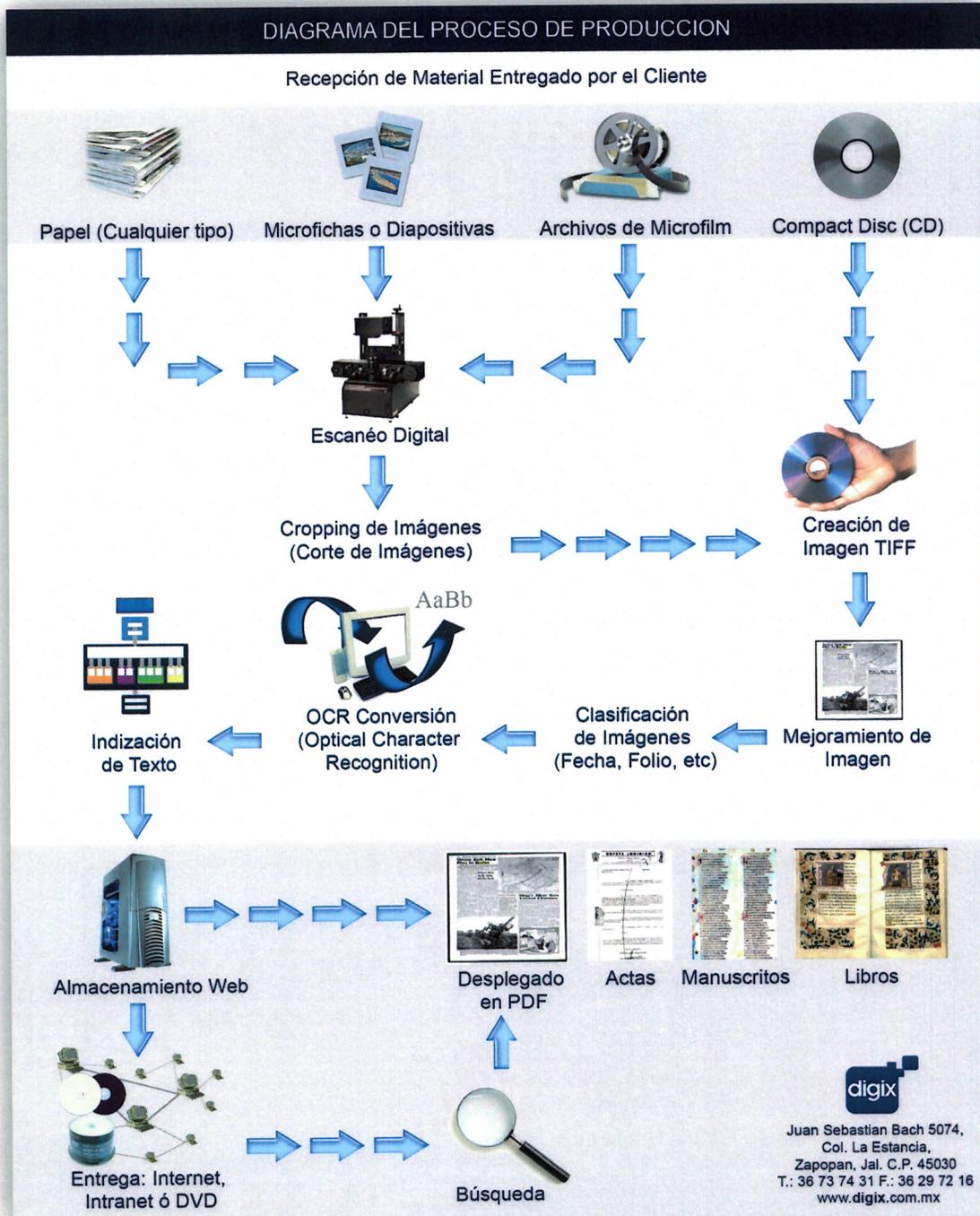
Esta solución de fases y de bajo costo permite la réplica de documentos como fueron creados originalmente. Además, **DIGIX** ofrece servicio de alojamiento para aquellos clientes que no tienen los recursos financieros, técnicos o humanos para administrar el contenido desde su oficina.

El proceso de digitalización de **DIGIX** fue diseñado para ser flexible, ofreciendo una gran variedad de opciones y requerimientos específicos de procesamiento, incluyendo:

Imágenes de dos tonos (Blanco y Negro)
OCR Reconocimiento Óptico de Caracteres
Indexación
Eliminación de Manchas
Enderezamiento
Recorte de Imágenes



El Proceso de Digitalización de DIGIX S.A. DE C.V. está representado en el diagrama siguiente:





El proceso de digitalización de **DIGIX** proporciona el barrido de imágenes en dos tonos, blanco y negro. Una imagen que es barrida en dos tonos puede ser perfeccionada de muchas maneras, incluyendo la eliminación de manchas (quitar manchas minúsculas de polvo o de suciedad), enderezamiento (endereza la imagen en base al texto y orientación de la foto) y el recorte de imágenes (elimina los bordes no necesarios alrededor de la imagen).

En algunos casos, las imágenes se pueden beneficiar de todos los aspectos de perfeccionamiento. Esto se logra al mejorar la imagen típicamente, después del grabado de TIFF. Un operador del dispositivo de barrido óptico sigue los procedimientos al barrer una imagen y al crear una imagen TIFF con el reconocimiento óptico de caracteres.

5.1 Barrido Óptico

El operador montará el libro en el dispositivo de barrido óptico. Creará una plantilla para ese libro. Varias opciones de perfeccionamiento serán aplicadas en esta etapa de barrido del libro. La plantilla variará de libro a libro. Después de la terminación de la plantilla, el libro será digitalizado y alojado en el servidor de TIFF Grupo 4 de archivos. El operador anotará sus comentarios sobre cada libro.

5.2 Control de calidad para páginas incompletas

Después de grabar las imágenes en el servidor deberán ser revisadas para ver que no estén incompletas o cortadas. Si varias imágenes están recortadas, el libro tendrá que ser barrido de nuevo. Si sólo una o dos imágenes están cortadas deberán ser grabadas y barridas de nuevo más adelante.

5.3 Recorte (manual y automático)

El recorte manual es usado cuando imágenes dobles necesitan ser divididas o cuando el estado del libro está tan deteriorado que el recorte automático no vale la pena. El recorte es hecho por un operador que tiene que definir los parámetros de cada imagen por separado.

El recorte automático es utilizado cuando las imágenes son individuales y el estado del libro es bueno. Los parámetros son definidos en el software y después aplicados a las imágenes en tandas. Los lados horizontales y verticales son recortados, eliminando así las imperfecciones innecesarias, reduciendo el tamaño del archivo.

5.4 Enderezamiento

Si las imágenes son de buena calidad, entonces el enderezamiento puede ser hecho al mismo tiempo que el recorte automático. Si las imágenes tienen que



ser recortadas a mano, entonces los parámetros de enderezamiento serán aplicados en las imágenes por tandas, después de haber completado el recorte.

5.5 Sorteo / Control de Calidad (páginas faltantes)

Después del recorte y enderezamiento, las imágenes tienen que ser clasificadas por número de obra. Este es el formato adecuado que el software de OCR necesita para el reconocimiento óptico de caracteres de las imágenes individuales. En esta etapa, el operador buscará también las páginas faltantes. Si se encuentra una página faltante, tendrá que ser apuntada en el registro de errores. El registro de errores abarcará los siguientes títulos:

- Numero de identificación – número de obra que se está procesando
- Número de página – el número de página y de las páginas faltantes
- Operador – el nombre del operador de sorteo

5.6 Nuevo barrido

Al completarse el proyecto, los nuevos barridos pueden efectuarse. El registro de errores que se ha llevado tiene que ser revisado por el operador de barrido, los libros localizados y el barrido tiene que ser comenzado de nuevo. Los nuevos barridos tienen que ser grabados en el servidor bajo un directorio llamado "Nuevos Barridos ". Luego tendrán que ser clasificados en el formato apropiado por número de obra y tendrán que ser sujetos de nuevo al proceso completo de barrido óptico.

5.7 Índice con dos campos llave

1. Número de obra
2. Fecha de publicación

Una vez barridas, las imágenes pueden ser tratadas con OCR para crear una versión ASCII del texto de la página. Si el OCR es ejecutado con una imagen de TIFF, un archivo de texto ASCII de la página será creado y grabado.

El proceso de OCR de texto completo será ejecutado después de haber mejorado y puestas en índice las imágenes o cuando, para ahorrar tiempo, las imágenes serán procesadas por OCR primero, para después estar listas para el último proceso.

El archivo de texto de OCR creado por el proceso añade 30 a 50 kb al archivo de Imágenes/ Registro de Base de Datos.



DIGIX emplea dos métodos para crear un índice de imágenes. Uno o ambos podrán ser utilizados. Si se ha ejecutado el IOC, todas las palabras extraídas de las imágenes son usadas para elaborar un índice que permite una búsqueda de texto sin límite de la colección de imágenes. También es posible añadir información descriptiva adicional a los registros, describiendo las páginas. Usualmente esto es hecho con XML o con los registros contenidos en una base de datos, describiendo la colección de imágenes. Esta descripción puede incluir el nombre de la obra, fecha, título, o cualquier información relevante que el cliente desee añadir a cada registro. Dependiendo del tipo de ésta información, puede ser añadida manualmente o automáticamente a la base de datos.

Una vez que la imagen ha pasado por el proceso de digitalización, un archivo de TIFF y uno de ASCII de texto son creados de la imagen. Esos archivos son alojados para el acceso del sitio de Web. Cuando un usuario de Web hace una consulta y obtiene un resultado de su búsqueda, la imagen del resultado es presentada en la pantalla con la(s) palabra(s) de búsqueda remarcadas en amarillo.



6. Las ventajas de los servicios de DIGIX son:

- La flexibilidad que presenta una solución de principio a final, utilizando la tecnología más reciente
- Tiempo de procesamiento rápido, debido a la tecnología y procesos de **DIGIX** avanzados con respecto a competidores tanto nacionales como internacionales
- Costos económicos, debido a la opción de fases ofrecida, para completar las especificaciones de los proyectos

DIGIX es una empresa 100% mexicana con operaciones Guadalajara, Jalisco, y con la capacidad de llevar todos los requisitos técnicos del proyecto a cabo. El lugar operacional de **DIGIX** se encuentra en un edificio en el centro histórico de la ciudad antes mencionada.

Dicho lugar tiene las siguientes ventajas: ofrece todas las facilidades porque cuenta con instalaciones de primer nivel y equipo de última generación, cuenta con alto control de tráfico de personas y personal lo que repercute en seguridad confiable las 24 horas del día durante todo el año.

Cabe mencionar que **DIGIX** tiene todo el equipo para el desarrollo de los procesos requeridos por el **CONEICC** y que en caso extraordinario es responsabilidad de **DIGIX** obtener el equipo necesario para tal fin (mediante la compra nacional o importación del equipo).

Lo anterior le permite a **DIGIX** tener una amplia gama de opciones para elegir el equipo que se adapte mejor a los requerimientos de los procesos ofrecidos y, a su vez, esto crea las condiciones para que **DIGIX** tenga acceso a equipos de almacenamiento, lo que hace a **DIGIX**, mucho más competitiva no solo a nivel nacional si no a nivel global.



7. Recomendaciones

- Analizar hasta agotar todos y cada uno de los puntos propuestos por **DIGIX** con el fin de que toda la propuesta sea clara y evitar interpretaciones erróneas
- Para cualquier aclaración sobre la información contenida en este documento dirigirse con el ejecutivo de **DIGIX**
- Si se requiriera de investigaciones adicionales de alguna otra colección fotográfica o de copias impresas de documentos; se tomarían en consideración para incorporarlas en un proyecto adicional o, en su defecto, en el mismo proyecto
- Este reporte fue preparado en base a la tecnología existente, sin previo aviso los equipos se mejorarán en la medida en que sean perfeccionados tanto software, en el área de barrido y de digitalización de libros y microfilmes

8. Características del proyecto

El proyecto completo estipula:

- Únicamente el proceso de Digitalización de una selección de aproximadamente **600,000 páginas** pertenecientes al acervo del obras del **CONEICC**
- Los servicios de digitalización se llevarán a cabo **FUERA** de las instalaciones del archivo del **CONEICC**
- El cliente deberá de proveer a **DIGIX** de la nomenclatura para identificar cada archivo

9. Plazos de entrega

Se propone digitalizar anualmente el material de un quinquenio, del material más reciente al más antiguo; es decir, este año digitalizar las tesis del 2005 a 2000, el año siguiente las tesis de 1999 a 1995 y así sucesivamente. El tiempo estimado se definirá para cada etapa.



10. Servicios incluidos

- Producción de imágenes digitales a partir de originales impresos
Formato: PDF multipágina bitonal a una resolución de 300dpi (un archivo por tesis)
- Aplicación de filtros gráficos para eliminación parcial de ruido de fondo y mejoramiento de caracteres
- Recorte de bordes negros o blancos
- Clasificación manual de imágenes de acuerdo al número de obra
- Respaldo en DVD

11. Costos

El presupuesto que a continuación se presenta se basa en la propuesta de digitalizar la totalidad del archivo del CONEICC, aunque como se comentó anteriormente, se realizará por etapas (material por periodos quinquenales).

PRODUCTO	Precio unitario	Cantidad	Total
<i>Digitalización de la obras del archivo del CONEICC</i>	US \$0.13	600,000 páginas	US\$ 78,000.00
	IVA		US\$ 11,700.00
	TOTAL		US\$ 89,700.00

Precios en dólares estadounidenses.

El pago de cada etapa se realizará de la siguiente forma:

- 25% de anticipo
- Resto al entregar el material digitalizado.

NOTA: Las 600,000 páginas son un aproximado, en caso de ser mayor o menor el costo total se ajustará de manera proporcional.