

Image conversion (imagemagick)

Description

ImageMagick is a tool that can read and write images in a variety of formats (over 200) including PNG, JPEG, JPEG-2000, GIF, TIFF, DPX, EXR, WebP, Postscript, PDF, and SVG. ImageMagick can also be used to resize, flip, mirror, rotate, distort, shear and transform images, adjust image colours, apply various special effects, or draw text, lines, polygons, ellipses and Bézier curves (e.g. set Command arguments to “-resample 90” to resize the image to 90 dpi). The results of conversion will be placed on a new representation under the same Archival Information Package (AIP) where the files were originally found. A PREMIS event is also recorded after the task is run. For a full list of supported formats, please visit <http://www.imagemagick.org/script/formats.php>

Parameters

Name	Description	Type	Default value	Possible values	Mandatory
Input format	Input file format to be converted (check documentation for list of supported formats). If the input file format is not specified, the task will run on all supported formats (check roda-core-formats.properties for list of supported formats).	STRING			true
Output format	Output file format to be converted (check documentation for list of supported formats).	STRING	jpg		true

Ignore other files	Do not process files that have a different format from the indicated.	BOOLEAN	true		false
Create dissemination	If this is selected then the plugin will convert the files and create a new dissemination. If not, a new representation will be created.	BOOLEAN	true		false
Dissemination title	If the 'create dissemination' option is checked, then this will be the respective dissemination title.	STRING	Low resolution image		false
Dissemination description	If the 'create dissemination' option is checked, then this will be the respective dissemination description.	STRING	JPEG format at 80% quality and a maximum of 1600 pixels.		false
Representation type	Attribute a type when creating a new representation	REPRESENTATION_TYPE			false
Command arguments	Command arguments that will be passed to the command of the tool as configured (advanced users	STRING	-geometry 1600x1600> -quality 80 -strip		true

	only!)				
--	--------	--	--	--	--

File not characterized risk assessment

Description

Check file formatName information (Mimetype, PRONOM and Format designation). If this information is missing, it creates a new risk called “File(s) not comprehensively characterized” and assigns the file to that risk in the Risk register.

Parameters

Name	Description	Type	Default value	Possible values	Mandatory
Mimetype	Check existence of a Mimetype?	BOOLEAN	true		false
PRONOM	Check existence of a PRONOM Unique Identifier (PUID)?	BOOLEAN	true		false
Format designation name	Check existence of a Format designation name?	BOOLEAN	true		false
Format designation version	Check existence of a Format designation version?	BOOLEAN	true		false

Missing representation information risk assessment

Description

Verifies if all intellectual entities, representations and files have at least one associated representation information. If there is not any association, a risk incidence is created in the risk register.

Parameters

Name	Description	Type	Default value	Possible values	Mandatory
Intellectual entity fields to validate	Check the fields you want to be related to a representation information.	AIP_FIELDS			false
Representation fields to validate	Check the fields you want to be related to a representation information.	REPRESENTATION_FIELDS			false
File fields to validate	Check the fields you want to be related to a representation information.	FILE_FIELDS			false

Representation information broken links verification

Description

It removes broken links to specific intellectual entities or other representation information

Parameters

No parameters.

AIP feature extraction (FITS)

Description

The File Information Tool Set (FITS) identifies, validates and extracts technical metadata for a wide range of file formats. It acts as a wrapper, invoking and managing the output from several other open source tools. Output from these tools are converted into a common format, compared to one another and consolidated into a single XML output file. The tools used in the latest version of FITS are: ADL Tool, Apache Tika, DROID, Exiftool, FFIdent, File Utility (windows port), Jhove, MediaInfo, National Library of New Zealand Metadata Extractor, OIS Audio Information, OIS

File Information, OIS XML Information. The task creates a new file under the [AIP_ID]/representation/metadata/other/FITS. This information is not yet added to PREMIS or indexed but it can be inspected by downloading the AIP. A PREMIS event is recorded after the task is run. More information on this tool can be found at <http://projects.iq.harvard.edu/fits>

Parameters

No parameters.

AIP feature extraction (ExifTool)

Description

ExifTool is a platform-independent application capable of reading technical metadata from a wide variety of file formats. ExifTool supports many different metadata formats including EXIF, GPS, IPTC, XMP, JFIF, GeoTIFF, ICC Profile, Photoshop IRB, FlashPix, AFCP and ID3, as well as the maker notes of many digital cameras by Canon, Casio, DJI, FLIR, FujiFilm, GE, HP, JVC/Victor, Kodak, Leaf, Minolta/Konica-Minolta, Motorola, Nikon, Nintendo, Olympus/Epson, Panasonic/Leica, Pentax/Asahi, Phase One, Reconyx, Ricoh, Samsung, Sanyo, Sigma/Foveon and Sony. The task creates a new file under the [AIP_ID]/representation/metadata/other/ExifTool. This information is not yet added to PREMIS or indexed but it can be inspected by downloading the AIP. A PREMIS event is recorded after the task is run. For a full list of supported file formats and metadata types, please visit <http://www.sno.phy.queensu.ca/~phil/exiftool/#supported>

Parameters

No parameters.

PDF/A format validation (VeraPDF)

Description

This action validates PDF files to make sure they comply to the PDF/A specification. PDF/A is an ISO-standardized version of the Portable Document Format (PDF) specialized for use in the archiving and long-term preservation of electronic documents. PDF/A differs from PDF by prohibiting features ill-suited to long-term archiving, such as font linking (as opposed to font embedding) and encryption. The specification for PDF/A is a set of restrictions and requirements applied to the “base” PDF standards (PDF 1.4 for PDF/A-1 and ISO 32000 for PDF/A-2 and PDF/A-3) plus a specific set of 3rd party standards. The outcome of this action is the creation of a new technical metadata file in the Archival Information Package (under the folder “metadata/other”) that records the output of the VeraPDF tool. A PREMIS event is also recorded after running this task. For more information about VeraPDF, please visit

Parameters

Name	Description	Type	Default value	Possible values	Mandatory
PDF Profile	Validates files according to the specified PDF/A profile (valid options include: 1a, 1b, 2a, 2b, 2u, 3a, 3b, 3u)	STRING	1b	[0, 1a, 1b, 2a, 2b, 2u, 3a, 3b, 3u, 4]	true
Ignore non PDF files	Ignore files that are not identified as PDF.	BOOLEAN	true		false

Create E-ARK AIP manifest files (METS.xml)

Description

Plugin that generates E-ARK AIP manifest files ("METS.xml") from existing AIP information in the storage layer. This plugin only works with filesystem as the storage service.

Parameters

No parameters.

Create E-ARK DIP manifest files (METS.xml)

Description

Plugin that generates E-ARK DIP manifest files ("METS.xml") from existing AIP information in the storage layer. This plugin only works with filesystem as the storage service.

Parameters

Name	Description	Type	Default value	Possible values	Mandatory
Include only selected descriptive metadata	Include only selected descriptive metadata on the following parameter.	BOOLEAN	false		false
Selected descriptive metadata (type_version)	The selected descriptive metadata to filter with types and versions.	STRING	ead3, dc2002, ead_2002		true
Include all preservation metadata	Include all preservation metadata inside the AIP.	BOOLEAN	true		false
Include only selected other metadata	Include only selected other metadata on the following parameter.	BOOLEAN	false		false
Selected other metadata (type)	The selected other metadata to filter with types.	STRING	Siegfried, Tika		true
Include only selected representations	Include only selected representations on the following parameter.	BOOLEAN	false		false
Selected representations (type)	The selected representations to filter with types.	STRING	pdfa, MIXED		true
Include schemas folder	Include schemas folder contained on AIP and representations.	BOOLEAN	true		false
Include documentation folder	Include documentation folder contained on AIP and representations.	BOOLEAN	true		false

GhostScript conversion

Description

Converts files using GhostScript.

Parameters

Name	Description	Type	Default value	Possible values	Mandatory
Input format	Input file format to be converted (check documentation for list of supported formats). If the input file format is not specified, the task will run on all supported formats (check roda-core-formats.properties for list of supported formats).	STRING			true
Output format	Output file format to be converted (check documentation for list of supported formats).	STRING			true
Ignore other files	Do not process files that have a different format from the indicated.	BOOLEAN	true		false
Create dissemination	If this is selected then the plugin will convert the files and create a new dissemination. If	BOOLEAN	true		false

	not, a new representation will be created.				
Dissemination title	If the 'create dissemination' option is checked, then this will be the respective dissemination title.	STRING	Dissemination title		false
Dissemination description	If the 'create dissemination' option is checked, then this will be the respective dissemination description.	STRING	Dissemination description		false
Representation type	Attribute a type when creating a new representation	REPRESENTATION_TYPE			false
Command arguments	Command arguments that will be passed to the command of the tool as configured (advanced users only!)	STRING			true

AIP feature extraction (MedialInfo)

Description

MedialInfo extracts technical and tag data for video and audio files. MedialInfo supports popular video formats (e.g. AVI, WMV, QuickTime, Real, DivX, XviD) as well as lesser known or emerging formats such as MKV including WebM. MedialInfo reveals information such as: Title, author, director, album, track number, date, duration, codec, aspect ratio, framerate, bitrate, Audio codec, sample rate, channels, language, bitrate, subtitle language, etc. The task creates a new file under the [AIP_ID]/representation/metadata/other/MedialInfo. This information is not yet added to

PREMIS or indexed but it can be inspected by downloading the AIP. A PREMIS event is recorded after the task is run.

Parameters

No parameters.

Feature extraction (Apache Tika)

Description

The Apache Tika tool extracts technical metadata and text from over a thousand different file types (such as PPT, XLS, and PDF). The task updates PREMIS objects metadata in the entity to store the results of the characterization process. A PREMIS event is also recorded after the task is run. For more information on this tool, please visit <https://tika.apache.org>

Parameters

Name	Description	Type	Default value	Possible values	Mandatory
Feature extraction	Perform feature extraction from files. This will extract properties such as number of pages, width, height, colour space, etc.	BOOLEAN	true		true
Full text extraction	Extracts full text from document/textual files. Extracted text is used to perform full-text searching on the catalogue.	BOOLEAN	true		true

AIP feature extraction (JHOVE)

Description

JHOVE, the JSTOR/Harvard Object Validation Environment, is an extensible software framework for performing format characterization of digital objects. JHOVE includes modules for arbitrary byte streams, ASCII and UTF-8 encoded text, TIFF, HTML, XML, JPEG, JPEG2000, PDF, AIFF, WAVE audio; and text and XML output handlers. The task creates a new file under the [AIP_ID]/representation/metadata/other/JHOVE. This information is not yet added to PREMIS or indexed but it can be inspected by downloading the AIP. A PREMIS event is recorded after the task is

run. For more information on this tool, please visit <http://jhove.openpreservation.org>

Parameters

No parameters.

General command based conversion

Description

Generates a new format file using a generic command line.

Parameters

Name	Description	Type	Default value	Possible values	Mandatory
Input format	Input file format to be converted (check documentation for list of supported formats). If the input file format is not specified, the task will run on all supported formats (check roda-core-formats.properties for list of supported formats).	STRING			true
Output format	Output file format to be converted (check documentation for list of supported formats).	STRING			true

Ignore other files	Do not process files that have a different format from the indicated.	BOOLEAN	true		false
Create dissemination	If this is selected then the plugin will convert the files and create a new dissemination. If not, a new representation will be created.	BOOLEAN	true		false
Dissemination title	If the 'create dissemination' option is checked, then this will be the respective dissemination title.	STRING	Dissemination title		false
Dissemination description	If the 'create dissemination' option is checked, then this will be the respective dissemination description.	STRING	Dissemination description		false
Representation type	Attribute a type when creating a new representation	REPRESENTATION_TYPE			false
Command arguments	Command arguments that will be passed to the command of the tool as configured (advanced users only!)	STRING			true

Video conversion (avconv)

Description

avconv is a very fast video and audio converter. It can also convert between arbitrary sample rates and resize video on the fly with a high quality polyphase filter. The results of conversion will be placed on a new representation under the same Archival Information Package (AIP) where the files were originally found. A PREMIS event is also recorded after the task is run. Each input or output file can in principle contain any number of streams of different types (video/audio/subtitle/attachment/data). Allowed number and/or types of streams can be limited by the container format. For more information about this tool, please visit <https://libav.org/documentation/avconv.html>

Parameters

Name	Description	Type	Default value	Possible values	Mandatory
Input format	Input file format to be converted (check documentation for list of supported formats). If the input file format is not specified, the task will run on all supported formats (check roda-core-formats.properties for list of supported formats).	STRING			true
Output format	Output file format to be converted (check documentation for list of supported formats).	STRING	mp4		true
	Do not process				

Ignore other files	files that have a different format from the indicated.	BOOLEAN	true		false
Create dissemination	If this is selected then the plugin will convert the files and create a new dissemination. If not, a new representation will be created.	BOOLEAN	true		false
Dissemination title	If the 'create dissemination' option is checked, then this will be the respective dissemination title.	STRING	MP4 video		false
Dissemination description	If the 'create dissemination' option is checked, then this will be the respective dissemination description.	STRING	MP4 format for web visualization.		false
Representation type	Attribute a type when creating a new representation	REPRESENTATION_TYPE			false
Command arguments before input	Command arguments that will be passed to the command of the tool as configured (advanced users only!)	STRING			true

Command arguments between input and output	Command arguments to modify the output type that will be passed to the command of the tool as configured (advanced users only!)	STRING	-c:v libx264 -crf 28 -c:a aac -b:a 128k -strict experimental	true
---	---	--------	--	------

AIP feature extraction (avprobe)

Description

avprobe extracts information from multimedia streams and prints it in human and machine-readable fashion. For example, it can be used to check the format of the container used by a multimedia stream and the format and type of each media stream contained in it. The task creates a new file under the [AIP_ID]/representation/metadata/other/Avprobe. This information is not yet added to PREMIS or indexed but it can be inspected by downloading the AIP. A PREMIS event is recorded after the task is run. For more information of the avprobe tool, please visit <https://libav.org/documentation/avprobe.html>

Parameters

No parameters.

Permissions management

Description

Manages AIP permissions giving the possibility to add or remove specific group or user permissions

Parameters

Name	Description	Type	Default value	Possible values	Mandatory
User permissions	A list of comma separated users to be added or removed to the AIPs permissions.	STRING			true
Group permissions	A list of comma separated groups to be added or removed to the AIPs permissions.	STRING			true
Permission types	Select the permission types you want to give or remove.	PERMISSION_TYPES			true
Remove permissions	If selected, the plugin will remove the permissions instead of adding it.	BOOLEAN	false		true

AIP file format identification (DROID)

Description

DROID (Digital Record Object Identification) is a software tool developed to perform automated batch identification of file formats. DROID is designed to meet the fundamental requirement of any digital repository to be able to identify the precise format of all stored digital objects, and to link that identification to a central registry of technical information about that format and its dependencies. DROID uses the PRONOM signature files to perform format identification. Like PRONOM, it was developed by the National Archives of the UK. The task creates a new file under the [AIP_ID]/representation/metadata/other/droid. This information is not yet added to PREMIS or indexed but it can be inspected by downloading the AIP. A PREMIS event is recorded after the task is run.

Parameters

No parameters.

Plugins

LTV extract signature dates plugin

Description

This plugin extracts the signature dates embedded on the signed PDF files. The dates extracted will be used to control when should the PDF be resigned again in order to maintain its validity

Parameters

Name	Description	Type	Default value	Possible values	Mandatory
Add preservation state to created representations	Add preservation state to created representations.	BOOLEAN	false		true

MLTV resign plugin

Description

This plugin executes a resigning of PDF files via an external request to specific endpoint responsible for it

Parameters

Name	Description	Type	Default value	Possible values	Mandatory
Delete last preservation representation	Deletes last preservation representation if not the original one.	BOOLEAN	true		true
Just create a to-sign report	Just create a to-sign report instead of actually sign the files.	BOOLEAN	false		true
Resign files with less than X days to be invalid	The plugin will resign files with less than X days to be invalid or already on that state.	INTEGER	null		false

Audio conversion (SoX)

Description

SoX is a cross-platform utility that can convert various formats of computer audio files in to other formats. It can also apply various effects to these sound files (see the “Command arguments” setting). The results of conversion will be placed on a new representation under the same Archival Information Package (AIP) where the files were originally found. A PREMIS event is also recorded after the task is run. For a full list of supported audio formats and effects, please visit <http://sox.sourceforge.net/Docs/Features>

Parameters

Name	Description	Type	Default value	Possible values	Mandatory
Input format	Input file format to be converted (check documentation for list of supported formats). If the input file format is not specified, the task will run on all supported formats (check roda-core-formats.properties for list of supported formats).	STRING			true
Output format	Output file format to be converted (check documentation for list of supported formats).	STRING	mp3		true
Ignore other files	Do not process files that have a different format from the indicated.	BOOLEAN	true		false

Create dissemination	If this is selected then the plugin will convert the files and create a new dissemination. If not, a new representation will be created.	BOOLEAN	true		false
Dissemination title	If the 'create dissemination' option is checked, then this will be the respective dissemination title.	STRING	MP3 audio		false
Dissemination description	If the 'create dissemination' option is checked, then this will be the respective dissemination description.	STRING	MP3 audio format for Web reproduction.		false
Representation type	Attribute a type when creating a new representation	REPRESENTATION_TYPE			false
Command arguments	Command arguments that will be passed to the command of the tool as configured (advanced users only!)	STRING			true

Office documents conversion (unoconv)

Description

Converts office files using the “unoconv” (Universal Office Converter). The results of conversion will be placed on a new representation under the same Archival Information Package (AIP) where the files were originally found. A PREMIS event is also recorded after the task is run. “unoconv” is a tool that converts between any document format that OpenOffice understands. It uses OpenOffice’s UNO bindings for non-interactive conversion of documents. Supported document formats include Open Document Format (odt), MS Word (doc), MS Office Open/MS OOXML (ooxml), Portable Document Format (pdf), HTML (html), XHTML (xhtml), RTF (rtf), Docbook (docbook), and more. The outcome of this task is the creation of a new OpenOffice (and thus unoconv) support various import and export formats. Not all formats that can be imported can be exported and vice versa. For a full list of supported formats, please visit - <http://dag.wiee.rs/home-made/unoconv/>

Parameters

Name	Description	Type	Default value	Possible values	Mandatory
Input format	Input file format to be converted (check documentation for list of supported formats). If the input file format is not specified, the task will run on all supported formats (check roda-core-formats.properties for list of supported formats).	STRING			true
Output format	Output file format to be converted (check documentation for list of supported formats).	STRING	pdf		true
	Do not process files that have a				

Ignore other files	different format from the indicated.	BOOLEAN	true		false
Create dissemination	If this is selected then the plugin will convert the files and create a new dissemination. If not, a new representation will be created.	BOOLEAN	true		false
Dissemination title	If the 'create dissemination' option is checked, then this will be the respective dissemination title.	STRING	PDF document		false
Dissemination description	If the 'create dissemination' option is checked, then this will be the respective dissemination description.	STRING	PDF document format for Web visualization.		false
Representation type	Attribute a type when creating a new representation	REPRESENTATION_TYPE			false
Command arguments	Command arguments that will be passed to the command of the tool as configured (advanced users only!)	STRING			true

Digital signature validation

Description

Checks if digital signatures embedded in files are valid. The task supports the following formats: PDF, Microsoft Office Formats (.docx, .xlsx, .pptx) and OpenDocument formats (.odt, .ods, .odp). The outcome of this action is three-fold: 1) the outcome of verification is stored in a PREMIS event; 2) the extracted digital signatures are stored within the AIP under the “metadata/other” folder; and 3) the files with the digital signature removed are stored under a new representation in the Archival Information Package (AIP).

Parameters

Name	Description	Type	Default value	Possible values	Mandatory
Verify digital signature	Verifies the digital signature of files.	BOOLEAN	true		true
Extract digital signature	Extracts the digital signature and stores it in the AIP under metadata/other folder.	BOOLEAN	false		true
Strip digital signature	Removes the digital signature from the original file.	BOOLEAN	true		true
Create dissemination	If this is selected then the plugin will strip the file to a new dissemination. If not, a new representation will be created.	BOOLEAN	true		false
Ignore non embedded digital signed format files	Ignore files that are not recognised as a embedded digital signed format.	BOOLEAN	true		false

Digital sign entities

Description

Creates a new Dissemination Information Package (DIP) for this AIP containing all the files in a given representation and appends a digital signature to each of these files. The digital signature (in PKCS#7 format) is an external file with

the same name as the original one but with a .p7s extension. Digital signatures are generated based on the digital certificate installed under “/config/certificates/”.

Parameters

No parameters.