

## Terceiro exercicio do proceso selectivo

### Bloque I

### Suposto 2

O Arquivo Histórico da USC ten a intención de presentarse a unha convocatoria de axudas da Axencia de Modernización Tecnolóxica de Galicia (AMTEGA) para a dixitalización de fondos de patrimonio documental galego. Dado que ten en custodia a documentación histórica do Concello de Santiago de Compostela, decide presentar o proxecto de dixitalización da seguinte documentación debido ao seu interese histórico e ao elevado número de consultas que tivo nos últimos anos:

Función: Seguridade e protección cidadá

Subfunción: Control de persoas detidas

Serie documental: **Expedientes de alteración da orde pública (1834-1916)**

Datas extremas: 1834-1916

Volume: 14 u.i.

Serie documental: **Expedientes de detención**

Datas extremas: 1917-1977

Volume: 43 u.i.

Serie documental: **Partes diarios de presos da cárcere**

Datas extremas: 1893-1930

Volume: 18 u.i.

Serie documental: **Licenzas de presos suxeitos a vixilancia**

Datas extremas: 1858-1922

Volume: 9 u.i.

Serie documental: **Informes de conducta**

Datas extremas: 1906-1976

Volume: 38 u.i.

Serie documental: **Testemuños de condena do xulgado municipal**

Datas extremas: 1883-1888

Volume: 5 u.i.

Esta documentación presenta as seguintes características:

- Non está organizada, polo tanto, non se atopa clasificada, ordenada nin descrita.
- Mantén a instalación orixinal coa que foi transferida desde o concello, en atados e carpetas de distinto formato polo que, nalgúns casos, os expedientes están dobrados.
- O estado de conservación é bo en liñas xerais, agás algúns expedientes de alteración da orde pública que presentan un deterioro do soporte a consecuencia dos efectos nocivos producidos pola humidade.
- Os expedientes de detención desde os anos 1960 a 1977 atópanse grampados, dificultando nalgúns casos a lectura do documento, e presentan un papel de mala calidade cun nivel de acidez moi elevado.
- Os informes de conducta teñen algúns documentos pegados con adhesivo. Tamén contan con fotografías dos presos, unhas grampadas e outras soltas. Ademais, intercaladas nestes informes, atopáronse 8 fotografías das instalacións da cárcere municipal realizadas en 1930.
- A maior parte desta documentación contén datos de carácter persoal pero, en concreto, os expedientes de detención dos anos 1936 a 1974 inclúen, ademais, información sobre a ideoloxía dos presos ou a actividade clandestina pola que foron apresados.

Tendo en conta toda esa información, **elabore un proxecto de dixitalización** que inclúa os seguintes aspectos:

- **Delimitación do proxecto e obxectivos** a acadar
  
- **Características** do tipo de documentación e do seu **interese como fonte documental** para a investigación.
  
- **Relación razoada das consideracións previas** a nivel de xestión administrativa e de carácter arquivístico que se deben ter en conta, **recursos materiais e persoais** necesarios e **traballos técnicos** que se deberán desenvolver para favorecer o proceso de dixitalización.
  
- **Definición dos requisitos técnicos e do traballo a realizar pola empresa adxudicataria** que deberá recoller o prego de prescricións técnicas do proxecto en relación ao tipo de documentación da que se trata. Para a redacción deste apartado pode consultar o documento Estándar técnico para a dixitalización do Patrimonio Documental de ARPAD. Arquivo Dixital de Galicia que figura como anexo I.
  
- Indicación de calquera outro aspecto que considere relevante nun proxecto destas características.

## Tercer ejercicio del proceso selectivo

### Bloque I

#### Supuesto 2

El Archivo Histórico de la USC tiene la intención de presentarse a una convocatoria de ayudas de la Agencia de Modernización Tecnológica de Galicia (AMTEGA) para la digitalización de fondos de patrimonio documental gallego. Dado que tiene en custodia la documentación histórica del Ayuntamiento de Santiago de Compostela, decide presentar el proyecto de digitalización de la siguiente documentación debido a su interés histórico y al elevado número de consultas que tuvo en los últimos años:

Función: Seguridad y protección ciudadana

Subfunción: Control de personas detenidas

Serie documental: **Expedientes de alteración del orden público (1834-1916)**

Fechas extremas: 1834-1916

Volumen: 14 u.i.

Serie documental: **Expedientes de detención**

Fechas extremas: 1917-1977

Volumen: 43 u.i.

Serie documental: **Partes diarios de presos de la cárcel**

Fechas extremas: 1893-1930

Volumen: 18 u.i.

Serie documental: **Licencias de presos sujetos a vigilancia**

Fechas extremas: 1858-1922

Volumen: 9 u.i.

Serie documental: **Informes de conducta**

Fechas extremas: 1906-1976

Volumen: 38 u.i.



Serie documental: **Testimonios de condena del juzgado municipal**

Fechas extremas: 1883-1888

Volumen: 5 u.i.

Esta documentación presenta las siguientes características:

- No está organizada, por lo tanto, no se encuentra clasificada, ordenada ni descrita.
- Mantiene la instalación original con la que fue transferida desde el ayuntamiento, en legajos y carpetas de distinto formato por lo que, en algunos casos, los expedientes están doblados.
- El estado de conservación es bueno en líneas generales, excepto algunos expedientes de alteración del orden público que presentan un deterioro del soporte a consecuencia de los efectos nocivos producidos por la humedad.
- Los expedientes de detención desde los años 1960 a 1977 se encuentran grapados, dificultando en algunos casos la lectura del documento, y presentan un papel de mala calidad con un nivel de acidez muy elevado.
- Los informes de conducta tienen algunos documentos pegados con adhesivo. También cuentan con fotografías de los presos, unas grapadas y otras sueltas. Además, intercaladas en estos informes, se encontraron 8 fotografías de las instalaciones de la cárcel municipal realizadas en 1930.
- La mayor parte de esta documentación contiene datos de carácter personal pero, en concreto, los expedientes de detención de los años 1936 a 1974 incluyen, además, información sobre la ideología de los presos o la actividad clandestina por la que fueron apresados.

Teniendo en cuenta toda esa información, **elabore un proyecto de digitalización** que incluya los siguientes aspectos:

- **Delimitación del proyecto y objetivos** a conseguir.
- **Características** del tipo de documentación y de su **interés como fuente documental** para la investigación.
- **Relación razonada de las consideraciones previas** a nivel de gestión administrativa y de carácter archivístico que se deben tener en cuenta, **recursos materiales y personales** necesarios y **trabajos técnicos** que se deberán desarrollar para favorecer el proceso de digitalización.
- **Definición de los requisitos técnicos y del trabajo a realizar por la empresa adjudicataria** que deberá recoger el pliego de prescripciones técnicas del proyecto en relación al tipo de documentación de la que se trata. Para la redacción de este apartado puede consultar el documento Estándar técnico para la digitalización del Patrimonio Documental de ARPAD. Archivo Digital de Galicia que figura como anexo I.
- Indicación de cualquier otro aspecto que considere relevante en un proyecto de estas características.

# Arquivo dixital e servizos sen papel



ESTÁNDAR TÉCNICO

Dixitalización do Patrimonio Documental

Condicións de dixitalización



PROGRAMA  
COOPERACIÓN TRANSFRONTERIZA  
ESPAÑA ~ PORTUGAL  
COOPERAÇÃO TRANSFRONTEIRA  
2 0 0 7 - 2 0 1 3



UNIÃO EUROPEA  
FONDO EUROPEO DE  
DESENVOLVIMENTO  
REGIONAL  
"Unha maneira de facer Europa"

## CONTROL DE VERSIÓNS E DISTRIBUCIÓN DO DOCUMENTO

<b>Nome</b>	Dixitalización do Patrimonio Documental	<b>Versión</b>	2
<b>Código</b>			
<b>Elaborado por</b>	Proxecto ARPAD	<b>Data</b>	15/10/2015
<b>Validado por</b>		<b>Data</b>	
<b>Aprobado por</b>		<b>Data</b>	

## REXISTRO DE CAMBIOS

Versión	Causa da nova versión	Data de aprobación
1	Versión inicial	06/02/2017
2	Corrección dos díxitos dos epígrafes pola Subdirección Xeral de Arquivos e modificación dos requisitos dos entregables	22/06/2018
3	Ampliación do detalle das entregas e actualización cos últimos estándares.	15/03/2019

## LISTA DE DISTRIBUCIÓN

Nome	Número de copia	Área / Centro / Localización

## DATOS BÁSICOS DO PROXECTO

<b>Nome</b>	Arquivo Electrónico do Patrimonio de Galicia
<b>Código</b>	ARPAD_PatArq



## Índice

<b>1. Introducción.....</b>	<b>4</b>
<b>2. Requisitos técnicos do proceso de dixitalización .....</b>	<b>6</b>
2.1. Proceso de captura .....	6
2.1.1. Manipulación de documentos .....	6
2.1.2. Equipamento técnico.....	6
2.1.3. Características técnicas da captura.....	7
2.1.4. Especificacións para cada tipo de material .....	9
2.1.4.1. Material textual.....	10
2.1.4.2. Fotografías.....	10
2.1.4.3. Materiais de grandes dimensións .....	11
2.1.4.4. Material gráfico.....	12
2.1.4.5. Cadro resumo para material textual, gráfico e de grandes dimensións .....	13
2.1.4.6. Cadros resumen para material fotográfico .....	14
2.1.5. Control de calidade no proceso de captura.....	16
2.2. Proceso de optimización da imaxe e segundo control de calidade.....	17
2.2.1. Optimización da imaxe.....	17
2.2.2. Segundo control de calidade: .....	17
2.2.3. Recoñecemento Óptico de Caracteres .....	18
2.2.4. Marca de auga.....	21
2.3. Proceso de creación de METS (Metadata Encoding and Transmission Standard).....	21
2.4. Proceso de validación de obxectos dixitais e metadatos .....	23
<b>3. Requisitos dos entregables dun proxecto de dixitalización .....</b>	<b>24</b>
3.1. Creación dos obxectos dixitais.....	24
3.2. Nomeado de ficheiros e cartafoles .....	24
3.3. Formatos de entrega.....	25
3.3.1. Copias para preservación .....	25
3.3.2. Copias para difusión .....	27
<b>4. Referencias bibliográficas .....</b>	<b>29</b>
<b>5. Anexo.....</b>	<b>30</b>
5.1. Termos técnicos.....	30
5.2. Normativa .....	31
5.3. Centros de arquivo integrados en Arquivo Dixital de Galicia.....	32

## 1. INTRODUCCIÓN

O presente documento ten como finalidade servir de guía de referencia a todas aquelas institucións que desexen emprender proxectos de dixitalización de patrimonio documental para a súa inclusión no Arquivo Electrónico de Patrimonio da Xunta de Galicia. Todas as recomendacións están enfocadas ao escenario "Dixitalización do patrimonio documental", polo que a súa elaboración está centrada no material textual e gráfico. O seu obxectivo é cubrir os aspectos máis importantes do proceso de dixitalización, desde as características técnicas á elaboración de metadatos, a xestión dos dereitos de autor, a xestión e o acceso á documentación dixitalizada, así como outros aspectos complementarios.

O Arquivo Electrónico do Patrimonio de Galicia está composto por un Sistema de Preservación, un Sistema de Difusión e Recolección e un Sistema de Xestión de Descricións Arquivísticas, obxectos dixitais e provisión de datos recolectables. Os obxectos dixitais obtidos nun proxecto de dixitalización inclúense nos tres subsistemas, aínda que tamén se contempla a posibilidade de que só se realicen entregas de obxectos dixitais ao sistema de preservación.

Neste documento faise un percorrido polas diferentes fases do proceso de dixitalización, desde os requisitos técnicos adecuados aos diferentes tipos de materiais para a captura de imaxes de calidade ata a entrega dos diferentes ficheiros nos formatos recomendados, pasando por procesos como o recoñecemento óptico de caracteres.

A dixitalización é un medio fundamental para asegurar a preservación, o acceso e o uso do patrimonio documental a condición de que se apliquen unha serie de recomendacións e boas prácticas que permitan garantir a longo prazo o contido dos obxectos dixitais e que poidan ser integrados no Arquivo Electrónico do Patrimonio de Galicia e os seus subsistemas. Estes condicionantes que se detallarán ao longo do documento teñen como finalidade definir as características dunha dixitalización de calidade.

As recomendacións e requisitos que seguen establecéronse tendo en conta os seguintes aspectos:

- Garantir a preservación dos contidos e dos obxectos dixitais a longo prazo
- Garantir que os documentos e os seus obxectos dixitais están perfectamente identificados e descritos e que, por tanto, dispoñen dos correspondentes esquemas de metadatos
- Garantir o acceso e uso aos documentos dixitalizados a través da web xa sexa en formato imaxe ou en formato texto
- Garantir as condicións de uso e reprodución dos documentos dixitalizados respectando os dereitos de propiedade intelectual a que haxa lugar
- Garantir a interoperabilidade e inxesta dos obxectos dixitais e dos seus metadatos no Sistema de Arquivo Electrónico do Patrimonio de Galicia

Todas as recomendacións que se establecen neste documento sobre dixitalización, xeración de metadatos e preservación, baséanse en diferentes guías de boas prácticas establecidas por organismos e proxectos nacionais e internacionais en materia de dixitalización de bibliotecas, arquivos e museos.

As institucións depositarias da documentación decidirán que material queren dixitalizar tendo en conta que dita documentación debe estar perfectamente cuantificada e identificada. Para iso deseñárase unha política de selección, na que se establecerán unha serie de criterios que determinarán a prioridade no proceso de dixitalización. Algúns criterios nos que se pode basear a institución para establecer a orde de dixitalización son:

- O valor do contido
- O valor histórico ou documental
- O estado de conservación
- Que a documentación estea completamente descrita
- Unha alta demanda da documentación
- A documentación é única ou rara.

Unha vez que se seleccionou a documentación para dixitalizar, a partir desta guía poderanse establecer os requisitos necesarios para a captura, o control de calidade, o proceso de recoñecemento óptico de caracteres, a xeración dos metadatos, a creación de ficheiros METS ([Metadata Encoding and Transmission Standard](#)) e as especificacións técnicas dos entregables.

É necesario mencionar que neste documento requírese o uso dunha normativa cando está plenamente definida e aceptada, mentres que noutros casos, como é o relativo ás características técnicas de dixitalización, recoméndase seguir determinadas prácticas xa que non existen unha normativa fixada. Por exemplo, o nivel de resolución dunha copia dixitalizada pode establecerse en función de diferentes parámetros como o estado de conservación da documentación, o tipo de material, o obxectivo da dixitalización, etc. A partir do estudo de todos estes factores sacaranse as conclusións que axudarán a tomar as decisións oportunas para cada proxecto de dixitalización.

## 2. REQUISITOS TÉCNICOS DO PROCESO DE DIXITALIZACIÓN

### 2.1. Proceso de captura

#### 2.1.1. Manipulación de documentos

- A institución depositaria, verificará que a documentación seleccionada para dixitalizar non conteña documentos duplicados, copias, borradores e demais documentación que careza de valor informativo.
- Atenderanse as indicacións e normas de seguridade que establezan as institucións depositarias en canto a procedementos de entrada e saída de persoas ou obxectos, petición e devolución de materiais, acceso a áreas restrinxidas ou de traballo, identificación, etc.
- En todos os procesos seguiranse as indicacións dos responsables da institución en canto á manipulación da documentación.
- A documentación hase de tratar co maior coidado posible, procurando non causar ningún dano durante a súa manipulación. No caso de documentación moi antiga recoméndase manipular con luvas de látex, vinilo ou algodón para non danar o documento.
- Por regra xeral, non se desencadernarán os documentos nin se adoptarán medidas que vaian contra a súa integridade física.
- Os orixinais encadernados e con problemas de conservación abríranse cun ángulo nunca superior a 120º.
- Manterase a zona de traballo ordenada e limpa, non dispoñendo nela de máis material que o necesario para o proceso de dixitalización.
- Dispoñeranse dos medios adecuados para a manipulación da documentación. Os operarios empregarán luvas de algodón, látex ou vinilo e irán equipados con batas e máscaras, cando así se considere necesario.

#### 2.1.2. Equipamento técnico

- Recoméndase que os escáneres se axusten ás especificacións dos estándares [TWAIN](#) (Technology Without An Interesting Name) e ISIS (Image and Scanner Interface Specification).

Para dixitalizar documentación histórica, material impreso en xeral, fotografías, material gráfico, ou documentación en grandes dimensións recoméndase a utilización de escáneres de captura

cenital. Estes escáneres permiten dixitalizar co mínimo contacto posible, evitando así posibles danos na documentación.

Os escáneres de captura cenital utilizan luz fría de alta frecuencia, o que permite que o proceso de escaneado non afecte o estado de conservación dos orixinais. Ademais permiten unha ampla variedade de tamaños chegando a dimensións A1 ou A2, segundo o modelo de escáner.

En canto á dixitalización de transparencias (diapositivas, negativos, placas de cristal) recoméndase utilizar escáneres planos específicos para este tipo de material.

De forma xeral, recoméndase dixitalizar páxina a páxina porque o tamaño das imaxes que se xera é máis pequeno e a visualización das imaxes é máis cómoda. Así mesmo, actualmente os escáneres de captura cenital a partir dunha única toma sobre dúas páxinas, permiten xerar dúas imaxes independentes, polo que o tempo de dixitalización é o mesmo. Doutra banda, se se desexa dixitalizar a dobre páxina recoméndase que o tamaño dos orixinais non sexa maior de A4, aínda que actualmente a tendencia e recomendación é dixitalización a páxina simple.

Para dixitalizar documentación histórica, fotografías, material gráfico, ou documentos de grandes dimensións recoméndase a utilización de escáneres de captura cenital. Estes escáneres permiten dixitalizar co mínimo contacto posible, evitando así posibles danos na documentación. Ademais utilizan luz fría de alta frecuencia, o que permite que o proceso de escaneado non afecte o estado de conservación dos orixinais. Por outro lado, permiten unha ampla variedade de tamaños chegando a dimensións A1 ou A2, segundo o modelo de escáner.

Para a dixitalización de transparencias (diapositivas, negativos, placas de cristal) recoméndase utilizar escáneres planos específicos para este tipo de material.

### **2.1.3. Características técnicas da captura**

- O produto final sempre será unha copia dixitalizada fiel ao orixinal (dixitalización facsimilar), de maneira tal que na copia dixitalizada manteranse as características da documentación orixinal. É dicir se o documento orixinal ten anotacións, rotos, manchas, etc., na copia dixitalizada non se eliminarán para que o orixinal e a copia sexan iguais.
- A dixitalización da documentación orixinal levarase a cabo no mesmo estado no que se presente; non se realizará intervención algunha sobre os exemplares, as súas encadernacións ou os seus soportes.
- Dixitalización a páxina simple (1 imaxe = 1 páxina). Poderá facerse unha única toma sobre dúas páxinas cando o orixinal non sexa maior do tamaño A4. Neste último caso deberán xerarse a continuación dúas imaxes independentes.

- Coa finalidade de conseguir a máxima calidade, o axuste dos escáneres (iluminación, contraste, brillo, resolución...) realizarase en función das características de cada documento, é dicir, tendo en conta o seu soporte, tamaño, grosor, tintas, tipo de encadernación e estado de conservación.
- Manterase unha constante limpeza e posta a punto dos escáneres.
- Os monitores manteranse limpos e calibraranse de acordo á iluminación da sala e ao material para dixitalizar.
- Na dixitalización é obrigatorio utilizar a carta de grises ou de cor, segundo proceda. Estas cartas serven para obter un perfil dos niveis de grises ou de cor que sexa fiel aos orixinais capturados. Para facer uso delas, se incluírá dentro da primeira imaxe reproducida dun documento, ou en todas, para conseguir unha referencia clara do aspecto real do orixinal.
- Cando as fotografías teñan información no dorso, tamén se dixitalizará esta parte.
- Por regra xeral dixitalizaranse tamén as páxinas en branco, a non ser que polas características da documentación decídase non facelo, por exemplo en casos de documentos que conteñan demasiadas páxinas en branco e se considere que non son necesarias para dar significado aos documentos.
- Establecerase máxima uniformidade en formatos, tamaños de marcos de dixitalización e alturas de cámara. Tamén as resolucións que se establezan serán homoxéneas.
- Despregaranse e estíranse os materiais co obxectivo de dixitalizar o maior texto posible. Pero en ningún caso forzaranse as páxinas para evitar o deterioro do material.
- Nos casos en que a encadernación, dobraduras das follas ou estado de conservación dificulte a dixitalización, estudarase a mellor forma de facer a copia coa menor perda de información.
- En documentos demasiado grandes para dixitalizar nunha única captura, recoméndase a dixitalización do documento por fragmentos. Cada un dos fragmentos deberá ser renomeado como nomeficheiro\_f01c01 (onde f representa a fila e c a columna; os dous díxitos establecen a orde de cada un dos elementos)

f01c01	f01c02	f01c03
f02c01	f02c02	f02c03
f03c01	f02c02	f03c03

#### 2.1.4. Especificacións para cada tipo de material

- De acordo ás directrices establecidas en NARA [Guidelines for Digitizing Archival Materials for Electronic Access](#), os requisitos técnicos para a dixitalización de documentación estableceranse a partir do tipo de material a dixitalizar, seguindo a seguinte clasificación:
  - **Documentos textuais:** material impreso, fotocopias, manuscritos.
  - **Fotografías:** branco e negro, cor, negativos, diapositivas, transparencias.
  - **Mapas, planos, e outros materiais de gran tamaño:** mapas, planos, carteis, etc.
  - **Material gráfico:** debuxos, ilustracións artísticas, litografías, etc.

Durante a dixitalización, e de acordo ás directrices mencionadas máis arriba e a norma *UNE-ISO/TR 13028 (2011) Información e documentación. Directrices para a implementación da dixitalización de documentos*, recoméndase xerar tres copias de cada imaxe dixitalizada: un ficheiro máster, un ficheiro de acceso e un ficheiro de miniaturas. Sendo o ficheiro máster o que se xere directamente na dixitalización, e os ficheiros de acceso e miniaturas, copias derivadas do ficheiro máster.

- **Ficheiro máster:** a finalidade do ficheiro máster é obter unha copia fiel do documento orixinal. Este ficheiro farase en alta resolución e será un ficheiro de gran tamaño a partir do cal se poderán obter reproducións impresas de calidade e outros ficheiros. O formato recomendado para este ficheiro é o formato TIFF sen comprimir, para asegurar que non se perde calidade.
- **Ficheiro de acceso:** imaxe que se obtén a partir do ficheiro máster, e cuxo obxectivo é facilitar o acceso aos usuarios finais a través dunha contorna web. Este ficheiro gardarase nunha resolución menor que o ficheiro máster e en formato comprimido, aínda que coa suficiente calidade para a súa correcta visualización. Para este tipo de ficheiro recoméndase o formato JPEG comprimido.
- **Ficheiro de miniaturas:** imaxe moi pequena para facilitar o acceso e referencia do obxecto dixital na internet, que funciona como previsualización. Este ficheiro tamén se obterá a partir do ficheiro máster, e tamén se recomenda gardar en formato JPEG comprimido. Xeraranse ficheiros de miniaturas de todos os ficheiros máster.

Como se mencionou, non existe unha normativa específica para indicar a resolución á que se debe dixitalizar, aínda que si existe documentación tanto nacional como internacional onde se recomenda e establecen os requisitos mínimos e recomendables para realizar unha dixitalización de calidade, tanto para a súa preservación como para a súa difusión.

#### 2.1.4.1. Material textual

Dentro de material textual inclúese tanto material impreso como manuscrito, e tanto documentación solta como documentación encadernada.

Á hora de dixitalizar material textual impreso, é importante ter en conta a posibilidade de realizar tamén un proceso de recoñecemento óptico de caracteres a posteriori (véxase o punto 2.2.3 Recoñecemento óptico de caracteres). Neste caso é importante que a imaxe que se obteña sexa de alta calidade e en escala de grises para minimizar o número de erros durante o recoñecemento óptico de caracteres.

##### ▪ Ficheiro máster

- Resolución: 400 ppp.
- Esquema de codificación: 8 bits (256 niveis de grises) ou 24 bits (16.777.216 millóns de cores ou cor verdadeira).
- Tamaño: real do orixinal.
- Formato de saída: TIFF 6.0 sen comprimir.

##### ▪ Ficheiro de acceso

- Resolución: 200 ppp.
- Esquema de codificación: 8 bits (256 niveis de grises) ou 24 bits (16.777.216 millóns de cores ou cor verdadeira).
- Tamaño: real do orixinal.
- Formato de saída: JPEG con comprensión media.

##### ▪ Ficheiro de miniaturas

- Resolución: 72 ppp
- Esquema de codificación: 8 bits (256 niveis de grises) ou 24 bits (16.777.216 millóns de cores ou cor verdadeira).
- Tamaño: máximo 200 ppp de altura ou anchura.
- Formato de saída: JPEG con comprensión media.

#### 2.1.4.2. Fotografías

O material fotográfico é moi delicado e agrupa tipos de materiais moi diferentes entre si. Para explicar as súas características, en primeiro lugar fíxose unha diferenciación entre materiais opacos e materiais transparentes. Dentro de materiais opacos atópanse as fotografías en papel, que poden ser en branco e negro, sepia ou cor. Mentres que dentro do material transparente ou transparencias, pódense atopar materiais como negativos, diapositivas, ou placas de cristal.



- **Ficheiro máster**
  - Resolución: segundo tamaño (véxase o punto 2.1.4.6 Cadros resumen para material fotográfico)
  - Esquema de codificación: 24 bits (16.777.216 millóns de cores ou cor verdadeira)
  - Tamaño: real do orixinal
  - Formato de saída: TIFF 6.0 sen comprimir
  
- **Ficheiro de acceso**
  - Resolución: 200 ppp
  - Esquema de codificación: 24 bits (16.777.216 millóns de cores ou cor verdadeira)
  - Tamaño: real do orixinal
  - Formato de saída: JPEG con comprensión media.
  
- **Ficheiro de miniaturas**
  - Resolución: 72 ppp
  - Esquema de codificación: 24 bits (16.777.216 millóns de cores ou cor verdadeira)
  - Tamaño: máximo 200 ppp de altura ou anchura.
  - Formato de saída: JPEG con comprensión media.

### 2.1.4.3. Materiais de grandes dimensións

Nesta clasificación inclúese cartografía, planos, carteis, fotografías aéreas de gran tamaño ou algunhas publicacións periódicas, en xeral todo o material maior do tamaño A3 (42 x 29,7 cm).

Naqueles documentos demasiado grandes para dixitalizar nunha única captura, débese proceder á dixitalización por fragmentos, seguindo as indicacións anotadas no epígrafe 2.1.3 Características técnicas da captura.

Os ficheiros máster incluírán cada un dos fragmentos dixitalizados e a imaxe final unida (composición de todos os fragmentos). Nos ficheiros de acceso incluírase unicamente a imaxe final unida.

- **Ficheiro máster**
  - Resolución: 400 ppp
  - Esquema de codificación: 8 bits (256 niveis de grises) ou 24 bits (16.777.216 millóns de cores ou cor verdadeira).
  - Tamaño: real do orixinal

- Formato de saída: TIFF 6.0 sen comprimir
- **Ficheiro de acceso**
  - Resolución: 200 ppp
  - Esquema de codificación: 8 bits (256 niveis de grises) ou 24 bits (16.777.216 millóns de cores ou cor verdadeira).
  - Tamaño: real do orixinal
  - Formato de saída: JPEG con comprensión media
- **Ficheiro de miniaturas**
  - Resolución: 72 ppp
  - Esquema de codificación: 8 bits (256 niveis de grises) ou 24 bits (16.777.216 millóns de cores ou cor verdadeira).
  - Tamaño: máximo 200 ppp de altura ou anchura.
  - Formato de saída: JPEG con comprensión media

#### 2.1.4.4. Material gráfico

No material gráfico inclúense debuxos, ilustracións artísticas, acuarelas, postais ou litografías.

- **Ficheiro máster**
  - Resolución: 400 ppp.
  - Esquema de codificación: 24 bits (16.777.216 millóns de cores ou cor verdadeira).
  - Tamaño: real do orixinal.
  - Formato de saída: TIFF 6.0 sen comprimir.
- **Ficheiro de acceso**
  - Resolución: 200 ppp.
  - Esquema de codificación: 24 bits (16.777.216 millóns de cores ou cor verdadeira).
  - Tamaño: real do orixinal.
  - Formato de saída: JPEG con comprensión media.
- **Ficheiro de miniaturas**
  - Resolución: 72 ppp.

- Esquema de codificación: 24 bits (16.777.216 millóns de cores ou cor verdadeira).
- Tamaño: máximo 200 ppp de altura ou anchura.
- Formato de saída: JPEG con compresión media.

#### 2.1.4.5. Cadro resumo para material textual, gráfico e de grandes dimensións

Neste cadro resúmense os requisitos mínimos de resolución e esquema de codificación establecidos para a dixitalización de material textual, material de grandes dimensións e material gráfico.

REQUISITO	MATERIAL TEXTUAL	GRANDES DIMENSIÓN	MATERIAL GRÁFICO
Resolución mínima	400 ppp	400 ppp	400 ppp
Esquema de codificación	8 bits (256 niveis de grises) ou 24 bits (16.777.216 millóns de cores ou cor verdadeira)	8 bits (256 niveis de grises) ou 24 bits (16.777.216 millóns de cores ou cor verdadeira)	24 bits (16.777.216 millóns de cores ou cor verdadeira)
Tamaño	Tamaño real do orixinal	Tamaño real do orixinal	Tamaño real do orixinal
Formato ficheiros Máster preservación	TIFF 6.0 sen comprimir	TIFF 6.0 sen comprimir	TIFF 6.0 sen comprimir
Formato ficheiros acceso na internet	JPEG con compresión media (200 ppp)	JPEG con compresión media (200 ppp)	JPEG con compresión media (200 ppp)
Formato ficheiros de acceso (miniaturas)	JPEG (72 ppp)	JPEG (72 ppp)	JPEG (72 ppp)
OCR	Si	Si	Non

#### 2.1.4.6. Cadros resumen para material fotográfico

Nos seguintes cadros resúmense os requisitos mínimos establecidos para a dixitalización de material fotográfico. Debido ao material tan diferente entre si, establecéronse dous cadros, un para o material opaco (en papel) e outro para o material transparente (negativos, diapositivas e placas de cristal).

MATERIAL OPACO			
DIMENSIÓNS	RESOLUCIÓN	ESQUEMA DE CODIFICACIÓN	TAMAÑO DE CAPTURA
9 x 13 cm	800 ppp	24 bits (16.777.216 millóns de cores ou cor verdadeira)	Tamaño real do orixinal
10 x 15 cm	600 ppp	24 bits (16.777.216 millóns de cores ou cor verdadeira)	Tamaño real do orixinal
13 x 18 cm	600 ppp	24 bits (16.777.216 millóns de cores ou cor verdadeira)	Tamaño real do orixinal
15 x 20 cm	600 ppp	24 bits (16.777.216 millóns de cores ou cor verdadeira)	Tamaño real do orixinal
Superior a 20 cm	400 ppp	24 bits (16.777.216 millóns de cores ou cor verdadeira)	Tamaño real do orixinal
TRANSPARENCIAS			
DIMENSIÓNS	RESOLUCIÓN	ESQUEMA DE CODIFICACIÓN	TAMAÑO DE CAPTURA
NEGATIVOS E DIAPOSITIVAS			
35 mm	2800 ppp	24 bits (16.777.216 millóns de cores ou cor verdadeira)	Tamaño real do orixinal
6 x 6 cm	1400 ppp	24 bits (16.777.216 millóns de cores ou cor verdadeira)	Tamaño real do orixinal
9 x 6 cms	1200 ppp	24 bits (16.777.216 millóns de cores ou cor verdadeira)	Tamaño real do orixinal
PLACAS DE CRISTAL			
6 x 6 cm	1400 ppp	24 bits (16.777.216 millóns de cores ou cor verdadeira)	Tamaño real do orixinal
9 x 6 cm	1200 ppp	24 bits (16.777.216 millóns de cores ou cor verdadeira)	Tamaño real do orixinal
A4	600 ppp	24 bits (16.777.216 millóns de cores ou cor verdadeira)	Tamaño real do orixinal
A3	400 ppp	24 bits (16.777.216 millóns de cores ou cor verdadeira)	Tamaño real do orixinal

**Ilustración 1:**

Textual Documents, Graphic Illustrations/Artwork, Maps, Plans, and Oversized		
Document Character - Original	Recommended Image Parameters	Alternative Minimum
Clean, high-contrast documents with printed type (e.g. laser printed or typeset)	1-bit bitonal mode or 8-bit grayscale - adjust scan resolution to produce a QI of 8 for smallest significant character or 1-bit bitonal mode - 600 ppi* for documents with smallest significant character of 1.0 mm or larger or 8-bit grayscale mode - 400 ppi for documents with smallest significant character of 1.0 mm or larger <b>NOTE:</b> Regardless of approach used, adjust scan resolution to produce a minimum pixel measurement across the long dimension of 6,000 lines for 1-bit files and 4,000 lines for 8-bit files  *The 600 ppi 1-bit files can be produced via scanning or created/derived from 400 ppi, 8-bit grayscale images.	1-bit bitonal mode - 300 ppi* for documents with smallest significant character of 2.0 mm or larger or 8-bit grayscale mode - 300 ppi for documents with smallest significant character of 1.5 mm or larger  *The 300 ppi 1-bit files can be produced via scanning or created/derived from 300 ppi, 8-bit grayscale images.
Documents with poor legibility or diffuse characters (e.g. carbon copies, Thermofax/ Verifax, etc.), handwritten annotations or other markings, low inherent contrast, staining, fading, halftone illustrations, or photographs	8-bit grayscale mode - adjust scan resolution to produce a QI of 8 for smallest significant character or 8-bit grayscale mode - 400 ppi for documents with smallest significant character of 1.0 mm or larger <b>NOTE:</b> Regardless of approach used, adjust scan resolution to produce a minimum pixel measurement across the long dimension of 4,000 lines for 8-bit files	8-bit grayscale mode - 300 ppi for documents with smallest significant character of 1.5 mm or larger
Documents as described for grayscale scanning and/or where color is important to the interpretation of the information or content, or desire to produce the most accurate representation	24-bit color mode - adjust scan resolution to produce a QI of 8 for smallest significant character or 24-bit RGB mode - 400 ppi for documents with smallest significant character of 1.0 mm or larger <b>NOTE:</b> Regardless of approach used, adjust scan resolution to produce a minimum pixel measurement across the long dimension of 4,000 lines for 24-bit files	24-bit RGB mode - 300 ppi for documents with smallest significant character of 1.5 mm or larger

Recomendacións do documento "*Technical Guidelines for Digitizing Cultural Heritage Materials: Creation of Raster Image Master Files*" p. 59. Agosto de 2010

## Ilustración 2:

U.S. National Archives and Records Administration - June 2004

Photographs - prints - black-and-white, monochrome, and color - reflection scanning:		
Format - Original	Recommended Image Parameters	Alternative Minimum
<b>Format range:</b> o 8"x10" or smaller  <b>Size range:</b> o Smaller than or equal to 80 square inches	<b>Pixel Array:</b> o 4000 pixels across long dimension of image area, excluding mounts and borders  <b>Resolution:</b> o Scan resolution to be calculated from actual image dimensions – approx. 400 ppi for 8"x10" originals and ranging up to the appropriate resolution to produce the desired size file from smaller originals, approx. 570 ppi for 5"x7" and 800 ppi for 4"x6" or 3.5"x5" originals  <b>Dimensions:</b> o Sized to match the original, no magnification or reduction  <b>Bit Depth:</b> o 8-bit grayscale mode for black-and-white, can be produced from a 16-bit grayscale file o 24-bit RGB mode for color and monochrome (e.g. albumen prints or other historic print processes), can be produced from a 48-bit RGB file	<b>Pixel Array:</b> o 3000 pixels across long dimension for all rectangular formats and sizes o 2700 pixels by 2700 pixels for square formats regardless of size  <b>Resolution:</b> o Scan resolution calculated from actual image dimensions – approx. 2100 ppi for 35mm originals and ranging down to the appropriate resolution to produce the desired size file from larger originals, approx. 600 ppi for 4"x6" and 300 ppi for 8"x10" originals  <b>Dimension:</b> o File dimensions set to 10" across long dimension at 300 ppi for rectangular formats and to 9"x9" at 300 ppi for square formats
<b>Format range:</b> o Larger than 8"x10" and up to 11"x14"  <b>Size range:</b> o Larger than 80 square inches and smaller than 154 square inches	<b>Pixel Array:</b> o 6000 pixels across long dimension of image area, excluding mounts and borders  <b>Resolution:</b> o Scan resolution to be calculated from actual image dimensions – approx. 600 ppi for originals approx. 8"x10" and ranging down to approx. 430 ppi for 11"x14" originals  <b>Dimensions:</b> o Sized to match the original, no magnification or reduction  <b>Bit Depth:</b> o 8-bit grayscale mode for black-and-white, can be produced from a 16-bit grayscale file o 24-bit RGB mode for color and monochrome (e.g. albumen prints or other historic print processes), can be produced from a 48-bit RGB file	<b>Dimension:</b> o File dimensions set to 10" across long dimension at 300 ppi for rectangular formats and to 9"x9" at 300 ppi for square formats
<b>Format range:</b> o Larger than 11"x14"  <b>Size range:</b> o Equal to or larger than 154 square inches	<b>Pixel Array:</b> o 8000 pixels across long dimension of image area, excluding mounts and borders  <b>Resolution:</b> o Scan resolution to be calculated from actual image dimensions – approx. 570 ppi for originals approx. 11"x14" and ranging down to the appropriate resolution to produce the desired size file from larger originals  <b>Dimensions:</b> o Sized to match the original, no magnification or reduction  <b>Bit Depth:</b> o 8-bit grayscale mode for black-and-white, can be produced from a 16-bit grayscale file o 24-bit RGB mode for color and monochrome (e.g. albumen prints or other historic print processes), can be produced from a 48-bit RGB file	<b>Bit Depth:</b> o 8-bit grayscale mode for black-and-white, can be produced from a 16-bit grayscale file o 24-bit RGB mode for color and monochrome (e.g. collodion wet-plate negative, pyro developed negatives, stained negatives, etc.) can be produced from a 48-bit RGB file

Recomendacións do documento "*Technical Guidelines for Digitizing Archival Materials for Electronic Access: Creation of Production Master Files – Raster Images*", p. 54, elaborado por Ou.S. National Archives and Records Administration (NARA), Xuño de 2004

### 2.1.5. Control de calidade no proceso de captura

Tanto no documento [NARA Technical Guidelines for Digitizing Archival Materials for Electronic Access: Creation of Production Master Files – Raster Images](#) como na norma *UNE-ISO/TR 13028 (2011) Información e documentación. Directrices para a implementación da dixitalización de documentos*, considérase moi importante que o proceso de dixitalización vaia acompañado dun plan de xestión de calidade.

Recoméndase que nos proxectos de dixitalización se faga un control de calidade en dúas fases, primeiro polo propio operario durante o proceso de dixitalización e despois, unha nova revisión de todas as imaxes por outra persoa diferente. Durante o proceso de captura se cotexarán todas as imaxes co orixinal para detectar as imaxes que non se dixitalizaron e as imaxes que se teñen que repetir.

Por tanto, este primeiro control de calidade consiste en:

- Cotexar o 100% das imaxes co documento orixinal á vista.

- Detectar as imaxes que non se dixitalizaron e as imaxes que se teñen que repetir.
- Detectar imaxes que non se dixitalizaron correctamente e se teñen que optimizar.

## 2.2. Proceso de optimización da imaxe e segundo control de calidade

### 2.2.1. Optimización da imaxe

Se as imaxes capturadas non teñen a calidade adecuada, é dicir, están desenfocadas, torcidas, teñen bordos negros, etc., recoméndase facer un proceso de optimización para que as imaxes queden o máis parecidas ao orixinal.

Para iso cada imaxe avaliarase de forma independente e decidirse que tipo de axuste necesita:

- Enfoque
- Axuste de cor
- Axuste de brillo/contraste
- Axuste de tamaño
- Endereitado
- Recorte de bordos negros

Algúns destes axustes poderanse realizar de forma xeral a todas as imaxes e outros serán específicos para cada imaxe.

O obxectivo final deste proceso é que a imaxe sexa o máis parecida posible ao orixinal.

### 2.2.2. Segundo control de calidade:

Unha vez que todas as páxinas están dixitalizadas, no seguinte control de calidade revisarase que:

- Foron dixitalizadas a totalidade das páxinas.
- As imaxes están correctamente encadradas e verticais.
- Non existen imaxes borrosas.
- Non se repetiu ningunha páxina.
- Os ficheiros son perfectamente lexibles.
- Non existe perda de información (salvo que sexa achacable ao estado de conservación do orixinal).
- A orde de páxinas é o correcto (salvo que sexa achacable ao orixinal).

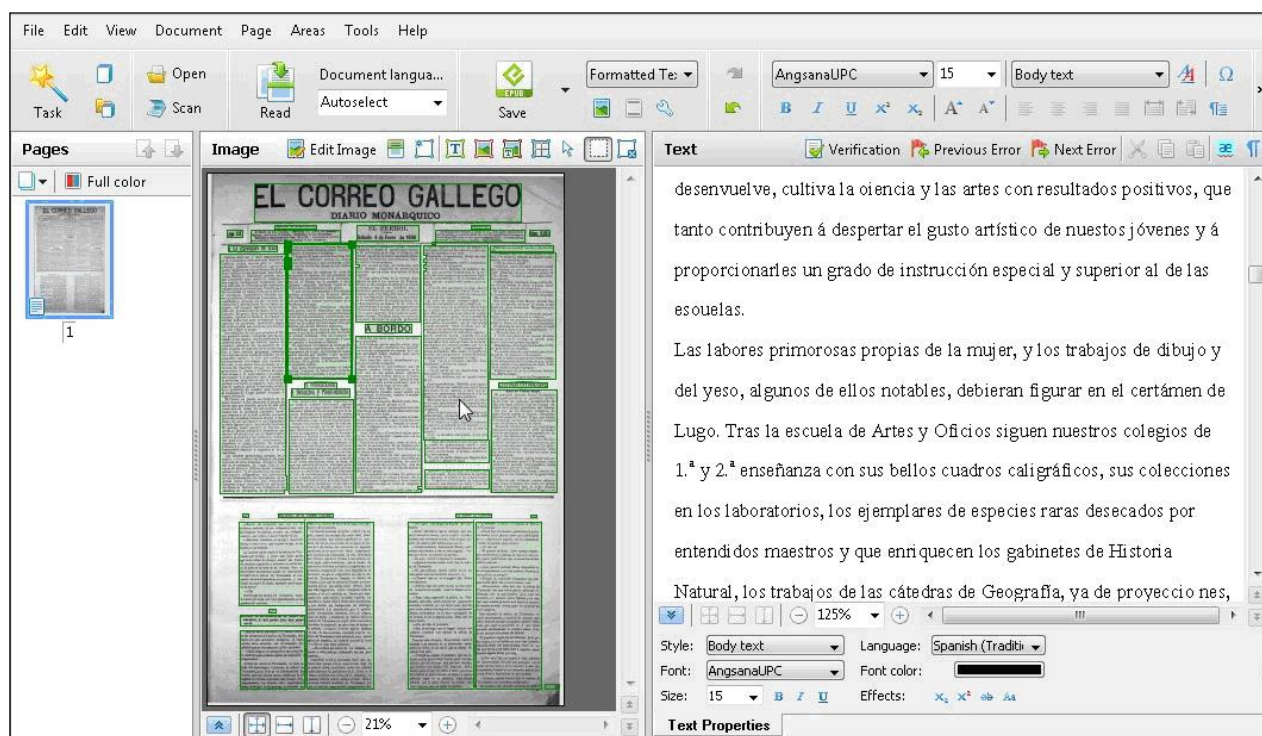
Ao realizar a validación das imaxes dixitalizadas, haberá incidencias que serán propias da documentación orixinal debido ao seu estado de conservación (falta de páxinas, falta parcial de texto, manchas de acidez, etc.). Recoméndase que estas incidencias rexístranse e documenten como parte do proceso de control de calidade, e mesmo se poidan incorporar ao rexistro de descrición.

### 2.2.3. Recoñecemento Óptico de Caracteres

Determinados proxectos de dixitalización tamén poden empregar procedementos adicionais, como o recoñecemento óptico de caracteres, destinado a garantir o acceso ao contido dos documentos en formato texto.

O recoñecemento óptico de caracteres é un proceso dirixido á dixitalización de textos, que consiste en identificar de forma automática a partir dunha imaxe, símbolos ou caracteres que pertencen a un determinado alfabeto, para logo almacenalos en forma de datos. É dicir, a finalidade deste proceso é converter as imaxes dun documento en texto editable.

#### Ilustración 3



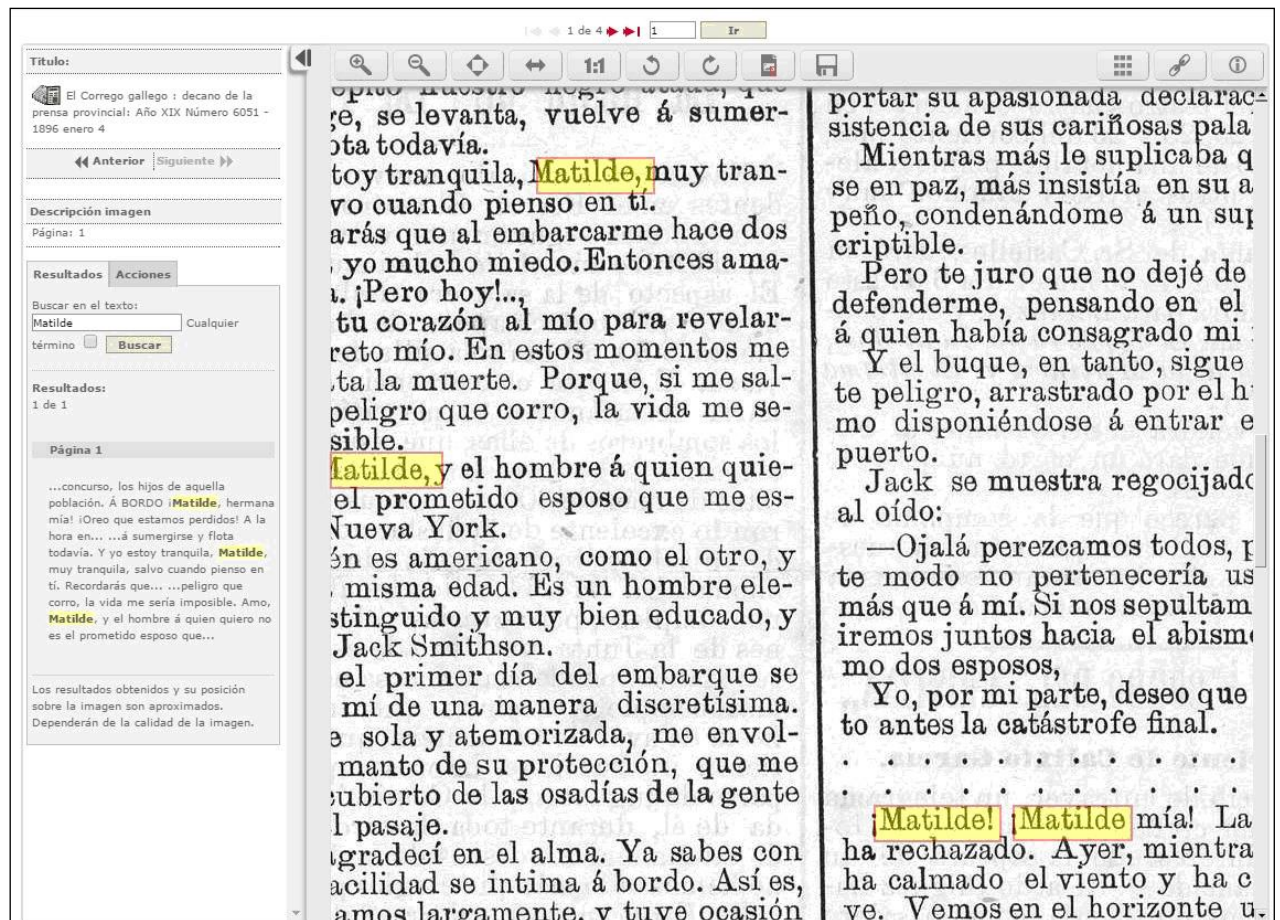
Exemplo de procesamento óptico de caracteres (OCR) sobre unha imaxe dunha publicación periódica

O proceso completo do recoñecemento óptico de caracteres pódese aplicar a toda a documentación impresa, permitindo así o acceso á devandita documentación a partir do seu contido, e facendo posible a súa accesibilidade a través de procuras a texto completo.



Xa que logo, recoméndase aplicar o recoñecemento óptico de caracteres a aquela documentación que estea impresa, que se conserve en boas condicións e que se considere importante ofrecer acceso a través do seu contido.

#### Ilustración 4



Exemplo de procura e marcación de texto sobre unha páxina con procesamento óptico de caracteres (OCR) oculto e sen editar

Unha vez realizado o recoñecemento óptico de caracteres poderase deixar o texto resultante sen editar, é dicir, sen corrixir os erros que xurdisen no proceso. Tamén é posible revisar o texto obtido e editalo ata deixar o texto exactamente igual que o orixinal. Esta última opción ten un custo máis alto xa que se necesita persoal que se dedique aos labores de edición do texto.

Con todo, os novos programas de recoñecemento óptico de caracteres, teñen ferramentas de capacitación e dicionarios de usuario, que permiten optimizar o proceso óptico de caracteres (OCR) e adaptalo ao tipo de documentación que se vai a procesar, podendo así corrixir os fallos máis comúns de forma automática. Desta maneira pódese establecer que o índice de erros sexa o máis baixo posible e evitar a corrección manual, sendo o resultado da procura a texto completo bastante satisfactorio.

Aínda que non é obxecto deste documento analizar as características dos diferentes softwares de recoñecemento óptico de caracteres, no sitio web de Impact: Centro da Competencia para a Dixitalización ([Centre of Competence in Digitisation](http://www.digitisation.eu)) pódese atopar unha listaxe de aplicacións de procesamento óptico de caracteres (OCR) dispoñibles no mercado <http://www.digitisation.eu/tools-resources/tools-for-text-digitisation/>.

Algunhas das características que se recomenda que cumpra o software de procesamento óptico de caracteres (OCR) elixido son as seguintes:

- Variedade de formatos de entrada: PDF, BMP, PCX, JPEG, JPEG2000, TIFF, GIF e PNG.
- Variedade de formatos de exportación de texto: RTF, HTML, XLS, DBF, texto plano, PPT, XML, ASCII XML, Word XML (OOXML), PDF e PDF/A.
- Idiomas: recoñecemento automático de palabras de todas as linguas europeas e asiáticas. O proceso de recoñecemento óptico de caracteres precisa que o idioma elixido no software sexa o mesmo que o idioma do documento procesado para que o índice de erro sexa mínimo.
- Dicionarios de usuario: ferramenta para mellorar a calidade do recoñecemento óptico de caracteres agregando palabras ao dicionario de usuario (termos específicos, abreviaturas, nomes propios).

O recoñecemento óptico de caracteres realízase páxina a páxina, sendo a unidade lóxica de tratamento o documento, a monografía, ou o número dunha publicación periódica. Como resultado deste proceso obtense os seguintes ficheiros:

- Ficheiros ALTO. [Analyzed Layout and Text Object](#) (ALTO) é un esquema XML que mantén a Library of Congress para describir a disposición e o contido dos recursos textuais, tales como a páxina dun libro, unha publicación periódica, ou un documento impreso. Os ficheiros resultantes do procesamento óptico de caracteres (OCR) débense codificar segundo o esquema ALTO que especifica a posición e disposición dun texto con respecto ao orixinal. Para iso utiliza un sistema de coordenadas que permite non só gardar a disposición da páxina orixinal senón tamén marcar os textos obtidos nunha procura a texto completo (véxase Ilustración 4). Estes ficheiros denominaranse co mesmo nome das imaxes en formato TIFF coa extensión 'xml'.
- Ficheiro PDF/A co texto oculto para cada unidade documental conforme á norma ISO 19005-1:2005 Document management—Electronic document file format for long-term preservation—Part 1: Use of PDF 1.4 (PDF/A-1).

Normativa e recomendacións:

- ALTO: Analyzed Layout and Text Object, Technical Metadata for Optical Character Recognition, versión 3.0 <http://www.loc.gov/standards/alto/v3/alto-3-0.xsd>
- Recomendacións “Using ALTO with METS” <http://www.loc.gov/standards/alto/techcenter/use-with-mets.php>
- Metadata Encoding & Transmission Standard (METS), 1.12 <http://www.loc.gov/standards/mets/mets.xsd>
- ISO 19005-1:2005 Document management—Electronic document file format for long-term preservation—Part 1: Use of PDF 1.4 (PDF/A-1)

#### 2.2.4. Marca de auga

A marca de auga ten como obxectivo identificar inequivocamente o propietario intelectual do obxecto dixital. Polo tanto, para os efectos de difusión e identificación da procedencia dos obxectos dixitais, nos ficheiros de consulta con formato JPEG, poderase inserir unha marca de auga transparente de tamaño reducido en calquera dos ángulos da imaxe dixital sempre que non obstaculice a súa visibilidade.

Pola contra, os ficheiros destinados a preservación (ficheiros TIFF) non levarán inserta ningunha marca de auga.

Como marca de auga recoméndase poñer as siglas do arquivo correspondente na marxe inferior dereito, fóra da área da imaxe.

#### 2.3. Proceso de creación de METS ([Metadata Encoding and Transmission Standard](#))

A dixitalización dos documentos en papel debe cumprir os requisitos necesarios para que os obxectos dixitais resultantes poidan integrarse no Sistema de Descricións Arquivísticas e no Sistema de Preservación.

Cada documento dixitalizado ten que xerar un obxecto dixital que se describirá nun ficheiro METS v.1.12 ([Metadata Encoding and Transmission Standard](#)). Estes metadatos son os que permiten identificar, describir e localizar os obxectos dixitais.

O esquema METS ([Metadata Encoding and Transmission Standard](#)) ofrece un medio flexible para codificar metadatos descritivos, administrativos e estruturais para un obxecto dixital, e expresar as complexas relacións entre estes tipos de metadatos, ofrecendo así un estándar útil para o intercambio de obxectos dixitais entre repositorios.

- Por tanto, todos os obxectos dixitais resultado da dixitalización deberán estar descritos nun ficheiro METS ([Metadata Encoding and Transmission Standard](#)). O ficheiro METS

incorpora diferentes seccións onde se especifican os metadatos descritivos, de dereitos de autor e propiedade intelectual, de preservación e técnicos en cada unha das seccións correspondentes.

Seguindo as directrices para proxectos de dixitalización de coleccións e fondos de dominio público, en particular para aqueles custodiados en bibliotecas e arquivos, creáronse os seguintes perfís de ficheiros METS ([Metadatas Encoding and Transmission Standard](#)):

- Para manuscritos, impresos, fotografías, música impresa e material cartográfico seguirase o perfil Model Imaged Object Profile:  
<http://www.loc.gov/standards/mets/profiles/00000004.xml>
- Para a prensa histórica seguirase o METS Profile for Historical Newspapers [Draft]  
<http://www.loc.gov/standards/mets/test/ndnp/00000010.html>
- Para material bibliográfico seguirase o Perfil METS-BVPB para inxesta e preservación de obxectos dixitais <http://www.loc.gov/standards/mets/profiles/00000044.xml>

Para permitir a codificación, transmisión e preservación dos obxectos dixitais xestionados no Sistema de Descricións Arquivísticas e no Sistema de Preservación do Arquivo Electrónico de Galicia creouse un perfil METS específico, dispoñible na ligazón:

<http://www.loc.gov/standards/mets/profiles/00000046.xml>

Este perfil está completamente descrito no documento "*Perfil METS. Arquivo Electrónico do Patrimonio de Galicia*", onde pode ser consultado. No mesmo documento descríbense as funcionalidades dos ficheiros METS.

Todos os obxectos dixitais que se cren en proxectos de dixitalización de patrimonio documental para a súa inclusión no Arquivo Electrónico de Patrimonio da Xunta de Galicia estarán descritos en ficheiros METS que sigan a estrutura do "*Perfil METS. Arquivo Electrónico do Patrimonio de Galicia*".

Na xeración dos METS terase en conta que o contido sexa o adecuado para cada tipo de entrega e se prestará especial atención ó contido do campo <recordid>, que se corresponderá co identificador do sistema do que se obtén a descrición e que se incluírá no METS sen alteracións para a vinculación do obxecto no Arquivo Dixital e no sistema de preservación.

Outros elementos do contido do METS que compre encher obrigatoriamente no caso da entrega de preservación son, entre outros <ead:filedesc>, <mix:ImageCaptureMetadata>, ou <ead:archdesc>.

Pódense descargar exemplos do eido arquivístico para difusión e preservación na seguinte ligazón:

[http://ficheiros-web.xunta.gal/memoria-dixital/arg\\_obxecto\\_mdg.zip](http://ficheiros-web.xunta.gal/memoria-dixital/arg_obxecto_mdg.zip)

## 2.4. Proceso de validación de obxectos dixitais e metadatos

Recoméndase realizar un proceso de validación de todos os tipos de ficheiros obtidos do tratamento técnico realizado. Así, validaránse as imaxes en formato TIFF, JPEG, ficheiros PDF e todos os ficheiros XML de metadatos.

Para iso pódese utilizar o software JHOVE v1.20 29-03-2018 (JSTOR/Harvard Object Validation Environment) ou a versión máis actualizada no momento da súa aplicación accesible no enderezo <http://jhove.openpreservation.org/>.

Con JHOVE poderase realizar a validación dos seguintes tipos de ficheiros:

- TIFF
- JPEG
- PDF
- ALTO
- METS e os seus compoñentes:
- PREMIS
- MIX
- TEXTMD
- METSRights

Os resultados da validación entregaranse nun único ficheiro con formato XML por cada formato. O resultado da validación rexistrárase ademais como un 'evento' PREMIS dentro do correspondente ficheiro METS, no elemento <digiprovMD> de <amdSec>.

```
<?xml version="1.0" encoding="UTF-8"?>
<jhove xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns="http://hul.harvard.edu/ois/xml/ns/jhove"
xsi:schemaLocation="http://hul.harvard.edu/ois/xml/ns/jhove
http://hul.harvard.edu/ois/xml/xsd/jhove/jhove.xsd" name="Jhove" release="1.11"
dáche="2015-05-15">
<dáche>2015-05-18T12:03:46+02:00</dáche>
<repInfo
uri="%5C%5Cgaliciana.gal%5Cdgb%5Cusuarios%5Cejemplo%5Cmecd%5Cdc%5Cdc_DGB
B20140083152.xml">
<reportingModule release="1.4" dáche="2007-01-08">XML-hul</reportingModule>
<lastModified>2015-05-18T11:45:31+02:00</lastModified>
<size>666</size>
<format>XML</format>
<version>1.0</version>
<status>Well-Formed and valid</status>
<sigMatch>
<module>XML-hul</module>
</sigMatch>
```

### 3. REQUISITOS DOS ENTREGABLES DUN PROXECTO DE DIXITALIZACIÓN

#### 3.1. Creación dos obxectos dixitais

Cos ficheiros das imaxes resultantes dun proxecto de dixitalización se crearán os obxectos dixitais, que incluírán os seus correspondentes ficheiros METS ([Metadata Encoding and Transmission Standard](#)).

Cada obxecto dixital estará composto polos seguintes formatos:

- Ficheiros máster en formato TIFF 6.0 sen comprimir
- Ficheiros de acceso con compresión media en formato JPEG
- Ficheiros de acceso (miniaturas) en formato JPEG
- Ficheiros METS correspondentes
- Ficheiros PDF/A, un por documento (opcional se non se realiza recoñecemento óptico de caracteres).<sup>1</sup>
- Ficheiros de validación con Jhove en formato XML.

Da documentación á que se realice Recoñecemento Óptico de Caracteres entregárase ademais:

- Ficheros ALTO v.3.0 (un por páxina) co resultado do Recoñecemento Óptico de Caracteres. Estes ficheiros denominaranse co mesmo nome das imaxes en formato TIFF, coa extensión 'xml'.
- Ficheros en formato PDF/A con o texto oculto conforme á norma ISO 19005-1 (un ficheiro PDF/A por documento).<sup>2</sup>

#### 3.2. Nomeado de ficheiros e cartafoles

Para nomear os ficheiros e directorios recoméndase:

- O nome de cada ficheiro debe ser único.
- A estrutura do nome debe ser consistente.
- Se se usa un esquema numérico no nomeado de ficheiros hai que ter en conta o número máximo de páxinas a dixitalizar en cada documento.

---

<sup>1</sup> Os ficheiros PDF son de utilidade para posibilitar a descarga do obxecto dixital completo e permitiránlle á persoa usuaria descargar dunha soa vez o documento e non ter que facelo páxina por páxina.

<sup>2</sup> Os ficheiros PDF con OCR oculto permiten mostrar a imaxe orixinal e buscar sobre o texto, o cal se mantén oculto á persoa usuaria.

- Non utilizar nomes longos e complexos que poidan inducir a erros durante a introdución de datos.
- Utilizar minúsculas e extensións.
- Utilizar números ou letras pero non símbolos nin espazos.
- Utilizar un cartafol por cada unidade documental cuxo nome será a súa signatura.
- Repetir nas imaxes o código utilizado para o cartafol máis un sufixo \_nnnnn.
- Crear cartafoles separados para cada formato de entrega, é dicir, un cartafol para os ficheiros TIFF, outra para os ficheiros JPEG, outra para os ficheiros PDF...
- Nomear todos os cartafoles e ficheiros en minúsculas, pois iso evita problemas en migracións entre contornas Windows e Unix.
- Normalizar a signatura, cambiando os caracteres acentuados por caracteres non acentuados, as eñes por enes e as cedillas por ces.
- Substituír na signatura os signos de puntuación por guións baixos. Poden deixarse tal cal os puntos e os guións medios, o resto de caracteres de puntuación deberían substituírse polo carácter guión baixo (\_).

### 3.3. Formatos de entrega

Realizárase a xeración de dous formatos de entrega, unha para preservación e outra para difusión.

#### 3.3.1. Copias para preservación

O formato de preservación terá como finalidade a carga dos obxectos dixitais no Sistema de Preservación. A inxesta no devandito sistema é responsabilidade da Axencia de Modernización Tecnolóxica de Galicia (AMTEGA) e outra para o Arquivo de Galicia.

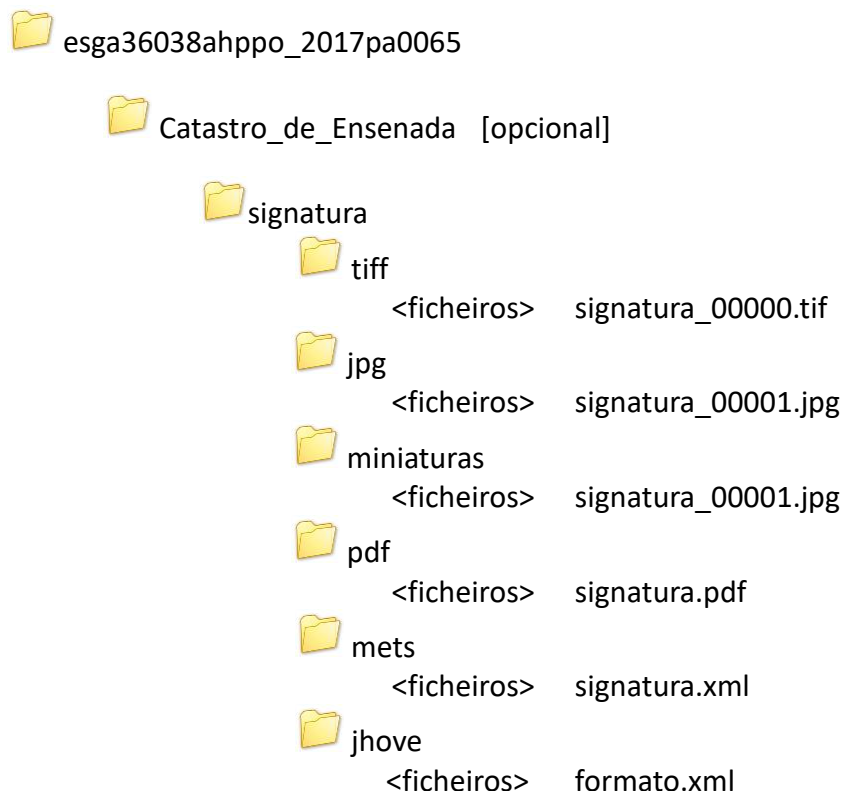
A entrega farase seguindo a seguinte **estrutura de directorios**:

- **Primeiro nivel:** Cartafol identificador do arquivo e do proxecto ao que corresponden as imaxes dixitalizadas. Para a designación do arquivo deberá utilizarse o seu código identificativo. Exem.: esga36038ahppo\_2017pa0065.  
  
Pódese engadir un subnivel adicional para a separación lóxica dos obxectos se o arquivo así o indica.
- **Segundo nivel:** Cartafoles identificadores dos obxectos dixitais. Para a súa designación utilizarase a signatura, engadindo opcionalmente algún prefixo definido pola institución coas súas iniciais e/ou a do fondo. Exem.: arg\_rig\_45793-30 ou adgd20180032806

- **Terceiro nivel:** Cartafoles identificativos dos formatos das imaxes que contén. Por cada obxecto dixital farase unha entrega dos seguintes formatos:
  - Ficheiro máster en formato TIFF 6.0 sen comprimir
  - Ficheiro de acceso con comprensión media en formato JPEG
  - Ficheiro de acceso (miniaturas) en formato JPEG
  - Ficheiro METS correspondente
  - Ficheiro PDF/A, un por documento
  - Ficheiros de validación con Jhove en formato XML.
- **Cuarto nivel:** Cando os obxectos dixitais estean formados por varias imaxes, a identificación do obxecto irá seguida do número de orde de cada imaxe dixitalizada.

Entregarase tamén por cada proxecto un **cartafol contendo os ficheiros resultado do proceso de validación** de todos os tipos de ficheiros obtidos.

Exemplo de Estrutura de entrega de **preservación**:





### 3.3.2. Copias para difusión

O formato de difusión terá como finalidade a carga dos obxectos dixitais no Arquivo Dixital de Galicia. A inxesta no devandito sistema é responsabilidade da Axencia de Modernización Tecnolóxica de Galicia (AMTEGA) e outra para o Arquivo de Galicia.

A entrega farase seguindo a seguinte estrutura de directorios:

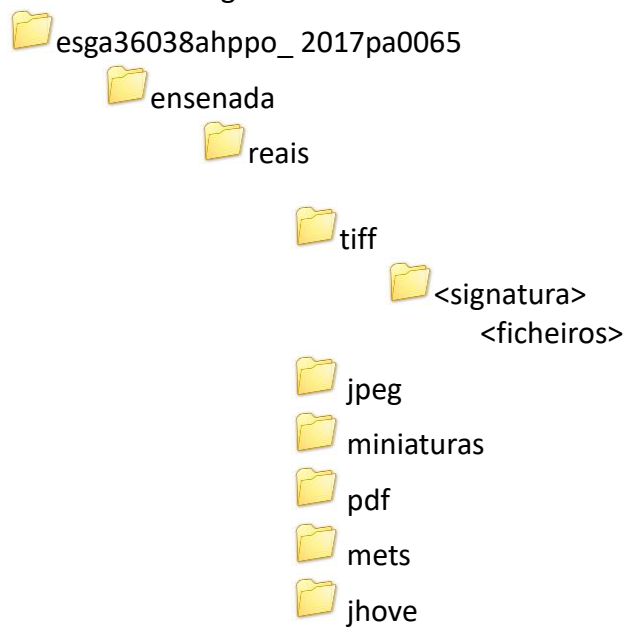
Entregarase una copia de difusión ao arquivo no que se custodien os documentos obxecto da dixitalización, que será o responsable da súa carga no Sistema de Difusión. Neste caso as imaxes entregaranse seguindo a seguinte **estrutura de directorios**:

- **Primeiro nivel:** Cartafol identificador do arquivo e do proxecto ao que corresponden as imaxes dixitalizadas. Para a designación do arquivo deberá utilizarse o seu código identificador. Exem.: esga36038ahppo\_2017pa0065  
  
Pódense engadir un ou varios subniveis adicionais para a separación lóxica dos obxectos se o arquivo así o indica.
- **Segundo nivel:** Cartafoles identificativos dos formatos das imaxes que contén. Por cada obxecto dixital farase unha entrega dos seguintes formatos:
  - Ficheiro máster en formato TIFF 6.0 sen comprimir
  - Ficheiro de acceso con comprensión media en formato JPEG
  - Ficheiro de acceso (miniaturas) en formato JPEG
  - Ficheiro METS correspondentes.
  - Ficheiro PDF/A, un por documento
  - Ficheiros de validación con Jhove en formato XML.
- **Terceiro nivel:** Cartafoles identificadores dos obxectos dixitais. Para a súa designación utilizarase a signatura, engadindo opcionalmente algún prefixo definido pola institución coas súas iniciais e/ou a do fondo. Exem.: arg\_rig\_45793-30 ou adgd20180032806
- **Cuarto nivel:** Cando os obxectos dixitais estean formados por varias imaxes, a identificación do obxecto irá seguida do número de orde de cada imaxe dixitalizada.

No caso das copia de difusión pódese engadir algún nivel intermedio entre o primeiro e o segundo coa identificación do fondo documental ao que pertencen as imaxes dixitalizadas.

Entregarase tamén por cada obxecto un **cartafol contendo os ficheiros resultado do proceso de validación** de todos os tipos de ficheiros obtidos.

Exemplo de Estrutura de entrega de difusión:



#### 4. REFERENCIAS BIBLIOGRÁFICAS

UNE-ISO/TR 13028 (2011) *Información y documentación. Directrices para la implementación de la digitalización de documentos*. Madrid : Aenor

Federal Agencies Digitization Initiative (FADGI). *Technical Guidelines for Digitizing Cultural Heritage Materials: Creation of Raster Image Master Files*, 2010. Disponible en: [http://www.digitizationguidelines.gov/guidelines/FADGI\\_Still\\_Image-Tech\\_Guidelines\\_2010-08-24.pdf](http://www.digitizationguidelines.gov/guidelines/FADGI_Still_Image-Tech_Guidelines_2010-08-24.pdf) [consultado 17 de junio de 2015]

U.S. National Archives and Records Administration (NARA). *Technical Guidelines for Digitizing Archival Materials for Electronic Access: Creation of Production Master Files – Raster Images*, 2004. Disponible en: <http://www.archives.gov/preservation/technical/guidelines.pdf> [consultado el 10 de junio de 2015]

IFLA. *Directrices para proyectos de digitalización : de colecciones y fondos de dominio público, en particular para aquellos custodiados en bibliotecas y archivos*, 2002. Disponible en: <http://www.ifla.org/files/assets/preservation-and-conservation/publications/digitization-projects-guidelines-es.pdf> [consultado el 12 de junio de 2015]

MATRIX: The Center for Humane Arts, Letters and Social Sciences Online at Michigan State University. *Digital Imaging for Archival Preservation and Online Presentation: Best Practices*, 2001. Disponible en: [http://www.historicalvoices.org/papers/image\\_digitization2.pdf](http://www.historicalvoices.org/papers/image_digitization2.pdf) [consultado el 8 de junio de 2015]

Biblioteca de la Universidad de Cornell. *Llevando la teoría a la práctica: tutorial de digitalización de imágenes*, 2000-2003. Disponible en: <http://preservationtutorial.library.cornell.edu/tutorial-spanish/contents.html> [consultado el 15 de marzo de 2019]

*National Archives Preservation Guidelines for Vendors Handling Records and Historical Materials*. Disponible en: <http://www.archives.gov/preservation/technical/vendor-training.html> [consultado el 16 de junio de 2015]

*Sustainability of Digital Formats*. Library of Congress, 2013. Disponible en: <http://www.digitalpreservation.gov/formats/index.shtml>

## 5. ANEXO

### 5.1. Termos técnicos

**Compresión:** utilízase para reducir o tamaño do arquivo de imaxe para o seu almacenamento, procesamento e transmisión. Pode ser con perda e sen perda.

**Copia dixitalizada:** duplicado dun documento mediante un proceso de dixitalización. Un mesmo documento pode ter diferentes copias dixitais: copia dixitalizada en formato tiff, copia dixitalizada en formato jpg, copia dixitalizada en formato pdf.

**Formatos de arquivo:** consisten tanto nos bits que comprende a imaxe como na información do encabezamento acerca de como ler e interpretar o arquivo. A continuación detállanse os formatos de arquivo normalizados máis comúns:

- **TIFF (Tagged Image File Format):** É un formato de ficheiro dixital que acepta a *compresión* sen perda, polo que se conserva a información orixinal completa. Trátase dun formato estándar e de uso libre, e é o máis utilizado con fins de preservación. Proporciona gran calidade para imaxes en cor, escala de grises e branco e negro. O seu principal inconveniente, o tamaño do arquivo que xera. (Norma [ISO 12234-2:2001](#)).
- **JPEG (Joint Photographic Experts Group):** Trátase dun formato de gran calidade para imaxes en cor que ofrece un alto grao de compresión, aínda que esta sempre é con perdas. É un formato estándar moi usado para a visualización en pantalla e o acceso rápido á imaxe. (Norma [ISO/IEC 10918-1:1994](#)).
- **PDF (Portable Document Format):** é un formato de almacenamento de documentos composto (imaxe vectorial, mapa de bits e texto), desenvolvido pola empresa Adobe Systems. Este formato permite distintos graos de compresión, aínda que sempre con perdas. É un estándar moi estendido e mantén unha calidade moi boa. Este formato é idóneo para imprimir e crear documentos secuenciais con páxinas múltiples. O ficheiro PDF/A, está baseado nun subconxunto do formato PDF, é o estándar ISO (ISO 19005-1:2005) para a preservación dos documentos electrónicos a longo prazo.

**Obxecto dixital:** Unidade de información específica en formato dixital. Un Obxecto dixital pode ser unha representación, un ficheiro, unha cadea de bits ou unha cadea de ficheiros. O obxecto dixital enténdese como a representación completa da unidade de información incluíndo todos os ficheiros e metadatos que permiten contextualizala e entendela.

**Píxel:** é a menor unidade que forma parte dunha imaxe dixital e pode estar representado por un número diferente de tons de grises ou de cores.

**Resolución:** a resolución dunha imaxe dixital é o número de píxeles ou puntos por polgada que hai nunha imaxe específica. A resolución determinará a claridade e o detalle da imaxe final. Así, pódese dicir que existe unha relación directa entre a resolución da imaxe e o seu número de cores coa calidade da imaxe e co tamaño do ficheiro que se xera no proceso de dixitalización. Cantos máis píxeles por polgada, a resolución do ficheiro será maior, do mesmo xeito que

aumentará o tempo de descarga. Por iso, os ficheiros en alta resolución teñen a finalidade de preservación e non de acceso.

**Profundidade de bits:** número de cores ou tons de grises que se definen nun único píxel. Os bits son as unidades máis pequenas que pode almacenar un computador, e a profundidade de bit refírese ao número de bits que se usan para representar cada píxel nunha imaxe dixital.

- Binario (1 bit): cada píxel está representado por un único dígito binario, que pode ser ou branco ou negro. Por tanto só hai dúas combinacións posibles.
- Escala de grises (8 bits, 256 niveis de grises): cada píxel está representado por oito díxitos binarios, o que significa que hai un total de 256 combinacións entre o branco e o negro, 256 niveis de grises, para representar un píxel.
- Cor (24 bits, 16.777.216 millóns de cores): neste caso 24 díxitos binarios representan cada píxel. Os píxeles de cada imaxe poden estar representados por máis de 16 millóns de cores. Estas imaxes son as que máis se achegan a representar unha copia fiel do orixinal, pola súa precisión no detalle.

**Tamaño:** é importante que nos parámetros de dixitalización se manteña o tamaño real do documento. A partir deste tamaño elixirase a resolución adecuada, para que o resultado final sexa de calidade.

## 5.2. Normativa

- Tagged Image File Format (TIFF). [Norma ISO 12234-2:2001. Electronic still-picture imaging -- Removable memory -- Part 2: TIFF/EP image data format](#)
- Joint Photographic Experts Group (JPEG). [Norma ISO/IEC 10918-1:1994. Information technology -- Dixital compression and coding of continuous-tone still images: Requirements and guidelines](#)
- Portable Document Format (PDF). [Norma ISO 32000-1:2008. Document management -- Portable document format -- Part 1: PDF 1.7](#)
- Portable Document Format for long-term preservation of electronic documents (PDF/A). [Norma ISO 19005-1:2005 Document management -- Electronic document file format for long-term preservation -- Part 1: Use of PDF 1.4 \(PDF/A-1\)](#)
- [Metadata Encoding and Transmission Standard](#) (METS)
- [Analyzed Layout and Text Object](#) (ALTO)

### 5.3. Centros de arquivo integrados en Arquivo Dixital de Galicia

<a href="#">ES.GA.15078.ACPL</a>	<a href="#">Arquivo Central da Secretaría Xeral de Política Lingüística</a>
<a href="#">ES.GA.15078.ACCU</a>	<a href="#">Arquivo Central de Cultura</a>
<a href="#">ES.GA.15078.ACI</a>	<a href="#">Arquivo Central de Economía e Industria de Santiago de Compostela</a>
<a href="#">ES.GA.15078.ACPS</a>	<a href="#">Arquivo Central de Política Social</a>
<a href="#">ES.GA.27028.ADLU</a>	<a href="#">Arquivo da Deputación Provincial de Lugo</a>
<a href="#">ES.GA.32054.ADOU</a>	<a href="#">Arquivo da Deputación Provincial de Ourense</a>
<a href="#">ES.GA.15078.AFLT</a>	<a href="#">Arquivo da Fundación Luís Tilve</a>
<a href="#">ES.GA.36057.AFP</a>	<a href="#">Arquivo da Fundación Penzol</a>
<a href="#">ES.GA.15065.AFCJC</a>	<a href="#">Arquivo da Fundación Pública Galega Camilo José Cela</a>
<a href="#">ES.GA.15030.ARAG</a>	<a href="#">Arquivo da Real Academia Galega</a>
<a href="#">ES.GA.15078.AG</a>	<a href="#">Arquivo de Galicia</a>
<a href="#">ES.GA.15030.ADTI</a>	<a href="#">Arquivo do Departamento Territorial de Industria da Coruña</a>
<a href="#">ES.GA.15030.ACD.EG</a>	<a href="#">Arquivo do Instituto de Ensino Medio Eusebio da Guarda</a>
<a href="#">ES.GA.27028.ACD.LA</a>	<a href="#">Arquivo do Instituto de Ensino Medio Lucas Augusti</a>
<a href="#">ES.GA.36038.ACD.OP</a>	<a href="#">Arquivo do Instituto de Ensino Medio Otero Pedrayo</a>
<a href="#">ES.GA.15030.ACD.SM</a>	<a href="#">Arquivo do Instituto de Ensino Medio Salvador de Madariaga</a>
<a href="#">ES.GA.15078.ACD.XI</a>	<a href="#">Arquivo do Instituto de Ensino Medio Xelmírez I</a>
<a href="#">ES.GA.15030.AIJCEC</a>	<a href="#">Arquivo do Instituto José Cornide de Estudos Coruñeses</a>
<a href="#">ES.GA.15030.ARG</a>	<a href="#">Arquivo do Reino de Galicia</a>
<a href="#">ES.GA.15078.AHDS</a>	<a href="#">Arquivo Histórico Diocesano de Santiago de Compostela</a>
<a href="#">ES.GA.36055.AHDT</a>	<a href="#">Arquivo Histórico Diocesano de Tui</a>
<a href="#">ES.GA.27028.AHPLU</a>	<a href="#">Arquivo Histórico Provincial de Lugo</a>
<a href="#">ES.GA.32054.AHPOU</a>	<a href="#">Arquivo Histórico Provincial de Ourense</a>
<a href="#">ES.GA.36038.AHPPQ</a>	<a href="#">Arquivo Histórico Provincial de Pontevedra</a>
<a href="#">ES.GA.15030.AM</a>	<a href="#">Arquivo Municipal da Coruña</a>
<a href="#">ES.GA.15001.AM</a>	<a href="#">Arquivo Municipal de Abegondo</a>
<a href="#">ES.GA.15002.AM</a>	<a href="#">Arquivo Municipal de Ames</a>
<a href="#">ES.GA.15008.AM</a>	<a href="#">Arquivo Municipal de Bergondo</a>
<a href="#">ES.GA.15009.AM</a>	<a href="#">Arquivo Municipal de Betanzos</a>
<a href="#">ES.GA.15901.AM</a>	<a href="#">Arquivo Municipal de Cariño</a>
<a href="#">ES.GA.15036.AM</a>	<a href="#">Arquivo Municipal de Ferrol</a>
<a href="#">ES.GA.27019.AM</a>	<a href="#">Arquivo Municipal de Foz</a>
<a href="#">ES.GA.27028.AM</a>	<a href="#">Arquivo Municipal de Lugo</a>
<a href="#">ES.GA.15054.AM</a>	<a href="#">Arquivo Municipal de Narón</a>
<a href="#">ES.GA.15061.AM</a>	<a href="#">Arquivo Municipal de Ortigueira</a>
<a href="#">ES.GA.15073.AM</a>	<a href="#">Arquivo Municipal de Ribeira</a>
<a href="#">ES.GA.36201.ATVI</a>	<a href="#">Arquivo Xeral Territorial de Vigo</a>
<a href="#">ES.GA.15030.AUDC</a>	<a href="#">Centro de Documentación e Arquivo da UDC</a>
<a href="#">ES.GA.0</a>	<a href="#">Patrimonio Documental de Galicia</a>

Consultar listado actualizado en:

<http://arquivo.galiciana.gal/arpadweb/gl/archivos/centrosArquivo.do>