PDFlib

Bavaria Report on PDF/A Validation Accuracy

The Bavaria[1] report reviews the precision of PDF/A-1 validation tools

This report provides a comparative review of PDF/A-1b validators. It assesses the output of PDF/A validation tools for conforming as well as for non-conforming documents. Our goal is to provide guidelines for users and to encourage vendors of PDF/A validation tools to optimize their products. We hope that increased accuracy of PDF/A validation will give users more precise means for checking their documents.

Evolution of PDF/A validation

The ISO 19005-1 standard for PDF/A-1 has been published in 2005. Early PDF/A adopters had a hard time regarding validation. Acrobat 7.0 introduced PDF/A validation in its Preflight plugin, but it predated the final standard and implemented only an earlier draft version which differed in important details from the final standard. Acrobat 7.07 incorporates modifications which reflect the final standard. As a consequence, documents created earlier were no longer considered valid. Acrobat 8 increased the breadth and depth of PDF/A validation checks and uncovered a lot of standard violations which haven't been noticed earlier. However, it did not thoroughly check XMP metadata and predated the TechNotes published by the PDF/A Competence Center. These TechNotes brought much needed clarifications in a lot of areas. The TechNotes and the Isartor test suite (which has been published in summer 2008) contributed significantly to a consolidation process regarding PDF/A validation. Acrobat 9 implements the majority of those clarifications, with a few important corrections and additions in Acrobat 9.1.

While we used Acrobat to outline the history of PDF/A validation, several third-party validators have had a similar history of increasingly stricter checks. This emphasis on providing more and stricter test criteria made users wonder whether PDF/A validation will ever stabilize or whether they'll be forced to constantly adjust their documents to the latest validation technology.

After conducting the Bavaria tests we can certainly say that the status of PDF/A validation is considerably better than in the early days, and the Isartor test suite and PDF/A TechNotes have significantly influenced the quality of validation. While the Bavaria report reveals shortcomings in the current product generation, we hope to contribute to further improvements and to help vendors of validation tools to enhance the accuracy of their products. This will ultimately increase the acceptance of PDF/A as a reliable standard for long-term document preservation.

PDFlib's motivation

PDFlib GmbH is interested in PDF/A validation tools for several reasons:

- As a member of the PDF/A Competence Center we are involved in a lot of technical and non-technical discussions regarding PDF/A conformance. The availability of reliable PDF/A validation tools is a crucial factor for the success of PDF/A. As co-authors of the Isartor test suite for PDF/A-1 we want to assess the effect of this test suite on validation tools.
- We use PDF/A validators to cross-check our own PDF/A output. As part of our internal QA process we regularly run PDF/A documents created with our own software through PDF/A validators to make sure that we don't miss any aspect of the standard. This process requires reliable validation tools, so we had to analyze a lot of documents and messages in order to make sure that our PDF output fully conforms to the standard.

1. In case you wonder about the name of this report: PDFlib GmbH's offices are located in Munich. From our offices you can see the Bavaria statue which is close to the area of the famous Oktoberfest.

- We often receive customer inquiries regarding PDF/A validation tools. Although we do not provide direct recommendations, we want to offer a solid basis for selecting a suitable tool.
- Some of our products (e.g. PDFlib+PDI 7, PLOP 4, PLOP DS 4) offer PDF/A-aware processing. This means that a processing step maintains PDF/A conformance: if the input conforms to PDF/A it is guaranteed that the output also conforms to PDF/A. While our products offer PDF/A-compliant features such as XMP metadata injection, digital signature, or page recombination they do not include full-blown PDF/A validation. Instead, they assume conforming input. Our customers need reliable PDF/A validation tools in order to check if this is true.

Therefore it is in the best interest of PDFlib GmbH and its customers to provide a comprehensive and balanced review of PDF/A validation tools.

Preparing the tests

As a basis for our tests we needed a set of »known good« conforming PDF/A documents. It took a lot of effort to classify real-world PDF/A documents. We did our best and ran multiple checks to make sure that no error went unnoticed. To the best of our knowledge, the accompanying set of test documents contains fully conforming PDF/A documents and can therefore be used as the basis for additional validation tests.

We tested only products which claim to support the Isartor test suite for PDF/A. The tool vendors were invited to comment our results before publication.

Disclosure

PDFlib GmbH is one of the founding members of the PDF/A Competence Center and was actively involved in the design and implementation of the Isartor test suite for PDF/A-1. PDFlib GmbH offers products for generating and processing PDF/A documents, but we do not currently offer any tools for validating PDF/A or for converting PDF documents to PDF/A.

Error Types and Rating Rules

The Bavaria test suite (which references the Isartor test suite produced by the PDF/A Competence Center) contains standard-conforming documents as well as documents which violate the standard. Based on the conformance status and the output of a validator we used the following distinction:

- False alarm: The validator wrongly rejects a conforming document. This is bad for those who generate the documents since they are asked to solve non-existing problems in their creation process, or have to argue the correctness of the validation result.
- Missing alarm: The validator wrongly accepts a non-conforming document because it overlooks one or more problems. This defeats the basic goals of PDF/A since documents which are not suited for long-term preservation will be archived without noticing any problem.
- All error messages wrong: The validator rejects a non-conforming document, but all messages on which its rejection is based are inappropriate. While rejecting the document is the proper action in this case, this type of error is critical nevertheless since the document producer will face wrong charges, and the wrong messages obscure the real problem.
- No result: the validator did not provide any usable output (e.g. crash, unreadable validation report, XML report not well-formed, etc.). This type of error obviously interrupts processing.

All other situations are classified as OK: the validator provides the correct decision. As long as the validator properly reports at least one real problem in a non-conforming document we will count its behavior as correct. For conforming documents the behavior is correct if no error is reported. Warnings are acceptable; we did not judge whether or not warnings are appropriate.

We will classify the behavior of PDF/A validators according to the four error classes described above. However, we will not take into account the following deviations from the ideal behavior:

- Some error messages missing: The validator detects some, but not all errors in a non-conforming document.
- Some error messages wrong: The validator correctly rejected a non-conforming document, but some of the error messages were wrong.

Accuracy of a validator

We call the ratio of OK tests (i.e. tests where the result does not fall in one of the four error classes described above) and the total number of tests the accuracy of a validator. The accuracy describes its ability to correctly distinguish conforming from non-conforming documents.

Our product-specific summary tables contain accuracy values for subsets of the test suite; these subsets are called categories. We do not calculate an overall accuracy value for each product since the relative weight of the categories is highly domain-specific. For example, an application without any metadata requirements may ignore XMP-related problems in the validator.

Aspects beyond the scope of this report

The following aspects are not addressed by the Bavaria report:

- PDF/A-1a validation has not been tested, but only PDF/A-1b. While PDF/A-1a can be considered the more advanced flavor of PDF/A-1, it requires the use of Tagged PDF and structure information which is not available in all workflows. PDF/A-1b is much more common than PDF/A-1a, and there is much larger software support available for 1b than for 1a. Another reason is that a test suite and clarifying TechNotes are currently only available for PDF/A-1b, but not for PDF/A-1a.
- PDF/A correction, i.e. converting non-conforming PDF documents to conforming PDF/A documents has not been tested.
- Other deployment issues, e.g. performance and memory consumption, are not a main target of this report, although it includes some notes on timing and robustness.

Although we focus on PDF/A validation accuracy, we realize that the items above are important questions which should be kept in mind when selecting a PDF/A validation tool.

The Bavaria Test Suite

Suitable test documents are the basis of PDF/A validation testing. In this section we describe our set of test documents, called the Bavaria test suite. The test documents are available as a separate package along with this report.

Non-conforming documents from the Isartor test suite

In order to get started with non-conforming test documents we used the Isartor test suite published by the PDF/A Competence Center in August 2008. This test suite contains more than 200 non-conforming PDF/A documents which can be used to check whether a validator finds all possible violations of PDF/A rules. Isartor has been designed in a systematic way along the requirements of PDF/A-1b and systematically violates all rules which are spelled out in the standard. Documents in the Isartor test suite are as simple as possible: in most cases they are single-page documents with very few page contents, where each document contains exactly one violation of the PDF/A standard if possible; in a few cases a problem can only be triggered in combination with others. Isartor has two main limitations (in addition to covering only PDF/A-1b and not PDF/A-1a):

- As a result of its design, it doesn't contain documents with combinations of problems. For example, a validator might detect a problem in an isolated test, but it may overlook the same problem in a larger test document because it is obscured by another problem or hidden in more complex PDF pages.
- Isartor violates only those rules which are explicitly spelled out in the PDF/A-1 standard. However, the standard refers to the PDF 1.4 reference and XMP 2004 specification documents which contain a very large number of additional requirements. These are not checked in Isartor.

Isartor can be used to find out whether all aspects of standard conformance testing are implemented at least in some way. Due to its design it is not sufficient for comprehensive PDF/A validation testing, but provides a solid starting point. Validators which have not been tested against Isartor during their development phase most probably don't match the quality of Isartor-proven products.

As an aid in judging the validation results we grouped the 204 Isartor files in categories such as file structure, fonts, metadata, etc. These categories are derived from the ISO 19005 document and may help in determining the areas where a particular validator has its strengths and weaknesses. The validation results are grouped by category.

Additional non-conforming real-world documents

In order to complement the Isartor test suite and compensate for the limitations mentioned above, we added other non-conforming PDF/A documents to the set of Bavaria test files. Unlike the artificial Isartor test files, this category contains real-world documents with a mixture of text, graphics, metadata, and other elements. Also unlike Isartor, these non-conforming real-world documents contain many PDF constructs in combination. The non-conforming documents have been classified as follows:

- ISO 19005-1 violations, i.e. one or more requirements of the PDF/A standard itself are not met. This category includes problems with XMP extension schemas since these are defined in the PDF/A standard.
- XMP 2004 violations, i.e. predefined XMP properties are not used according to the XMP 2004 specification which is referenced by PDF/A. XMP syntax problems also fall in this category.
- PDF 1.4 violations, i.e. some requirements of the PDF reference are not met. This includes various implementation limits which are documented in the PDF 1.4 reference and must adhered to by conforming PDF/A documents. While blatant violations could result in unusable files (e.g. a page cannot be displayed), there are many subtle requirements in PDF 1.4 which must be met for PDF/A conformance and should therefore be checked by a validator.

Unlike PDF/A violations in the Isartor test suite, XMP 2004 and PDF 1.4 violations are not tested in any systematic way in the Bavaria collection; we simply classified non-conforming real-world documents according the categories described above.

Obviously, a document may contain multiple errors in any combination of the categories above. In some cases we modified a real-world document in order to reduce the number of PDF/A violations. For example, a document contained a very subtle problem in combination with a very common problem. In order to isolate the subtle problem we fixed the second one.

Conforming PDF/A documents

Since the variations of conforming PDF/A documents are infinite, and constructing even a minimal set of test files which contains every conceivable PDF construct (not to mention combinations of those constructs) is a considerable challenge, we again used real-world PDF/A documents for our testing. In order to check as many variations of standard-conforming PDF/A documents as possible, we tried to get hold of documents from different application domains such as academic publishing, transactional documents, scans, presentation slides, etc. The files in our test collection have been created with a variety of PDF producer software which makes validation more interesting (in a sense).

In situations where we couldn't get hold of a publishable real-world sample (e.g. because of confidentiality requirements) we constructed artificial documents if this was necessary to model a particular aspect of PDF/A validation, or to represent a certain PDF construct in the test suite.

In some cases documents have been modified to internally flag PDF/A-1b instead of PDF/A-1a (i.e. we changed the PDF/A conformance identification in the document). This was necessary because some widely used products support only PDF/A-1a generation and we wanted to include sample output because it contains interesting constructs. Since PDF/A-1a is a superset of PDF/A-1b except for the conformance identification entry, this modification doesn't affect our test results, but allows us to include these documents in the PDF/A-1b validation test.

The conforming PDF/A test documents fall in the following categories:

- Real-world documents, mostly collected on the Web.
- Test documents created with software from PDFlib GmbH. These documents contain various constructs regarding fonts, color, and other aspects which are relevant for PDF/A. Although there is nothing special about these documents, we wanted to distinguish our in-house test material from real-world documents.
- XMP test documents: these have also been created with PDFlib GmbH software. While the main section of these documents is trivial (a single page with small amounts of text), they contain XMP metadata which exercises many of the more interesting facets of PDF/A rules for XMP. With this group we wanted to check the support for XMP validation in a more systematic way.

Unclear clauses in the standard

Unfortunately, there are still a few aspects of the PDF/A-1 standard which require clarification since the standard is not as clear as we would like it to be. In certain situations a validator may be more strict or more liberal than others with respect to these unclear clauses. We realize that this is a very unfortunate situation which should be rectified as soon as possible, e.g. with additional TechNotes or another Technical Corrigendum issued by ISO. For the time being, we don't have the authority to decide whether or not a validation result for these clauses is correct or not.

In order to provide at least some guidelines to users and hints for vendors we included sample documents for several unclear clauses. However, for lack of authoritative sources we do not take any position regarding the correctness of the result. While we state our own opinion, it is important for readers to understand that a few topics are currently undecided because the corresponding clause in the PDF/A standard requires clarification.

For this reason we summarize only the results of each product for unclear cases, but do not calculate any accuracy values for documents in the »unclear« category.

Representative documents or not?

Whether or not the Bavaria test suite can be considered representative is hard to say since it depends on the typical PDF producer software, your application requirements, and other factors. In assembling the collection of test files we tried not to overrepresent certain kinds of errors, but rather tried to cover a broad spectrum of characteristics (conforming PDF constructs as well as standard violations). One aspect in the document selection was whether or not a test document would help us in finding differences in the tested products. When judging the accuracy values also keep in mind that the absolute numbers of test documents per category is rather small.

Summary of test documents

The following table lists the number of documents in the Isartor and Bavaria test suites along with some interesting properties. A detailed listing of test documents along with comments regarding their PDF/A conformance is available in the test suite package.

Test suite	number of documents	total size	total number of pages	total number of messages produced by 7 validators
Isartor (included in Bavaria)	204	8 MB	10.203 (all Isartor test files contain a single page, except for one test with 10.000 pages)	2.759
Bavaria (in addition to Isartor)	85	50 MB	1830	22.816

When is a Document considered Standard-conforming?

A crucial question, of course, is how can we determine whether or not a document conforms to the PDF/A-1 standard while we are still in the process of testing the validators? First of all, let's clarify our basis for standard conformance (see references at the end of this report):

- ISO 19005-1 including Technical Corrigendum 1, i.e. the actual PDF/A standard;
- XMP 2004 specification which is referenced in the PDF/A standard;
- PDF 1.4 reference which is also referenced in the PDF/A standard;
- TechNotes published by the PDF/A Competence Center before the end of 2008.

Note that we do not fully take into account the PDF/A- 1 Application Notes published by AIIM. While the Application Notes contain a lot of useful material, they haven't been updated since January 2007 and unlike the TechNotes have not been discussed within the community of PDF/A product vendors.

Equipped with these reference documents we engage in the following procedure in order to determine whether or not a document conforms to the standard:

- If all validators accept the document as valid, it is assumed to conform to the standard unless we identify an obvious problem upon manual inspection of the document.
- If one or more validators reject the document we thoroughly check the aspects mentioned in the corresponding error messages. This usually involves the use of suitable tools for PDF analysis and debugging, including free and commercial Acrobat plugins, home-made tools, hex editors and other software.
- If we cannot detect any standard violation based on the validation reports and our own analysis we contact the vendor to ask for explanation.
- If the vendor couldn't explain to us why the file wouldn't conform to the standard we will still treat it as standard-conforming.

In other words: we consider documents as »conforming« if either none of the validators finds a problem, or the problems found by a validator cannot be located and understood based on our experience with the PDF/A standard. This definition implies that the classification of a document may have to be adjusted in the future. However, we tried to minimize the risk of wrong classifications. Based on our PDF development experience we scrutinized the documents more thoroughly than it would be possible for end users.

Testing Methodology

All tested PDF/A validators provide more or less detailed validation results in different output formats: some create plain text output, others generate comprehensive machine-readable XML reports, and still others present their results in human-readable HTML pages. While the reports sometimes exposed secondary problems (e.g. the report doesn't contain enough detail to analyze a problem or special characters are not correctly expressed in XML), our challenge was generating, collecting, assessing, and grouping thousands of messages generated by the validators. We devised a framework which automates the following tasks:

- Assemble a list of test files with the expected validation result and auxiliary information such as comments about interesting aspects of a test document. The XML input file with this information is called the Bavaria list. It is included in the test suite package.

- Run the validator against all documents in the Bavaria list. In order to work around crashes of some validators we had to include information in the Bavaria list regarding which tests must be skipped by a particular validator to avoid the crash. Skipped tests are marked as »no result« in the product summary.
- Collect the output generated by the validator. The output consists of the classification »conforms to the standard« / »does not conform«, the actual error message plus additional messages which are supposed to aid in locating the problem in the PDF document. This may include a page number, font name, PDF object number, etc. This output is converted from the validator's native format to a normalized intermediate XML file.
- The intermediate XML is converted to the final XML report which contains all relevant information. Since a lot of detail is included in these reports they are generally quite large. A readable shorter HTML version of the product-specific reports is therefore generated. The full XML and summarized HTML tables for each product are available along with this report.
- A summary of the product-specific report is generated which lists the results and accuracy values in each test category. The summary tables are reproduced in the product-specific sections of this report.

The full XML reports have been sent to the vendors, and feedback from the vendors has been reviewed. After sending the results no new software versions have been accepted from the vendors for regression testing. However, we may issue updated editions of the Bavaria report in the future.

Adobe Acrobat 9.0

Vendor	Adobe Systems (San Jose, CA, USA)
Web	www.adobe.com
Product	Acrobat
Tested version	9.0 for Windows

Adobe Systems is a member of the PDF/A Competence Center.

Performance and robustness

- We used Acrobat's integrated batch processing facility to validate all documents in the test suite. However, batch processing doesn't work reliably for larger numbers of documents. Acrobat 9.0 reported 45.000 documents for the Bavaria test suite instead of ca. 300 and generated dozens of identical reports for each input document. Processing took more than 6 hours, i.e. unacceptable performance. We found that manually separating the test suite in smaller portions works better. Although it required a lot of mouse clicks this method was faster. Obviously, such a workaround defeats the purpose of batch processing, especially since after processing the first batch Acrobat refused to process another batch and had to be restarted.
- Acrobat 9.0 crashed for some of the test documents.
- Acrobat 9.0 did not provide any validation result for some of our test documents.

Validation accuracy

- Except for two glitches, Isartor coverage is almost complete.
- Conforming documents have a high false alarm rate which in almost all cases is caused by the wrong treatment of the InstanceID XMP property.
- While advanced XMP coverage including extension schemas is very good, many basic XMP checks are missing.
- Checks for PDF 1.4 violations are not complete.

Validation results

The table below lists the results of validating the Bavaria test suite, summarized per category. More detailed validation results including the expected messages and actual error messages are available in an HTML table which is attached to this report.

	number of PDFs	false alarm	missing alarm	all messages wrong	no result	total OK	accuracy
Isartor testsuite (non-conforming)							
6.1 File structure	31	n/a	-	-	1	30	97%
6.2 Graphics	47	n/a	-	-	-	47	100%
6.3 Fonts	28	n/a	-	-	-	28	100%
6.4 Transparency	6	n/a	-	-	-	6	100%
6.5 Annotations	25	n/a	1	-	-	24	96%
6.6 Actions	37	n/a	-	-	-	37	100%
6.7 Metadata	27	n/a	-	-	-	27	100%
6.9 Interactive Forms	3	n/a	-	-	-	3	100%
Other non-conforming							
ISO 19005 violations	9	n/a	-	1	-	8	89%
XMP 2004 violations	5	n/a	2	2	-	1	20%
PDF 1.4 violations	8	n/a	1	4	-	3	38%
Conforming							
Real world	34	19	n/a	n/a	2	13	38%
PDFlib samples	8	1	n/a	n/a	-	7	88%
Advanced XMP	16	-	n/a	n/a	1	15	94%
Unclear in the standard							
probably conforming	2	1	-	-	1	0	n/a
probably non-conforming	3	1	-	-	-	2	n/a

Adobe Acrobat 9.1

Vendor	Adobe Systems (San Jose, CA, USA)
Web	www.adobe.com
Product	Acrobat
Tested version	9.1.0 for Windows

Adobe Systems is a member of the PDF/A Competence Center.

Several changes in Acrobat 9.1 regarding PDF/A validation are documented in *www.adobe.com/devnet/livecycle/pdfs/readercomp_pdfa.pdf*.

Performance and robustness

- We used the integrated batch processing facility to validate all documents in the test suite. The Acrobat 9.0 problems with batch processing have been fixed in Acrobat 9.1: the reported number of documents is correct now, and processing the Bavaria test suite took 9 minutes, i.e. medium performance.
- Acrobat 9.1 did not provide any validation result for some test documents.

Validation accuracy

- Except for two glitches, Isartor coverage is almost complete.
- Most problems found in Acrobat 9.0 have been fixed in Acrobat 9.1, especially the InstanceID problem which caused a lot of false alarms.
- While advanced XMP coverage including extension schemas is very good, many basic XMP checks are missing.
- Checks for PDF 1.4 violations are not complete.
- While none of the conforming real world documents has been rejected, Acrobat 9.1 did not provide any validation result for some documents in this category.

Validation results

The table below lists the results of validating the Bavaria test suite, summarized per category. More detailed validation results including the expected messages and actual error messages are available in an HTML table which is attached to this report.

	number of PDFs	false alarm	missing alarm	all messages wrong	no result	total OK	accuracy
Isartor testsuite (non-conforming)							
6.1 File structure	31	n/a	-	-	1	30	97%
6.2 Graphics	47	n/a	-	-	-	47	100%
6.3 Fonts	28	n/a	-	-	-	28	100%
6.4 Transparency	6	n/a	-	-	-	6	100%
6.5 Annotations	25	n/a	1	-	-	24	96%
6.6 Actions	37	n/a	-	-	-	37	100%
6.7 Metadata	27	n/a	-	-	-	27	100%
6.9 Interactive Forms	3	n/a	-	-	-	3	100%
Other non-conforming							
ISO 19005 violations	9	n/a	1	-	-	8	89%
XMP 2004 violations	5	n/a	3	1	-	1	20%
PDF 1.4 violations	8	n/a	5	-	-	3	38%
Conforming							
Real world	34	-	n/a	n/a	4	30	88%
PDFlib samples	8	-	n/a	n/a	-	8	100%
Advanced XMP	16	-	n/a	n/a	-	16	100%
Unclear in the standard							
probably conforming	2	2	-	-	-	0	n/a
probably non-conforming	3	1	1	-	-	1	n/a

Adobe LiveCycle PDF Generator

Vendor	Adobe Systems (San Jose, CA, USA)
Web	www.adobe.com
Product	LiveCycle PDF Generator
Tested version	8.2.1.1 for Windows Server

Adobe Systems is a member of the PDF/A Competence Center.

We gratefully acknowledge the assistance of FileAffairs GmbH (www.fileaffairs.de) who ran the actual tests and provided the XML files containing the PDF/A validation output produced by LiveCycle.

According to Adobe, PDF/A validation in LiveCycle is based on a different implementation than the validation in Acrobat Preflight. This is confirmed by our accuracy results which differ significantly from Acrobat 9.0/9.1.

Performance and robustness

- Since the actual validation runs have been performed on different hardware than all other tests, performance information is not available.
- We didn't experience any crashes.

Validation accuracy

- Although Isartor coverage is complete in the graphics, transparency, annotation, and forms categories, there are significant gaps in several other categories.
- XMP 2004 and PDF 1.4 coverage are weak.
- There was a significant number of false alarms for our conforming real world documents.
- The validator rejected almost all of our conforming documents in the Advanced XMP category. In all cases the same inappropriate error message was provided.

Validation results

The table below lists the results of validating the Bavaria test suite, summarized per category. More detailed validation results including the expected messages and actual error messages are available in an HTML table which is attached to this report.

	number of PDFs	false alarm	missing alarm	all messages wrong	no result	total OK	accuracy
Isartor testsuite (non-conforming)							
6.1 File structure	31	n/a	9	-	-	22	71%
6.2 Graphics	47	n/a	-	-	-	47	100%
6.3 Fonts	28	n/a	1	-	-	27	96%
6.4 Transparency	6	n/a	-	-	-	6	100%
6.5 Annotations	25	n/a	-	-	-	25	100%
6.6 Actions	37	n/a	12	-	-	25	68%
6.7 Metadata	27	n/a	10	-	-	17	63%
6.9 Interactive Forms	3	n/a	-	-	-	3	100%
Other non-conforming							
ISO 19005 violations	9	n/a	-	-	-	9	100%
XMP 2004 violations	5	n/a	3	1	-	1	20%
PDF 1.4 violations	8	n/a	6	1	-	1	13%
Conforming							
Real world	34	6	n/a	n/a	-	28	82%
PDFlib samples	8	1	n/a	n/a	-	7	88%
Advanced XMP	16	15	n/a	n/a	-	1	6%
Unclear in the standard							
probably conforming	2	1	-	-	-	1	n/a
probably non-conforming	3	-	1	1	-	1	n/a

Apago PDF Appraiser

Vendor	Apago (Alpharetta, GA, USA)
Web	www.apago.com
Product	PDF Appraiser
Tested version	2.0 alpha for Windows

Apago is a member of the PDF/A Competence Center.

When judging the validation results keep in mind that testing has been done with an Alpha version.

Performance and robustness

- Appraiser was the slowest validator in our test.
- The Bavaria test suite triggered a few crashes.

Validation accuracy

- Isartor coverage is almost complete.
- While none of the remaining non-conforming test documents was accepted as conforming, we experienced several crashes and cases where a non-conforming document was rejected with inappropriate messages (»all messages wrong«).
- The main problem is that Appraiser accepted none of our conforming real-world test documents, nor any of our conforming XMP tests: according to its validation results conforming PDF/A documents don't exist. Regardless of the type of document (simple scan, office document, academic paper, etc.) Appraiser wrongly rejected all documents with a variety of error messages. Upon examining the messages in detail we were unable to locate any non-conforming PDF constructs in the test documents.

Validation results

The table below lists the results of validating the Bavaria test suite, summarized per category. More detailed validation results including the expected messages and actual error messages are available in an HTML table which is attached to this report.

	number of PDFs	false alarm	missing alarm	all messages wrong	no result	total OK	accuracy
Isartor testsuite (non-conforming)							
6.1 File structure	31	n/a	2	-	-	29	94%
6.2 Graphics	47	n/a	-	-	-	47	100%
6.3 Fonts	28	n/a	-	-	-	28	100%
6.4 Transparency	6	n/a	-	-	-	6	100%
6.5 Annotations	25	n/a	-	-	-	25	100%
6.6 Actions	37	n/a	-	-	-	37	100%
6.7 Metadata	27	n/a	-	-	-	27	100%
6.9 Interactive Forms	3	n/a	-	-	-	3	100%
Other non-conforming							
ISO 19005 violations	9	n/a	-	1	-	8	89%
XMP 2004 violations	5	n/a	-	-	2	3	60%
PDF 1.4 violations	8	n/a	-	2	2	4	50%
Conforming							
Real world	34	34	n/a	n/a	-	0	0%
PDFlib samples	8	8	n/a	n/a	-	0	0%
Advanced XMP	16	16	n/a	n/a	-	0	0%
Unclear in the standard							
probably conforming	2	2	-	-	-	0	n/a
probably non-conforming	3	1	-	1	-	1	n/a

Callas pdfaPilot

Vendor	callas software GmbH (Berlin, Germany)
Web	www.callassoftware.com
Product	pdfaPilot Server
Tested version	1.2.077 for Windows

Callas software GmbH is one of the founding members of the PDF/A Competence Center and participated in the design and implementation of the Isartor test suite for PDF/A.

Since Acrobat 6 callas has been Adobe's OEM partner and provided the Preflight plugin for Acrobat. In parallel callas offers their PDF/A validation technology as stand-alone products.

Performance and robustness

- Performance is medium.
- We didn't experience any crashes.

Validation accuracy

- Isartor coverage is complete.
- While advanced XMP coverage including extension schemas is very good, many basic XMP checks are missing.
- Checks for PDF 1.4 violations are not complete.

Unique features

We didn't experience a single false alarm in all our testing (and only very few missing alarms). pdfaPilot was the only validator which correctly accepted all our conforming test documents.

Validation results

The table below lists the results of validating the Bavaria test suite, summarized per category. More detailed validation results including the expected messages and actual error messages are available in an HTML table which is attached to this report.

	number of PDFs	false alarm	missing alarm	all messages wrong	no result	total OK	accuracy
Isartor testsuite (non-conforming)							
6.1 File structure	31	n/a	-	-	-	31	100%
6.2 Graphics	47	n/a	-	-	-	47	100%
6.3 Fonts	28	n/a	-	-	-	28	100%
6.4 Transparency	6	n/a	-	-	-	6	100%
6.5 Annotations	25	n/a	-	-	-	25	100%
6.6 Actions	37	n/a	-	-	-	37	100%
6.7 Metadata	27	n/a	-	-	-	27	100%
6.9 Interactive Forms	3	n/a	-	-	-	3	100%
Other non-conforming							
ISO 19005 violations	9	n/a	-	-	-	9	100%
XMP 2004 violations	5	n/a	3	-	-	2	40%
PDF 1.4 violations	8	n/a	5	-	-	3	38%
Conforming							
Real world	34	-	n/a	n/a	-	34	100%
PDFlib samples	8	-	n/a	n/a	-	8	100%
Advanced XMP	16	-	n/a	n/a	-	16	100%
Unclear in the standard							
probably conforming	2	-	-	-	-	2	n/a
probably non-conforming	3	-	2	-	-	1	n/a

Intarsys PDF/A Live

Vendor	intarsys consulting GmbH (Karlsruhe, Germany)
Web	www.intarsys.de
Product	PDF/A Live!
Tested version	4.0.7 on Windows

Intarsys is a partner member of the PDF/A Competence Center and participated in the design and implementation of the Isartor test suite for PDF/A. The product is available on the vendor's Web site for free online PDF/A validation.

Performance and robustness

- The Java-based validator is very fast. Using the integrated batch facility it processed the full Bavaria test suite in 2 minutes.
- We didn't experience any crashes.

Validation accuracy

- Isartor coverage is complete.
- Checks for PDF 1.4 violations are not complete.
- There was a significant number of false alarms for our conforming real world documents.
- Although the validator passed all XMP tests in the Isartor test suite, it wrongly rejected the majority of conforming documents in the Advanced XMP category.

Validation results

The table below lists the results of validating the Bavaria test suite, summarized per category. More detailed validation results including the expected messages and actual error messages are available in an HTML table which is attached to this report.

	number of PDFs	false alarm	missing alarm	all messages wrong	no result	total OK	accuracy
Isartor testsuite (non-conforming)							
6.1 File structure	31	n/a	-	-	-	31	100%
6.2 Graphics	47	n/a	-	-	-	47	100%
6.3 Fonts	28	n/a	-	-	-	28	100%
6.4 Transparency	6	n/a	-	-	-	6	100%
6.5 Annotations	25	n/a	-	-	-	25	100%
6.6 Actions	37	n/a	-	-	-	37	100%
6.7 Metadata	27	n/a	-	-	-	27	100%
6.9 Interactive Forms	3	n/a	-	-	-	3	100%
Other non-conforming							
ISO 19005 violations	9	n/a	1	2	-	6	67%
XMP 2004 violations	5	n/a	-	3	-	2	40%
PDF 1.4 violations	8	n/a	4	2	-	2	25%
Conforming							
Real world	34	15	n/a	n/a	-	19	56%
PDFlib samples	8	2	n/a	n/a	-	6	75%
Advanced XMP	16	12	n/a	n/a	-	4	25%
Unclear in the standard							
probably conforming	2	1	-	-	-	1	n/a
probably non-conforming	3	-	1	1	-	1	n/a

PDF Tools: 3Heights PDF Validator Shell

Vendor	PDF Tools AG (Winkel, Switzerland)
Web	www.pdf-tools.com
Product	3-Heights PDF Validator Shell
Tested version	1.8.32.1 on Windows

PDF Tools AG is one of the founding members of the PDF/A Competence Center and participated in the design and implementation of the Isartor test suite for PDF/A.

Performance and robustness

- The validator is extremely fast and at the same time delivers very accurate results. Running PDF Validator Shell on the full Bavaria test suite required only 21 seconds!
- We didn't experience any crashes.

Validation accuracy

- Isartor coverage is complete.
- PDF 1.4 coverage is strong.
- While advanced XMP coverage including extension schemas is very good, many basic XMP checks are missing.
- The accuracy for conforming documents including advanced XMP is high.

Unique features

Although the validator did not catch all PDF 1.4 violations, it found several PDF 1.4 problems which went unnoticed in all other products. The validator provides by far the fastest PDF/A validation available.

Validation results

The table below lists the results of validating the Bavaria test suite, summarized per category. More detailed validation results including the expected messages and actual error messages are available in an HTML table which is attached to this report.

	number of PDFs	false alarm	missing alarm	all messages wrong	no result	total OK	accuracy
Isartor testsuite (non-conforming)							
6.1 File structure	31	n/a	-	-	-	31	100%
6.2 Graphics	47	n/a	-	-	-	47	100%
6.3 Fonts	28	n/a	-	-	-	28	100%
6.4 Transparency	6	n/a	-	-	-	6	100%
6.5 Annotations	25	n/a	-	-	-	25	100%
6.6 Actions	37	n/a	-	-	-	37	100%
6.7 Metadata	27	n/a	-	-	-	27	100%
6.9 Interactive Forms	3	n/a	-	-	-	3	100%
Other non-conforming							
ISO 19005 violations	9	n/a	1	-	-	8	89%
XMP 2004 violations	5	n/a	4	-	-	1	20%
PDF 1.4 violations	8	n/a	3	-	-	5	63%
Conforming							
Real world	34	5	n/a	n/a	-	29	85%
PDFlib samples	8	1	n/a	n/a	-	7	88%
Advanced XMP	16	-	n/a	n/a	-	16	100%
Unclear in the standard							
probably conforming	2	1	-	-	-	1	n/a
probably non-conforming	3	1	1	-	-	1	n/a

Seal Systems: PDF Longlife Suite/PDF Checker

Vendor	SEAL Systems AG (Röttenbach, Germany)
Web	www.sealsystems.de, www.pdfforever.com
Product	PDF Longlife Suite/PDF Checker
Tested version	2.1.1.2 Beta for Windows

Seal Systems is a partner member of the PDF/A Competence Center and participated in the design and implementation of the Isartor test suite for PDF/A. The vendor commented our test results as follows: »*One of the main concerns of the PDF-Checker from SEAL Systems is to guarantee the correct reproduction of the content over the life time of a document and the delivery of practical solutions. Therefore discrepancies of the XMP specifications are marked as warnings and not as errors. This can be changed by a simple modification of the standard configuration and will lead to nearly 100% in the XMP tests. The results of the Bavaria test suite with adapted configuration are available at* www.pdfforever.com.«

Performance and robustness

- PDF Checker is fast: processing the full Bavaria test suite required little more than 4 minutes.
- We didn't experience any crashes.

Validation accuracy

- The validator passed several Isartor categories with 100% accuracy. However, in the categories file structure, graphics, and fonts it missed a significant number of standard violations.
- In the Isartor metadata category it did not report all errors. However, PDF Checker correctly identified all remaining Isartor XMP problems, but listed them only as warnings instead of errors.
- PDF 1.4 coverage is weak.
- The accuracy for conforming documents including advanced XMP is high.

Unique features

The validator detected several subtle XMP problems (e.g. wrong value types for predefined properties) which went unnoticed by all other validators. However, these problems have been listed as warnings only instead of as errors. PDF Checker was the only validator which detected a problem with an invalid ICC output intent profile.

Validation results

The table below lists the results of validating the Bavaria test suite, summarized per category. More detailed validation results including the expected messages and actual error messages are available in an HTML table which is attached to this report.

	number of PDFs	false alarm	missing alarm	all messages wrong	no result	total OK	accuracy
Isartor testsuite (non-conforming)							
6.1 File structure	31	n/a	18	-	-	13	42%
6.2 Graphics	47	n/a	8	-	-	39	83%
6.3 Fonts	28	n/a	14	-	-	14	50%
6.4 Transparency	6	n/a	-	-	-	6	100%
6.5 Annotations	25	n/a	-	-	-	25	100%
6.6 Actions	37	n/a	-	-	-	37	100%
6.7 Metadata	27	n/a	16	-	-	11	41%
6.9 Interactive Forms	3	n/a	-	-	-	3	100%
Other non-conforming							
ISO 19005 violations	9	n/a	6	-	-	3	33%
XMP 2004 violations	5	n/a	2	1	-	2	40%
PDF 1.4 violations	8	n/a	4	2	-	2	25%
Conforming							
Real world	34	6	n/a	n/a	-	28	82%
PDFlib samples	8	-	n/a	n/a	2	6	75%
Advanced XMP	16	-	n/a	n/a	-	16	100%
Unclear in the standard							
probably conforming	2	-	-	-	-	2	n/a
probably non-conforming	3	-	2	-	-	1	n/a

Solid Documents: Solid Framework

Vendor	Solid Documents, LLC (Redmond, WA, USA)
Web	www.soliddocuments.com
Product	Solid Framework
Tested version	5.1.168 for Windows

Solid Documents is a member of the PDF/A Competence Center. The product is available at *www.validatepdfa.com* for free online PDF/A validation. The vendor's comments on Bavaria test results are available at *www.soliddocuments.com/bavaria.htm*.

Performance and robustness

- Performance is medium.
- We didn't experience any crashes.

Validation accuracy

- Isartor coverage is complete except for the metadata category where the validator missed the majority of problems.
- The accuracy for non-conforming documents is low.
- The accuracy for conforming documents is very low.
- XMP and PDF 1.4 coverage are weak.

Validation results

The table below lists the results of validating the Bavaria test suite, summarized per category. More detailed validation results including the expected messages and actual error messages are available in an HTML table which is attached to this report.

	number of PDFs	false alarm	missing alarm	all messages wrong	no result	total OK	accuracy
Isartor testsuite (non-conforming)							
6.1 File structure	31	n/a	-	-	-	31	100%
6.2 Graphics	47	n/a	-	-	-	47	100%
6.3 Fonts	28	n/a	-	-	-	28	100%
6.4 Transparency	6	n/a	-	-	-	6	100%
6.5 Annotations	25	n/a	-	-	-	25	100%
6.6 Actions	37	n/a	-	-	-	37	100%
6.7 Metadata	27	n/a	17	-	-	10	37%
6.9 Interactive Forms	3	n/a	-	-	-	3	100%
Other non-conforming							
ISO 19005 violations	9	n/a	2	1	-	6	67%
XMP 2004 violations	5	n/a	-	3	-	2	40%
PDF 1.4 violations	8	n/a	1	4	-	3	38%
Conforming							
Real world	34	20	n/a	n/a	-	14	41%
PDFlib samples	8	7	n/a	n/a	-	1	13%
Advanced XMP	16	16	n/a	n/a	-	0	0%
Unclear in the standard							
probably conforming	2	2	-	-	-	0	n/a
probably non-conforming	3	1	1	-	-	1	n/a

Appendix

References

- Bavaria Report on PDF/A Validation Accuracy (this document and the Bavaria test suite):
  www.pdflib.com/developer/pdfa/validation-report
- Isartor test suite published by the PDF/A Competence Center:
  www.pdfa.org/doku.php?id=pdfa:en:isartor_test_suite
- ISO 19005-1: Document management – Electronic document file format for long-term preservation – Part 1: Use of PDF 1.4 (PDF/A-1)
  www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=38920
- ISO 19005-1: Document management – Electronic document file format for long-term preservation – Part 1: Use of PDF 1.4 (PDF/A-1)
  Technical Corrigendum 1, published 2007-04-01
  www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=45613
- TechNotes published by the PDF/A Competence Center:
  www.pdfa.org/doku.php?id=pdfa:en:techdoc
- ISO 19005-1 (PDF/A-1) Application Notes published by AIIM:
  www.aiim.org/documents/standards/PDF-A/ISO19005AppNotes.pdf

Copyright and redistribution

PDFlib GmbH
Franziska-Bilek-Weg 9
80339 München, Germany
phone +49 • 89 • 452 33 84-0
pdfa@pdflib.com
www.pdflib.com

PDFlib GmbH is completely focused on PDF technology. Customers worldwide use PDFlib products since 1997. The company closely follows development and market trends, such as ISO standards for PDF. PDFlib GmbH products are distributed all over the world with major markets in North America, Europe, and Japan.

PDFlib GmbH is one of the founding members of the PDF/A Competence Center and participated in the design and implementation of the Isartor test suite for PDF/A.

PDF/A Competence Center
www.pdfa.org

The ISO standard PDF/A is generating enormous interest in the market. The PDF/A Competence Center was founded in 2006 as an international association to address the high demand for information about PDF/A and to promote the exchange of ideas and experience. The combined knowledge of numerous PDF/A experts is available to interested parties. The association conducts seminars and conferences and contributes to further standardization.