Deliverable

Grant Agreement Number	270137
Full Project Title	Scalable Preservation Environments
Project Acronym	**SCAPE**
Title of Deliverable	Identification and selection of large-scale migration tools and services
Deliverable Number	D10.1
Work-package	WP10
Dissemination Level	**PU** = Public
Deliverable Nature	**R** = Report
Contractual Delivery Date	2011-07-31
Actual Delivery Date	2011-07-31
Author(s)	KEEPS, EA, KB, MSR, ONB, STFC, TUW

Abstract

This document reports on the task of identifying, evaluating and selecting available preservation action tools and services to be used in the SCAPE project. The report presents related work in the context of action tools, namely the work developed under the Planets, CRiB and RODA projects, and introduces models for assessing the quality of software. The evaluation procedure follows a well-defined methodology supported by an evaluation framework based on an international standard. Domain-specific metrics have been created to address the particular problem of selecting tools for the SCAPE project. An overall of 40 tools, ranging from image converters to database migration suites, have been identified and evaluated. After the assessment procedure, 16 of these tools have been considered inadequate to be used in the SCAPE project (40% rejection rate).

Keyword list

Digital preservation, Action tool, Action services, Conversion, Migration, Evaluation framework

Authors

Person	Role	Partner	Contribution
Rui Castro	Author	KEEPS	Main authoring
Luís Faria	Author	KEEPS	Metrics
Christopher Becker	Author	TUW	Quality model
Markus Hamm	Author	TUW	Quality model
Miguel Ferreira	Owner	KEEPS	Related work and revision

Document Approval

Person	Role	Partner	
Hannes Kulovits	Reviewer	TUW	Review
Paul Wheatley	SP lead	BL	Review

Distribution

Person	Role	Partner	
Hannes Kulovits	Reviewer	TUW	Review
Paul Wheatley	SP lead	BL	Review

Revision History

Version	Status	Author	Date	Changes
0.1	Draft	Rui Castro	2011-04-04	Initial document draft
0.2	Draft	Luís Faria, Miguel Ferreira	2011-04-05	Additions to initial draft
0.3	Draft	Miguel Ferreira	2011-04-10	Revision work
0.4	Draft	Rui Castro, Miguel Ferreira	2011-04-13	Integrated SP lead comments
0.5	Draft	Rui Castro, Luís Faria	2011-05-19	Integrated Unified Quality Model and Testbeds Scenarios
0.6	Draft	Miguel Ferreira	2011-05-19	Revision work
0.7	Draft	Rui Castro, Luís Faria	2011-05-21	Revision and formating
0.8	Draft	Miguel Ferreira	2011-05-22	Revision work
0.9	Draft	Rui Castro	2011-05-22	Revision and formating
0.10	Draft	Miguel Ferreira	2011-05-22	Revision work
0.11	Draft	Miguel Ferreira	2011-05-26	Incorporated Paul Wheatley suggestions
0.12	Draft	Luís Faria, Rui Castro	2011-05-30	Addressed Christoph Becker and Hannes Kulovits comments

0.13	Draft	Miguel Ferreira	2011-06-02	Revision work
0.14	Draft	Rui Castro, Luís Faria	2011-06-27	Addressed Christoph Becker and Paul Wheatley newest comments
0.15	Draft	Miguel Ferreira	2011-06-28	Revision work

References

Ref.	Document	Date	Details and Version

Executive Summary

A preservation action is a concrete action, usually implemented by a software tool, that is performed on digital content in order to achieve some preservation goal. The execution of a preservation action has the purpose of supporting the continuous access to content or to make sure that digital preservation is being carried out effectively. For example, a migration of content from an obsolete format to a more up-to-date one, the replacement of a viewer application in some rendering environment, the execution of legacy software in an emulation environment, etc.

This document reports on the task of identifying, evaluating and selecting available preservation action tools and services to be used in the SCAPE project. The report presents related work in the context of action tools, namely the work developed under the Planets, CRiB and RODA projects, and introduces models for assessing the quality of software tools.

The tools have been evaluated using an evaluation framework based on an international standard for software quality assessment (i.e. the ISO/IEC 25010:2011) that focus on the specific requirements of the project, namely, 1) the suitability of the tool to solve the problems depicted in testbed scenarios; 2) its compatibility with the SCAPE parallel execution platform; and 3) the license that accompanies the tool is well-suited and compatible with the purposes of the project.

An overall of 40 tools, ranging from image converters to database migration suites, have been identified and assessed. After the assessment procedure, 16 of these tools have been considered inadequate to be used in the SCAPE project (40% rejection rate). The main reasons for rejection were licensing constraints and format incompatibility with the testbed scenarios.

The selected tools are considered appropriate for solving the problems depicted by the testbed scenarios as they meet the minimum requirements to run on the SCAPE parallel execution platform and can be used free of charge without breaking any legal constraints.

This deliverable will have a second version to be published in project month 24 and will include: a) a new evaluation framework with criteria and metrics for gauging the improvement of existing action tools during the overall time of the project; b) the results of evaluating selected tools according to

the new evaluation framework; and c) a list of additional criteria that may be used for evaluating action tools in the context of digital preservation.

Table of Contents

1 Introduction

A preservation action is a concrete action, usually implemented by a software tool, that is performed on digital content in order to achieve some preservation goal. The execution of a preservation action has the purpose of supporting the continuous access to content or to make sure that digital preservation is being carried out effectively (Ferreira, Baptista, & Ramalho, 2006). For example, a migration of content from an obsolete format to a more up-to-date one, the replacement of a viewer application in some rendering environment, the execution of legacy software in an emulation environment, etc.

Preservation actions have been extensively analysed and employed in experimental digital preservation systems (Eld Zierau & Caroline van Wijk, 2008; Ferreira, Baptista, & Ramalho, 2007). However, current approaches are not capable of coping with real-size collections of today's preservation environments, e.g. an e-journal library may contain several thousands of documents adding up to several terabytes of information. A national-wide Web archiving service will contain a myriad of unrelated formats and can easily host Petabytes of data.

The Work package WP10 of SCAPE project (i.e. Action Services Components) is focused on the applicability of action tools and services to large collections of complex digital objects by analysing and improving interfaces and internal functionality of existing software tools, extending them and creating new functionality to cope with large-scale digital preservation scenarios. Within this work package, tools will be enhanced to deal with, not only single file formats, but also compound objects, large-sized objects and large collections of digital objects.

This document is structured as follows: Section 1 provides an introduction to this report; Section 2 presents related work in the context of preservation action tools and software quality models; Section 4 thoroughly describes the software quality model that served as a foundation for the development of the action tools evaluation framework; Section 4 presents the methodology used to identify, evaluate and select action tools; Section 5 outlines the requirements of the SCAPE project that have influenced the definition of the tool evaluation framework; Section 6 presents the evaluation framework including the criteria, metrics and selection conditions that were used to select action tools; Section 7 outlines the results of evaluating the tools; and finally, Section 8 draws the main conclusions of this report and sheds light on the future tasks to be developed in work package WP10.

1.1 Scope of this document

This deliverable constitutes a report on Task 1 of the work package WP10 of the SCAPE project – *Identification, evaluation and selection of large scale action tools & services*. The goal of this task is to compile and evaluate a list of existing action tools that are likely to be used in the SCAPE parallel execution platform, and suitable for solving to problems raised by the Testbed scenarios.

Due to the dependencies that this task maintains with other work packages, namely the Testbeds[1] (work packages WP15, WP16 and WP17) and the Platform architecture design[2] (WP4) this report will be written in 2 versions during the overall time of the project:

1. The first version, published in month 6, includes a list of action tools, an evaluation framework and the results of assessing those tools according to the selection conditions defined in the evaluation framework;
2. The second version of this report will be published in project month 24 and will include: a) a new evaluation framework with criteria and metrics for gauging the improvement of existing action tools during the overall time of the project; b) the results of evaluating selected tools according to the new evaluation framework (this will serve as input for deliverable D10.2 - *Gap analysis on action services tools and SCAPE platform and testbeds requirements*); and c) a list of additional criteria that may be used for evaluating action tools in the context of digital preservation. The later criteria will not be evaluated within work package WP10 but will be published as a possible evaluation framework for preservation action tools.

2 Related work

Over the last two decades, the digital preservation research community has come up with a considerable number of strategies aiming at solving the problem of digital preservation and technological obsolescence. Among these is format migration.

Format migration consists of a "(…) set of organized tasks designed to achieve the periodic transfer of digital materials from one hardware/software configuration to another or from one generation of computer technology to a subsequent generation." (Task Force on Archiving of Digital Information., 1996).

Contrary to other preservation techniques, migration strategies do not attempt to preserve digital objects in their original formats. Alternatively, they intentionally transform objects from near obsolete formats into up-to-date encodings that most users are able to interpret using their personal computers. In order to accomplish that, preservation systems resort to off-the-shelf conversion tools to re-encode digital content into formats that more likely to be supported in the long run (Ferreira, 2005).

The major drawback in this approach is that whenever an object is converted to a new format, some of its original properties may not be adequately transferred to the target format. This may occur due to incompatibilities between the source and target formats or because the application used to do the conversion is not capable of carrying out its tasks correctly (Ferreira et al., 2006). In this context, it is important to highlight the importance of software quality models, i.e. a framework of reasoning that allows one to assess the quality of a software tool.

[1] The testbeds first definition, including representative datasets, will only by available by month 9.
[2] The platform work package will only deliver guidelines for deploying preservation tools and environments (D5.1) in month 14.

To better understand what software quality models are, it is necessary to address the question: *what is software quality*? (R. W. Hoyer & B. B. Y. Hoyer, 2001) There are two major paths to choose from when discussing the meaning and definition of software quality:

1. **Conformance to specification** – in this context, software quality is defined as a matter of products whose measurable characteristics satisfy a fixed specification, i.e. conformance to an in beforehand defined specification.
2. **Meeting customer needs** – software quality is identified independent of any measurable characteristics, i.e. quality is defined as the product capability to meet customer expectations (explicit or not).

The evaluation framework included in this report is focused on the first definition of software quality, i.e. *Conformance to specification*. In the context of SCAPE, the *specification*, i.e. the set of requirements that are expected to be fulfilled by the preservation action, are determined by the constrains of the project itself and, further on, by the outcome of preservation planning activities.

The following sections depict related work in the context of preservation action tools and quality assessment models.

2.1 Preservation action tools

Several projects have made use of off-the-shelf migration software as core basis for implementing preservation strategies based on format migration. In this section we present an overview of action tools used in other digital preservation projects, namely, PLANETS, CRiB and RODA, that will serve as a starting point for the work presented in this report.

2.1.1 PLANETS

Planets - Preservation and Long-term Access through Networked Services - was a four-year project co-funded by the European Union under the Sixth Framework Programme to address core digital preservation challenges. Planets started on 1st June 2006 and ended on 31 May 2010. One objective of Planets was to create an Interoperability Framework that integrated tools and services in a distributed service network (Jackson, Andrew Lindley, & Fabian Steeg, 2010).

A compendium of tools identified and integrated in the Planets project is depicted in Table 1. This table includes the tools identified in the Planets deliverable D3 – Gap analysis: a survey of PA tool provision (KB-NL, 2009) - and tools that that were published in the Planets Interoperability Framework Service Registry (ARC, 2009). However, the following list of exceptions apply:

- The table will only include tools that have a particular focus on format migration. A few exceptions have been made for the sake of completeness (e.g. Microsoft Word, Photoshop).
- Tools that belong to the same software package or that depend of a common software library have been added as a single entry.
- Tools that are not attainable due to lack of support or project termination have not been included.
- Tools that depend of external software or services to perform their primary function (e.g. virtual printers, online services, add-ins) have been excluded.

- Tools that merely wrap functionality of other tools, which have already been included in the table, have been suppressed.
- Tools that are bundled with their own application environment and cannot be separated from it (e.g. tools bundled as a whole server).

Table 1 - Tools and services deployed in Planets.

Program name	Description	Web site
ImageMagick	Software suite to create, edit, compose, or convert bitmap images	http://www.imagemagick.org
GraphicsMagick	Software suite to create, edit, compose, or convert bitmap images, forked from the ImageMagick project.	http://www.graphicsmagick.org
Sanselan	Pure-Java library reads and writes a variety of image formats, including fast parsing of image info and metadata	http://commons.apache.org/sanselan
Inkscape	Vector graphics editor, with capabilities similar to Illustrator, CorelDraw, or Xara X, using the W3C standard Scalable Vector Graphics (SVG)	http://inkscape.org
SoX	Command line utility that can convert various formats of computer audio files in to other formats	http://sox.sourceforge.net
Gimp	GNU Image Manipulation Program. It is a freely distributed piece of software for such tasks as photo retouching, image composition and image authoring.	http://www.gimp.org
Java ImageIO	Java library to read and write to image formats	http://java.net/projects/imageio
Kakadu	JPEG2000 developer toolkit	http://www.kakadusoftware.com
JTidy	Java port of HTML Tidy, a HTML syntax checker and pretty printer	http://jtidy.sourceforge.net
JasPer	Software-based implementation of the codec specified in the JPEG-2000 Part-1 standard	http://www.ece.uvic.ca/~mdadams/jasper
OpenJPEG	JPEG 2000 codec written in C language	http://www.openjpeg.org
Avidemux	Video editor designed for simple cutting, filtering and encoding tasks	http://avidemux.sourceforge.net
b2xtranslator	Software tool to convert documents written in Binary Formats (doc, xls, ppt) to the Office Open XML format	http://b2xtranslator.sourceforge.net
Microsoft Word	Text processing software	http://office.microsoft.com/word
Abiword	Word processing program similar to Microsoft Word	http://www.abisource.com
ACDSee	Photo processing software	http://www.acdsee.com
CZ-Doc2Pdf 2.0	Batch PDF Converter that convert word to PDF, DOC to PDF, html to PDF, text to PDF and RTF to PDF	http://www.convertzone.com/doc2pdf/help.htm
Dia	Diagram creation program	http://projects.gnome.org/dia/
Document2PDF Pilot	Converts JPEG, GIF, TIFF, BMP, PNG, EMF, PPT, POT, PPS, XLT, XLS, XLW, DOC, DOT, WPS, WRI, RTF, HTML documents into PDF.	http://www.colorpilot.com/document2pdf.html
EscapeE	Software to view and convert print streams, like PCL, PDF, Postscript, TIFF, DCX, and others.	http://escapee.redtitan.fr
GraphicConverter	Picture converter, editor and manager for Mac OS X	http://www.lemkesoft.com/content/188/graphicconverter.html
MsgText	Software for converting .msg-files to .txt-files and extracting the attachments.	http://www.enterag.ch/enterag/downloads/msgtext.xhtml
Netpbm	Netpbm is a toolkit for manipulation of graphic images, including conversion of images between a variety of different formats.	http://netpbm.sourceforge.net/

PDF Version Converter	Convert PDF file between different versions for compatibility purpose.	http://www.nicepdf.com/products.html
PDF/A converter	PDF/A Converter Service converts PDF files to PDF/A conform and compliant documents, based on the international Norm ISO 19005-1.	http://www.pdfa.at/en/pdfa-converter.html
Apache PDFBox	Apache PDFBox™ is an open source Java PDF library for working with PDF documents.	http://pdfbox.apache.org/
Photoshop	Photo editing software suite.	http://www.photoshop.com/
SIARD Suite	Software package for converting relational databases into the SIARD format.	http://www.bar.admin.ch/dienstleistungen/00823/00825/index.html?lang=en
VisualIntegrity	Software suite to convert PDF into bitmap image, vector image or CAD.	http://www.visual-integrity.com/
FFmpeg	Complete, cross-platform solution to record, convert and stream audio and video.	http://ffmpeg.org/
MEncoder	Underlying framework of the media player viewer mplayer	http://www.mplayerhq.hu

2.1.2 RODA

RODA - Repository of Authentic Digital Objects - is a digital repository capable of ingesting, managing and providing continuous access to various types of digital objects. It was developed as part of a project, with the same name, co-founded by the Portuguese National Archives with the objective of identifying and bringing together all the necessary technology, human resources and political support to carry out long-term preservation of digital materials (Faria et al., 2009).

RODA is able to ingest and preserve text documents, images, video, audio and relational databases available in several distinct formats. It automatically normalizes ingested data to formats more adequate for long-term preservation and makes data available to its consumers in various dissemination formats (including the original ingested format). To accomplish this, RODA resorts to several action tools to carry out all the necessary format migrations. The action tools deployed in RODA are outlined in Table 2.

Table 2 - Action tools deployed in RODA.

Program name	Description	Web site
ImageMagick	Software suite to create, edit, compose, or convert bitmap images	http://www.imagemagick.org
Gstreamer	Open source multimedia framework	http://gstreamer.freedesktop.org
DBML import/export module	Library of the RODA project that allows database migration of information. It supports several database management systems and also a created standard format called DBML (Marta, Librelotto, Ramalho, & P. R. Henriques, 2002).	http://redmine.keep.pt/projects/show/roda-public
JODConverter	The Java OpenDocument Converter, converts documents between different office formats using OpenOffice.org.	http://sourceforge.net/projects/jodconverter
MEncoder	Underlying framework of the media player viewer mplayer	http://www.mplayerhq.hu
Ghostscript	Interpreter for the PostScript language and for PDF	http://pages.cs.wisc.edu/~ghost/
RODA's doc2pdf for	Executable that allows converting word	http://redmine.keep.pt/projects/show/roda-

Microsoft Word	documents to PDF using Microsoft Word by a command line interface.	public

2.1.3 CRiB

The CRiB platform is a set of services designed to assist cultural heritage institutions in the implementation of migration-based preservation interventions. CRiB works by assessing the quality of distinct migration paths tools and services to produce recommendations of optimal migration paths. The recommendations produced by the system take into account the specific preservation requirements of each client institution (Ramalho et al., 2008). The action tools deployed in CRiB are depicted in Table 3.

Table 3 - Action tools deployed in CRiB.

Program name	Description	Web site
ImageMagick	Software suite to create, edit, compose, or convert bitmap images	http://www.imagemagick.org
OpenOffice	Open-source office software suite for word processing, spreadsheets, presentations, graphics, databases and more	http://www.openoffice.org
Microsoft Word	Text processing software	http://office.microsoft.com/word
Sam2p	Command line utility written in ANSI C++ that converts many raster (bitmap) image formats into Adobe PostScript or PDF files and several other formats	http://pts.szit.bme.hu/sam2p

2.2 Software quality models

To evaluate the quality of software one must perform a systematic assessment of the capability that a software infolds to implement a given set of requirements. In other words, an adequate quality model should be defined that will guide the process of evaluation a given software application.

A quality model consists of several quality attributes that are used as a checklist for determining software quality (*ISO 9126-1*, 2001). Additionally, to properly evaluate software one will also need an evaluation framework composed of metrics, measurements and, if possible, supporting software tools to facilitate the evaluation process (Beus-Dukic & Bøegh, 2003).

There are several examples of quality models available in research literature:

1. ***Factors in Software Quality*** – McCall, J.A., Richards, P.K., Walters, G.F., *Factors in Software Quality,* RADC TR-77-369, 1977;
2. ***Characteristics of Software Quality*** – Boehm, B.W, Brown, J.R., Kaspar, J.R., et.al, *Characteristics of Software Quality*, TRW Series of Software Technology, Amsterdam, North Holland, 1978;
3. ***Specification of software quality attributes*** – Bowen, T. P., Wigle, G. B., Tsai, J. T. 1985. *Specification of software quality attributes*. Tech. Rep. RADC-TR- 85-37, Rome Air Development Center;
4. **ISO standard 25010:2011** – *Systems and software engineering, Systems and software Quality Requirements and Evaluation (SQuaRE);*

These examples of software quality models are independent of the application domain. The ISO standard 25010:2011 is the most recent of these initiatives, and can be considered as a superset of the other three.

An additional framework of thought that should be considered is the taxonomy of decision criteria compiled by the Plato Preservation Planning tool (Becker & Rauber, 2011). This taxonomy of decision factors is based on quality requirements of real-world digital preservation case studies and aims at assessing the *quality* of preservation action tools.

The last two examples will be analysed in more detail in the following sections. The rest of the models were not further described in this report because they have already served as a basis for the development of the two examples selected.

2.2.1 ISO standard 25010

The ISO standard 25010 (*ISO 25010*, 2011) is based on the earlier ISO 9126 family (*ISO 9126-1*, 2001). The ISO/IEC 9126 standards provide guidance for quality models and define a hierarchy of high-level quality attributes. Quality measures are based on measurement procedures recommended in ISO 15939 (*ISO 15939*, 2007).

The ISO 25000 standards for Software Product Quality Requirements and Evaluation (SQuaRE) supersede the ISO 9126 models and combine a revised quality model with evaluation procedures based on ISO 14598 (*ISO 14598-1*, 1999). They also define requirements on the specification of software product quality criteria.

ISO/IEC 25010:2011 defines a *product quality model composed of eight characteristics (which are further subdivided into sub-characteristics) that relate to static properties of software and dynamic properties of the computer system. The model is applicable to both computer systems and software products* - Figure 1 (*ISO 25010*, 2011).

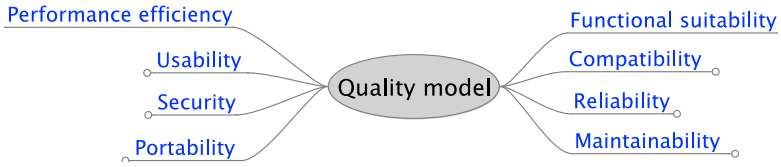

Figure 1 - ISO 25010 top product quality characteristics.

2.2.2 Decision criteria in digital preservation

In the digital preservation domain, a preservation planning process is responsible for defining criteria and evaluating preservation actions for a given identified risk or to achieve a certain preservation goal. These decision criteria have been used by a tool called Plato to assess the *quality* of competing preservation action options and to rank them according to the preservation objectives of the person or organization managing a particular preservation project. Although these criteria have been designed to assess the quality of preservation action tools, they should not be considered as a software quality model *per se* as they also take into consideration factors that are outside of the scope of software quality assessment, e.g. organisation policies, costs, etc.

In (Becker & Rauber, 2011) over 500 criteria where obtained from different case studies, and a classification of these criteria was derived (Figure 2).

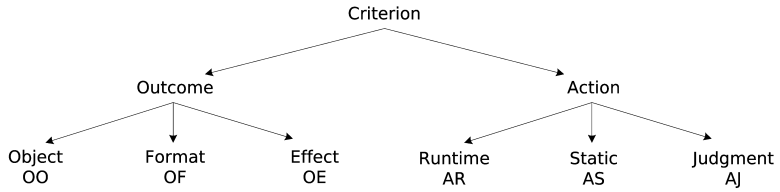

Figure 2 - Taxonomy of decision factors in digital preservation.

All criteria requiring measurement refer to either an **Action** or the **Outcome** of that action. Action criteria exhibit properties that are somewhat related to the preservation action itself, while outcome criteria can describe general effects of the outcome of the action.

A preservation objective can be composed of several measurable criteria belonging to different categories. For instance, the general objective of minimizing costs may include both a criterion for assessing the migration cost per object and one for specifying runtime characteristics such as memory usage or processor time.

Sub-characteristics of the action criterion:

- ***Runtime*** *- This category entails runtime properties of action tools such as performance, throughput, and memory utilization. Since these properties are highly dynamic and depend on a number of factors, measurements need to be taken in a controlled environment. Examples of this category include Peak memory usage, average processing cycles consumed per MB and average memory consumed per MB* (Becker & Rauber, 2011).
- ***Static*** *- Criteria of this category refer to properties of the action components that do not vary per execution run nor show differences when evaluated by different users; i.e., they are not subject to the evaluator's perception and can be determined objectively. These criteria can thus often be obtained from trusted sources. For example, the question whether a component is open source or not should be documented in tool registries* (Becker & Rauber, 2011).
- ***Judgment*** *- This category is sometimes relevant, but decision criteria in this category should be kept to a minimum. It comprises criteria that cannot be objectively determined with reasonable effort. Usability is a prime example where judgment may be necessary. In digital preservation this does not have high influence on the decision, since the components to be evaluated are not to be applied by an end user. In other cases, this has more importance; but in any case, proper documentation of evaluation values is essential. Examples of criteria in this category include Ease of component integration into existing workflow environment and process log output is human readable* (Becker & Rauber, 2011).

Sub-characteristics of the outcome criterion:

- ***Object*** *- This category entails all desired properties of digital objects. This includes desired properties of the objects and properties that have to be kept unchanged compared to the original object* (Becker & Rauber, 2011).
- ***Format -*** *This category comprises criteria that specify desirable characteristics of the formats that are used for representing digital content. As a significant portion of the risks to digital content lies in the form of representation and its understandability, this is often a central decision criterion. Typical criteria include standardization (e.g. Format is standardized by ISO), format complexity, or openness of formats* (Becker & Rauber, 2011).
- ***Effect of outcome*** *- This refers to any other effects caused by the application of a certain action. Typically, these effects are calculated by organization-specific models or recognized cost models such as LIFE* (Ayris et al., 2008) *based on measures as model inputs. For example, storage costs will depend on organizational cost structures, but strongly correlate with the file size of objects* (Becker & Rauber, 2011).

The presented taxonomy has proven to be complete in its expressiveness to cover all the criteria encountered in the case studies evaluated so far, since it models all relevant entities encountered in the decision process (Becker & Rauber, 2011).

3 Criteria for evaluating action tools

After careful examination of the quality models described in the previous sections, namely, the ISO standard 25010 and the decision criteria used by the Plato preservation planning tool, we have elected the ISO standard as being the most suitable to serve as a basis for the creation of an evaluation framework for action tools (used to aid in the objective selection of action tools to be included in the SCAPE project.). The rationale behind this decision was threefold:

1. The ISO standard is based on a well defined international standard, allowing others to easily relate to it;
2. The ISO standard is very up-to-date, as it has been published in 2011; and
3. The quality models described in Section 2.2 have already served as a basis to the development of the ISO standard, so in a way this standard reflects the vision of most quality models previously published.

The ISO 25010 is generic in the sense that it defines characteristics that relate to static properties of software products and dynamic properties of computer systems independently of its application domain. The ISO model defines eight top criteria:

- **Functional suitability** - "degree to which a product or system provides functions that meet stated and implied needs when used underspecified conditions" (*ISO 25010*, 2011)*;*
- **Performance efficiency** - "performance relative to the amount of resources used under stated conditions" (*ISO 25010*, 2011);
- **Compatibility** - "degree to which a product, system or component can exchange information with other products, systems or components, and/or perform its required functions, while sharing the same hardware or software environment" (*ISO 25010*, 2011);

- **Usability** - "degree to which a product or system can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use" (*ISO 25010*, 2011);
- **Reliability** - "degree to which a system, product or component performs specified functions under specified conditions for a specified period of time" (*ISO 25010*, 2011);
- **Security** - "degree to which a product or system protects information and data so that persons or other products or systems have the degree of data access appropriate to their types and levels of authorization" (*ISO 25010*, 2011);
- **Maintainability** - "degree of effectiveness and efficiency with which a product or system can be modified by the intended maintainers" (*ISO 25010*, 2011);
- **Portability** - "degree of effectiveness and efficiency with which a system, product or component can be transferred from one hardware, software or other operational or usage environment to another" (*ISO 25010*, 2011).

Some of the previously described criteria have been subdivided one step further making the quality model more domain-specific and appropriate to address the specific goal of evaluating action tools. The fully expanded taxonomy of criteria is depicted in Figure 3. The third-level leaf nodes represent the domain-specific refinements, i.e. accuracy, suitability for testbeds, suitability for platform, composability, technical instalability and legal constraints.

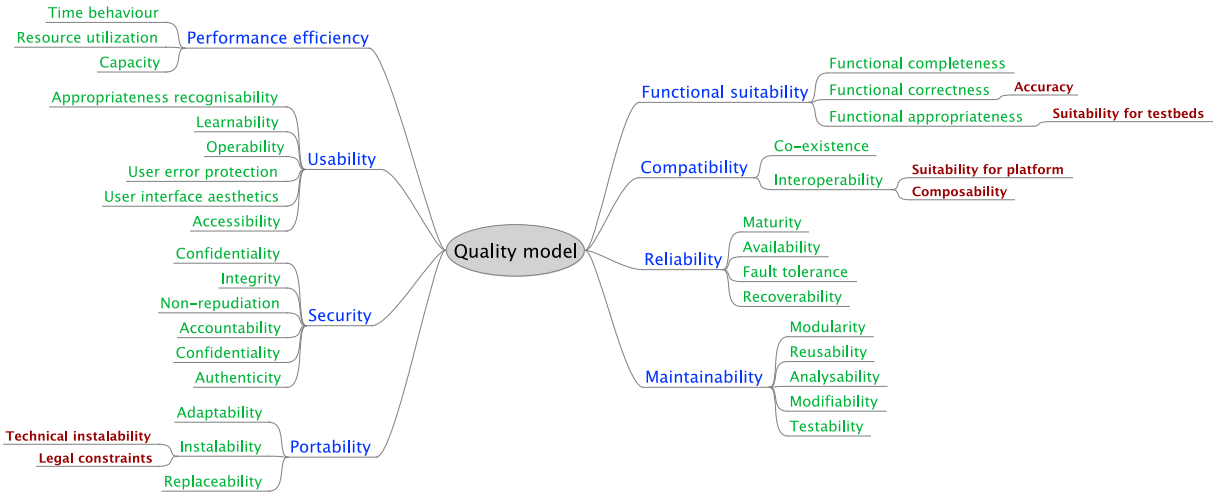

Figure 3 – Quality model with domain-specific criteria.

All of the criteria included in the quality model are explained in the following sections.

3.1 Functional suitability

Functional suitability is "degree to which a product or system provides functions that meet stated and implied needs when used underspecified conditions" (*ISO 25010*, 2011). It has the following sub-characteristics:

- **Functional completeness** - "degree to which the set of functions covers all the specified tasks and user objectives" (*ISO 25010*, 2011);

- **Functional correctness** - "degree to which a product or system provides the correct results with the needed degree of precision" (*ISO 25010*, 2011);
    - **Accuracy** - in the context of preservation action tools, accuracy is the degree of similarity between the input and the output of the action, and directly relates to the quality assurance problem (domain-specific criterion).
- **Functional appropriateness** - "degree to which the functions facilitate the accomplishment of specified tasks and objectives" (*ISO 25010*, 2011).
    - **Suitability for testbeds** - degree to which the functionality of a preservation action tool fulfils the needs of the testbeds scenarios (domain-specific criterion).

3.2 Performance efficiency

Performance efficiency is the "performance relative to the amount of resources used under stated conditions" (*ISO 25010*, 2011). It has the following sub-characteristics:

- **Time behaviour** - "degree to which the response and processing times and throughput rates of a product or system, when performing its functions, meet requirements" (*ISO 25010*, 2011);
- **Resource utilization - "**degree to which the amounts and types of resources used by a product or system, when performing its functions, meet requirements" (*ISO 25010*, 2011);
- **Capacity** - "degree to which the maximum limits of a product or system parameter meet requirements" (*ISO 25010*, 2011);

3.3 Compatibility

Compatibility is the "degree to which a product, system or component can exchange information with other products, systems or components, and/or perform its required functions, while sharing the same hardware or software environment" (*ISO 25010*, 2011). It has the following sub-characteristics:

- **Co-existence** - "degree to which a product can perform its required functions efficiently while sharing a common environment and resources with other products, without detrimental impact on any other product" (*ISO 25010*, 2011);
- **Interoperability** - "degree to which two or more systems, products or components can exchange information and use the information that has been exchanged" (*ISO 25010*, 2011);
    - **Suitability for platform** - degree of effort needed to make a tool compatible with the SCAPE platform, namely the workflow system (domain-specific criterion).
    - **Composability** - degree to which the output of an action tool is compatible with the input of another (domain-specific criterion).

3.4 Usability

Usability is the "degree to which a product or system can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use" (*ISO 25010*, 2011). It has the following sub-characteristics:

- **Appropriateness recognisability** - "degree to which users can recognize whether a product or system is appropriate for their needs" (*ISO 25010*, 2011);
- **Learnability -** "degree to which a product or system can be used by specified users to achieve specified goals of learning to use the product or system with effectiveness, efficiency, freedom from risk and satisfaction in a specified context of use" (*ISO 25010*, 2011);
- **Operability -** "degree to which a product or system has attributes that make it easy to operate and control" (*ISO 25010*, 2011);
- **User error protection -** "degree to which a system protects users against making errors" (*ISO 25010*, 2011);
- **User interface aesthetics -** "degree to which a user interface enables pleasing and satisfying interaction for the user" (*ISO 25010*, 2011);
- **Accessibility -** "degree to which a product or system can be used by people with the widest range of characteristics and capabilities to achieve a specified goal in a specified context of use" (*ISO 25010*, 2011).

3.5 Reliability

Reliability is the "degree to which a system, product or component performs specified functions under specified conditions for a specified period of time" (*ISO 25010*, 2011). It has the following sub-characteristics:

- **Maturity -** "degree to which a system, product or component meets needs for reliability under normal operation" (*ISO 25010*, 2011);
- **Availability -** "degree to which a system, product or component is operational and accessible when required for use" (*ISO 25010*, 2011);
- **Fault tolerance -** "degree to which a system, product or component operates as intended despite the presence of hardware or software faults" (*ISO 25010*, 2011);
- **Recoverability -** "degree to which, in the event of an interruption or a failure, a product or system can recover the data directly affected and re-establish the desired state of the system" (*ISO 25010*, 2011).

3.6 Security

Security is the "degree to which a product or system protects information and data so that persons or other products or systems have the degree of data access appropriate to their types and levels of authorization" (*ISO 25010*, 2011). It has the following sub-characteristics:

- **Confidentiality** - "degree to which a product or system ensures that data are accessible only to those authorized to have access" (*ISO 25010*, 2011);
- **Integrity** - "degree to which a system, product or component prevents unauthorized access to, or modification of, computer programs or data" (*ISO 25010*, 2011);
- **Non-repudiation** - "degree to which actions or events can be proven to have taken place, so that the events or actions cannot be repudiated later" (*ISO 25010*, 2011);
- **Accountability** - "degree to which the actions of an entity can be traced uniquely to the entity" (*ISO 25010*, 2011);

- **Confidentiality** - "degree to which a product or system ensures that data are accessible only to those authorized to have access" (*ISO 25010*, 2011);
- **Authenticity** - "degree to which the identity of a subject or resource can be proved to be the one claimed" (*ISO 25010*, 2011).

3.7 Maintainability

Maintainability is the "degree of effectiveness and efficiency with which a product or system can be modified by the intended maintainers" (*ISO 25010*, 2011). It has the following sub-characteristics:

- **Modularity -** "degree to which a system or computer program is composed of discrete components such that a change to one component has minimal impact on other components" (*ISO 25010*, 2011);
- **Reusability -** "degree to which an asset can be used in more than one system, or in building other assets" (*ISO 25010*, 2011);
- **Analysability -** "degree of effectiveness and efficiency with which it is possible to assess the impact on a product or system of an intended change to one or more of its parts, or to diagnose a product for deficiencies or causes of failures, or to identify parts to be modified" (*ISO 25010*, 2011);
- **Modifiability -** "degree to which a product or system can be effectively and efficiently modified without introducing defects or degrading existing product quality" (*ISO 25010*, 2011);
- **Testability -** "degree of effectiveness and efficiency with which test criteria can be established for a system, product or component and tests can be performed to determine whether those criteria have been met" (*ISO 25010*, 2011).

3.8 Portability

Portability is the "degree of effectiveness and efficiency with which a system, product or component can be transferred from one hardware, software or other operational or usage environment to another" (*ISO 25010*, 2011). It has the following sub-characteristics:

- **Adaptability -** "degree to which a product or system can effectively and efficiently be adapted for different or evolving hardware, software or other operational or usage environments" (*ISO 25010*, 2011);
- **Instalability -** "degree of effectiveness and efficiency with which a product or system can be successfully installed and/or uninstalled in a specified environment" (*ISO 25010*, 2011);
    - **Technical instalability** - degree of effectiveness and efficiency with which an action tool can be installed in the specified software and hardware (domain-specific criterion);
    - **Legal constraints** - degree to which the legal constraints of a tool are compatible with a specified environment and use (domain-specific criterion);
- **Replaceability** - "degree to which a product can replace another specified software product for the same purpose in the same environment" (*ISO 25010*, 2011).

4 Methodology

Current preservation action tools used in other digital preservation projects like PLANETS and CRiB do not cope well with all aspects of large-scale processing. Because of that, they will be evaluated and targeted for improvement in the subsequent tasks of work package WP10.

In this section we present the methodology used to identify, evaluate and select existing action tools to be improved and then used in the SCAPE platform. The methodology is composed of the following steps:

1. Analysis of requirements
2. Definition of an evaluation framework
3. Identification, evaluation and selection of action tools

The following sections provide a brief description of each of these steps.

4.1 Analysis of requirements

This step consists of the identification of requirements and constraints that should be part of the evaluation framework that will be used to assess action tools that are to be included in the SCAPE project. The two major items that need to be analysed are the requirements of the SCAPE parallel execution platform (where the action tools are expected to run) and the needs of the testbed scenarios (whose preservation problems the tools should be able to address).

4.2 Definition of the evaluation framework

This step consists of the definition of an evaluation framework based on quality criteria obtained from the software quality model previously described on Section 3 (i.e. the ISO standard 25010). This evaluation framework will be used to assess and select the action tools that will be included in the SCAPE project.

4.3 Identification, evaluation and selection of action tools

This step consists in the identification and assembly of action tools considered to be good candidates to integrate the SCAPE project. As a starting point, tools were collected by inspecting the deliverables of the digital preservation projects described in Section 2.

In addition to that, a community-driven list was created so that action tools could be registered by all partners involved in the work package based on their own background experience. The information collected about each tool is summarized in Table 4.

Table 4 – Information collected for each action tool.

Column name	Description	Example
Tool name	The name of the tool and a link to the tool site.	ImageMagick
Tool version	The current version of the tool.	6.6.7
License	The tool license	GPL V3 compatible
Interface	The type of interfaces provided by the tool: *Command line,*	Command line

	API, Service, GUI, …	
Operating system	The operating systems supported: *GNU/Linux, Microsoft Windows, Mac OS X, …*	GNU/Linux; Windows; Mac OS X; Other
Open-source	Is the tool open-source?	Yes
Object class	The type of files that the tool can handle: images, video, audio, word processor documents, spread-sheets, CAD, relational databases, …	Images
Input formats	Formats that the tool can handle	JPG, TGA, PNG, ...
Output formats	Formats that tool can produce	PDF, ODT, DOCX, ...

After collecting an assortment of action tools, these were assessed using the evaluation framework. The goal of this activity was to determine if a given tool was a good candidate for integration in the SCAPE project. The evaluation framework, described in Section 6, was created on top of the quality model described in Section 3.

Some sub-criteria have been added to the quality model to better evaluate preservation action tools in the context of SCAPE. However, only a subset of relevant criteria has been ported to the final evaluation framework.

For each selection criteria we have identified metrics (i.e. ways of measuring the criteria). These are described by the following attributes:

1. **Metric ID** – a unique identifier attributed to each metric;
2. **Description** – a description of the metric;
3. **Unit** – the type of the measurement values of the metric;
4. **Evaluation effort** – how much effort is needed to actually perform the measurement of a metric, or the dependencies on components provided by other work packages that hinder the assessment of the metric.

For each metric, a *selection condition* has also been defined, i.e. if a tool satisfies all the conditions, it is selected for adaptation and improvement in the SCAPE platform, otherwise it is rejected.

5 Analysis of requirements

The goals of the SCAPE project form a set of requirements that play an important role in the definition of the evaluation framework. Only tools capable of coping with those requirements should be selected for inclusion in the project. Fundamentally, these requirements are of two types:

1. Requirements of the SCAPE platform; and
2. Requirements of the testbed scenarios.

Each of these types of requirements are analysed in the following sections.

5.1 Requirements for the SCAPE platform

The SCAPE parallel execution platform has certain requirements that condition the deployment and execution of certain software tools. During the first year of the project, the Platform sub-project (PT) will develop guidelines to assist in the development and deployment of scalable preservation tools. At the time of writing, these guidelines are still at a very early stage. Nevertheless, in this report we present some information that should be taken into account when evaluating and selecting action tools for the SCAPE platform.

Existing tools may be deployed in SCAPE platform as long as they comply with the following list of requirements[3].

- Graphical environments are not supported (e.g. XServer);
- Java-based tools and wrapped Linux binaries may run but must be preinstalled on the cluster nodes;
- It may be possible to run Windows applications, however this will not be investigated in the first year of the project;

5.2 Requirements of the testbed scenarios

In order to select the tools that will integrate the SCAPE platform, it was necessary to do a comprehensive analysis of the testbed scenarios. We were particularly interested in knowing what formats were inherent to the testbed scenarios to make sure that all the relevant action tools were properly evaluated.

At the time of writing of this report, only a preliminary version of the testbed scenarios was available. From the analysis of that document, assumptions on which file formats were necessary to be handled by action tools were made. In month 9, the testbed scenarios will be fully defined and representative datasets will be created. By then it will be possible to determine more accurately which formats will play an important role in the SCAPE project. For the time being we will evaluate tools based on the assumptions outlined in the following draft scenarios.

5.2.1 Scenario 1: Normalize document formats contained in the web archive

Document instances of many different file formats are referenced in web content. Many of these formats might not be renderable in a web archive viewer in the future. This relates especially to older versions of text document formats. A possible solution could be to normalize formats by migrating the document instances into an agreed standard format. For example, an institutional decision could be to migrate all document formats (Plain text, DjVu, PS, ODF, DOC, DOCX, RTF, etc.) to PDF. In this context, quality assurance could play a major role. Also on-the-fly/on demand migration for document formats is an option in order to avoid changing the original web archive content.

[3] Please note that these restrictions only apply to preservation tools that are intended to run on the platform scalable environment. They do not apply to tools that will be deployed in other application contexts (e.g. rendering environments for quality assurance).

Related file formats: Plain text, DjVu, PS, ODF, DOC, DOCX, RTF, PDF, HTML, ARC, WARC

5.2.2 Scenario 2: Deep characterisation of huge media files

This scenario is about determining technical dependencies of large MPEG2 streams in order to identify preservation risks and ensure render/extraction services can continue to be supported.

At SB, data from broadcasters contain huge media files like MPEG2 videos, for example. There is an agreement that only allows streaming this data, but not distribution of copies of the archived content. SB captures broadcast radio and television as complex MPEG2 streams. The video content is accompanied by metadata, typically used to support the production of TV guides. SB preserves the MPEG2 streams as the preservation masters. Chunks of this data that relate to specific programmes are extracted, migrated and served to users as streaming Flash video.

Related file formats: MPEG2, MP3, AAC, Flash Video, possibly other video and audio encodings

5.2.3 Scenario 3: Migrate digitised TIFFs to JPEG2000

The master files from legacy digitised image collections are typically TIFF files that can be costly to store due to large file sizes. Preservation planning exercises have indicated that migration to JPEG2000 would reduce storage size and costs while at the same time facilitating enhanced user access via JPEG2000 functionality. It is therefore assumed that preservation planning has been completed, and this scenario focuses on the migration of TIFF files to JPEG2000 and associated quality assurance of the migration.

Related file formats: TIFF, JPEG2000

5.2.4 Scenario 4: Migrate archive to new archiving system?

A typical characteristic of digital archives that aim for “long-term preservation” is that the life cycle of the technical infrastructure on which they are based is much shorter than the period for which their contained materials should be preserved. This means that migrations from one archival system to another are inevitable. In the simplest case this could be nothing more than a migration of AIPs from one storage medium to another. However, in most cases this will also involve the migration of metadata, and the contents of each AIP from the source system may need to be taken apart and re-assembled on the destination system. This will result in changes to the AIP’s internal structure that must be accounted for in the migrated (structural) metadata. Finally, such migrations may also involve one or more metadata enrichment steps (for example, because the availability of new or improved characterisation tools makes it possible to automatically extract technical and preservation metadata that couldn’t be established within the old system).

At the most basic level this scenario would like to ensure that the system migration does not result in the loss or alteration of any archived objects. In the case of a pure medium migration this could be realised very easily using checksums. More sophisticated mechanisms are needed for migrations where, as an example, AIPs that are held together in a physical container (e.g. a TAR file) on the source system need to be taken apart and subsequently re-assembled on the destination system. In

that case we will need to check the integrity of each single file within the AIP, before and after the migration.

Related file formats: TAR, ZIP

5.2.5 Scenario 5: RAW to NEXUS migration

AT STFC, scientific data sets have been stored in the RAW format, and there is a method to migrate the RAW format to the XML-based NEXUS format. Currently, there is only a simple format conversion in place that does not record any provenance information about why and by whom information items have been changed.

Related file formats: RAW, XML/Nexus

6 Evaluation framework

In this section, we describe the evaluation framework that has been devised to assess and select action tools to be adapted and improved within the SCAPE project. To enable a systematic evaluation of the tools, metrics and selection conditions have been defined for each quality criterion included in the evaluation framework.

Figure 4 depicts the complete taxonomy of criteria that compose the evaluation framework.

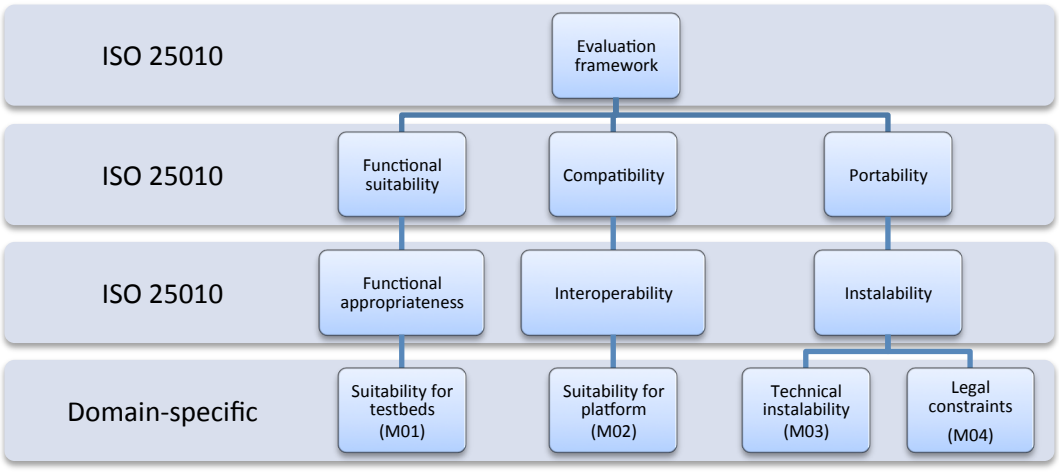

Figure 4 - Evaluation framework.

Not all of the criteria in the ISO 25010 were included in the evaluation framework. Some will be used afterwards, on version 2 of this report, to appraise the need for improvement and serve as a basis for validating task 3 of this work package - *Improve functionality and coverage of preservation action tools*. Such criteria will be used to assess the *dynamic qualities* of tools before and after the improvement activities.

The criteria, metrics and conditions that are part of this evaluation framework are described in the following sections:

6.1 Suitability for testbeds

In order to be selected, a tool should be aligned with the objectives of the SCAPE project. This relates particularly to its suitability to solve the problems depicted by the testbed scenarios and is determined by the quality criterion **suitability for testbeds**, under **functional appropriateness** (Section 3.1)**.**

For a tool to be selected, at least one of the formats it supports should be part of the testbed scenarios related file formats.

Metric ID	M01
Criterion	Suitability for testbeds
Description	Number of input or output formats the tool supports (directly or transitively by composing other tools) that exist in the testbed scenarios related file formats.
Unit	Number
Evaluation effort	Low

6.2 Suitability for platform

A second requirement is based on the fact that the tool should be compatible with the SCAPE parallel execution platform. By compatible we mean that the effort needed to install and make the tool run in the SCAPE platform, and in the workflow engine, is not beyond the level of resources available in the project. This requirement is determined by the criterion **suitability for platform,** under **interoperability**.

For a tool to be selected, it must provide a way to access its functionality programmatically or the effort to make it so should be equal or less than medium.

Metric ID	M02
Criterion	Suitability for platform
Description	Effort necessary to make the tool interface compatible with the SCAPE platform: 1. the tool has a command-line interface, effort is close to none 2. a small effort is necessary to adapt the tool (e.g. a wrapper needs to be developed) 3. a medium effort is necessary (e.g. the tool is a software library or has an API) 4. a high effort is needed to adapt the tool 5. the effort is very high or nearly impossible (e.g. the tool has a graphical user interface and no command-line or programming interface)
Unit	1 to 5 scale
Evaluation effort	Low

6.3 Technical instalability

One must ensure that it is possible to install the tool in the SCAPE parallel execution platform. This relates to the **instalability** criterion, which has been further refined by **technical instalability.**

For a tool to be selected, it must be feasible, using available resources in SCAPE, to make it runnable in the SCAPE platform. This means that the effort necessary to install the tool or to make it compatible with platform should be equal or less than medium.

Metric ID	M03
Criterion	Technical instalability
Description	Effort necessary to make the tool supported by the SCAPE platform. The level of effort to install a tool is defined by: 1. the tool is directly supported (an installer is available for the Operating System) 2. a small effort is necessary to make the tool supported (the tools requires some configuration) 3. a medium effort is necessary to make the tool supported 4. a high effort is necessary to make the tool supported 5. it is nearly impossible to support the tool
Unit	1 to 5 scale
Evaluation effort	Low

6.4 Legal constrains

For a tool to be selected its license must enable it to be used in the context of the SCAPE platform and it should be free of charge since the SCAPE project does not include a budget for acquiring software licenses. This means that the end-user license agreement that accompanies the tool should allow one to run it on a server environment and be invoked by a random number of anonymous users, and the tool should be open-source or free to be used in a research context.

In order to be selected a tool should not have any constraints that hinder its use from the legal point-of-view.

Metric ID	M04
Criterion	Legal constraints
Description	Whereas the license of the tool allows it to be run as a server-side service and the tool is free of charge for use in a research context.
Unit	Boolean
Evaluation effort	Low

6.5 Summary

A summary of the metrics and conditions included in the evaluation framework is presented in Table 5.

Table 5 - Metrics and conditions for tool selection.

Criterion	Metric	Selection condition
Suitability for testbeds	M01 - Number of input or output formats the tool supports (directly or transitively by composing other tools) that exist in the testbed scenarios.	Must be greater than zero (>0)
Suitability for platform	M02 - Level of effort needed to make the tool interface compatible with the ones supported by the SCAPE platform.	Effort must be equal or less than medium (<=3)
Technical instalability	M03 - Level of effort to make the tool supported in the SCAPE platform.	Effort must be equal or less than medium (<=3)
Legal constraints	M04 – Whereas the license of the tool allows it to be run as a server-side service and the tool is free of charge for use in a research context..	Must be true (=true)

7 Results

7.1 Identification of candidate tools

As previously mentioned, a list of candidate action tools has been assembled by inspecting the outcomes of reference digital preservation projects and by collecting information from preservation experts involved in work package activities.

This section presents a summary of all the off-the-shelf action tools that have been identified. The complete list of tools with detailed information is presented in Appendix 10.1.

Table 6 - Identified action tools.

Tool	Interface	Operating system	Open-source	Object class
ImageMagick	Command-line	Linux; Win; Mac; Other	Yes	Bitmap
Mencoder	Command-line	Linux; Win; Mac	Yes	Video; Audio
HandBrake	Command-line;GUI	Linux; Win; Mac	Yes	Video
FFMpeg	Command-line;API	Linux; Win; Mac	Yes	Video; Audio
SoX	Command-line	Linux; Win; Mac	Yes	Audio
GStreamer	Command-line;API	Linux; Win	Yes	Video; Audio
Microsoft Office	API;GUI	Win; Mac	No	Text; Spreadsheet
OpenOffice (JODConverter)	Command-line	Linux; Win; Mac; Other	Yes	Text; Spreadsheet
Teigha File Converter	Command-line;GUI	Linux; Win; Mac	No	CAD
Kakadu	API	Linux; Win	Yes	Bitmap
Apache Sanselan	API	Linux; Win; Mac; Other	Yes	Bitmap
OpenJPEG	API	Linux; Win; Mac	Yes	Bitmap
SIARD Suite	Command-line	Linux; Win; Mac	No	Database
GIMP	Command-line;GUI	Linux; Win; Mac	Yes	Bitmap
Inkscape	Command-line;GUI	Linux; Win; Mac	Yes	Bitmap; Vector
Avidemux	Command-line;GUI	Linux; Win	Yes	Video
PDFBox	API	Linux; Win; Mac; Other	Yes	Other

JTidy	Command-line;API	Linux; Win; Mac; Other	Yes	Other
JasPer	API	Linux; Win; Mac; Other	Yes	Bitmap
Aware AccuRad J2KSuite	API	Linux; Win; Mac	No	Bitmap
Luratech LuraWave	Command-line	Linux; Win	No	Bitmap
b2xtranslator	Command-line	Linux; Win; Mac; Other	Yes	Text; Spreadsheet
warc-tools	Command-line	Linux; Win; Mac; Other	Yes	Web archive
NedlibToArc	Command-line; API	Linux; Win; Mac; Other	Yes	Web archive
Heritrix (org.archive.io)	API	Linux; Win; Mac; Other	Yes	Web archive
Java Image I/O	API	Linux; Win; Mac; Other	Yes	Bitmap
GraphicsMagick	Command-line	Linux; Win; Mac; Other	Yes	Bitmap
ACDSee	GUI	Win; Mac	No	Bitmap; Audio; Video
CZ-Doc2Pdf	GUI	Win	No	Text
Dia	Command-line; GUI	Linux; Win	Yes	Vector
Document2PDF Pilot	GUI	Win	No	Text
EscapeE	Command-line; GUI	Win	No	Text
GraphicConverter X	API; GUI	Mac	No	Bitmap
MsgText	Command-line	Win	Yes	Email
Tesseract-ocr	Command-line	Linux; Win; Mac	Yes	Bitmap
AbiWord	Command-line; GUI	Linux; Win; Mac	Yes	Text
VisualIntegrity	Command-line	Linux; Win; Mac	No	Text
Photoshop	GUI	Win; Mac	No	Bitmap; Vector
PDF Version converter	Command-line; GUI	Win	No	Text
Netpbm	Command-line	Linux; Win; Mac	Yes	Bitmap

7.2 Evaluation and selection of tools

This section presents the results of the tool evaluation (Table 7). For each tool, the selection conditions presented in Section 6 are evaluated.

Table 7 is composed of the following information:

1. **Tool name** - the name of the tool;
2. **Criterion M01** - Number of input or output formats the tool supports (directly or transitively by composing other tools) that exist in the testbed scenarios. Number must be greater than 0;
3. **Criterion M02** - Level of effort needed to make the tool interface compatible with the ones supported by the SCAPE platform. Number must be less or equal to 3;
4. **Criterion M03** - Level of effort to make the tool supported in the SCAPE platform. Number must be less or equal to 3;
5. **Criterion M04** – Whereas the license of the tool allows it to be run as a server-side service and the tool is free of charge for use in a research context.. Value must be *True*;
6. **Selected?** - The value is *Yes* if the tool is selected for use in the SCAPE project or *No* otherwise.

Table 7 - Evaluation of action tools.

Criterion	M01	M02	M03	M04	Selected?
Selection condition	(>0)	(<=3)	(<=3)	(True/False)	(Yes/No)
ImageMagick	4 (JP2, HTML, TIFFF, DjVu, TXT)	1 (Command line)	1	True	Yes
Mencoder	1 (MPEG2)	1 (Command line)	1	True	Yes
HandBrake	4 (MPEG2, FLV, AAC, MP3)	1 (Command line)	1	True	Yes
FFMpeg	7 (FLV, MP3, AAC, MPEG2, TXT, JP2, TIFF)	1 (Command line)	1	True	Yes
SoX	3 (MP3, AAC, MPEG2)	1 (Command line)	1	True	Yes
GStreamer	8 (JP2, FLV, MP3, AAC, MPEG2, TXT, JP2, TIFF)	1 (Command line)	1	True	Yes
Microsoft Office	9 (DOCX, DOC, RTF, TXT, ODF, HTML, XML, PDF, TIFF)	3 (API)	3	False	No
OpenOffice (JODConverter)	6 (ODF, DOC, RTF, TXT, HTML, PDF)	1 (Command line)	5	True	Yes
TeighaFileConverter	0	1 (Command line)	3	True	No
Kakadu	2 (JP2, TIFF)	3 (API)	1	False	No
Apache Sanselan	1 (TIFF)	3 (API)	1	True	Yes
OpenJPEG	1 (JP2, TIFF)	3 (API)	1	True	Yes
SIARD Suite	0	1 (Command line)	3	True	No
GIMP	5 (PS, TIFF, PDF, HTML, TXT)	1 (Command line)	1	True	Yes
Inkscape	1 (PDF, TIFF, PS)	1 (Command line)	1	True	Yes
Avidemux	4 (FLV, MPEG2, MP3, AAC)	1 (Command line)	1	True	Yes
PDFBox	2 (PDF, TXT)	3 (API)	1	True	Yes
JTidy	1 (HTML)	1 (Command line)	1	True	Yes
JasPer	1 (JP2)	3 (API)	1	True	Yes
Aware AccuRad J2KSuite	2 (JP2, TIFF)	3 (API)	3	False	No
Luratech LuraWave	2 (JP2, TIFF)	1 (Command line)	3	False	No
b2xtranslator	2 (DOC, DOCX)	1 (Command line)	1	True	Yes
warc-tools	2 (ARC, WARC)	1 (Command line)	1	False	No
NedlibToArc	1	1	1	True	Yes

	(ARC)	(Command line)			
Heritrix (org.archive.io)	2 (ARC, WARC)	3 (API)	1	True	Yes
Java Image I/O	1 (TIFF)	3 (API)	1	True	Yes
GraphicsMagick	6 (HTML, JP2, PDF, PS, TIFF, TXT)	1 (Command line)	1	True	Yes
ACDSee	4 (AAC, DJVU, JP2, TIFF)	5 (GUI)	3	False	No
CZ-Doc2Pdf	5 (DOC, HTML, PDF, TXT, RTF)	5 (GUI)	5	False	No
Dia	0	1 (Command line)	1	True	No
Document2PDF Pilot	5 (TXT, RTF, HTML, DOC, PDF)	5 (GUI)	5	False	No
EscapeE	5 (PDF, PS, TIFF, RTF, PDF)	1 (Command line)	3	False	No
GraphicConverter X	2 (PDF, TIFF)	3 (API)	5	False	No
MsgText	2 (PDF, TIFF)	1 (Command line)	3	True	Yes
Tesseract-ocr	2 (TIFF, TXT)	1 (Command line)	1	True	Yes
AbiWord	6 (DOC, RTF, TXT, HTML, ODF, DOCX)	1 (Command line)	1	True	Yes
VisualIntegrity	3 (PDF, TIFF, TXT)	1 (Command line)	1	False	No
Photoshop	3 (JP2, PDF, TIFF)	5 (GUI)	5	False	No
PDF Version converter	1 (PDF)	1 (Command line)	3	False	No
Netpbm	1 (TIFF)	1 (Command line)	1	True	Yes

8 Conclusions

In the SCAPE description of work it was indicated that action tools would be evaluated against four criteria: 1) Suitability for deployment on the distributed SCAPE platform; 2) the effort required to make them compatible with the use case scenarios determined by the testbeds; 3) their scalability in terms of their ability to migrate a large set of objects, complex objects composed of multiple bit streams of different types and/or large-sized files and 4) their robustness in terms of reliability and accuracy.

After a careful analysis of the objectives of this task, we came to conclusion that its main goal was to identify and select action tools and services that were suitable for SCAPE integration. That being said, we believe that no tools should be discarded based on the fact that they are not capable of coping with large-sized digital objects or complex objects. Some tools may not be able to handle large-sized files, but be extremely efficient in handling small objects – and that may be exactly what an institution is looking for when planning the preservation of a certain collection of files.

The decision on whether an action tool is suitable or not to preserve a collection of objects is something that should be addressed by a Preservation Planning service. This type of judgment varies according to the preservation needs of the collection, the stakeholders and/or the infrastructure in place (just to name a few of the possible influence factors). Any attempt to evaluate tools on the grounds that their scalability, reliability or accuracy would potentially result in an inaccurate decision based on poor assumptions or incomplete testing results.

The tools have been evaluated using an evaluation framework based on an international standard for software quality assessment (i.e. the ISO/IEC 25010:2011) that focus on the requirements of suitability for the testbed scenarios, compatibility with the SCAPE parallel execution platform and the license of use that accompanies the tool. The selected tools are appropriate for solving the problems depicted by the testbed scenarios, meet the minimum requirements to run on the SCAPE platform and can be used freely without breaking any legal constraints established by the SCAPE agreement or the tools licenses. An overall of 40 tools, ranging from image converters to database migration suites, have been identified and evaluated. After the assessment procedure, 16 of these tools have been considered inadequate to be used in the SCAPE project (40% rejection rate). The main reasons for rejection were licensing constraints and format incompatibility with the testbed scenarios.

Following the work developed in task 1 and reported here, task 2 will focus on adapting the selected action tools to fully support the SCAPE platform and take full advantage of the benefits that a scalable computational platform has to offer, such as enhanced computational power and fault tolerance. Additionally, the tools will need to be adapted to play along with the workflow engine being developed, which will be responsible for executing the plans produced by the Preservation Planning service.

9 References

ARC. (2009). Guidelines for Creating and Installing IF Preservation Workflows and Templates (Deliverable No. IF5-D1). Planets. Obtido de http://www.planets-project.eu/docs/reports/Planets_IF5-D1_Creating&Install_IF_Pres_Workflows.pdf

Ayris, P., Davies, R., McLeod, R., Miao, R., Shenton, H., & Wheatley, P. (2008). The LIFE2 final project report. Obtido de http://eprints.ucl.ac.uk/11758/

Becker, C., & Rauber, A. (2011). Decision criteria in digital preservation: What to measure and how. Journal of the American Society for Information Science and Technology, n/a-n/a. doi:10.1002/asi.21527

Beus-Dukic, L., & Boegh, J. (2003). COTS Software Quality Evaluation. ICCBSS (pp 72-80).

Eld Zierau, & Caroline van Wijk. (2008). The Planets Approach to Migration Tools. Apresentado na IS&T Archiving 2008, Bern, Switzerland.

Faria, L., Ferreira, M., Castro, R., Barbedo, F., Henriques, C., Corujo, L., & Ramalho, J. C. (2009). RODA : a service-oriented repository to preserve authentic digital objects. Apresentado na

International Conference on Open Repositories, Atlanta, USA. Obtido de http://hdl.handle.net/1822/9408

Ferreira, M. (2005). Automatic evaluation of migration quality in distributed networks of converters. Apresentado na European Conference on Digital Libraries (ECDL), Vienna. Obtido de http://hdl.handle.net/1822/3340

Ferreira, M., Baptista, A. A., & Ramalho, J. C. (2006, Julho). A foundation for automatic digital preservation, (48). Obtido de http://hdl.handle.net/1822/5571

Ferreira, M., Baptista, A. A., & Ramalho, J. C. (2007). CRiB : preservation services for Digital Repositories. Apresentado na International Conference Open Repositories, San Antonio, Texas, United States of America. Obtido de http://hdl.handle.net/1822/6195

Hoyer, R. W., & Hoyer, B. B. Y. (2001, Julho). What Is Quality?, 34(7), 10.

ISO/IEC 14598-1:1999, Information technology -- Software product evaluation -- Part 1: General overview. (1999). (p 19). International Organization for Standardization. Obtido de http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=24902

ISO/IEC 15939:2007, Systems and software engineering -- Measurement process. (2007). . International Organization for Standardization. Obtido de http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=44344

ISO/IEC 25010:2011, Systems and software engineering - Systems and software Quality Requirements and Evaluation (SQuaRE) - System and software quality models. (2011). (p 34). International Organization for Standardization. Obtido de http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=35733

ISO/IEC 9126-1:2001, Software engineering - Product quality - Part 1: Quality model. (2001). . International Organization for Standardization. Obtido de http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=22749

Jackson, A., Andrew Lindley, & Fabian Steeg. (2010). Consolidated Release and Documentation (Deliverable No. IF-D11). Planets. Obtido de http://www.planets-project.eu/docs/reports/Planets_IF-D11_ConsolidatedReleaseDocumentation.pdf

KB-NL. (2009). Gap analysis: a survey of PA tool provision (Deliverable No. D3). Planets. Obtido de http://www.planets-project.eu/docs/reports/PA2D3gapanalysis.pdf

Marta, M. H., Librelotto, G. R., Ramalho, J. C., & Henriques, P. R. (2002). Bidirectional conversion between XML documents and relational data bases, 6.

Ramalho, J. C., Ferreira, M., Faria, L., Castro, R., Barbedo, F., & Corujo, L. (2008). RODA and CRiB a service-oriented digital repository (p 7). Apresentado na iPRESS 2008, London. Obtido de http://hdl.handle.net/1822/8226

Task Force on Archiving of Digital Information. (1996). Preserving digital information : report of the Task Force on Archiving of Digital Information. Washington D.C.: Commission on Preservation and Access.

10 Appendix

10.1 List of identified action tools

Tool	ImageMagick[4]
Tool version	6.6.7
License	GPL V3 compatible[5]
Interface	Command-line
Operating system	GNU/Linux; Windows; MacOS X; Other
Open-source	Yes
Object class	Bitmap image
Input formats	AAI, ART, ARW, AVI, AVS, BMP, CALS, CGM, CIN, CMYK, CMYKA, CR2, CRW, CUR, CUT, DCM, DCR, DCX, DIB, DJVU, DNG, DOT, DPX, EMF, EPDF, EPI, EPS, EPS2, EPS3, EPSF, EPSI, EPT, EXR, FAX, FIG, FITS, FPX, GIF, GPLT, GRAY, HPGL, HRZ, HTML, ICO, INFO, INLINE, JBIG, JNG, JP2, JPC, JPEG, MAN, MAT, MIFF, MONO, MNG, M2V, MPEG, MPC, MPR, MRW, MSL, MTV, MVG, NEF, ORF, OTB, P7, PALM, PAM, PBM, PCD, PCDS, PCL, PCX, PDB, PDF, PEF, PFA, PFB, PFM, PGM, PICON, PICT, PIX, PNG, PNG8, PNG24, PNG32, PNM, PPM, PS, PS2, PS3, PSB, PSD, PTIF, PWP, RAD, RAF, RGB, RGBA, RLA, RLE, SCT, SFW, SGI, SHTML, SID, MrSID, SUN, SVG, TGA, TIFF, TIM, TTF, TXT, UIL, UYVY, VICAR, VIFF, WBMP, WMF, WPG, X, XBM, XCF, XPM, XWD, X3F, YCbCr, YCbCrA, YUV
Output formats	AAI, AVS, BMP, CIN, CMYK, CMYKA, DCX, DIB, DPX, EPDF, EPI, EPS, EPSF, EPSI, EPT, EXR, FAX, FITS, FPX, GIF, GRAY, HRZ, HTML, INFO, JBIG, JNG, JP2, JPC, JPEG, MIFF, MONO, MNG, M2V, MPEG, MPC, MPR, MSL, MTV, MVG, OTB, P7, PALM, PBM, PCD, PCDS, PCX, PDB, PDF, PFM, PGM, PICON, PICT, PNG, PNG8, PNG24, PNG32, PNM, PPM, PS, PS2, PS3, PSB, PSD, PTIF, RGB, RGBA, SGI, SHTML, SUN, SVG, TGA, TIFF, TXT, UIL, UYVY, VICAR, VIFF, WBMP, X, XBM, XPM, XWD, YCbCr, YCbCrA, YUV

Tool	Mencoder[6]
Tool version	1.0rc4-4.4.5
License	GPL V2[7]
Interface	Command-line
Operating system	GNU/Linux; Windows; MacOS X
Open-source	Yes
Object class	Video; Audio
Input formats	(S)VCD (Super Video CD), CDRwin's .bin image file, DVD, including encrypted DVD, MPEG-1/2 (ES/PS/PES/VOB), AVI file format, ASF/WMV/WMA format, QT/MOV/MP4 format, RealAudio/RealVideo format, Ogg/OGM files, Matroska, NUT, NSV (Nullsoft Streaming Video), VIVO format, FLI format, NuppelVideo format, yuv4mpeg format, FILM (.cpk) format, RoQ format, PVA format, streaming via HTTP/FTP, RTP/RTSP, MMS/MMST, MPST, SDP, TV grabbing
Output formats	avi - Microsoft Audio/Video Interleaved; mpeg - MPEG-1/2 system stream format; lavf - FFmpeg libavformat muxers; rawvideo - (video only, one stream only) raw stream, no muxing; rawaudio - (audio only, one stream only) raw stream, no muxing;

Tool	HandBrake[8]
Tool version	0.9.5

[4] http://www.imagemagick.org/
[5] http://www.imagemagick.org/script/license.php
[6] http://www.mplayerhq.hu/
[7] http://www.gnu.org/licenses/old-licenses/gpl-2.0.html
[8] http://handbrake.fr

License	GPL V2[9]
Interface	Command-line
Operating system	GNU/Linux; Windows; MacOS X
Open-source	Yes
Object class	Video
Input formats	Any DVD or Bluray-like source: VIDEO_TS folder, DVD image, real DVD or bluray (unencrypted -- removal of copy protection is not supported), and some .VOB, .TS and M2TS files; Most any multimedia file it can get libavformat to read and libavcodec to decode.
Output formats	File format: MP4(M4V) and MKV; Video: MPEG-4(ffmpeg), H.264(x264), or Theora(libtheora); Audio: AAC, CoreAudio AAC (OS X Only), MP3, or Vorbis. AC-3 pass-through, DTS pass-thorugh (MKV only)

Tool	FFMpeg[10]
Tool version	0.6.1
License	GPL v2[11] or GPL v2.1[12]
Interface	Command-line; API
Operating system	GNU/Linux; Windows; MacOS X
Open-source	Yes
Object class	Video; Audio
Input formats	**File formats:** 4xm 8088flex TMV Adobe Filmstrip Audio IFF (AIFF) American Laser Games MM 3GPP AMR Apple HTTP Live Streaming ASF AVI AVISynth AVS Beam Software SIFF Bethesda Softworks VID Bink Bitmap Brothers JV Brute Force & Ignorance Interplay C93 Delphine Software International CIN CD+G Core Audio Format Creative Voice CRYO APC D-Cinema audio Deluxe Paint Animation DFA DV video

[9] http://www.gnu.org/licenses/old-licenses/gpl-2.0.html
[10] http://www.ffmpeg.org
[11] http://www.gnu.org/licenses/old-licenses/gpl-2.0.html
[12] http://www.gnu.org/licenses/old-licenses/lgpl-2.1.html

	DXA Electronic Arts cdata Electronic Arts Multimedia FFM (FFserver live feed) Flash (SWF) Flash 9 (AVM2) FLI/FLC/FLX animation Flash Video (FLV) FunCom ISS GXF id Quake II CIN video id RoQ IEC61937 encapsulation IFF Interplay MVE IV8 IVF (On2) LMLM4 LXF Matroska FFmpeg metadata MAXIS XA MD Studio Mobotix .mxg Monkey's Audio Motion Pixels MVI MOV/QuickTime/MP4 MP2 MP3 MPEG-1 System MPEG-PS (program stream) MPEG-TS (transport stream) MPEG-4 MSN TCP webcam MTV Musepack Musepack SV8 Material eXchange Format (MXF) Material eXchange Format (MXF), D-10 Mapping NC camera feed NTT TwinVQ (VQF) Nullsoft Streaming Video NuppelVideo NUT Ogg Playstation Portable PMP TechnoTrend PVA QCP raw ADTS (AAC) raw AC-3 raw Chinese AVS video

	raw CRI ADX raw Dirac raw DNxHD raw DTS raw E-AC-3 raw FLAC raw GSM raw H.261 raw H.263 raw H.264 raw Ingenient MJPEG raw MJPEG raw MLP raw MPEG raw MPEG-1 raw MPEG-2 raw MPEG-4 raw video raw Shorten raw TrueHD raw VC-1 raw PCM A-law raw PCM mu-law raw PCM signed 8 bit raw PCM signed 16 bit big-endian raw PCM signed 16 bit little-endian raw PCM signed 24 bit big-endian raw PCM signed 24 bit little-endian raw PCM signed 32 bit big-endian raw PCM signed 32 bit little-endian raw PCM unsigned 8 bit raw PCM unsigned 16 bit big-endian raw PCM unsigned 16 bit little-endian raw PCM unsigned 24 bit big-endian raw PCM unsigned 24 bit little-endian raw PCM unsigned 32 bit big-endian raw PCM unsigned 32 bit little-endian raw PCM floating-point 32 bit big-endian raw PCM floating-point 32 bit little-endian raw PCM floating-point 64 bit big-endian raw PCM floating-point 64 bit little-endian RDT REDCODE R3D RealMedia Redirector Renderware TeXture Dictionary RL2 RPL/ARMovie Lego Mindstorms RSO RTMP RTP

	RTSP SAP SDP Sega FILM/CPK Sierra SOL Sierra VMD Smacker Sony OpenMG (OMA) Sony PlayStation STR Sony Wave64 (W64) SoX native format SUN AU format Text files THP Tiertex Limited SEQ True Audio VC-1 test bitstream WAV WavPack WebM Windows Televison (WTV) Wing Commander III movie Westwood Studios audio Westwood Studios VQA xWMA YUV4MPEG pipe Psygnosis YOP **Image formats:** .Y.U.V animated GIF BMP DPX JPEG JPEG 2000 JPEG-LS PAM PBM PCX PGM PGMYUV PIC PNG PPM PTX SGI Sun Rasterfile TIFF Truevision Targa **Video codecs:**

	4X Movie 8088flex TMV 8SVX exponential 8SVX fibonacci American Laser Games MM AMV Video ANSI/ASCII art Apple MJPEG-B Apple QuickDraw Asus v1 Asus v2 ATI VCR1 ATI VCR2 Auravision Aura Auravision Aura 2 Autodesk Animator Flic video Autodesk RLE AVS (Audio Video Standard) video Beam Software VB Bethesda VID video Bink Video Bitmap Brothers JV video Brute Force & Ignorance C93 video CamStudio CD+G Chinese AVS video Delphine Software International CIN video Cinepak Cirrus Logic AccuPak Creative YUV (CYUV) DFA Dirac Deluxe Paint Animation DNxHD Duck TrueMotion 1.0 Duck TrueMotion 2.0 DV (Digital Video) Feeble Files/ScummVM DXA Electronic Arts CMV video Electronic Arts Madcow video Electronic Arts TGV video Electronic Arts TGQ video Electronic Arts TQI video Escape 124 FFmpeg video codec #1 Flash Screen Video v1 Flash Video (FLV) Fraps H.261 H.263 / H.263-1996

	H.263+ / H.263-1998 / H.263 version 2 H.264 / AVC / MPEG-4 AVC / MPEG-4 part 10 H.264 / AVC / MPEG-4 AVC / MPEG-4 part 10 (VDPAU acceleration) HuffYUV HuffYUV FFmpeg variant IBM Ultimotion id Cinematic video id RoQ video IFF ILBM IFF ByteRun1 Intel H.263 Intel Indeo 2 Intel Indeo 3 Intel Indeo 5 Interplay C93 Interplay MVE video Karl Morton's video codec Kega Game Video (KGV1) Lagarith LCL (LossLess Codec Library) MSZH LCL (LossLess Codec Library) ZLIB LOCO lossless MJPEG Microsoft RLE Microsoft Video 1 Mimic Miro VideoXL MJPEG (Motion JPEG) Mobotix MxPEG video Motion Pixels video MPEG-1 video MPEG-1/2 video XvMC (X-Video Motion Compensation) MPEG-1/2 video (VDPAU acceleration) MPEG-2 video MPEG-4 part 2 MPEG-4 part 2 Microsoft variant version 1 MPEG-4 part 2 Microsoft variant version 2 MPEG-4 part 2 Microsoft variant version 3 Nintendo Gamecube THP video NuppelVideo/RTjpeg On2 VP3 On2 VP5 On2 VP6 VP8 planar RGB Q-team QPEG QuickTime 8BPS video QuickTime Animation (RLE) video QuickTime Graphics (SMC) QuickTime video (RPZA) R10K AJA Kona 10-bit RGB Codec

	R210 Quicktime Uncompressed RGB 10-bit Raw Video RealVideo 1.0 RealVideo 2.0 RealVideo 3.0 RealVideo 4.0 Renderware TXD (TeXture Dictionary) RL2 video Sierra VMD video Smacker video SMPTE VC-1 Snow Sony PlayStation MDEC (Motion DECoder) Sorenson Vector Quantizer 1 Sorenson Vector Quantizer 3 Sunplus JPEG (SP5X) TechSmith Screen Capture Codec Theora Tiertex Limited SEQ video V210 Quicktime Uncompressed 4:2:2 10-bit VMware Screen Codec / VMware Video Westwood Studios VQA (Vector Quantized Animation) video Windows Media Video 7 Windows Media Video 8 Windows Media Video 9 Wing Commander III / Xan Wing Commander IV / Xan Winnov WNV1 WMV7 YAMAHA SMAF Psygnosis YOP Video ZLIB Zip Motion Blocks Video **Audio codecs:** 8SVX audio AAC AC-3 ADPCM 4X Movie ADPCM CDROM XA ADPCM Creative Technology ADPCM Electronic Arts ADPCM Electronic Arts Maxis CDROM XS ADPCM Electronic Arts R1 ADPCM Electronic Arts R2 ADPCM Electronic Arts R3 ADPCM Electronic Arts XAS ADPCM G.722 ADPCM G.726 ADPCM IMA AMV ADPCM IMA Electronic Arts EACS

	ADPCM IMA Electronic Arts SEAD ADPCM IMA Funcom ADPCM IMA QuickTime ADPCM IMA Loki SDL MJPEG ADPCM IMA WAV ADPCM IMA Westwood ADPCM ISS IMA ADPCM IMA Duck DK3 ADPCM IMA Duck DK4 ADPCM Microsoft ADPCM MS IMA ADPCM Nintendo Gamecube THP ADPCM QT IMA ADPCM SEGA CRI ADX ADPCM Shockwave Flash ADPCM SMJPEG IMA ADPCM Sound Blaster Pro 2-bit ADPCM Sound Blaster Pro 2.6-bit ADPCM Sound Blaster Pro 4-bit ADPCM Westwood Studios IMA ADPCM Yamaha AMR-NB AMR-WB Apple lossless audio Atrac 1 Atrac 3 Bink Audio CELT (Opus) Delphine Software International CIN audio COOK DCA (DTS Coherent Acoustics) DPCM id RoQ DPCM Interplay DPCM Sierra Online DPCM Sol DPCM Xan DSP Group TrueSpeech DV audio Enhanced AC-3 FLAC (Free Lossless Audio Codec) GSM GSM Microsoft variant IMC (Intel Music Coder) MACE (Macintosh Audio Compression/Expansion) 3:1 MACE (Macintosh Audio Compression/Expansion) 6:1 MLP (Meridian Lossless Packing) Monkey's Audio MP1 (MPEG audio layer 1) MP2 (MPEG audio layer 2) MP3 (MPEG audio layer 3) MPEG-4 Audio Lossless Coding (ALS)

	Musepack SV7 Musepack SV8 Nellymoser Asao PCM A-law PCM mu-law PCM 16-bit little-endian planar PCM 32-bit floating point big-endian PCM 32-bit floating point little-endian PCM 64-bit floating point big-endian PCM 64-bit floating point little-endian PCM D-Cinema audio signed 24-bit PCM signed 8-bit PCM signed 16-bit big-endian PCM signed 16-bit little-endian PCM signed 24-bit big-endian PCM signed 24-bit little-endian PCM signed 32-bit big-endian PCM signed 32-bit little-endian PCM signed 16/20/24-bit big-endian in MPEG-TS PCM unsigned 8-bit PCM unsigned 16-bit big-endian PCM unsigned 16-bit little-endian PCM unsigned 24-bit big-endian PCM unsigned 24-bit little-endian PCM unsigned 32-bit big-endian PCM unsigned 32-bit little-endian PCM Zork QCELP / PureVoice QDesign Music Codec 2 RealAudio 1.0 (14.4K) RealAudio 2.0 (28.8K) RealAudio 3.0 (dnet) RealAudio SIPR / ACELP.NET Shorten Sierra VMD audio Smacker audio SMPTE 302M AES3 audio Sonic Sonic lossless Speex True Audio (TTA) TrueHD TwinVQ (VQF flavor) Vorbis WavPack Westwood Audio (SND1) Windows Media Audio 1 Windows Media Audio 2 Windows Media Audio Pro Windows Media Audio Voice
Output formats	**File formats:**

	Adobe Filmstrip Audio IFF (AIFF) 3GPP AMR ASF AVI Core Audio Format CRC testing format Creative Voice D-Cinema audio DV video FFM (FFserver live feed) Flash (SWF) Flash 9 (AVM2) Flash Video (FLV) framecrc testing format GIF Animation GXF id RoQ IEC61937 encapsulation IVF (On2) Matroska Matroska audio FFmpeg metadata MOV/QuickTime/MP4 MP2 MP3 MPEG-1 System MPEG-PS (program stream) MPEG-TS (transport stream) MPEG-4 MIME multipart JPEG Material eXchange Format (MXF) Material eXchange Format (MXF), D-10 Mapping NUT Ogg raw ADTS (AAC) raw AC-3 raw Chinese AVS video raw CRI ADX raw Dirac raw DNxHD raw DTS raw E-AC-3 raw FLAC raw H.261 raw H.263 raw H.264 raw MJPEG raw MPEG-4 raw NULL raw video

	raw id RoQ raw TrueHD raw PCM A-law raw PCM mu-law raw PCM signed 8 bit raw PCM signed 16 bit big-endian raw PCM signed 16 bit little-endian raw PCM signed 24 bit big-endian raw PCM signed 24 bit little-endian raw PCM signed 32 bit big-endian raw PCM signed 32 bit little-endian raw PCM unsigned 8 bit raw PCM unsigned 16 bit big-endian raw PCM unsigned 16 bit little-endian raw PCM unsigned 24 bit big-endian raw PCM unsigned 24 bit little-endian raw PCM unsigned 32 bit big-endian raw PCM unsigned 32 bit little-endian raw PCM floating-point 32 bit big-endian raw PCM floating-point 32 bit little-endian raw PCM floating-point 64 bit big-endian raw PCM floating-point 64 bit little-endian RealMedia Lego Mindstorms RSO RTMP RTP RTSP SAP SoX native format SUN AU format VC-1 test bitstream WAV WebM YUV4MPEG pipe **Image formats:** .Y.U.V animated GIF BMP DPX JPEG JPEG 2000 JPEG-LS LJPEG PAM PBM PCX PGM PGMYUV PNG PPM

	SGI TIFF Truevision Targa **Video codecs:** A64 multicolor Asus v1 Asus v2 Chinese AVS video Dirac DNxHD DV (Digital Video) FFmpeg video codec #1 Flash Screen Video v1 Flash Screen Video v2 Flash Video (FLV) H.261 H.263 / H.263-1996 H.263+ / H.263-1998 / H.263 version 2 H.264 / AVC / MPEG-4 AVC / MPEG-4 part 10 H.264 / AVC / MPEG-4 AVC / MPEG-4 part 10 (VDPAU acceleration) HuffYUV HuffYUV FFmpeg variant id RoQ video LCL (LossLess Codec Library) ZLIB lossless MJPEG MJPEG (Motion JPEG) MPEG-1 video MPEG-2 video MPEG-4 part 2 MPEG-4 part 2 Microsoft variant version 2 MPEG-4 part 2 Microsoft variant version 3 VP8 QuickTime Animation (RLE) video Raw Video RealVideo 1.0 RealVideo 2.0 Snow Sorenson Vector Quantizer 1 Theora V210 Quicktime Uncompressed 4:2:2 10-bit Windows Media Video 7 Windows Media Video 8 WMV7 YAMAHA SMAF ZLIB Zip Motion Blocks Video **Audio codecs:** AAC AC-3

	ADPCM G.722 ADPCM G.726 ADPCM IMA QuickTime ADPCM IMA WAV ADPCM Microsoft ADPCM MS IMA ADPCM QT IMA ADPCM SEGA CRI ADX ADPCM Shockwave Flash ADPCM Yamaha AMR-NB AMR-WB Apple lossless audio DCA (DTS Coherent Acoustics) DPCM id RoQ Enhanced AC-3 FLAC (Free Lossless Audio Codec) GSM GSM Microsoft variant MP2 (MPEG audio layer 2) MP3 (MPEG audio layer 3) Nellymoser Asao PCM A-law PCM mu-law PCM 32-bit floating point big-endian PCM 32-bit floating point little-endian PCM 64-bit floating point big-endian PCM 64-bit floating point little-endian PCM D-Cinema audio signed 24-bit PCM signed 8-bit PCM signed 16-bit big-endian PCM signed 16-bit little-endian PCM signed 24-bit big-endian PCM signed 24-bit little-endian PCM signed 32-bit big-endian PCM signed 32-bit little-endian PCM unsigned 8-bit PCM unsigned 16-bit big-endian PCM unsigned 16-bit little-endian PCM unsigned 24-bit big-endian PCM unsigned 24-bit little-endian PCM unsigned 32-bit big-endian PCM unsigned 32-bit little-endian PCM Zork RealAudio 1.0 (14.4K) RealAudio 3.0 (dnet) Sonic Sonic lossless Vorbis Windows Media Audio 1 Windows Media Audio 2

Tool	SoX[13]
Tool version	14.3.2
License	GPL v2[14]
Interface	Command-line
Operating system	GNU/Linux; Windows; MacOS X
Open-source	Yes
Object class	Audio
Input formats	Raw files in various binary formats; Raw textual data; Amiga 8svx files; Apple/SGI AIFF files; SUN .au files: PCM, u-law, A-law, G7xx ADPCM files, mutant DEC .au files, NeXT .snd files; AVR files, CDDA (Compact Disc Digital Audio format); CVS and VMS files (continuous variable slope); Grandstream ring-tone files; GSM files; HTK files; LPC-10 files; Macintosh HCOM files; Amiga MAUD files; AMR-WB & AMR-NB (with optional libamrwb & libamrnb libraries); MP3 (with optional libmad and libmp3lame libraries); MP4, AAC, AC3, WAVPACK, AMR-NB files (with optional ffmpeg library); AVI, WMV, Ogg Theora, MPEG video files (with optional ffmpeg library); Ogg Vorbis files (with optional Ogg Vorbis libraries); FLAC files (with optional libFLAC); IRCAM SoundFile files; NIST SPHERE files; Turtle beach SampleVision files; Sounder & Soundtool (DOS) files; Yamaha TX-16W sampler files; SoundBlaster .VOC files; Dialogic/OKI ADPCM files (.VOX); Microsoft .WAV files: PCM, u-law, A-law, MS ADPCM, IMA ADPCM, GSM, RIFX (big endian); WavPack files (with optional libwavpack library); Psion (palmtop) A-law WVE files and Record voice notes; Maxis XA Audio files: EA ADPCM
Output formats	Raw files in various binary formats; Raw textual data; Amiga 8svx files; Apple/SGI AIFF files; SUN .au files: PCM, u-law, A-law, mutant DEC .au files, NeXT .snd files; AVR files, CDDA (Compact Disc Digital Audio format); CVS and VMS files (continuous variable slope); Grandstream ring-tone files; GSM files; HTK files; LPC-10 files; Macintosh HCOM files; Amiga MAUD files; AMR-WB & AMR-NB (with optional libamrwb & libamrnb libraries); MP3 (with optional libmad and libmp3lame libraries); MP4, AAC, AC3, WAVPACK, AMR-NB files (with optional ffmpeg library); AVI, WMV, Ogg Theora, MPEG video files (with optional ffmpeg library); Ogg Vorbis files (with optional Ogg Vorbis libraries); FLAC files (with optional libFLAC); IRCAM SoundFile files; NIST SPHERE files; Turtle beach SampleVision files; Sounder & Soundtool (DOS) files; Yamaha TX-16W sampler files; SoundBlaster .VOC files; Dialogic/OKI ADPCM files (.VOX); Microsoft .WAV files: PCM, u-law, A-law, MS ADPCM, IMA ADPCM, GSM, RIFX (big endian); WavPack files (with optional libwavpack library); Psion (palmtop) A-law WVE files and Record voice notes

Tool	GStreamer[15]
Tool version	0.10.28
License	LGPL v3[16]
Interface	Command-line; API
Operating system	GNU/Linux; Windows
Open-source	Yes
Object class	Bitmap; Vector; Video; Audio
Input formats	NSF, PNM, Musepack, BZ2, DTS, SVG, CELT,

[13] http://sox.sourceforge.net/
[14] http://www.gnu.org/licenses/old-licenses/gpl-2.0.html
[15] http://gstreamer.freedesktop.org
[16] http://www.fsf.org/licenses/lgpl.html

	JPEG2000, TTA audio, Bayer, XviD, ADPCM, AAC, VMnc, GSM, AMR-WB, MPEG1, MPEG 2, DVD LPCM, AMR-NB, Sid, MP3, ATSC A/52, Vorbis, A Law, DV, PNG, Mu Law, Speex, Wavpack, FLAC, Theora, CMML, JPEG, Smoke, FLX, Bluetooth SBC **From FFMpeg library:** **File formats:** 4xm 8088flex TMV Adobe Filmstrip Audio IFF (AIFF) American Laser Games MM 3GPP AMR Apple HTTP Live Streaming ASF AVI AVISynth AVS Beam Software SIFF Bethesda Softworks VID Bink Bitmap Brothers JV Brute Force & Ignorance Interplay C93

	Delphine Software International CIN CD+G Core Audio Format Creative Voice CRYO APC D-Cinema audio Deluxe Paint Animation DFA DV video DXA Electronic Arts cdata Electronic Arts Multimedia FFM (FFserver live feed) Flash (SWF) Flash 9 (AVM2) FLI/FLC/FLX animation Flash Video (FLV) FunCom ISS GXF id Quake II CIN video id RoQ IEC61937 encapsulation IFF Interplay MVE IV8 IVF (On2) LMLM4 LXF Matroska FFmpeg metadata MAXIS XA MD Studio Mobotix .mxg Monkey's Audio Motion Pixels MVI MOV/QuickTime/MP4 MP2 MP3 MPEG-1 System MPEG-PS (program stream) MPEG-TS (transport stream) MPEG-4 MSN TCP webcam MTV Musepack Musepack SV8 Material eXchange Format (MXF) Material eXchange Format (MXF), D-10 Mapping NC camera feed NTT TwinVQ (VQF) Nullsoft Streaming Video

	NuppelVideo NUT Ogg Playstation Portable PMP TechnoTrend PVA QCP raw ADTS (AAC) raw AC-3 raw Chinese AVS video raw CRI ADX raw Dirac raw DNxHD raw DTS raw E-AC-3 raw FLAC raw GSM raw H.261 raw H.263 raw H.264 raw Ingenient MJPEG raw MJPEG raw MLP raw MPEG raw MPEG-1 raw MPEG-2 raw MPEG-4 raw video raw Shorten raw TrueHD raw VC-1 raw PCM A-law raw PCM mu-law raw PCM signed 8 bit raw PCM signed 16 bit big-endian raw PCM signed 16 bit little-endian raw PCM signed 24 bit big-endian raw PCM signed 24 bit little-endian raw PCM signed 32 bit big-endian raw PCM signed 32 bit little-endian raw PCM unsigned 8 bit raw PCM unsigned 16 bit big-endian raw PCM unsigned 16 bit little-endian raw PCM unsigned 24 bit big-endian raw PCM unsigned 24 bit little-endian raw PCM unsigned 32 bit big-endian raw PCM unsigned 32 bit little-endian raw PCM floating-point 32 bit big-endian raw PCM floating-point 32 bit little-endian raw PCM floating-point 64 bit big-endian raw PCM floating-point 64 bit little-endian RDT

	REDCODE R3D RealMedia Redirector Renderware TeXture Dictionary RL2 RPL/ARMovie Lego Mindstorms RSO RTMP RTP RTSP SAP SDP Sega FILM/CPK Sierra SOL Sierra VMD Smacker Sony OpenMG (OMA) Sony PlayStation STR Sony Wave64 (W64) SoX native format SUN AU format Text files THP Tiertex Limited SEQ True Audio VC-1 test bitstream WAV WavPack WebM Windows Televison (WTV) Wing Commander III movie Westwood Studios audio Westwood Studios VQA xWMA YUV4MPEG pipe Psygnosis YOP **Image formats:** .Y.U.V animated GIF BMP DPX JPEG JPEG 2000 JPEG-LS PAM PBM PCX PGM PGMYUV PIC

	PNG PPM PTX SGI Sun Rasterfile TIFF Truevision Targa **Video codecs:** 4X Movie 8088flex TMV 8SVX exponential 8SVX fibonacci American Laser Games MM AMV Video ANSI/ASCII art Apple MJPEG-B Apple QuickDraw Asus v1 Asus v2 ATI VCR1 ATI VCR2 Auravision Aura Auravision Aura 2 Autodesk Animator Flic video Autodesk RLE AVS (Audio Video Standard) video Beam Software VB Bethesda VID video Bink Video Bitmap Brothers JV video Brute Force & Ignorance C93 video CamStudio CD+G Chinese AVS video Delphine Software International CIN video Cinepak Cirrus Logic AccuPak Creative YUV (CYUV) DFA Dirac Deluxe Paint Animation DNxHD Duck TrueMotion 1.0 Duck TrueMotion 2.0 DV (Digital Video) Feeble Files/ScummVM DXA Electronic Arts CMV video Electronic Arts Madcow video Electronic Arts TGV video

	Electronic Arts TGQ video Electronic Arts TQI video Escape 124 FFmpeg video codec #1 Flash Screen Video v1 Flash Video (FLV) Fraps H.261 H.263 / H.263-1996 H.263+ / H.263-1998 / H.263 version 2 H.264 / AVC / MPEG-4 AVC / MPEG-4 part 10 H.264 / AVC / MPEG-4 AVC / MPEG-4 part 10 (VDPAU acceleration) HuffYUV HuffYUV FFmpeg variant IBM Ultimotion id Cinematic video id RoQ video IFF ILBM IFF ByteRun1 Intel H.263 Intel Indeo 2 Intel Indeo 3 Intel Indeo 5 Interplay C93 Interplay MVE video Karl Morton's video codec Kega Game Video (KGV1) Lagarith LCL (LossLess Codec Library) MSZH LCL (LossLess Codec Library) ZLIB LOCO lossless MJPEG Microsoft RLE Microsoft Video 1 Mimic Miro VideoXL MJPEG (Motion JPEG) Mobotix MxPEG video Motion Pixels video MPEG-1 video MPEG-1/2 video XvMC (X-Video Motion Compensation) MPEG-1/2 video (VDPAU acceleration) MPEG-2 video MPEG-4 part 2 MPEG-4 part 2 Microsoft variant version 1 MPEG-4 part 2 Microsoft variant version 2 MPEG-4 part 2 Microsoft variant version 3 Nintendo Gamecube THP video NuppelVideo/RTjpeg On2 VP3 On2 VP5

	On2 VP6 VP8 planar RGB Q-team QPEG QuickTime 8BPS video QuickTime Animation (RLE) video QuickTime Graphics (SMC) QuickTime video (RPZA) R10K AJA Kona 10-bit RGB Codec R210 Quicktime Uncompressed RGB 10-bit Raw Video RealVideo 1.0 RealVideo 2.0 RealVideo 3.0 RealVideo 4.0 Renderware TXD (TeXture Dictionary) RL2 video Sierra VMD video Smacker video SMPTE VC-1 Snow Sony PlayStation MDEC (Motion DECoder) Sorenson Vector Quantizer 1 Sorenson Vector Quantizer 3 Sunplus JPEG (SP5X) TechSmith Screen Capture Codec Theora Tiertex Limited SEQ video V210 Quicktime Uncompressed 4:2:2 10-bit VMware Screen Codec / VMware Video Westwood Studios VQA (Vector Quantized Animation) video Windows Media Video 7 Windows Media Video 8 Windows Media Video 9 Wing Commander III / Xan Wing Commander IV / Xan Winnov WNV1 WMV7 YAMAHA SMAF Psygnosis YOP Video ZLIB Zip Motion Blocks Video **Audio codecs:** 8SVX audio AAC AC-3 ADPCM 4X Movie ADPCM CDROM XA ADPCM Creative Technology ADPCM Electronic Arts

	ADPCM Electronic Arts Maxis CDROM XS ADPCM Electronic Arts R1 ADPCM Electronic Arts R2 ADPCM Electronic Arts R3 ADPCM Electronic Arts XAS ADPCM G.722 ADPCM G.726 ADPCM IMA AMV ADPCM IMA Electronic Arts EACS ADPCM IMA Electronic Arts SEAD ADPCM IMA Funcom ADPCM IMA QuickTime ADPCM IMA Loki SDL MJPEG ADPCM IMA WAV ADPCM IMA Westwood ADPCM ISS IMA ADPCM IMA Duck DK3 ADPCM IMA Duck DK4 ADPCM Microsoft ADPCM MS IMA ADPCM Nintendo Gamecube THP ADPCM QT IMA ADPCM SEGA CRI ADX ADPCM Shockwave Flash ADPCM SMJPEG IMA ADPCM Sound Blaster Pro 2-bit ADPCM Sound Blaster Pro 2.6-bit ADPCM Sound Blaster Pro 4-bit ADPCM Westwood Studios IMA ADPCM Yamaha AMR-NB AMR-WB Apple lossless audio Atrac 1 Atrac 3 Bink Audio CELT (Opus) Delphine Software International CIN audio COOK DCA (DTS Coherent Acoustics) DPCM id RoQ DPCM Interplay DPCM Sierra Online DPCM Sol DPCM Xan DSP Group TrueSpeech DV audio Enhanced AC-3 FLAC (Free Lossless Audio Codec) GSM GSM Microsoft variant

	IMC (Intel Music Coder) MACE (Macintosh Audio Compression/Expansion) 3:1 MACE (Macintosh Audio Compression/Expansion) 6:1 MLP (Meridian Lossless Packing) Monkey's Audio MP1 (MPEG audio layer 1) MP2 (MPEG audio layer 2) MP3 (MPEG audio layer 3) MPEG-4 Audio Lossless Coding (ALS) Musepack SV7 Musepack SV8 Nellymoser Asao PCM A-law PCM mu-law PCM 16-bit little-endian planar PCM 32-bit floating point big-endian PCM 32-bit floating point little-endian PCM 64-bit floating point big-endian PCM 64-bit floating point little-endian PCM D-Cinema audio signed 24-bit PCM signed 8-bit PCM signed 16-bit big-endian PCM signed 16-bit little-endian PCM signed 24-bit big-endian PCM signed 24-bit little-endian PCM signed 32-bit big-endian PCM signed 32-bit little-endian PCM signed 16/20/24-bit big-endian in MPEG-TS PCM unsigned 8-bit PCM unsigned 16-bit big-endian PCM unsigned 16-bit little-endian PCM unsigned 24-bit big-endian PCM unsigned 24-bit little-endian PCM unsigned 32-bit big-endian PCM unsigned 32-bit little-endian PCM Zork QCELP / PureVoice QDesign Music Codec 2 RealAudio 1.0 (14.4K) RealAudio 2.0 (28.8K) RealAudio 3.0 (dnet) RealAudio SIPR / ACELP.NET Shorten Sierra VMD audio Smacker audio SMPTE 302M AES3 audio Sonic Sonic lossless Speex True Audio (TTA) TrueHD

	TwinVQ (VQF flavor) Vorbis WavPack Westwood Audio (SND1) Windows Media Audio 1 Windows Media Audio 2 Windows Media Audio Pro Windows Media Audio Voice
Output formats	PNM, BZ2, CELT, Jasper JPEG2000, XviD video, ADPCM, GSM audio, mp3, AMR-NB audio, mp2, Vorbis, A Law, YUV4MPEG, PNG, Mu Law, Speex, Wavpack, FLAC, Cairo, Theora, CMML streams, JPEG, Smoke video, Bluetooth SBC **From FFMpeg library:** **File formats:** Adobe Filmstrip Audio IFF (AIFF) 3GPP AMR ASF AVI Core Audio Format CRC testing format Creative Voice D-Cinema audio DV video FFM (FFserver live feed) Flash (SWF) Flash 9 (AVM2) Flash Video (FLV) framecrc testing format

	GIF Animation GXF id RoQ IEC61937 encapsulation IVF (On2) Matroska Matroska audio FFmpeg metadata MOV/QuickTime/MP4 MP2 MP3 MPEG-1 System MPEG-PS (program stream) MPEG-TS (transport stream) MPEG-4 MIME multipart JPEG Material eXchange Format (MXF) Material eXchange Format (MXF), D-10 Mapping NUT Ogg raw ADTS (AAC) raw AC-3 raw Chinese AVS video raw CRI ADX raw Dirac raw DNxHD raw DTS raw E-AC-3 raw FLAC raw H.261 raw H.263 raw H.264 raw MJPEG raw MPEG-4 raw NULL raw video raw id RoQ raw TrueHD raw PCM A-law raw PCM mu-law raw PCM signed 8 bit raw PCM signed 16 bit big-endian raw PCM signed 16 bit little-endian raw PCM signed 24 bit big-endian raw PCM signed 24 bit little-endian raw PCM signed 32 bit big-endian raw PCM signed 32 bit little-endian raw PCM unsigned 8 bit raw PCM unsigned 16 bit big-endian raw PCM unsigned 16 bit little-endian raw PCM unsigned 24 bit big-endian

	raw PCM unsigned 24 bit little-endian raw PCM unsigned 32 bit big-endian raw PCM unsigned 32 bit little-endian raw PCM floating-point 32 bit big-endian raw PCM floating-point 32 bit little-endian raw PCM floating-point 64 bit big-endian raw PCM floating-point 64 bit little-endian RealMedia Lego Mindstorms RSO RTMP RTP RTSP SAP SoX native format SUN AU format VC-1 test bitstream WAV WebM YUV4MPEG pipe **Image formats:** .Y.U.V animated GIF BMP DPX JPEG JPEG 2000 JPEG-LS LJPEG PAM PBM PCX PGM PGMYUV PNG PPM SGI TIFF Truevision Targa **Video codecs:** A64 multicolor Asus v1 Asus v2 Chinese AVS video Dirac DNxHD DV (Digital Video) FFmpeg video codec #1 Flash Screen Video v1 Flash Screen Video v2

	Flash Video (FLV) H.261 H.263 / H.263-1996 H.263+ / H.263-1998 / H.263 version 2 H.264 / AVC / MPEG-4 AVC / MPEG-4 part 10 H.264 / AVC / MPEG-4 AVC / MPEG-4 part 10 (VDPAU acceleration) HuffYUV HuffYUV FFmpeg variant id RoQ video LCL (LossLess Codec Library) ZLIB lossless MJPEG MJPEG (Motion JPEG) MPEG-1 video MPEG-2 video MPEG-4 part 2 MPEG-4 part 2 Microsoft variant version 2 MPEG-4 part 2 Microsoft variant version 3 VP8 QuickTime Animation (RLE) video Raw Video RealVideo 1.0 RealVideo 2.0 Snow Sorenson Vector Quantizer 1 Theora V210 Quicktime Uncompressed 4:2:2 10-bit Windows Media Video 7 Windows Media Video 8 WMV7 YAMAHA SMAF ZLIB Zip Motion Blocks Video **Audio codecs:** AAC AC-3 ADPCM G.722 ADPCM G.726 ADPCM IMA QuickTime ADPCM IMA WAV ADPCM Microsoft ADPCM MS IMA ADPCM QT IMA ADPCM SEGA CRI ADX ADPCM Shockwave Flash ADPCM Yamaha AMR-NB AMR-WB Apple lossless audio DCA (DTS Coherent Acoustics) DPCM id RoQ

	Enhanced AC-3 FLAC (Free Lossless Audio Codec) GSM GSM Microsoft variant MP2 (MPEG audio layer 2) MP3 (MPEG audio layer 3) Nellymoser Asao PCM A-law PCM mu-law PCM 32-bit floating point big-endian PCM 32-bit floating point little-endian PCM 64-bit floating point big-endian PCM 64-bit floating point little-endian PCM D-Cinema audio signed 24-bit PCM signed 8-bit PCM signed 16-bit big-endian PCM signed 16-bit little-endian PCM signed 24-bit big-endian PCM signed 24-bit little-endian PCM signed 32-bit big-endian PCM signed 32-bit little-endian PCM unsigned 8-bit PCM unsigned 16-bit big-endian PCM unsigned 16-bit little-endian PCM unsigned 24-bit big-endian PCM unsigned 24-bit little-endian PCM unsigned 32-bit big-endian PCM unsigned 32-bit little-endian PCM Zork RealAudio 1.0 (14.4K) RealAudio 3.0 (dnet) Sonic Sonic lossless Vorbis Windows Media Audio 1 Windows Media Audio 2

Tool	Microsoft Office[17]
Tool version	2010
License	Comercial[18]
Interface	API; GUI
Operating system	Windows; MacOS X
Open-source	No
Object class	Word Processor; Documents; Spreadsheets
Input formats	Text Formats Word Documents (*.docx)

[17] http://office.microsoft.com

[18] http://www.microsoft.com/downloads/en/details.aspx?FamilyID=2b37ad5e-dc64-400b-a00d-88d982292b07

	Word 97-2003 Documents (*.doc) Rich Text Format (*.rtf) Text Files (*.txt) Open Document Text (*.odt) Word Macro-Enabled Documents (*.docm) All Web Pages (*.html) WordPerfect 5.x (*.wp5) WordPerfect 6.x (*.wp6) XML (*.xml) Works 6-9 Document (*.wps) Spreadsheet Formats Excel Workbook (*.xlsx) Excel Macro-Enabled Workbook (*.xlsm) Excel Binary Workbook (*.xlsb) Excel 97-2003 Workbook (*.xls) Web Pages (*.htm) Text Files (*.txt) CSV (Comma delimited) (*.csv) DIF (Data Interchange Format) (*.dif) SYLK (Symbolic Link) (*.slk) Open Document Spreadsheet (*.ods) XML (*.xml) Access (*.accdb) dBase Files (*.dbf) Presentation Formats All PowerPoint Presentations (*.pptx, *.pptm, *.ppt) Presentations and Shows (*.ppsx, *.ppsm, *.pps) All Web Pages (*.html) OpenDocument Presentation (*.odp) - may not support all features

Output formats	Text Formats Word Documents (*.docx) Word 97-2003 Documents (*.doc) Rich Text Format (*.rtf) Text Files (*.txt) Open Document Text (*.odt) Word Macro-Enabled Documents (*.docm) Web Pages (*.htm) WordPerfect 5.x (*.wp5) WordPerfect 6.x (*.wp6) Word XML Document (*.xml) Works 6-9 Document (*.wps) PDF (*.pdf) XPS Document (*.xps) Spreadsheet Formats Excel Workbook (*.xlsx) Excel Macro-Enabled Workbook (*.xlsm) Excel Binary Workbook (*.xlsb) Excel 97-2003 Workbook (*.xls) Web Pages (*.htm) Text Files (*.txt) CSV (Comma delimited) (*.csv) DIF (Data Interchange Format) (*.dif) SYLK (Symbolic Link) (*.slk) PDF (*.pdf) XPS Document (*.xps) Open Document Spreadsheet (*.ods) Presentation Formats PowerPoint Presentation (*.pptx) PowerPoint Macro-Enabled Presentation (*.pptm) PowerPoint 97-2003 Presentation (*.ppt) PDF (*.pdf) XPS Document (*.xps) PowerPoint Show (*.ppsx) PowerPoint Macro-Enabled Show (*.ppsm) PowerPoint 97-2003 Show (*.pps) Windows Media Video (*.wmv) GIF Graphics Interchange Format (*.gif) JPEG File Interchange Format (*.jpg) PNG Portable Network Graphics Format (*.png) TIFF Tag Image File Format (*.tiff) Device Independent Bitmap (*.bmp) Windows Metafile (*.wmf) Enhanced Windows Metafile (*.emf) Outline/RTF (*.rtf) OpenDocument Presentation (*.odp) - may not support all features

Tool	OpenOffice[19] with JODConverter[20]
Tool version	3.2.1 (JODConverter 2.2.2)
License	LGPL v3[21]
Interface	Command-line
Operating system	GNU/Linux; Windows; MacOS X; Other
Open-source	Yes
Object class	Word Processor; Documents; Spreadsheets
Input formats	Text Formats OpenDocument Text (*.odt) OpenOffice.org 1.0 Text (*.sxw) Rich Text Format (*.rtf) Microsoft Word (*.doc) WordPerfect (*.wpd) Plain Text (*.txt) HTML (*.html) Spreadsheet Formats OpenDocument Spreadsheet (*.ods) OpenOffice.org 1.0 Spreadsheet (*.sxc) Microsoft Excel (*.xls) Comma-Separated Values (*.csv) Tab-Separated Values (*.tsv) Presentation Formats OpenDocument Presentation (*.odp) OpenOffice.org 1.0 Presentation (*.sxi) Microsoft PowerPoint (*.ppt) Drawing Formats OpenDocument Drawing (*.odg)
Output formats	Text Formats Portable Document Format (*.pdf) OpenDocument Text (*.odt) OpenOffice.org 1.0 Text (*.sxw) Rich Text Format (*.rtf) Microsoft Word (*.doc) Plain Text (*.txt) HTML2 (*.html) MediaWiki wikitext (*.wiki) Spreadsheet Formats Portable Document Format (*.pdf) OpenDocument Spreadsheet (*.ods)

[19] http://openoffice.org
[20] http://www.artofsolving.com/opensource/jodconverter
[21] http://www.gnu.org/licenses/lgpl.html

	OpenOffice.org 1.0 Spreadsheet (*.sxc) Microsoft Excel (*.xls) Comma-Separated Values (*.csv) Tab-Separated Values (*.tsv) HTML (*.html) Presentation Formats Portable Document Format (*.pdf) Macromedia Flash (*.swf) OpenDocument Presentation (*.odp) OpenOffice.org 1.0 Presentation (*.sxi) Microsoft PowerPoint (*.ppt) HTML (*.html) Drawing Formats Scalable Vector Graphics (*.svg) Macromedia Flash (*.swf)

Tool	TeighaFileConverter[22]
Tool version	3.4.1
License	Comercial[23]
Interface	Command-line; GUI
Operating system	GNU/Linux; Windows; MacOS X
Open-source	No
Object class	CAD
Input formats	DWG, DXF
Output formats	DWG, DXF

Tool	Kakadu
Tool version	6.4.1
License	Kakadu SDK Evaluation/Commercial/Public/Non Commercial Licence
Interface	API
Operating system	GNU/Linux; Windows
Open-source	Yes
Object class	Bitmap images
Input formats	JP2, TIFF (if compiled against libtiff)
Output formats	JP2, TIFF (if compiled against libtiff)

Tool	Apache Sanselan[24]
Tool version	0.97
License	Apache 2
Interface	API
Operating system	GNU/Linux; Windows; Other

[22] http://www.opendesign.com/
[23] http://www.opendesign.com/join
[24] http://commons.apache.org/sanselan/

Open-source	Yes
Object class	Bitmap image
Input formats	PNG, GIF, TIFF, JPEG/JFIF, JPEG/JFIF EXIF Metadata, JPEG/JFIF IPTC Metadata, BMP, ico, PNM/PGM/PBM/PPM Portable Pixmap, PSD/Photoshop, XMP
Output formats	PNG, GIF, TIFF, JPEG/JFIF, JPEG/JFIF EXIF Metadata, BMP, PNM/PGM/PBM/PPM Portable Pixmap, XMP

Tool	OpenJPEG[25]
Tool version	1.4
License	BSD[26]
Interface	
Operating system	GNU/Linux; Windows; MacOS X
Open-source	Yes
Object class	Bitmap image
Input formats	J2K, J2C, JP2, JPT, bmp, tif, raw, tga, pnm, pgm, ppm, pgx
Output formats	j2k, jp2, j2c, BMP, TIF, RAW, TGA, PGM, PPM, PNM, PGX

Tool	SIARD Suite[27]
Tool version	
License	"SIARD Suite" Software License Agreement[28]
Interface	Command-line
Operating system	GNU/Linux; Windows; MacOS X
Open-source	No
Object class	Relational database
Input formats	Oracle, SQL Server, Microsoft Access
Output formats	SIARD

Tool	GIMP[29]
Tool version	2.6.11
License	GPL V3[30]
Interface	Command-line; GUI
Operating system	GNU/Linux; Windows; MacOS X
Open-source	Yes
Object class	Bitmap image

[25] http://www.openjpeg.org/
[26] http://www.openjpeg.org/BSDlicense.txt
[27] http://www.bar.admin.ch/dienstleistungen/00823/00825/index.html?lang=en
[28] http://www.bar.admin.ch/dienstleistungen/00823/00825/index.html?lang=en&download=M3wBPgDB/8ull6Du36WenojQ1NTTjaXZnqWfVp3Uhmfhnapmmc7Zi6rZnqCkkIN1gXt+bKbXrZ6lhuDZz8mMps2gpKfo
[29] http://www.gimp.org/
[30] http://www.gnu.org/licenses/gpl-3.0.html

Input formats	GIMP XCF, the native format (.xcf, or compressed as .xcf.gz or .xcf.bz2) GIMP brush (.gbr, .gpb, and animated as .gih) GIMP pattern (.pat) GIMP compressed XJT image (.xjt, .xjtgz, .xjtbz2) Alias | Wavefront PIX image (.pix, .matte, .mask, .alpha, .als) Scalable vector graphics for exporting paths (.svg) Autodesk flic animations (.fli) Digital Imaging + Communications in Medicine (.dcm or .dicom) PostScript documents (.ps, .eps, or compressed as .ps.gz) FITS astronomical images (.fits, or .fit) Scalable vector graphics for exporting paths (.svg) Microsoft Windows icon (.ico) Microsoft Uncompressed AVI Video (.avi) Windows bitmap (.bmp) Paintshop Pro image (.psp or .tub) Adobe Photoshop PSD (.ps) PNM image (.pnm, .ppm, .pgm, and .pbm) Compuserve GIF images and animations (.gif) JPEG photos (.jpeg, .jpg, or .jpe) PNG (.png) KISS CELL (.cel) Tagged Image File Format (.tiff or .tif) TARGA (.tga) Silicon Graphics IRIS image (.sgi, .rgb, .bw, .icon) Sun Rasterfile image (.im1, .im8, .im24, .im32, .rs, .ras) X bitmap image (.xpm, .icon, or .bitmap) X pixmap image (.xpm) X window dump (.xwd) Zsoft PCX (.pcx) PDF document (.pdf) G3 Fax (.g3) Windows WMF file (.wmf, .apm)
Output formats	GIMP XCF, the native format (.xcf, or compressed as .xcf.gz or .xcf.bz2) GIMP brush (.gbr, .gpb, and animated as .gih) GIMP pattern (.pat) GIMP compressed XJT image (.xjt, .xjtgz, .xjtbz2) Alias | Wavefront PIX image (.pix, .matte, .mask, .alpha, .als) Scalable vector graphics for exporting paths (.svg) Autodesk flic animations (.fli) Digital Imaging + Communications in Medicine (.dcm or .dicom) PostScript documents (.ps, .eps, or compressed as .ps.gz) FITS astronomical images (.fits, or .fit) Scalable vector graphics for exporting paths (.svg) Microsoft Windows icon (.ico) Microsoft Uncompressed AVI Video (.avi) Windows bitmap (.bmp) Paintshop Pro image (.psp or .tub) Adobe Photoshop PSD (.ps) PNM image (.pnm, .ppm, .pgm, and .pbm) Compuserve GIF images and animations (.gif) JPEG photos (.jpeg, .jpg, or .jpe) PNG (.png) KISS CELL (.cel)

	Tagged Image File Format (.tiff or .tif) TARGA (.tga) Silicon Graphics IRIS image (.sgi, .rgb, .bw, .icon) Sun Rasterfile image (.im1, .im8, .im24, .im32, .rs, .ras) X bitmap image (.xpm, .icon, or .bitmap) X pixmap image (.xpm) X window dump (.xwd) Zsoft PCX (.pcx) HTML as a table with coloured cells (.html) C source files as an array (.c or .h) MNG animations (like animated GIFs, but using PNGs) (.mng) gzip compressed image (.xcf, .gz.gz, .xcfgz) bzip compressed image (.xcf, bz2.bz2, .xcfbz2) ascii image (if aalib is present) (.txt, .ansi, .text) If Perl support is present the following: DATAURL (embed image into HTML, best looking of the embedded variety, works great in Firefox and Opera, but not IE) (.dataurl) COLORHTML, COLORXHTML (interesting but useless) (.colorhtml, .colorxhtml)

Tool	Inkscape[31]
Tool version	0.48.1
License	GPL v2[32]
Interface	Command-line; GUI
Operating system	GNU/Linux; Windows; MacOS X
Open-source	Yes
Object class	Bitmap image; Vector image
Input formats	Scalable Vector Graphics (*.svg) Compressed Inkscape SVG (*.svgz) Adobe PDF (*.pdf) Adobe Illustrator 9.0 and above (*.ai) Enhanced Metafiles (*.emf) Windows Metafiles (*.wmf) WordPerfect Graphics (*.wpg) GIMP Gradient (*.ggr) Adobe Illustrator SVG (*.ai.svg) Adobe Illustrator 8.0 and bellow (*.ai) Corel DRAW Compressed Exchange files (*.ccx) Corel DRAW 7-X4 files (*.cdr) Corel DRAW 7-13 template files (*.cdt) Computer Graphics Metafile files (*.cgm) Corel DRAW Presentation Exchange files (*.cmx) AutoCAD DXF R13 (*.dxf) HP Graphics Language Plot file (*.plt) sK1 vector graphics files (*.sk1) Microsoft XAML (*.xaml) ani (*.ani) png (*.png) wbmp (*.wbmp) pnm (*.pnm *.pbm *.pgm *.ppm)

[31] http://inkscape.org/
[32] http://www.gnu.org/licenses/gpl-2.0.html

	ras (*.ras) xpm (*.xpm) xbm (*.xbm) tga (*.tga *.targa) pcx (*.pcx) icns (*.icns) ico (*.ico) cur (*.cur) Aldus Placeable Metafiles (*.apm) bmp (*.bmp) gif (*.gif) jpeg (*.jpeg *.jpe *.jpg) tiff (*.tiff *.tif)
Output formats	Inkscape SVG (*.svg) Plain SVG (*.svg) Compressed Inkscape SVG (*.svgz) Compressed Plain SVG (*.svgz) Adobe PDF (*.pdf) Cairo PNG (*.png) PostScript (*.ps) Encapsulated PostScript (*.eps) Enhanced Metafiles (*.emf) PovRay (*.pov) (paths and shapes only) JavaFX (*.fx) OpenDocument drawing (*.odg) LaTeX With PS Tricks macros (*.tex) Desktop Cutting Plotter (R13) (*.dxf) GIMP Palette (*.pgl) HP Graphics Language file (*.hpgl) Jessylnk zipped pdf or png output (*.zip) HP Graphics Language Plot file [AutoCAD] (*.plt) Optimized SVG (*.svg) sK1 vector graphics file (*.sk1) Microsoft XAML (*.xaml) Compressed Inkscape SCG with media (*.zip) Windows Metafile (*.wmf)

Tool	Avidemux[33]
Tool version	2.5.4
License	GPL v2[34]
Interface	Command-line
Operating system	GNU/Linux; Windows
Open-source	Yes
Object class	Video
Input formats	AVI, OpenDML, MPEG, ASF, NuppelVideo, Images, H.263(+),

[33] http://fixounet.free.fr/avidemux/
[34] http://www.gnu.org/licenses/gpl-2.0.html

	MPEG-4, QuickTime, 3GP, MP4, OGM, Matroska, Flash Video Video decoders: Cinepak DV CRAM FFV1 H.263 H.264 (MPEG-4 AVC) HuffYUV MJPEG MPEG-1, MPEG-2 MPEG-4 SP/ASP (encoded e.g. with DivX, Xvid, FFmpeg MPEG-4 etc.) MS MPEG-4 (used in DivX ;-) 3.11 Alpha) Raw RGB Raw YV12 SVQ3/Sorenson Video Codec 3 VP3 VP6 WMV2, WMV3 (VC-1), both in ASF and AVI containers Y800 Audio decoders: MP3, MP2 A52 aka AC3 DTS aka DCA Vorbis AMR Narrow Band AAC
Output formats	AVI, MPEG video, MPEG PS A+V, MPEG TS A+V, OGM, MP4, Matroska, Flash Video Video encoders: libavcodec MPEG-1 (VCD) libavcodec MPEG-2 (DVD, SVCD) mpeg2enc MPEG-2 (DVD, SVCD) libavcodec FFHuffYUV libavcodec FFV1 libavcodec H.263 libavcodec H.263+ libavcodec HuffYUV libavcodec MJPEG libavcodec MPEG-4 libavcodec Snow x264

	Xvid Y800 Audio encoders: LAME, libvorbis, FAAC, FFmpeg MP2, FFmpeg AC3, TwoLAME, WAV PCM, WAV LPCM

Tool	PDFBox[35]
Tool version	1.5.0
License	Apache License v2[36]
Interface	API
Operating system	GNU/Linux; Windows; MacOS X; Other
Open-source	Yes
Object class	Other
Input formats	PDF, plain text
Output formats	PDF, plain text, image

Tool	JTidy[37]
Tool version	r938
License	Custom license[38]
Interface	Command-line; API
Operating system	GNU/Linux; Windows; MacOS X; Other
Open-source	Yes
Object class	Other
Input formats	HTML
Output formats	HTML

Tool	JasPer[39]
Tool version	1.900.1
License	JasPer License Version 2.0[40]
Interface	API
Operating system	GNU/Linux; Windows; MacOS X; Other
Open-source	Yes
Object class	Bitmap image

[35] http://pdfbox.apache.org/
[36] http://www.apache.org/licenses/LICENSE-2.0
[37] http://jtidy.sourceforge.net/
[38] http://jtidy.sourceforge.net/license.html
[39] http://www.ece.uvic.ca/~mdadams/jasper/
[40] http://www.ece.uvic.ca/~mdadams/jasper/LICENSE

Input formats	bmp Windows BMP jp2 JPEG-2000 JP2 jpc JPEG-2000 Code Stream jpg JPEG pgx PGX pnm PNM/PGM/PPM mif My Image Format ras Sun Rasterfile
Output formats	bmp Windows BMP jp2 JPEG-2000 JP2 jpc JPEG-2000 Code Stream jpg JPEG pgx PGX pnm PNM/PGM/PPM mif My Image Format ras Sun Rasterfile

Tool	Aware AccuRad J2KSuite
Tool version	3.19.0
License	Commercial
Interface	Command-line; API
Operating system	GNU/Linux; Windows; MacOS X
Open-source	No
Object class	Bitmap image
Input formats	JP2, J2K (JPEG 2000 code stream), JPG,PPM, PGM, PGX, BMP, TGA, TIFF, DICOM
Output formats	JP2, J2K (JPEG 2000 code stream),JPG, PPM, PGM, PGX, BMP, TGA, TIFF, DICOM, PDF, RAW

Tool	Luratech LuraWave Command Line Tool
Tool version	
License	Commercial
Interface	Command-line
Operating system	GNU/Linux; Windows
Open-source	No
Object class	Bitmap image
Input formats	TIFF, PNM (PPM+PGM), BMP, JPG, LWF (LuraWave file format),RAW, JP2, JPX, J2K (JPEG 2000 codestream)
Output formats	JP2, JPX, J2K (JPEG 2000 codestream), TIFF, PNM (PPM+PGM), BMP, JPG, RAW

Tool	Office Binary (doc, xls, ppt) Translator to Open XML[41] (b2xtranslator)
Tool version	Phase III Milestone 5
License	BSD-like[42]
Interface	Command-line
Operating system	GNU/Linux; Windows; MacOS X; Other
Open-source	Yes
Object class	Word Processor Document; Spreadsheet
Input formats	DOC, XLS, PPT
Output formats	DOCX, XLSX, PPTX

[41] http://b2xtranslator.sourceforge.net/
[42] http://b2xtranslator.sourceforge.net/index.html#license

Tool	warc-tools[43]
Tool version	
License	Unknown
Interface	Command-line
Operating system	GNU/Linux; Windows; MacOS X; Other
Open-source	Yes
Object class	Web archive
Input formats	ARC
Output formats	WARC

Tool	NedlibToARC[44]
Tool version	1.0
License	LGPL 2.1[45]
Interface	Command-line; API
Operating system	GNU/Linux; Windows; MacOS X; Other
Open-source	Yes
Object class	Web archive
Input formats	Nedlib
Output formats	ARC

Tool	Heritrix (org.archive.io)[46]
Tool version	1.14.4
License	LGPL 2.1[47]
Interface	API
Operating system	GNU/Linux; Windows; MacOS X; Other
Open-source	Yes
Object class	Web archive
Input formats	ARC,WARC
Output formats	ARC,WARC

Tool	Java Image I/O with JAI[48]
Tool version	Java SE 1.4+ (JAI 1.1)
License	Oracle JDK BCL[49]
Interface	API
Operating system	GNU/Linux; Windows; MacOS X; Other
Open-source	Yes
Object class	Bitmap image

[43] http://code.hanzoarchives.com/warc-tools
[44] http://nwatoolset.sourceforge.net/docs/NedlibToARC/
[45] http://www.gnu.org/licenses/lgpl-2.1.html
[46] http://crawler.archive.org/
[47] http://www.gnu.org/licenses/lgpl-2.1.html
[48] http://download.oracle.com/javase/6/docs/technotes/guides/imageio/index.html
[49] http://www.oracle.com/technetwork/java/javase/downloads/jdk-6u21-license-159167.txt

Input formats	JPEG, PNG, BMP, WBMP, GIF, FlashPix, PNM, TIFF, WBMP
Output formats	JPEG, PNG, BMP, WBMP, GIF, PNM, TIFF, WBMP

Tool	GraphicsMagick[50]
Tool version	1.3.12
License	MIT style license and other mixed open-source licenses[51]
Interface	Command-line
Operating system	GNU/Linux; Windows; MacOS X; Other
Open-source	Yes
Object class	Bitmap image
Input formats	ART, AVI, AVS, BMP, CALS, CGM, CMYK, CUR, CUT, DCM, DCX, DIB, DPX, EMF, EPDF, EPI, EPS, EPSF, EPSI, EPT, FAX, FIG, FITS, FPX, GIF, GPLT, GRAY, HPGL, HTML, ICO, JBIG, JNG, JP2, JPC, JPEG, MAN, MAT, MIFF, MONO, MNG, MPEG, M2V, MPC, MSL, MTV, MVG, OTB, P7, PALM, PAM, PBM, PCD, PCDS, PCX, PDB, PDF, PFA, PFB, PGM, PICON, PICT, PIX, PNG, PNM, PPM, PS, PS2, PS3, PSD, PTIF, PWP, RAS, RAD, RGB, RGBA, RLA, RLE, SCT, SFW, SGI, SUN, SVG, TGA, TIFF, TIM, TTF, TXT, UYVY, VICAR, VIFF, WBMP, WPG, XBM, XCF, XPM, XWD, YUV
Output formats	ART, AVS, BMP, CMYK, DCX, DIB, DPX, EPDF, EPI, EPS, EPS2, EPS3, EPSF, EPSI, EPT, FAX, FITS, FPX, GIF, GRAY, HTML, JBIG, JNG, JP2, JPC, JPEG, MAT, MIFF, MONO, MNG, MPEG, M2V, MPC, MSL, MTV, MVG, OTB, P7, PALM, PAM, PBM, PCD, PCDS, PCX, PDB, PDF, PGM, PICON, PICT, PNG, PNM, PPM, PS, PS2, PS3, PSD, PTIF, RGB, RGBA, SGI, SHTML, SUN, SVG, TGA, TIFF, TXT, UIL, UYVY, VICAR, VIFF, WBMP, XBM, XPM, XWD, YUV

Tool	ACDSee[52]
Tool version	Pro 4
License	Comercial
Interface	GUI
Operating system	Windows; MacOS X
Open-source	No
Object class	Bitmap images; Audio; Video
Input formats	3G2, 3GP, 3GP2, 3GPP, AAC, ABR, AC3, ADTS, AIF, AIFC, AIFF, AMC, AMR, ANI, APD, ARW, ASF, AU, AVI, BMP, BW, BWF, CAF, CDDA, CEL, CNV, CR2, CRW, CS1, CUR, DCR, DIF, DJVU, DNG, DVM, EMF, EPS, ERF, FLC, FLI, FPX, GIF, GSM, HDR, ICL, ICN, ICO, IFF, JBR, JFIF, JIF, JP2, JPC, JPE, JPEG, KAR, KDC, M15, M1A, M1V, M2A, M3U, M4A, M4B, M4P, M4V, M75, MEF, MIDI, MOV, MOS, MP2, MP2V, MPV, MPV2, MRW, NEF, NRW, ORF, PBM, PBR, PCD, PCT, PCX, PEF, PGM, PIC, PICS, PICT, PIX, PNG, PPM, PSD, PSPBRUSH, PSPIMAGE, QCP, QT, QTPF, RAF, RAS, RAW, RGB, RGBA, RSB, RW2, RWL, SFIL, SGI, SMI, SMIL, SML, SR2, SRF, SRW, SWA, TGA, THM, TIF, TIFF, TTC, TTF, ULW, V40PO, V40PP, V40PPF, VFW, WBM, WAV, WBMP, WMF, WMV, XBM, XIF, XMP
Output formats	BMP, GIF, IFF, JP2, JPEG, PCX, PNG, PSD, RAS, RSB, SGI, TGA, TIFF, WBMP

Tool	CZ-Doc2Pdf[53]
Tool version	2.0
License	Comercial
Interface	GUI
Operating system	Windows
Open-source	No

[50] http://www.graphicsmagick.org
[51] http://www.graphicsmagick.org/Copyright.html
[52] http://dioscuri.sourceforge.net/
[53] http://www.convertzone.com/doc2pdf/help.htm

Object class	Word processing documents
Input formats	DOC, HTML, TXT, RTF
Output formats	PDF

Tool	Dia[54]
Tool version	0.96.1
License	GPL v2[55]
Interface	Command-line; GUI
Operating system	GNU/Linux; Windows
Open-source	Yes
Object class	Vector image
Input formats	XML
Output formats	XML, EPS, SVG, XFIG, WMF, PNG

Tool	Document2PDF Pilot[56]
Tool version	2.16.108
License	Comercial
Interface	GUI
Operating system	Windows
Open-source	No
Object class	Word processing documents
Input formats	TXT, RTF, HTML, SHTML, CHM, DOC, MCW, XLS, XLW, WRI, WPS, WPT
Output formats	PDF

Tool	EscapeE[57]
Tool version	9.10
License	Comercial
Interface	Command-line; GUI
Operating system	Windows
Open-source	No
Object class	Word processing documents
Input formats	PCL, PDF, PS, TIFF, DCX, RTF
Output formats	PDF

Tool	GraphicConverter X[58]
Tool version	7
License	Comercial
Interface	API, GUI
Operating system	Mac OS X

[54] http://projects.gnome.org/dia/
[55] http://www.gnu.org/licenses/old-licenses/gpl-2.0.html
[56] http://www.colorpilot.com/document2pdf.html
[57] http://escapee.redtitan.fr
[58] http://www.lemkesoft.com/content/188/graphicconverter.html

Open-source	No
Object class	Bitmap Images
Input formats	.D , 8BIM, Acorn Sprite, AFP, ALIAS, AMBER ARR, ANI, ANI, ANPA, Apple Preferred, ART, ASCII, Atari Portfolio, B3D, BioRad, BLD, BMP, BUF, BUM, c4, CAM, CALS, CEL, CGM, CRW, CR2, csource, , CLP, CT, CVG, DCR, DCX, DDS, DeltaVision, DESR VFF, DICoM, DJ100DL, Doodle, DPX, Dr. Halo, DrawIT, ECW, ElectronicImage, EPSF, ESM, FAX, FaxSTF, FireViewer PDB, FITS, FLH/FLI/FLC, FPX, Freehand 1FUJI, GATAN, GEM, GFX, GIF, GrayPaint, GRP, HAM, HDR, HP-GL/2, IBM - PIC, IC ? - Imagic, ICo/ICN, IFF/LBM, IM, , ImageLab/PrintTechnic, IMG/XIMG, IMQ, IPLab, IRIS, ISS, j6i, JBI, JIF, JPEG/JFIF, JPEG 200KDC, KISS CEL, Koala, KoNTRoN, LDF, Lotus-PIC, LWF (LuraWave), MacDraw I, MacPaint, MAG, MAYA-IFF, MBM (Psion 5), Meteosat 5, MDC, MHT, MonkeyLogo, MonkeyCard, Mov, MRC, MrSID, MRW, MSP, MSX - MSX2, MSX+ - MSX2+, NASA Raster Metafile, NEF, NEO, NGG/NCG, NIF, NoL, oNCoR, ORF, OTA, P ? ? - Degas, , PAC - STAD, PaperPort (MAX), PBM/PGM/PPM, PCD, PCX, PIC (32K), PIC, PIC, PIC, PICS, PICT, PICT resource, PDB, PDF, PEF, PGPF, PGC/PGF, Photo Raw, PM, PNG, PoRST, Ppat, PSD, PSP, QDV, QNT, QTIF, RAF, RAW, RDC, RIFF, RLA, RLE, RSB, , SCR, SCX, SFF, SFW, SGI, SHP, Sinclair QL, SIXEL, SKETCH, SMV, SNX, SoFTIMAGE, SPC, SRF, ST - X - SBIG, StartupScreen, SUN, Super-Hi-Res 320SVG, TBC, TCL, TealPaint PDB, TGA, TIFF, TIM PSX, Tiny Viewer PDB, TN ?, TRS-8VBM, VFF, VGS-8, VITRoNIC, VoXEL, VPB, VPM, WBIN, WBMP, Winfax (FXM), WMF, WPG, X11, X3F, XBM, XCF, XFIG, X-Face, XPM, XWD, YUV
Output formats	ASCII, Atari Portfolio, BLD, BMP, BUM, CEL, csource, System 7 clip, DDS, ECW, ElectronicImage, EPSF, FITS, GIF, HP-GL/2, ICo/ICN, IFF/LBM, Image ->Finder, IMG/XIMG, JIF, JPEG/JFIF, JPEG 200, KISS CEL, LWF (LuraWave), MacPaint, MBM (Psion 5), MonkeyLogo, MonkeyCard, Mov, NGG/NCG, NoL, OTA, Palette, PBM/PGM/PPM, PCX, PIC, PICS, PICT, PICT resource, PDB, PDF, PGPF, PNG, Ppat, PSD, RAW, RSB, RTF, SCR, SFW, SoFTIMAGE, StartupScreen, SVG, TGA, TIFF, TRS-8, VPB, WBIN, WBMP, WMF, XBM, XCF, XFIG, X-Face, XPM, XWD

Tool	MsgText[59]
Tool version	1.0.0
License	GNU GPL
Interface	Command-line
Operating system	Windows
Open-source	Yes
Object class	Email
Input formats	MSG
Output formats	TXT, HTML (and any attachments)

Tool	tesseract-ocr[60]
Tool version	3.00
License	Apache license 2.0[61]
Interface	Command-line
Operating system	GNU/Linux; Windows; MacOS X
Open-source	Yes
Object class	Bitmap image
Input formats	TIFF, PNG, JPEG
Output formats	TXT

[59] http://www.enterag.ch/enterag/downloads/msgtext.xhtml
[60] http://code.google.com/p/tesseract-ocr/
[61] http://www.apache.org/licenses/LICENSE-2.0

Tool	AbiWord[62]
Tool version	2.8.6
License	GPL[63]
Interface	Command-line; GUI
Operating system	GNU/Linux; Windows; MacOS X
Open-source	Yes
Object class	Word Processor Document
Input formats	AbiWord Documents (.abw, .awt, .azbw) Microsoft Word (.doc, .dot) Rich Text Format (.rtf) Text (.txt, .text) Encoded Test (.txt, .text) HTML (.html, .htm, .xhtml) Applix Word (.aw) ClarisWorks/AppleWorks 5 (.cwk) Collaborative File Descriptor (.abicollab) DocBook (.dbk, .xml) ISCII Text (.isc, .iscii) OpenDocument (.odt, .ott) OpenOffice Writer (.stw, .sxw) Office Open XML (.docx, .dotx, .docm, .dotm) OPML (.opml) StarWriter up to 5.x (.sdw) T602 (.602, .txt) WML (.wml) WordPerfect (.wpd, .wp) XSL-FO (.fo)
Output formats	AbiWord (.abw, .zabw, .abw.gz) AbiWord Template (.awt) Microsoft Word (.doc) HTML/XHTML (.html) Multipart HTML (.mht) Rich Text Format (.rtf) Rich Text Format for old apps (.rtf) Text (.txt, .text) Encoded Test (.txt, .text) Applix Word (.aw) DocBook (.dbk, .xml) Outlook Express Email (.eml) Newsgroup Formatted Text (.nws) ISCII Text (.isc, .iscii) LaTeX (.latex) OpenDocument (.odt) OpenOffice Writer (.sxw) Office Open XML (.docx) WML (.wml) XSL-FO (.fo)

Tool	VisualIntegrity[64]

[62] http://www.abisource.com/
[63] http://www.fsf.org/copyleft/gpl.html
[64] http://www.flysdk.com/

Tool version	8.0
License	Commercial[65]
Interface	Command-line
Operating system	GNU/Linux; Windows; MacOS X
Open-source	No
Object class	Word Processor Document
Input formats	PDF
Output formats	JPEG, GIF, PNG, BMP, WMF, EMF, SVG, TIFF, TXT, DXF

Tool	Photoshop[66]
Tool version	8.0
License	Commercial[67]
Interface	GUI
Operating system	Windows; MacOS X
Open-source	No
Object class	Bitmap image
Input formats	3D Studio Max (.3DS) Adobe Illustrator (.ai) Alias PIX* (.pix) Amiga IFF* (.iff, .tdi) AVI (.avi) BMP (.bmp, .rle, .dib) Camera Raw (.tif, .crw, .nef, .raf, .orf, .mrw, .dcr, .mos, .raw, .pef, .srf, .dng, .x3f, .cr2, .erf, .sr2, .kdc, .mfw, .mef, .arw, .srw) Cineon (.cin, .spdx, .dpx, .fido) Collada (.dae) CompuServe GIF (.gif) Dicom (.dcm, .dc3, .dic) Digital Negative (.dng) ElectricImage* (.img, .ei, .eiz, .eizz) EPS (.eps, .epsf, .epsp) EPS with JPEG Preview (.eps) EPS with PICT Preview (.eps) EPS with TIFF Preview (.eps) Filmstrip (.flm) Google Earth 4 (.kmz) JPEG (.jpg, .jpeg, .jpe) JPEG 2000* (.jpf, .jpx, .jp2, .j2c, .j2k, .jpc) Kodak Photo CD (.pcd) Large Document Format (.psb) MacPaint* (.mpt, .mac) OpenEXR (.exr) PCX (.pcx) PDF (.pdf, .pdp) Photoshop (.psd, .pdd) Photoshop DCS 1.0 and 2.0 (.eps) Photoshop Raw (.raw) PICT (.pct, .pict)

[65] http://www.flysdk.com/fly-sdk-01.htm
[66] http://www.photoshop.com/
[67] http://www.adobe.com/products/eulas/pdfs/gen_wwcombined_20091001_1604.pdf

	Pixar (.pxr) PixelPaint* (.px1) PNG (.png) Portable Bitmap (.pbm, .pgm, .ppm, .pnm, pfm, .pam) QuickTime Movie** (.mov, .avi, .mpg, .mpeg, .mp4, .m4v) Radiance (.hdr, .rgbe, .xyze) Scitex CT (.sct) SGI RGB* (.sgi, .rgb, .rgba, .bw) SoftImage* (.pic) Targa (.tga, .vda, .icb, .vst) TIFF (.tif) U3D (.u3d) Wavefront|OBJ (.obj) Wavefront RLA* (.rla) Wireless Bitmap (.wbm,.wbmp)
Output formats	Adobe Illustrator Paths (.ai) Alias PIX* (.pix) Amiga IFF* (.iff, .tdi) AVI (.avi) BMP (.bmp, .rle, .dib) Cineon (.cin, .spdx, .dpx, .fido) CompuServe GIF (.gif) Dicom (.dcm, .dc3, .dic) Digital Negative (.dng) ElectricImage* (.img, .ei, .eiz, .eizz) EPS (.eps, .epsf, .epsp) EPS with TIFF Preview (.eps) Filmstrip (.flm) Google Earth 4 (.kmz) JPEG (.jpg, .jpeg, .jpe) JPEG 2000* (.jpf, .jpx, .jp2, .j2c, .j2k, .jpc) Large Document Format (.psb) OpenEXR (.exr) PCX (.pcx) PDF (.pdf, .pdp) Photoshop (.psd, .pdd) Photoshop DCS 1.0 and 2.0 (.eps) Photoshop Raw (.raw) PICT (.pct, .pict) Pixar (.pxr) PNG (.png) Portable Bitmap (.pbm, .pgm, .ppm, .pnm, pfm, .pam) QuickTime Movie** (.mov, .avi, .mpg, .mpeg, .mp4, .m4v) Radiance (.hdr, .rgbe, .xyze) Scitex CT (.sct) SGI RGB* (.sgi, .rgb, .rgba, .bw) SoftImage* (.pic) Targa (.tga, .vda, .icb, .vst) TIFF (.tif) Wavefront|OBJ (.obj) Wavefront RLA* (.rla) Wireless Bitmap (.wbm,.wbmp) ZoomView (.mtx)

Tool	PDF Version converter[68]
Tool version	2.01
License	Commercial[69]
Interface	Command-line; GUI
Operating system	Windows
Open-source	No
Object class	Word Processor Document
Input formats	PDF (1.0 to 1.7)
Output formats	PDF (1.0 to 1.7)

Tool	Netpbm[70]
Tool version	10.35.80
License	Artistic license[71], GPL[72]
Interface	Command-line
Operating system	GNU/Linux; Windows; MacOS X; Other
Open-source	Yes
Object class	Bitmap image
Input formats	Portable Bitmap File Format (PBM), Portable Graymap File Format (PGM), Portable Pixmap File Format (PPM), Portable Arbitrary map File Format (PAM), JPEG, BMP, ICO, GIF, PNG, Palm pixmap (PALM), JBIG BIE (compressed bitmap), Fiasco highly compressed format (WFA), Photo CD (HPCD), Wireless Bitmap (WBMP), Atari Neochrome (NEO), Microdesign (for Amstrad PCWs) (MDA), Andrew Toolkit raster object (ATK), Xerox doodle brushes (BRUSH), CMU Window Manager format (CMUWM), Group 3 FAX (G3), Sun icon (ICON), GEM .img format (GEM), MacPaint (MACP), MGR format (MGR), Atari Degas .pi3 (PI3), X10 or X11 bitmap (XBM), Bennet Yee "Face" (YBM), ASCII character graphic (TXT), HP PaintJet file (PJ),

[68] http://www.nicepdf.com/products.html
[69] http://www.nicepdf.com/purchase.html
[70] http://netpbm.sourceforge.net/
[71] http://www.opensource.org/licenses/artistic-license-1.0
[72] http://www.gnu.org/licenses/gpl.html

	HP Thinkjet printer stream (thinkjet), FITS format (FITZ), USENIX FaceSaver(tm) format (FS), HIPS format (HIPS), Lisp Machine bitmap (LISPM), Postscript (PS), Postscript "image" data (PSID), RAW grayscale bytes (RAW), Gould scanner file (GOULD), IFF ILBM (ILBM), Img-whatnot (IMG), MTV ray-tracer output (MTV), PC Paintbrush format (PCX), Atari Degas .pi1 (PI1), Macintosh PICT (PICT), QRT ray-tracer output (QRT), RAY RGB bytes (RAW), AutoCAD slide file (SLD), Atari compressed Spectrum (SPC), Atari uncompressed Spectrum (SPU), TrueVision Targa file (TGA), Xim (XIM), XPM format (XPM), Abekas YUV format (YUV), Encoder/Berkeley YUV format (EYUV), 3 subsampled ray YUV files (YUVSPLIT), Sun raster file (RAST), TIFF file (TIFF), X10 or X11 window dump (XWD), 411 (Sony Mavica) (411), Santa Barbara Instrument Group CCD (SBIG), Parallax XVideo JPEG (VID), UTAH Raster Toolkit (urt/rle) (RLE), Interleaf (LEAF), Biorad confocal image (BIORAD), Packed format font (PK)
Output formats	Portable Bitmap File Format (PBM), Portable Graymap File Format (PGM), Portable Pixmap File Format (PPM), Portable Arbitrary map File Format (PAM), MPEG, JPEG, BMP, ICO, GIF, PNG, Palm pixmap (PALM), JBIG BIE (compressed bitmap), Fiasco highly compressed format (WFA), Nokia Smart Messaging Format (SMF), Wireless Bitmap (WBMP), Atari Neochrome (NEO), Microdesign (for Amstrad PCWs) (MDA), Andrew Toolkit raster object (ATK), CMU Window Manager format (CMUWM), Group 3 FAX (G3), Sun icon (ICON),

	MacPaint (MACP), MGR format (MGR), Atari Degas .pi3 (PI3), X11 bitmap (XBM), X10 bitmap (X10BM), Bennet Yee "Face" (YBM), Gemini 10x printer graphics (10X), ASCII Graphic form (TXT), BBN BitGraphic graphics (BBNBG), Epson printer graphics (EPSON), GEM .img file (GEM), GraphOn graphics (GO), HP LaserJet black and white graphics (LJ), HP LaserJet color graphics (PCL) (LJ), HP PaintJet file (PJ), UNIX plot (PLOT), Printronix graphics (PTX), Zinc interface library icon (ZINC), FITS format (FITZ), USENIX FaceSaver(tm) format (FS), HIPS format (HIPS), Lisp Machine bitmap (LISPM), PostScript using lines (LPS), Encapsulated Postscript preview bitmap (EPSI), IFF ILBM (ILBM), Atari Degas .pi1 (PI1), Macintosh PICT (PICT), TrueVision Targa file (TGA), XPM format (XPM), Abekas YUV format (YUV), Encoder/Berkeley YUV format (EYUV), 3 subsampled ray YUV files (YUVSPLIT), AutoCAD database or slide (ACAD), NCSA ICR graphics (ICR), PC Paintbrush format (PCX), X11 "puzzle" file (PUZZ), Sun raster file (RAST), TIFF RGB file (TIFF), TIFF CMYK file (TIFFCMYK), X11 window dump (XWD), DEC sixel format (SIXEL), Motif UIL icon (UIL), UTAH Raster Toolkit (urt/rle) (RLE), Interleaf (LEAF), Dec LN03+ Sixel image (LN03), Packed format font (PK)