

DOKUMENDIHALDUS
Digidokumendi pikaajalise säilitamise
vorming
Osa 1: PDF 1.4 (PDF/A-1) kasutamine

Document management
Electronic document file format for long-term
preservation
Part 1: Use of PDF 1.4 (PDF/A-1)

EESTI STANDARDI EESSÕNA

Käesolev Eesti standard on rahvusvahelise standardi ISO 19005-1:2005 “Document management – Electronic document file format for long-term preservation – Part 1: Use of PDF 1.4 (PDF/A-1)” ingliskeelse teksti tõlge eesti keelde.

Eesti standardi koostas standardimise tehniline komitee EVS/TK 22 “Informatsioon ja dokumentatsioon”. Standardi tõlkis: Veiko Berendsen.

Tõlkesse lisatud selgitavad märkused on standardi tekstis eristatud fraasiga “Eesti standardi märkus”.

Standardis käsitletakse pikaajaliseks säilitamiseks sobivat failivormingut PDF/A, sellise faili tehnilisi nõudeid ja nõuete tasandeid. Samuti määratletakse tehnilised piirangud, mida PDF/A võrreldes PDFiga sisaldada ei tohi.

Rahvusvaheline standard ISO 19005-1:2005 on avaldatud Eesti standardina EVS-ISO 19005-1:2006, mis on kinnitatud Standardikeskuse 15.12.2006. a käskkirjaga nr 185 ning jõustub selle kohta EVS Teataja 2007. aasta jaanuarikuu numbris teate avaldamisega.

This standard contains an Estonian translation of the English version of the International Standard ISO 19005-1:2005 “Document management – Electronic document file format for long-term preservation – Part 1: Use of PDF 1.4 (PDF/A-1)”. Explanatory notes added to the original text are distinguished in this text with a phrase “Eesti standardi märkus”.

The International Standard ISO 19005-1:2005 has the status of an Estonian National Standard.

Standardite reprodutseerimis- ja levitamiseõigus kuulub Eesti Standardikeskusele

SISUKORD

	SISSEJUHATUS	INTRODUCTION	V
1	KÄSITLUSALA	1 SCOPE	1
2	NORMATIIVVIITED	2 NORMATIVE REFERENCES	1
3	TERMINID JA MÄÄRATLUSED	3 DEFINITIONS	3
4	ESITUSVIIS	4 NOTATION	6
5	NÕUETELE VASTAVUSE TASANDID	5 CONFORMANCE LEVELS	7
5.1	Üldist	5.1 General	7
5.2	Nõuetele vastavuse tasand A	5.2 Level A conformance	8
5.3	Nõuetele vastavuse tasand B	5.3 Level B conformance	8
5.4	Nõuetega vastavuses olevad lugemisvahendid	5.4 Conforming readers	9
6	TEHNILISED NÕUDED	6 TECHNICAL REQUIREMENTS	10
6.1	Faili struktuur	6.1 File structure	10
6.2	Graafika	6.2 Graphics	14
6.3	Fondid	6.3 Fonts	19
6.4	Läbipaistvus	6.4 Transparency	24
6.5	Lisamärked	6.5 Annotations	24
6.6	Tegevused	6.6 Actions	26
6.7	Metaandmed	6.7 Metadata	28
6.8	Loogiline struktuur	6.8 Logical structure	40
6.9	Interaktiivsed vormid	6.9 Interactive Forms	45
	Lisa A (teatmelisa) PDF/A-1 nõuetelevastavuse kokkuvõte	Annex A (informative) PDF/A-1 conformance summary	46
	Lisa B (teatmelisa) PDF/A parim praktika	Annex B (informative) Best practices for PDF/A	53
	Kasutatud kirjandus	Bibliography	56

EESSÕNA

ISO (Rahvusvaheline Standardiorganisatsioon) on ülemaailmne riikide standardiorganisatsioonide (ISO liikmete) liit. Rahvusvahelised standardid koostatakse tavaliselt ISO tehnilistes komiteedes. Igal ISO liikmel, keda huvitab tehnilise komitee töövaldkond, on õigus olla esindatud selles komitees. Standardite koostamisest võtavad osa ka valitsustevahelised ja valitsusvälised rahvusvahelised organisatsioonid – ISO koostööpartnerid. Kõigis elektrotehnika standardimise küsimustes teeb ISO tihedat koostööd Rahvusvahelise Elektrotehnikakomisjoniga (IEC).

Rahvusvahelised standardid kavandatakse nõuetelevastavalt ISO/IEC direktiivide 2. osas toodud reeglitele.

Tehniliste komiteede põhiülesanne on rahvusvaheliste standardite koostamine. Tehnilistes komiteedes heaks kiidetud rahvusvaheliste standardite kavandid esitatakse ISO täisliikmetele hääletamiseks. Rahvusvahelise standardina avaldamiseks peab kavand saama vähemalt 75 % hääletanud liikmete heakskiidu.

Selle standardi mõningad elemendid võivad olla patendiõiguse objektiks. ISO ei ole kohustatud mingeid selletaolisi õigusi välja selgitama.

Rahvusvahelise standardi ISO 19005-1 koostas tehniline komitee ISO/TC 171 “Document management applications” alamkomitee SC 2 “Application issues” koostöös tehniliste komiteede ISO/TC 130 “Graphic technology” ja ISO/TC 42 “Photography” ning ISO/TC 46 “Information and documentation” alamkomiteega SC 11 “Records management/Archives”.

ISO 19005 koosneb järgmistest osadest üldpealkirjaga “Document management – Electronic document file format for long-term preservation”:

- Part 1: Use of PDF 1.4 (PDF/A-1).

SISSEJUHATUS

PDF on dokumentide esitamise digivorming. PDF faile võib kohe alguses luua PDF vormingus, teisendada neid muust elektroonilisest vormingust või digiteerida paberilt, mikrovormilt või muult püsivalt vormingult. Ettevõtted, valitsused, raamatukogud, arhiivid ning teised asutused ja eraisikud üle maailma kasutavad PDFi, et esitada arvestatavat kogust vajalikku infot. Suurt osa sellest infost peab alal hoidma pikemat aega; osa säilitama alatiselt. Neid PDF-faile peab saama kasutada mitme tehnoloogia-generatsiooni kestel. Nende objektide kasutamine ja juurdepääs neile tulevikus sõltub väljanägemise säilitamisest, samuti nende kõrgemat järku omadustest nagu lehekülgede, osade ja lõikude loogiline ülesehitus, masinloetav tekstivoog loomulikult loetavana ning administratiivne, säilitamise ja kirjel-damise metaandmestik.

Adobe Systems Incorporated teeb PDFi spetsifikatsiooni kättesaadavaks avalikult. Siiski, vormingu eriline rohkete tunnustega loomus nõuab, et selle kasutamisele seatakse lisatingimusi, mis teevad selle sobivaks digidokumentide pikaajaliseks säilitamiseks.

Käesoleva ISO 19005 osa põhieesmärk on määratleda PDFil põhinev failivorming, mida nimetatakse PDF/A ning mis näitab, kuidas esitada digidokumenti nii, et säiliks selle väljanägemine läbi aja sõltumatuna faili loomiseks, talletamiseks ja esitamiseks kasutatud vahenditest ja süsteemist.

ISO 19005 selle osa teisene eesmärk on anda raamistik, kuidas talletada digidokumentide konteksti ja ajalugu nõuetega vastavuses oleva faili metaandmetesse.

INTRODUCTION

PDF is a digital format for representing documents. PDF files may be created natively in PDF form, converted from other electronic formats or digitized from paper, microform, or other hard copy format. Businesses, governments, libraries, archives and other institutions and individuals around the world use PDF to represent considerable bodies of important information. Much of this information must be kept for substantial lengths of time; some must be kept permanently. These PDF files must remain useable and accessible across multiple generations of technology. The future use of, and access to, these objects depends upon maintaining their visual appearance as well as their higher-order properties, such as the logical organization of pages, sections, and paragraphs, machine recoverable text stream in natural reading order, and a variety of administrative, preservation and descriptive metadata.

Adobe Systems Incorporated makes the PDF specification publicly available. However, the inclusive, featurerich nature of the format requires that additional constraints be placed on its use to make it suitable for the long-term preservation of electronic documents.

The primary purpose of this part of ISO 19005 is to define a file format based on PDF, known as PDF/A, which provides a mechanism for representing electronic documents in a manner that preserves their visual appearance over time, independent of the tools and systems used for creating, storing or rendering the files.

A secondary purpose of this part of ISO 19005 is to provide a framework for recording the context and history of electronic documents in metadata within conforming files.

Lisaks on käesoleva ISO 19005 osa eesmärgiks määratleda nõuetega vastavuses olevas failis sisalduvate digidokumentide loogilise struktuuri ja muu semantilise informatsiooni raamistik.

Need eesmärgid on saavutatud PDFi kasutatavate osade kindlaksmääramisega ning nende osade kasutuspiirangutega nõuetega vastavuses olevas PDF/A failis.

PDF/A iseenesest ei taga tingimata, et mistahes originaalsest allikmaterjalist loodud nõuetega vastavuses olev fail kuvab algset väljanägemist täpselt; st nõuetega vastavuses oleva faili loomisprotsessis võivad fondid olla asendatud, teksti asetus muutunud, kujutise kvaliteet vähenenud või kasutatud kadudega pakkimist. Organisatsioonidel, kui nad peavad olema kindlad, et nõuetega vastavuses olev fail esitab allikmaterjali täpselt, on vaja kehtestada nõuetega vastavuses olevate failide loomiseks lisanõudmisi kõrvuti käesolevas ISO 19005 osas toodule. Peale selle on neile organisatsioonidele tähtis juurutada nõuetega vastavuses olevate failide korrektse kuva tagamise kava ja praktika.

Käesolevat ISO 19005 osa tuleb kasutada osana organisatsiooni dokumentide digiarhiivis pikaajalise säilitamise keskkonnast. Käesoleva ISO 19005 osa edukas rakendamine arhiivinduslikel eesmärkidel sõltub järgnevast:

- organisatsiooni arhiivi säilitusnõuetest, dokumendihalduspoliitikast ja -protse-duuridest nagu need on esitatud ISO 15489-1s [9];

- mistahes vajalikest lisanõuetest ja -tingimustest, mis tagavad digidokumentide püsivuse ja nende omadused läbi aja. Sealhulgas need, mis on määratletud järgnevates standardites,

Another purpose of this part of ISO 19005 is to define a framework for representing the logical structure and other semantic information of electronic documents within conforming files.

These goals are accomplished by identifying the set of PDF components that may be used, and restrictions on the form of their use, within conforming PDF/A files.

By itself, PDF/A does not necessarily ensure that the visual appearance of the content accurately reflects any original source material used to create the conforming file; e.g. the process used to create a conforming file might substitute fonts, reflow text, downsample images or use lossy compression. Organizations that need to ensure that a conforming file is an accurate representation of original source material may need to impose additional requirements on the processes that generate the conforming file beyond those imposed by this part of ISO 19005. In addition, it is important for those organizations to implement policies and practices regarding the inspection of conforming files for correct visual appearance.

This part of ISO 19005 should be used as one component of an organization's electronic archival environment for long-term retention of documents. Successful implementation of this part of ISO 19005 for archival purposes depends upon:

- the retention requirements of an organization's archival environment, records management policies and procedures as specified in ISO 15489-1, [9];

- any additional requirements and conditions necessary to ensure the persistence of electronic documents and their characteristics over time, including, but not limited to, those defined by:

kuid ei piirdu selle loetletuga:

- ISO 14721;
- ISO/TR 15801 [10];
- ISO/TR 18492 [12];
- ISO 18509-1 [13];
- ISO 18509-2 [14];

– kvaliteedikindlustamise protsessist, et kinnitada vastavust kehtestatud nõudmistele ja tingimustele; näiteks konverteeritud algandmete kvaliteedi ja tervikluse kontrolli kord.

Käesolev ISO 19005 osa peaks viima erinevate rakenduste väljaarendamiseni, mis loevad, esitavad, kirjutavad ja kontrollivad nõuetelevastavaid faile. Erinevad rakendused sisaldavad erinevaid võimalusi nõuetega vastavuses olevate failide ettevalmistamiseks, tõlgendamiseks ja töötlemiseks lähtuvalt sellest, mida nõuetega vastavuses oleva rakenduse pakkujad on pidanud vajalikuks. Siiski on tähtis märkida, et nõuetega vastavuses olevad rakendused peavad suutma nõuetekohaselt lugeda ja töödelda kõiki faile, millele rakendub kindlaksmääratud nõuetega vastavuse tasand.

Käesolev dokument on koostatud kui ISO 19005 osa 1, et oleks võimalik tulevikus luua teisi osasid, mis oleksid ühilduvad PDF spetsifikatsiooni tulevikuversioonidega ilma, et seda dokumenti või PDF versioonil 1.4 baseeruvaid rakendusi tuleks ümber töötada.

Kui kasutatud pole ISO täieliku nimetust, on soovitatav kasutada termineid järgmistest spetsifikatsioonidest või nende osadest:

- “PDF/A” – sünonüüm ISO 19005 standardiperekonnale;
- “PDF/A-1” – sünonüüm ISO 19005-1le;

- ISO 14721;
- ISO/TR 15801 [10];
- ISO/TR 18492 [12];
- ISO 18509-1 [13];
- ISO 18509-2 [14];

– quality assurance processes necessary to verify conformance with applicable requirements and conditions; e.g. an inspection regime to verify the quality and integrity of converted source data.

This part of ISO 19005 should lead to the development of various applications that read, render, write and validate conforming files. Different applications will incorporate various capabilities to prepare, interpret and process conforming files based on needs as perceived by the suppliers of those applications. However, it is important to note that a conforming application must be able to read and process appropriately all files complying with a specified conformance level.

This document has been created as Part 1 of ISO 19005 to allow the creation of future parts, which can provide compatibility with future versions of the underlying PDF specification without rendering this document or applications based on PDF Version 1.4 obsolete.

The following terms, referring to this specification or parts thereof, are recommended when referring to this specification when the full ISO name is not being used:

- “PDF/A” – a synonym for the ISO 19005 family of standards;
- “PDF/A-1” – a synonym for ISO 19005-1;

– “PDF/A-1a” – sünonüüm ISO 19005-1 nõuetelevastavustasandmele A;

– “PDF/A-1b” – sünonüüm ISO 19005-1 nõuetelevastavustasandmele B.

Käesolev ISO 19005 osa seotuna *PDF Reference*'i ja *XMP Specification*'iga jaanuarist 2004 annab piisavalt informatsiooni selleks, et tõlgendada mistahes nõuetega vastavuses olevat PDF/A faili. *PDF Reference* sisaldab Adobe Systems Incorporated intellektuaalset omandit puudutavat deklaratsiooni ja ettevõtte valmisolekut lubada alaliselt, tasuta ja eranditeta selle omandi kasutamist, et edendada PDFi kasutamist. Adobe on esitanud ISOLE samasuguse *XMP Specification*'i käsitleva deklaratsiooni. Üldiselt võib igäüks kasutada *PDF Reference*'i ja *XMP Specification*'i selleks, et luua rakendusi, mis loevad, kirjutavad või töötlevad muul viisil PDF/A faile.

Selliste rakenduste patendinõuded, mis loevad, esitavad, kirjutavad või töötlevad muul viisil PDF/A faile, on väljaspool käesoleva ISO 19005 osa käsitusala.

NPES ja AIIM (tunnustatud standardimisorganisatsioonid) jälgivad käesoleva ISO 19005 osa arendajatele ja kasutajatele suunatud rakenduste juhiseid. Need rakenduste juhiseid on kättesaadavad: <http://www.npes.org/standards/toolspdfa.html> ja <http://www.aiim.org/pdfa/app-notes>. Nii NPES kui ka AIIM säilitavad ka koopiaid spetsiifilistest mitte-ISO normatiivviidetest käesolevale ISO 19005 osale. Need on avalikult kättesaadav digidokumendid.

– “PDF/A-1a” – a synonym for ISO 19005-1 Level A conformance;

– “PDF/A-1b” – a synonym for ISO 19005-1 Level B conformance.

This part of ISO 19005, in conjunction with *PDF Reference* and *XMP Specification*, January 2004, provides sufficient information to interpret any conforming PDF/A file. *PDF Reference* contains a statement from Adobe Systems Incorporated concerning its intellectual property and its willingness to allow perpetual, royalty-free, non-exclusive use of that property in order to promote the use of PDF. Adobe has provided ISO with a similar statement relating to *XMP Specification*. In general, anyone may use *PDF Reference* and *XMP Specification* to create applications that read, write or otherwise process PDF/A files.

Patent claims regarding applications that read, render, write or otherwise process PDF/A files are outside the scope of this part of ISO 19005.

NPES and AIIM (accredited standards developing organizations) maintain an ongoing series of application notes for guiding developers and users of this part of ISO 19005. These application notes are available at <http://www.npes.org/standards/toolspdfa.html> and <http://www.aiim.org/pdfa/app-notes>. Both NPES and AIIM will also retain copies of the specific non-ISO normative references of this part of ISO 19005 which are publicly available electronic documents.

DOKUMENDIHALDUS

Digidokumendi pikaajalise säilitamise vorming

Osa 1: PDF 1.4 (PDF/A-1) kasutamine

Document management

Electronic document file format for long-term preservation

Part 1: Use of PDF 1.4 (PDF/A-1)

Tõlgendamise erimeelsuste korral on kehtiv ingliskeelne tekst.	In case of interpretation disputes the English text applies.
----------------------------------------------------------------	--------------------------------------------------------------

1 KÄSITLUSALA

Käesolev ISO 19005 osa täpsustab, kuidas kasutada Portable Document Format (PDF) 1.4 digidokumentide pikaajaliseks säilitamiseks. See on rakendatav dokumentidele, mis sisaldavad tähemärke, raster- ja vektorandmeid.

See ISO 19905 osa ei rakendu:

– konkreetsetele protsessidele, mille abil konverditakse paber- või digidokument PDF/A vormingusse;

– konkreetse tehnilise ülesehituse, kasutajaliidese, rakenduse või esitamise üksikasjadele;

– konkreetsetele dokumentide hoidmise füüsilistele meetoditele nagu meediumid ja hoiutingimused;

– arvuti riist- ja/või tarkvara nõuetele.

2 NORMATIIVVIITED

Käesoleva dokumendi rakendamisel on hädavajalikud järgnevalt toodud dokumendid.

1 SCOPE

This part of ISO 19005 specifies how to use the Portable Document Format (PDF) 1.4 for long-term preservation of electronic documents. It is applicable to documents containing combinations of character, raster and vector data.

This part of ISO 19005 does not apply to:

– specific processes for converting paper or electronic documents to the PDF/A format;

– specific technical design, user interface, implementation, or operational details of rendering;

– specific physical methods of storing these documents such as media and storage conditions;

– required computer hardware and/or operating systems.

2 NORMATIVE REFERENCES

Dateeritud viidete puhul kasutatakse ainult viidatud väljaannet. Dateerimata viidete puhul kasutatakse viidatud dokumendi kõige hilisemat väljaannet (koos kõigi parandustega).

ISO/IEC 646 *Information technology – ISO 7-bit coded character set for information interchange*

Märkus 1. ISO/IEC 646s määratletud kodeering on samaväärne ANSI X3.4 (ASCII) [1] ja ECMA-6 [2] kodeeringuga.

NOTE 1 The character encoding defined in ISO/IEC 646 is equivalent to ANSI X3.4 (ASCII) [1] and ECMA-6 [2].

ISO/IEC 9541-1 *Information technology – Font information interchange – Part 1: Architecture*

ISO/IEC 10646-1 *Information technology – Universal Multiple-Octet Coded Character Set (UCS) – Part 1: Architecture and Basic Multilingual Plane*

Märkus 2. Märkide koodväärtused, mis on määratletud ISO/IEC 10646-1s on võrdsed Unicode'iga [22].

NOTE 2 The character code values defined in ISO/IEC 10646-1 are equivalent to those of Unicode [22].

ISO 14721 *Space data and information transfer systems – Open archival information system – Reference model*

ISO 15930-4 *Graphic technology – Prepress digital data exchange using PDF – Part 4: Complete exchange of CMYK and spot colour printing data using PDF 1.4 (PDF/X-1a)*

Date and Time Formats, W3C Note, 15 September 1997
<http://www.w3.org/TR/NOTE-datetime/>

Errata for PDF Reference, third edition, 18 June 2003
<http://partners.adobe.com/asn/acrobat/docs/PDF14errata.txt>

Extensible Markup Language (XML) 1.0 (Third Edition), W3C Recommendation, 4 February 2004
<http://www.w3.org/TR/2004/REC-xml-20040204>

ICC.1:1998-09 *File Format for Color Profiles*, International Color Consortium
http://www.color.org/ICC-1_1998-09.PDF

ICC.1A:1999-04, *Addendum 2 to Spec. ICC.1:1998-09*, International Color Consortium
http://www.color.org/ICC-1A_1999-04.PDF

PDF Reference: Adobe Portable Document Format, Version 1.4, Adobe Systems Incorporated – 3rd ed. (ISBN 0-201-75839-3)
http://partners.adobe.com/asn/acrobat/docs/File_Format_Specifications/PDFReference.pdf

RDF/XML Syntax Specification (Revised), W3C Recommendation, 10 February 2004
<http://www.w3.org/TR/2004/REC-rdf-syntax-grammar-20040210/>

Tags for the Identification of Languages, RFC 1766, March 1995
<http://www.ietf.org/rfc/rfc1766.txt>

XMP Specification, January 2004, Adobe Systems Incorporated
<http://partners.adobe.com/asn/tech/xmp/pdf/xmpspecification.pdf>

Märkus 3. NPES ja AIIM (tunnustatud standardimisorganisatsioonid) säilitavad ka koopiaid spetsiifilistest mitte-ISO normatiivviidetest, mis on avalikult kättesaadavad digidokumentid.

NOTE 3 AIIM and NPES (accredited standards developing organizations) maintain copies of the non-ISO references that are publicly available electronic documents.

3 TERMINID JA MÄÄRATLUSED

3 DEFINITIONS

Käesolevas standardis kasutatakse järgnevaid termineid ja määratlusi.

For the purposes of this standard, the following definitions apply.

3.1 nõuetele vastavuse tasand

3.1 conformance level

määratletud piirangute ja nõudmiste hulk millele failid ja lugemisvahendid peavad vastama [ISO 15930-4]

identified set of restrictions and requirements to which files and readers must comply [ISO 15930-4]

3.2 ristviitetabel

3.2 cross reference table

PDF andmestruktuur, mis sisaldab faili vahendatud objekti alguses baiditõmmist

PDF data structure that contains the byte offset of the start of indirect objects within the file

3.3 sõnastik

3.3 dictionary

võtmepaari sisaldav seostetabel, mis täpsustab objektide omaduste nimetust ja väärtust ning mida üldiselt kasutatakse, et koguda ja seostada kompleksseid objekte [ISO 15930-4]

associative table containing key-value pairs, specifying the name and value of an attribute for objects, which is generally used to collect and tie together the attributes of a complex object [ISO 15930-4]

3.4 digidokument

3.4 electronic document

lehele vormistatud tekstilise ja graafilise andmehulga elektrooniline esitus koos metaandmetega, mida saab kasutada nende andmete identifitseerimiseks, mõistmiseks ja esitamiseks, et neid taasesitada paberil või optilisel mikroformil märkimisväärsete infosisu kadudeta

electronic representation of a page-oriented aggregation of text and graphic data, and metadata useful to identify, understand and render that data, that can be reproduced on paper or optical microform without significant loss of its information content