

Czech University of Life Sciences

Department of Information Technologies



DIPLOMA THESIS

PORTABLE DOCUMENT FORMAT

Author: Bc. Gianluca Turturro

Supervisor: Doc. Ing. Zdeněk Havlíček, CSc.

© 2010

Statutory Declaration

I declare, that I elaborated this diploma thesis on the theme “Portable Document Format” independently, using the listed resources, my knowledge, experience and consultations with the diploma thesis leader.

Prague, 24.3.2010

Gianluca Turturro

Acknowledgement

I thank Mr. Doc. Ing. Zdeněk Havlíček, CSc. for his time and precious advices he has given me

Portable Document Format

Summary

This work presents the Portable Document Format, or PDF. It describes the principles and purposes of its existence. The description touches its history, the modern PDF and the standards it has developed into, the security features and some of the tools for working with the format. Multiple descriptions about mechanisms working in PDF are used to illustrate its properties and suitability for various purposes. The work shows the applicability of the PDF format in praxis on an example company and analyses the adoption of the PDF format in its workflow. The example company is an architectural studio, which is competing in a selection procedure. It then has to perform multiple document exchanges and for the purpose uses multiple methods. These methods are analyzed, compared and commented.

Keywords

Portable Document Format, PDF, Adobe, Adobe Acrobat, Adobe Reader, document management, document storage

Portable Document Format

Souhrn

V této práci je představen Portable Document Format, tedy PDF. Práce popisuje principy a důvody jeho existence. Popis sahá od historie, až po moderní PDF formát a standardy, ve které se vyvinul, možnosti zabezpečení a některé z nástrojů pro práci s tímto formátem. Popisy mechanismů způsobu práce PDF formátu jsou použity k lepšímu vysvětlení jeho vlastností a vhodnosti použití v různých podmínkách. Práce ukazuje použitelnost PDF formátu v praxi na modelové společnosti a analyzuje adopci PDF formátu do jejích workflow. Modelová společnost je architektonické studio, které soutěží ve výběrovém řízení. Společnost musí pak provést vícero výměn dokumentů a pro tento úkol využije vícero metod. Tyto metody jsou analyzovány, porovnány a komentovány.

Klíčová slova

Portable Document Format, PDF, Adobe, Adobe Acrobat, Adobe Reader, document management, repositář dokumentů

Index

1.	Introduction	6
2.	Objectives and Methodology	7
2.1.	Objectives	7
2.2.	Methodology	7
3.	Introduction to the PDF Format	8
3.1.	The History.....	8
4.	Portable Document Format Specification	12
4.1.	Overview.....	12
4.2.	Imaging Model.....	13
4.2.1.	Page Description Languages.....	13
4.2.2.	Adobe Imaging Model	14
4.3.	PDF file security	15
4.3.1.	Digital signatures in PDF.....	18
5.	The standardization	20
5.1.	PDF standard subsets	21
5.1.1.	PDF/X.....	22
5.1.2.	PDF/A.....	24
5.1.3.	PDF/E	26
6.	Acrobat.com	28
6.1.	Online cooperation on a PDF document	29
7.	The case study	33
7.1.	Situation.....	33
7.2.	The approach of the competing company	35
7.2.1.	Using prevalently printed solutions	35

7.2.2.	Using prevalently PDF solutions.....	38
7.2.3.	Using prevalently conventional electronic solutions	40
8.	Analysis of the case study.....	42
8.1.	Prevalently printed solutions	42
8.1.1.	SWOT	42
8.1.2.	The Costs	43
8.2.	Using prevalently PDF solutions	46
8.2.1.	SWOT	46
8.2.2.	The Costs	48
8.3.	Using prevalently conventional electronic solutions	49
8.3.1.	SWOT	50
8.3.2.	The Costs	51
8.4.	Final Outcome	53
9.	Conclusions.....	55
10.	Bibliography.....	56
11.	Appendixes	A
11.1.	Calculation of the print cost per page	A

Images

Image 1 – Acrobat.com – Main screen	28
Image 2 – Acrobat.com – Meetings	30
Image 3 – Acrobat.com – ConnectNow 1	31
Image 4 – Acrobat.com – ConnectNow 2	31
Image 5 – Acrobat.com – ConnectNow Whiteboard	32

Tables

Table 1 – Ballot results	21
Table 2 – PDF/X types overview	22
Table 3 – Monthly Fixed Costs (Printing)	44
Table 4 – Variable Costs – 1st portfolio for the selection procedure	44
Table 5 – Variable Costs – 2nd portfolio with the documentation for New Buildings Corp.	44
Table 6 – Variable Costs – The permits	45
Table 7 – Variable Costs – The construction process.....	45
Table 8 – Variable Costs – handover of the final documentation	45
Table 9 – Cost Grand Totals	45
Table 10 – Monthly Fixed Costs (PDF solution).....	48
Table 11 – Cost Grand Totals	49
Table 12 – Monthly Fixed Costs (conventional electronic solutions)	51
Table 13 – Variable Costs – handover of the final documentation	52
Table 14 – Cost Grand Totals	52

1. Introduction

The need to transport various kinds of documents instantly across various platforms at the speed of light stood at the beginning of a challenging task: inventing a technology, capable of containing all the needed information, readable for everyone. However, just the invention of such technology wasn't enough; the second step was the acceptance of this technology by the public and the introduction of it into normal workflow of common people, firms and governments.

The Portable Document Format (or PDF) is a standard for transportation, delivery and archiving documents. The PDF format offers tools to protect their content at various levels. It offers also various options of presentation, interaction, data collection and distribution.

Portable document format can be a substitute to a wide range of technologies, commonly used today as distribution and presentation tools, such as web presentations or printouts, while offering similar or better performance to cost ratio.

This document will introduce the reader to the PDF format, its history, the development in time, and its properties. It will also demonstrate its practical application on a case study, where it will be compared to other applicable methods, available to date.

2. Objectives and Methodology

2.1. Objectives

The main goal of this thesis is to present the PDF format and its main features. The first part of this work is theoretical and it describes the PDF file format as such. It also introduces the reader to various aspects of the PDF format, such as its ISO standards, some of the features commonly used to date. The theoretical part shows the online tool for PDF conversion from Adobe: Acrobat.com.

The practical part is a case study that shows one of the many practical usages of the PDF file format. It puts it then in comparison with other methods of handling documents: printing and other electronic approach. There is a calculation of costs for each of this method and a final outcome of the analyzed methods.

Finally, the conclusion wraps up the thesis with final thoughts and knowledge.

2.2. Methodology

The theoretical part results mainly from the study of specialized material, such as ISO standards, the Project Camelot (1) and thematic web portals.

The practical part is based upon a case study and it contains the description of the case study itself, where the requirements are set. Then there is the analysis of the case study, where the author is proposing three ways of dealing with the requirements of the case study, including the usage of PDF, printing and other electronic solutions. The three ways are compared both by economical points of view and by their strengths, weaknesses, opportunities and threats. In the annex, there are the calculations of one printed page that are used in the evaluation of the case study – printing solution.

3. Introduction to the PDF Format

3.1. The History

The Portable Document Format (or PDF) appeared for the first time as project codename Camelot written by John Warnock, now chairman of the Board at Adobe Systems Incorporate (2). It made its first apparition in 1992 and the main goal of this project, according to “The Camelot Project” paper (1) was to solve a fundamental problem that confronts today’s companies - the ability to communicate visual material between different computer applications and systems. The specific problem was that most programs printed to a wide range of printers, but there was no universal way to communicate and view this printed information electronically.

The ideator of the project Camelot, John Warnock, spotted a gap in the market in the early nineties, and explained the need of a new way to communicate through a variety of platforms, including various operating systems, networks and computer configurations. The documents exchanged in that way had to be printable on all of the printers of that time. The Camelot Project had to provide a collection of utilities, applications and system software, so that corporation could effectively capture documents from any application, send electronic versions of these documents anywhere, and view and print these documents on any machines. (1)

According to the Camelot Project (1), there were at least two different technical approaches to solve the question mentioned above. Both solutions depended on Adobe’s PostScript technology. One approach was to try to make Display PostScript and PostScript implementations smaller and faster so that they could run on the vast majority of the machines at that time. This approach had been tried and was extremely difficult.

A second approach was to divide the problem into smaller problems. That approach would allow each piece to run independently on smaller machines while achieving acceptable performance and a solution for the complete problem. This latter approach required that the problem was divided in a way that is natural for users, and provided

a solution for every user, an approach to the Camelot Project that depends on a unique property of the PostScript language.

PostScript, as an interpretive language, has some properties that other interpretive languages do not have. In particular, the semantics of operators is not fixed. Operators can be redefined to have any desired behavior. This property of PostScript allowed the execution of a PostScript file to have side effects that are very different from the normal printing of a page.

The execution of the PostScript file has allowed a derivative file to be generated. In some sense this derivative file is simpler and uses fewer operators than the original PostScript file but has the same net effect. This operation of processing one PostScript file into another form of PostScript file is called “rebinding”.

Some examples in the Camelot Project (1) illustrate the “rebinding” capability of the PostScript language, which is unique to this language. This property is extremely valuable and the Camelot Project depends on the variations of this idea.

The approach that was taken with Camelot (1) was to define a new language of operators and conventions. For the purposes of the project, this language was called “Interchange PostScript” or IPS. IPS would primarily contain the graphics and imaging operators of PostScript. The language would be defined so that any IPS file was a valid PostScript file. The file would have the appropriate baggage so that it was a valid EPS file. IPS files would print on PostScript printer and would be able to be used by applications that accept EPS files. IPS would also be structured so that the complete PostScript parser was not necessary to read any file written in IPS. IPS would have an adequate set of operators so that any practical document expressed in PostScript can be represented in IPS. There would be situations in IPS where the IPS file couldn't represent visual situations that could be theoretically generated in PostScript. These situations were believed to be extremely rare and all practical application documents could be represented efficiently in IPS. The right way to think about IPS is as it relates

to English. No person in the world knows every English word, but a small subset of the English words, and certain usage patterns enable people to consistently communicate.

Once the IPS format was defined, there would be build a version of the PostScript interpreter (IPS binder) that would read any PostScript file and rebind that file into an IPS file. The IPS binder could be quite small in that it did not need the graphics, font or device machinery contained in full PostScript interpreter. Another function of the IPS binder would be to include reconstituted fonts into the IPS file. The idea here was to include just the characters of a font that are actually used in the document. A result of including the necessary characters from the fonts used is that an IPS file will be completely self contained. In other words, when a file was sent around the country, there was no need to worry about whether the receiving location had all the fonts required by the document. The situation at that time was that complex font substitution schemes were used to deal with locations not having the appropriate fonts. (1)

Once IPS was defined and the IPS binder implemented, then users could capture any PostScript file emitted by a PostScript driver, and convert that file to a self contained IPS file. This file could be shipped anywhere around the network and printed on any PostScript machine.

In addition to the IPS binder, a viewer and browser would be written that would read IPS files, and render those files on displays or to dumb raster printers. It was believed that IPS interpreters could be substantially simpler, and smaller than full PostScript interpreters. It was also believed that an IPS interpreter could have acceptable performance on small machines. The real hope was to make the IPS viewer and browser small enough so that it could co-exist with other applications. It was interesting to think about what those applications could be.

One obvious application for the IPS viewer was in its use in electronic mail systems. The goal was to being able to send full text and graphics documents (newspapers, magazine articles, technical manuals etc.) over electronic mail distribution networks.

These documents could be viewed on any machine and any selected document could be printed locally. Large centrally maintained databases of documents could be accessed remotely and selectively printed remotely. This would save considerable amounts of money in document inventory costs.

Specific large visual data bases like the value-line stock charts, encyclopedias, atlases, Military maps, Service Manuals, Time-Life Books etc. could be shipped on CD-ROM's with a viewer. This would allow full publication (text, graphics, images and all) to be viewed and printed across a very large base of machines.

Another goal according to the Camelot Project (1) paper was the option for the IPS viewer to be also equipped with text searching capabilities. In this case the user could find all documents that contained a certain word or phrase, and then view that word or phrase in context within the document. Entire libraries could be archived in electronic form, and since IPS files were planned to be self-contained, they would be printable at any location.

One of the central requirements of the Camelot Project (1) was that the IPS file format was device independent. This was essential because it was necessary to be able to print the documents on color or black and white machines — on low or high resolution machines. This requirement was also essential in order to visualize the documents at various magnifications on the screen. For example, it was imperative that the user were able to magnify portions of complex maps, so that sub-portions of the image were easy to read even on low resolution displays.

To accomplish the above requirement it was necessary that consistent font rendering machinery was available to the viewer. For this reason the viewers would need to contain the full ATM implementations as part of each system.

In considering all the requirements of corporations regarding documents, it was important to structure Camelot components so that they can be sold in ways that were useful to the corporations. Several ideas have come to mind.

Components of Camelot were generally not interesting to single users. The exception to this was in the distribution of large generally useful databases. If someone produced a CD-ROM with “maps of the world” on it, then one could imagine selling a retail package with one viewer and the CD-ROM.

In most other applications, the distribution of information was to many people. In these latter cases a corporation would like a copy of the viewer for every PC. One can imagine viewers integrated into mail systems, or as general stand-alone browsing systems. In any event corporations should be interested in site-licensing arrangements.
(1)

4. Portable Document Format Specification

4.1. Overview

The PDF Reference (3) states that PDF is a file format for representing documents in a manner independent of the application software, hardware, and operating system used to create them and of the output device on which they are to be displayed or printed. A *PDF document* consists of a collection of *objects* that together describe the appearance of one or more *pages*, possibly accompanied by additional interactive elements and higher-level application data. A *PDF file* contains the objects making up a PDF document along with associated structural information, all represented as a single self-contained sequence of bytes.

A document’s pages (and other visual elements) can contain any combination of text, graphics, and images. A page’s appearance is described by a *PDF content stream*, which contains a sequence of *graphics objects* to be painted on the page. This appearance is fully specified; all layout and formatting decisions have already been made by the application generating the content stream¹.

¹ For example Microsoft Word or Autodesk AutoCAD

In addition to describing the static appearance of pages, a PDF document can contain interactive elements that are possible only in an electronic representation. PDF supports *annotations* of many kinds for such things as text notes, hypertext links, markup, file attachments, sounds, and movies. A document can define its own user interface; keyboard and mouse input can trigger *actions* that are specified by PDF objects. The document can contain *interactive form* fields to be filled in by the user, and can export the values of these fields to or import them from other applications.

Finally, a PDF document can contain higher-level information that is useful for interchange of content among applications. In addition to specifying appearance, a document's content can include identification and logical structure information that allows it to be searched, edited, or extracted for reuse elsewhere. PDF is particularly well suited for representing a document as it moves through successive stages of a prepress production workflow. (3)

The PDF format is suited for carrying various types of information, making their delivery easier and giving the user more power over their aspect and contents.

4.2. Imaging Model

The PDF format is based on the ability to deliver the exact appearance of sophisticated graphics and typography. According to the PDF Reference (3), this is achieved using the "Adobe imaging model". The task of the imaging model is to represent text and illustration in the best possible quality. This is achieved by containing text and images in an abstract form, rather than directly in terms of device pixels. Such a description is economical and device independent and can be used to produce high-quality output on a broad range of printers, displays, and other output devices. (3)

4.2.1. Page Description Languages

PDF contains a *page description language* that is used to describe the graphical appearance of pages with respect to an imaging model. An application program produces output through two-stage process:

1. The application generates a device-independent description of the desired output in the page description language.
2. A program controlling a specific output device interprets the description and *renders* it on that device.

These two steps can be executed on different devices at different times and are used to facilitate the interchange standard for transmission and storage of printable and displayable documents. (3)

4.2.2. Adobe Imaging Model

The PDF Reference (3) states, that the Adobe imaging model is a simple and unified view of two-dimensional graphics borrowed from the graphic arts. In this model, “paint” is placed on a page in selected areas:

- The painted figures can be in the form of character shapes (*glyphs*), geometric shapes, lines, or sampled images such as digital representations of photographs.
- The paint may be in color or in black, white, or any shade of gray. It may also take the form of a repeating *pattern* (PDF 1.2) or a smooth transition between colors (PDF 1.3).
- Any of these elements may be *clipped* to appear within other shapes as they are placed onto the page.

“A page’s content stream contains operands and operators describing a sequence of graphics objects. A PDF consumer application maintains an implicit current page that accumulates the marks made by the painting operators. Initially, the current page is completely blank. For each graphics object encountered in the content stream, the application places marks on the current page, which replace or combine with any previous marks they may overlay. Once the page has been completely composed, the accumulated marks are rendered on the output medium and the current page is cleared to blank again.” (3)

Before PDF 1.4 made appearance, all the earlier versions used an *opaque imaging model*. This model made so, that each new graphics object painted onto a page completely obscured anything that was in the same place before. The object is placed on the page as it was applied with opaque paint. With the introduction of version 1.4, it was possible to use *transparent imaging model* in which objects painted on the page are not required to be fully opaque. Instead, newly painted objects are *composited* with the previously existing contents of the page, producing results that combine the colors of the object and its backdrop according to their respective opacity characteristics. (3)

4.3. PDF file security

In many cases, securing a document is essential either because of government policy or because of company policy. Securing a document can be crucial, when trying to protect sensitive data, such as employees' data, upcoming product images or blueprints of a new car.

According to Shea (4), there are four levels of protection:

- 1) No security
- 2) Password security
- 3) Certificate security
- 4) Policy server

The basic setting of the PDF file does not include any security – apart from the impossibility to edit the present text. This is the step that is most commonly used, when saving in PDF format.

The next level is the password protection. The document can be saved with password protecting multiple events that can occur. This includes password protecting the opening of the document (so nobody can see its content without first inputting the right password). The document can be also protected against copying (anybody can

see the content, but nobody can copy it, unless the document is unlocked), printing and so on.

The certificate security identifies authorized users by the certificates they hold. When the document is secured, user certificates are assigned to a list identifying who's authorized to access the document. (4)

The last and most powerful tool for administration of the rights of a certain file is the policy server. By using the policy server, a PDF document can be managed remotely (i.e. even switched off, so nobody can open it). This allows extensive possibilities in managing security and access rights to the information contained in the PDF file. The rights management server can be for instance the Adobe LiveCycle Rights Management ES2.

The features of securing the content of a PDF file extend beyond the four features mentioned above. In fact, there are possibilities of setting various permissions using various passwords for one file. The file can be visualized with one password, printed in lo-res with no password, printed in hi-res with another password and modified with yet another password. Or the printing can be done only with a watermark, otherwise invisible in the electronic form of the document. This is useful for large printing jobs, where the comp can be printed with a watermark to see, if the result is ok and can be purchased, or if it needs some adjustments.

According to Shea (4), there are four permission settings:

1. Printing Allowed:

- *None*: Prohibits all printing.
- *Low resolution (150 DPI)*: Allows printing, but only at a low resolution.
- *High resolution*: Allows full printing privileges.

2. Changes Allowed:

- *None*: Prohibits all changes to the document.
- *Inserting, deleting and rotating pages*: Allows the insertion, deletion, and rotation of pages.
- *Filling in form fields and signing existing signature fields*: Prohibits changes to the document, but allows the addition to the addition of form data, and digital signatures.
- *Commenting, filling in form fields and signing existing signature fields*: Allows for the addition of comments, form data, and digital signatures.
- *Any except extracting pages*: Allows all changes except for the extraction of pages.

3. Enable copying of text, images, and other content

- *Off*: Prohibits readers selecting text or graphical objects, which would allow copying and pasting of content.
- *On*: Enables text and graphics selection, allowing for the extraction of this content.

4. Enable text access for screen readers for the visually impaired

- *Off*: Allows text to be accessed by screen-reading software for the visually impaired.
- *On*: Allows text to be accessed by screen-reading software for the visually impaired.

4.3.1. Digital signatures in PDF

PDF files can be secured by adding a digital signature. This signature is a unique code, which authenticates the document and its content. When distributing documents electronically, it is often important that recipients can verify:

- That the content has not been altered (integrity)
- That the document is coming from the actual person who sent it (authenticity)
- That an individual who has signed the document cannot deny the signature (non-repudiation)

Digital signatures address these security requirements by providing greater assurances of document integrity, authenticity, and non-repudiation. (5)

Integrity

Digital signatures enable recipients to verify the integrity of an electronic document that is used in one-way or round-trip workflows. Methods for maintaining integrity include:

- Parity bits or cyclical redundancy checking (CRC) functions – CRC functions work well for unintentional modifications, such as wire interference, but they can be circumvented by a clever attacker.
- One-way hash – A one-way hash creates a fixed-length value, called the hash value or message digest for a message of any length. A hash is like a unique fingerprint. With a hash attached to the original message, a recipient can determine if the message was altered by re-computing the hash and comparing his or her answer to the attached hash. Common hashing algorithms are MD5, SHA-1, and SHA-256. Adobe has adopted the SHA-1 and SHA-256 algorithms because of their wide acceptance as a security standard.
- Message Authentication Codes (MAC) – A MAC prevents an attacker from obtaining the original message, modifying it, and attaching a new hash. In this case, a symmetric

key is connected to the MAC and then hashed (HMAC). Without the key, an attacker cannot forge a new message. Adobe uses HMACs where appropriate. (5)

Authenticity

Digital signatures provide document authenticity by verifying a signer's digital identity.

Digital signatures are created using asymmetric key cryptography. For document encryption, a document's author encrypts a document using a public key. Because the recipient is the only person with the private key, he or she is the only one who can decrypt the message. Digital signatures reverse the use of public and private keys for document authenticity. The author encrypts the hash of the message with a private key. Only the public key can correctly decrypt the hash and use it to see if it matches a new hash of the document. Because recipients of the document have the author's public key, they gain greater assurances that the individual who signed the document was the person who encrypted the original hash.

The process that constitutes a digital signature is as follows:

1. A hash is created of the original document.
2. The digital signature is created, which encrypts the hash with a private key.
3. The signature is included with the document.

Adobe Acrobat supports multiple digital signatures placed anywhere in the document for proper presentation. In fact, Adobe Acrobat tracks all previously "signed" versions within the document for easy verification of changes made during the document's lifecycle. (5)

Non-repudiation

Non-repudiation is a document security service that prevents the signor of the document from denying that they signed the document. Support for this service is often driven by authentication and time-stamping capabilities. (5)

5. The standardization

The PDF format has been approved by the International Standard Organization as a standard for digital representation of documents. PDF version 1.7 became the ISO 32000 Standard. According to the ISO 32000 paper (6):

“ISO 32000 specifies a digital form for representing documents called the Portable Document Format or usually referred to as PDF. PDF was developed and specified by Adobe Systems Incorporated beginning in 1993 and continuing until 2007 when this ISO standard was prepared. The Adobe Systems version PDF 1.7 is the basis for this ISO 32000 edition. The specifications for PDF are backward inclusive, meaning that PDF 1.7 includes all of the functionality previously documented in the Adobe PDF Specifications for versions 1.0 through 1.6. It should be noted that where Adobe removed certain features of PDF from their standard, they too are not contained herein.”

In order to be approved as ISO, the PDF specification had to pass an international ballot, where the positive votes had to be more than 66,6% of and the negative should be less than 25% (from the present). The overall results were: 13 positive (93%) and 1 negative (7%). In Table 1, there is an overview of the results of the ballot, including the number of comments.

Table 1 – Ballot results

Country	Vote	Comments
Australia	Positive	0
Bulgaria	Positive	0
China	Positive	0
Japan	Positive	0
Poland	Positive	0
South Africa	Positive	0
Spain	Positive	0
Sweden	Positive	0
Ukraine	Positive	0
United Kingdom	Positive	13
USA	Positive	125
Germany	Positive	11
Switzerland	Positive	19
France	Negative	37
Russia	Abstained	-

Source of data: ISO Ballot for PDF 1.7 Passed! (7)

The PDF format became ISO 32000-1:2008 on the 1st of July, 2008 and was released as an open format.

5.1. PDF standard subsets

The PDF format has four subsets. Each of these subsets has its own purpose. Some of them are a series of ISO standards, some other are separate ISO standards. Because PDF is a very versatile file format, it makes it unsuitable for certain applications to make use of all of its features. The subsets listed below, define a set of rules that make the PDF document suitable for the specific purpose, meant by the standard subset (such as printing and so on) and forbid the other features of PDF in that particular type of document. Some of the most common subsets of the PDF format are listed and described below.

5.1.1. PDF/X

PDF/X is a series of PDF standards made especially for prepress production and printing. In Table 2, they are listed as the following ISO standards under ISO 15930 about Graphic technology :

Table 2 – PDF/X types overview

PDF/X format	Contents	Use	Standard	Compatibility	Transparency support on output
PDF/X-1a	CMYK colors, named spot colors. (No RGB or device-independent (color-managed) data.)	Blind exchange of documents; streamlined and validated for use in high-end press workflows.	PDF/X-1a: 2001	Acrobat 4.0/ PDF 1.3	Transparency must be flattened
			PDF/X-1a: 2003	Acrobat 5.0/ PDF 1.4	
PDF/X-3 (All PDF/X-1a files meet the requirements for PDF/X-3 files.)	CMYK, spot, calibrated RGB and Lab colors, ICC-based color.	ICC color-managed workflows. ICC color profiles are included in PDF/X file. Allows RGB images that include enough information to be treated as device-independent.	PDF/X-3: 2002	Acrobat 4.0/PDF 1.3	Transparency must be flattened
			PDF/X-3: 2003	Acrobat 5.0/ PDF 1.4	
PDF/X-4	CMYK, spot, calibrated RGB and Lab colors, and ICC. Can contain live (unflattened) transparency and layers.	Color-managed workflows and workflows using Adobe PDF Print Engine. Transparency in artwork stays live.	PDF/X-4: 2007	Acrobat 7.0/ PDF 1.6	Live transparency and layers supported

Source of data: Adobe.com (8)

A characteristic of PDF/X subset, according to Adobe (8) is that using PDF/X-compliant files eliminates the most common errors in file preparation: fonts that aren't embedded, incorrect color spaces, missing images, and overprinting and trapping issues. These issues don't arise with PDF/X files because, by definition, these standards

require that all fonts be embedded, appropriate PDF bounding boxes be specified, and color be correctly and consistently defined. PDF/X-compliant files also must describe the printing condition for which they are prepared.

PDF/X files include:

- Fonts and images are embedded.
- Page geometry (the trim box and bleed box) defined.
- Intended printing condition (output intent).
- Trapping key defined.
- Title, creator, producer, creation, and modification dates.

PDF/X files prohibit:

- Live transparency, except in PDF/X-4 files.
- Layers, except in PDF/X-4 files.
- Encryption (security).
- Form fields.
- Interactive elements including movies, sounds, buttons, and hyperlinks.
- Annotations within the bleed box.
- Preseparated PDFs.
- Transfer functions.
- Actions and JavaScripts.
- Embedded PostScript®.

5.1.2. PDF/A

PDF/A is a subset of the PDF standard, designed for long-term file archiving and document preservation. The files to be archived could be of all different sorts, like texts, pictures, spreadsheets, CAD drawings and such can be difficult because of the possible specific association with the program, where they were created. If such program ceases to exist, the file could be no longer readable and the information which is contained within becomes unusable.

In order to preserve the information in its digital form for long periods of time and make it readable across different platforms, there are certain requirements to be fulfilled. According to the PDF/A whitepaper (9), these rules are:

1. *Authentic*

It must be possible to prove that a record is what it purports to be, that it has been created or sent by the person who claims to have created or sent it, and that it was sent at the time alleged. This can be accomplished by use of metadata, which is data about the data. In the case of PDF files, metadata can be programmatically embedded inside of the file, thereby ensuring that it is what it purports to be. The creation, receipt, and transmission of records need to be controlled to ensure that record creators are authorized and identified. While this is usually a function of the overall electronic records management system, there are certain features of PDF files, such as security settings, that support the establishment of authenticity. Electronic signatures are an additional level of authenticity that can be applied to PDF files.

In electronic transactions, a PDF file can combine data on who used the system, when they used it, what they did while using it, and the results of the transaction. PDF files can be used to capture and preserve as many elements of the electronic transaction as possible, specifically, the complete “visual presentation” of the transaction to the user. Sometimes this visual presentation is shown back to the user for confirmation. This can significantly improve an organization’s ability to meet the tests of admissibility, since the

record of the transaction will be the same as what the user confirmed to be acceptable at the time of the transaction.

2. *Reliable*

It must be possible to trust that the content of a record is an accurate representation of the transaction to which it attests. It should be created and captured in a timely manner by an individual who has direct knowledge of the event, or it should be generated automatically by processes routinely used by the organization to conduct the transaction. This is particularly true for records of electronic transactions. Using the PDF specification, a system integration firm can automate the capture of digital records in PDF directly from the source application, whether it is a database, a word processor, or a spreadsheet program.

3. *Complete and unaltered*

It must be possible to protect a record against unauthorized alteration and to monitor and track any authorized annotation, addition, or deletion. Records management policies and procedures should specify what, if any, additions or annotations may be made to a record after it is created, under what circumstances additions or annotations may be authorized, and who is authorized to make them. This is typically in the realm of metadata changes not changes to the record itself. The records management system where PDF files are stored will typically provide a high level of security. PDF files can also be secured with password protection and encryption. Additionally, there are third-party digital signature and public key infrastructure (PKI) solutions for PDF documents from companies such as Entrust, Inc., and VeriSign, Inc. Their products work within the Adobe Acrobat® application as plug-ins.

4. *Usable*

It must be possible to locate, retrieve, render, and interpret a record and understand the sequence of activities in which it was created and used for as long as such evidence is required. The newest feature of encapsulating XML metadata into the PDF file ensures that record-quality metadata will be

readable and accessible into the future. Additionally, PDF files are available for fulltext search. Many well-known software vendors have integrated PDF files into their full-text search engines for many years. This is because the PDF file specification and software development kit (SDK) are publicly available, and a complete PDF library used for PDF software development and integration is available for a nominal fee.

5. System integrity

It must be possible to implement control measures such as access monitoring, user verification, authorized destruction, security, and disaster mitigation to ensure the integrity of the records. The reliability of systems is important to ensuring integrity, and records management applications excel at this because they are designed with this in mind. Using a well-documented file format as an archival standard is also important to ensuring individual document integrity. The PDF specification has been used by scores of vendors to create unique applications that help ensure integrity. The direct integration of third-party digital signature and PKI solutions is one example.

The PDF/A standard has the above mentioned characteristics. As in the case of PDF/X, PDF/A applies a set of rules, when creating a PDF document meant for archiving. These rules are defined by the ISO 19005-1:2005 standard.

5.1.3. PDF/E

PDF/E format is suited for use in the engineering field. It is specified as ISO standard 24517-1:2008 and it is based on PDF 1.6. The need to create a universal format for engineer-data exchange is explained in the introduction of the “Creating PDF/E-ready files” document (10): PDF/E (ISO 24517-1:2008), a document standard ratified by ISO in 2007, evolved from the need for an open, neutral exchange format for engineering and technical documentation. While multiple proprietary formats exist, they each have their own viewers, making it difficult to repurpose 3D and engineering data for downstream uses. The cost of distributing and storing paper contributes to the high

cost of managing distribution and change throughout the project for product development teams as well as for extended supply chains.

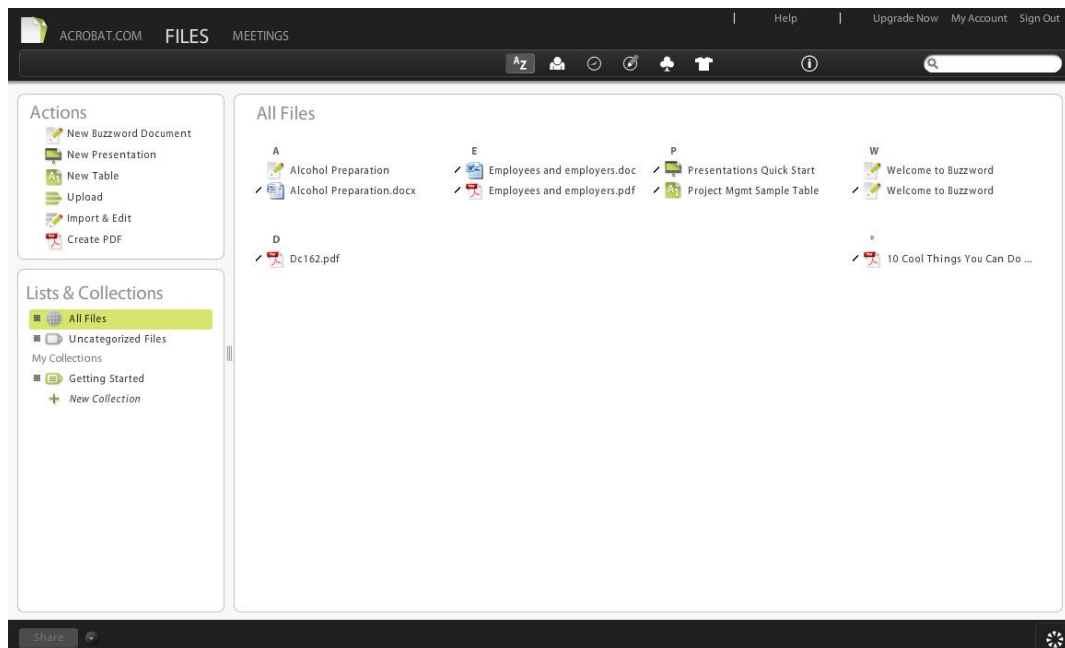
Like PDF, PDF/E is a digital container which supports a wide variety of content and can be viewed and marked up using free and widely available Adobe Reader® software. PDF/E can help support the secure distribution of sensitive information and reduce the complexity and costs associated with distributing and storing paper. While PDF/E is an open standard developed and maintained by an ISO working group, it also leverages U3D, another open standard, for the representation of 3D content.

6. Acrobat.com

For better cooperation on various projects, there is an online web application, dedicated especially to work with PDF files. This application is the Adobe's Acrobat.com and it comes for free online at WWW.ACROBAT.COM with some usage limitations. In order to access the application and use it, an Adobe ID is needed which can be created using any e-mail address and a password.

Once the Adobe ID is created, the user can access the interface – *Image 1 – Acrobat.com main screen.*

Image 1 – Acrobat.com – Main screen



Acrobat.com offers a 5 GB of online disk space to store the user's documents. In the main screen, the user has the overview of his documents stored online, listed alphabetically. There are more options of listing the files, such as:

- By author
- By last view date
- By last change date
- By file type

- By file size

The files can be also categorized, labeled and sorted into collections.

The owner of the repository at Acrobat.com has the option to share some of the files with other users. There is no need for the other users to have an active account on Acrobat.com or an Adobe ID though. The distribution is provided via e-mail with a generated link to the file (no attachments). This enables the user of the Acrobat.com application to quickly distribute the material he owns and have everything under control using only one application.

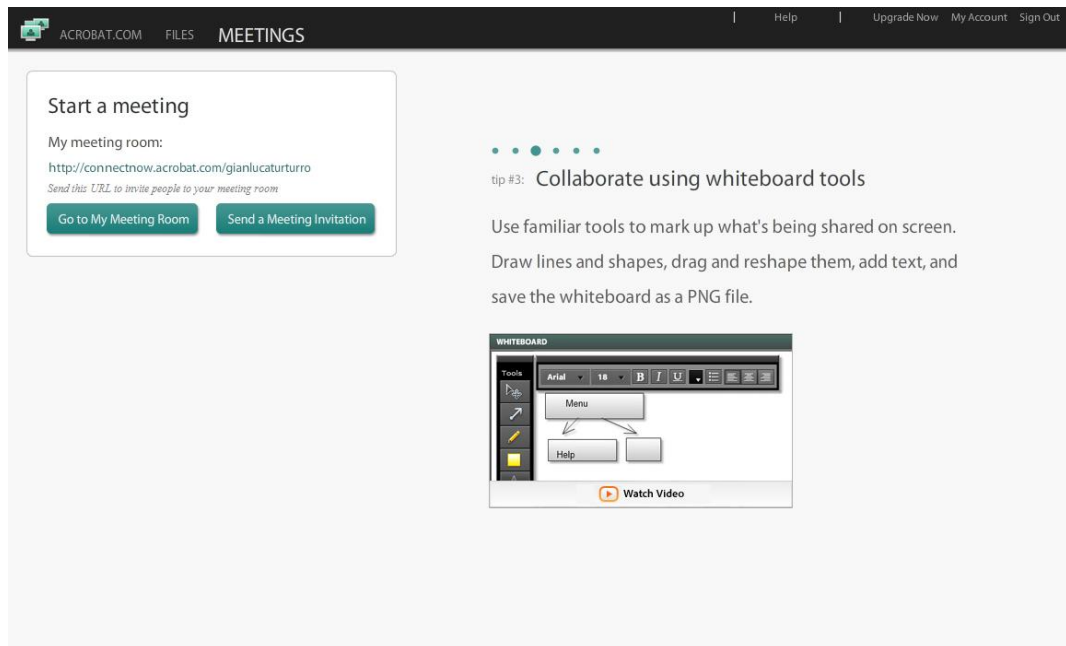
Acrobat.com can be used not only for the purpose of storing and sharing PDF files, but also for their creation; in fact, for creation of content as such. There is a section of applications meant for content creation. This section includes:

- A text editor
- A tool for creating presentations
- A table processor
- An upload tool
- An import and edit tool
- A PDF creation tool

6.1. Online cooperation on a PDF document

Acrobat.com provides also the option to collaborate on a pdf document via the internet connection. Using the function “Meetings” seen in *Image 2 – Acrobat.com – Meetings*, the user has the ability to set up a meeting with another person.

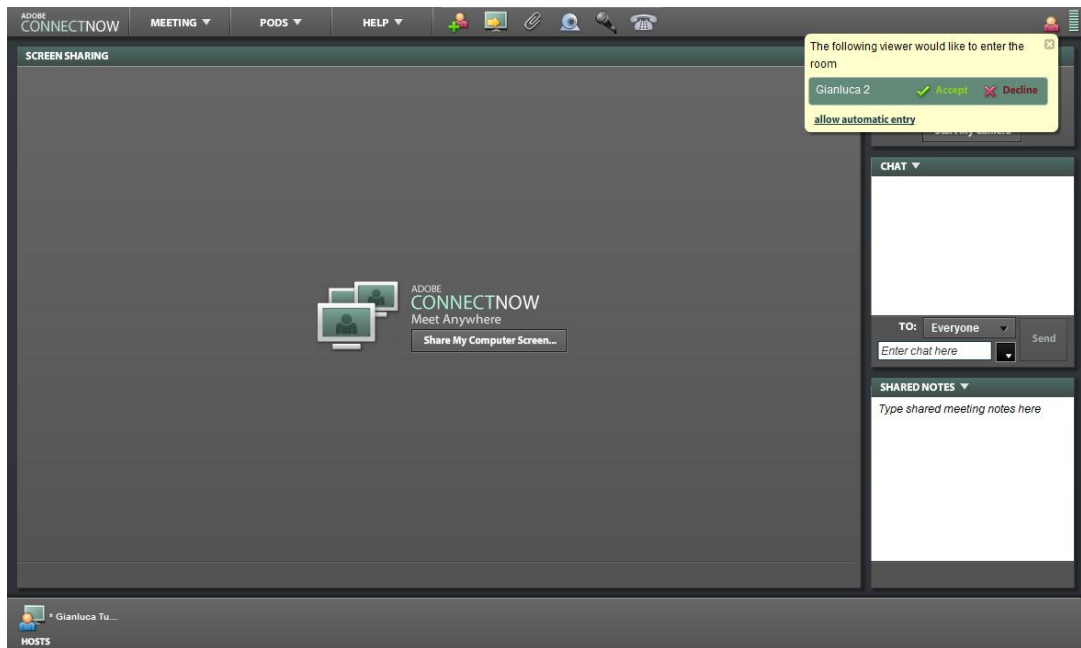
Image 2 – Acrobat.com – Meetings



The guest receives an e-mail invitation to the meeting, and clicking on the link immediately gains access to the meeting, with no need of any Adobe ID.

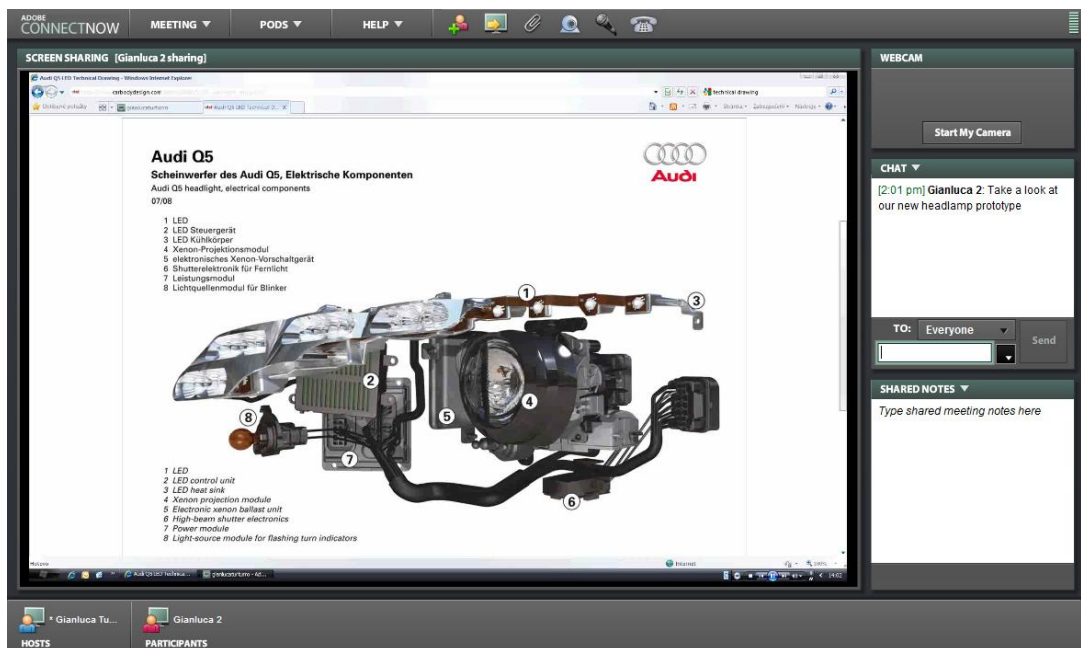
There is an option of screen-sharing during the online meeting in Acrobat.com. This is possible thanks to the function "ConnectNow". In *Image 3 – Acrobat.com – ConnectNow 1* can be seen, that the login screen contains the screen sharing area, the chat area, the shared notes area and an ongoing request of entering the "room" in order to collaborate on a project by the user "Gianluca 2".

Image 3 – Acrobat.com – ConnectNow 1



When the requesting user is allowed to enter the room, he can share the screen with the owner of the Acrobat.com account, so that the two users see exactly the subject of interest – see *Image 4 – Acrobat.com – ConnectNow2*.

Image 4 – Acrobat.com – ConnectNow 2

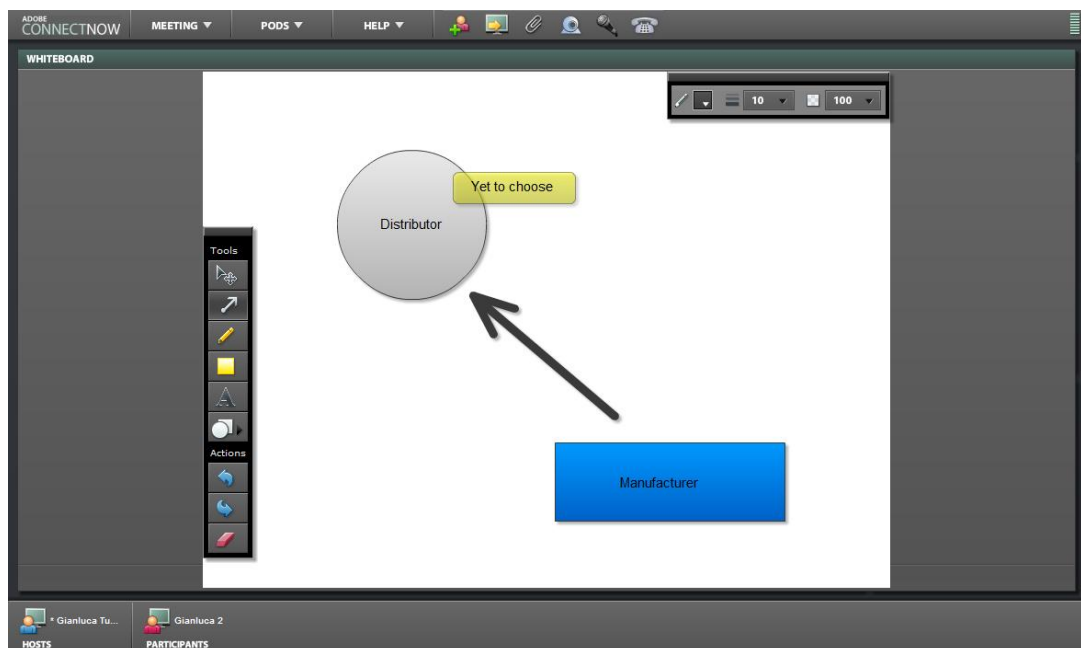


The users can decide what to share and allow or disallow accordingly the access of the other user to the computer resources. The users can also magnify the display in order to compensate for eventual disorders in screen resolution.

Acrobat.com can be also used to send files on a P2P basis, directly through the environment of the web application.

If any of the participants would want to draw some simple objects, without the need of installing specific programs for this purpose, ConnectNow has a tool to do that. It is called the Whiteboard and it allows the users to draw simple vector objects to better illustrate their point. The Whiteboard tool can be seen in *Image 5 – Acrobat.com – ConnectNow Whiteboard*.

Image 5 – Acrobat.com – ConnectNow Whiteboard



7. The case study

To better demonstrate the purpose and the versatility of the PDF format, a case study will be used.

Note: all the characters in the case study are fictive subjects. This case study supposes the winning of the company Angle Shape Studio.

7.1. Situation

Angle Shape Studio, an architectural studio from Prague, is going to compete in a selection procedure, held by New Buildings Corp. – an American developer company which wants to build a new office and commercial complex in the proximity of the Prague city center.

In order to win the job order, they have to pass successfully the selection procedure. Angle Shape Studio is competing with several other architectural studios.

The selection procedure is held in two stages. In the first stage New Buildings Corp. is requesting from all the competitors to handle a portfolio of their creations of the past four years, including a description and the cost of the presented projects. This would help New Buildings Corp. to filter out those companies that have worked on smaller projects before, rather than extensive projects, such the one New Buildings Corp. is planning to undertake.

In the second stage all of the competitors that passed the primary selection will be introduced to the requirements of the project itself.

The new building is situated near a tram stop. It has to be capable of hosting at least:

- 13 000 m² of office area,
- 7 000 m² of commercial area,
- a reception,
- 2 entrances (each one from a different side of the building),
- 300 underground parking places,
- 2x services on each floor,
- 3x elevators,
- 3x stairs.

The building has to be projected according to all applicable laws and regulations about the construction itself, health and safety, fire-protection and other rules that apply.

New Buildings Corp. requires having all the plans of the building, to be able to carefully inspect them. The competitors also have to deliver a static 3D visualization of the building in the existent environment from at least 3 view angles. They are also required to handle the situation of the building, the section and the top view.

After the winning competitor is awarded the job order, there are several other steps to perform in order to begin with the construction. The planned building is situated in a position, where multiple permits are needed. The architects have to gather the permits from several subjects – beginning with the public sphere, ending with the owners of the neighboring real estates. All of these subjects have the right to see part of the documentation – namely the situation and the top view.

Because of the vastness of the building operation, good cooperation and coordination of all the taskforce will be needed. There should be multiple controlling teams that will perform checks on a daily basis, how the construction works. Everything in the construction should be as drawn in the plans and every significant change must be added and described.

After the building is finished, the architects must hand all the needed documentation, including plans, drawings, project documentation and photographic documentation of the structure in multiple phases of its construction (i.e. main walls, hydraulic installation, electric installation, air conditioning and so on).

7.2. The approach of the competing company

There are several steps throughout the process of winning a selection procedure, communicating with the client and the public sphere, building well an office/commercial complex and closing the whole job by handing the final documentation to the client. Angle Shape Studio has several possibilities, how to facilitate the flow of documents and information between it and the other side – client, building company, construction supervisors or various commissions. Next, there will be an analysis of some of these options with their advantages, disadvantages and costs.

Note: Angle Shape Studio will be presenting itself in the first step of the selection procedure with 6 projects.

7.2.1. Using prevalently printed solutions

7.2.1.1. Selection procedure 1st stage

In order to prepare the requested material for the first step of the selection procedure, Angle Shape Studio architects have to search in their paper archives to carefully select their realizations, they want to present.

After the selection is made, they have to find the files, relating to the presented projects, on the computer and make a portfolio with descriptions. Because of the separate file formats of the drawings and the descriptions, each one of them has to be printed separately and then manually reassembled.

Assuming, that all the prints were correct at the first time they were printed and that they were assembled together in the portfolio with no error, the company would have

to pay the paper, the print and the postal delivery of the project to New Buildings Corp.

7.2.1.2. Selection Procedure 2nd stage

Angle Shape Studio receives the instructions regarding all the aspects of the intended construction. According to them, the architects will prepare the draft plan of the building. The plan will be provided with some detailing documentation, including a rough calculation of the project and some explanatory documentation.

According to the request of New Building Corp. they have to provide a static 3D visualization from at least three view points. Angle Shape Studio architects decide to make an all-3D visualization², that includes not only the building itself, but the environment around the building with some human silhouettes and vehicles will be shown in the images as well.

Because they use a 3D rendering software³, they can model pretty much everything they want in very good quality and detail. After the 3D model is finished, they select three viewpoints and render the final high resolution images which are then printed.

After finishing all the steps above, Angle Shape Studio has to put together their project by assembling it into various folders and tubes and handle it to the client either by mailing it or by personal delivery.

7.2.1.3. The Job Order

Each subject that needs to sign the permit will have to receive the part of the documentation that is needed for the process. The architects will then print multiple instances of one document unit meant for the permitting process. There are however multiple permitting processes that differ one from another. This means, that the architects have to prepare document units for various permitting processes (neighbors, conservationists, telephone line providers...). Each of these permits has

² A 3D visualization with no elements other than 3D rendered

³ Usually 3D Studio Max, Rhinoceros or Cinema 4D

then to be signed and collected for further usage, in the process of requesting the building license.

The process of obtaining the building license provides multiple document exchanges between the Studio and the bureau of construction. Because this process requires usually multiple sessions of various subjects and commissions to be brought to an end, there are also inevitable costs of printing and posting the required documentation. In 2009 however, there has been a change in the law. Nowadays data boxes are mandatory for any type of communication between a corporation and the state. Angle Shape Studio would use only this type of communication with the bureau of construction.

When the building process begins, Angle Shape Studio is in charge of administering and controlling the correctness of the process itself. This requires daily inspections in the field and handling multiple plans and drawings. These have to be printed out and transported on site, were going to be handled in various situations including low light areas, places with missing or insufficient support, wet areas or areas with high risk of fire (like welding sites). The on-location inspector is only limited to the material he has brought alongside. It is therefore very difficult for him to obtain any additional information in case of necessity.

The last step for Angle Shape Studio, after the successful completion of the building process, is the handover of the documentation to the submitter. The studio is going to create a final folder with all the documentation, plans and blueprints, permits and photographs of the completed structure, including the various phases of its construction. Finally the shipment will be performed by one of the Angle Shape Studio employees.

7.2.2. Using prevalently PDF solutions

7.2.2.1. Selection procedure 1st stage

In order to prepare the requested material for the first step of the selection procedure, Angle Shape Studio architects perform a quick search in their PDF library, stored on their server.

Because of the character of PDF files, there is no need to perform another search and selection in order to find additional information, relating to the presented projects. It is sufficient to combine the required PDF files containing complete information about each project to create a portfolio.

To deliver the portfolio to New Buildings Corp., Angle Shape Studio can send it via e-mail. If the throughput of the e-mail was not enough for the PDF portfolio to be delivered, the contractor can set up an FTP account on the server and hand the login credentials to New Buildings Corp for the download.

7.2.2.2. Selection Procedure 2nd stage

According to the instructions provided by New Buildings Corp., the architects of Angle Shape Studio will prepare the draft plan of the building. The plan will be provided with some detailing documentation, including a rough calculation of the project and some explanatory documentation.

Angle Shape Studio architects will prepare an all-3D model of the building and the surroundings. Instead of choosing 3 angles and make static renders, the Angle Shape Studio architects will include the whole 3D scene into the PDF file, with some camera points already suggested. The recipient of the PDF file will however be able to navigate himself in 3D space and get to the point of interest. The 3D object is also layered, so it is easy to get to a specific area that is otherwise covered by some other objects.

After finishing all the steps above, Angle Shape Studio can create a PDF container to embed the documentation, the plans and the 3D scenes. The finished document can

then be opened using just Adobe Reader, which is freely available on the internet for download. The delivery of the document can be performed via e-mail.

7.2.2.3. The Job Order

Each of the subjects that need to sign the permit will receive the part of the documentation that is needed for the process. Because of the documentation being administered electronically, the architects will just assemble one instance of one document unit meant for the permitting process. This instance will then be sent to the subjects in question using either e-mail or a document management system.

In order to distribute and then collect the documents that are needed for the obtainment of the building license, Angle Shape Studio can employ e-mail. The returning documents can be signed digitally or can be printed out, signed and stamped and then posted traditionally by mail.

The process of obtaining the building license provides multiple document exchanges between the Studio and the bureau of construction. Because in the Czech Republic it is mandatory to use data boxes for communication between corporation and the state organs, the sending of multiple notices is all performed online via data boxes.

When the building process begins, Angle Shape Studio is in charge of all the daily inspections in the field and handling the needed documentation for the purpose. The plans and drawings can be embedded in a PDF file and can be visualized on a laptop or a handheld device, such as a tablet PC with multitouch user interface. Because of the various precarious situations of the building site, using such device would be more practical, due to its compact size. It is also possible to connect to the data repository of the studio to have all the documentation regarding a certain situation, even an unplanned one. This would only require a mobile access to the internet, providing a reasonable speed of download.

The last step for Angle Shape Studio, after the successful completion of the building process, is the handover of the documentation to the submitter. The studio is going to create a final PDF container with all the documentation, plans and blueprints, permits

and photographs of the completed structure, including the various phases of its construction. Due to the characteristics of the PDF file, there could be one PDF container that contains separately savable PDF files of various standards – A PDF/A for all the written documentation and the photographs, a PDF/X for all the drawings and a normal PDF for the 3D visualization of the building. The delivery of the file would be done by e-mail and on two optical media with a major protective coating assuring endurance and readability in time.

7.2.3. Using prevalently conventional electronic solutions

7.2.3.1. Selection procedure 1st stage

The studio keeps all of its original files in their original proprietary file formats. They have a central replicated repository with all their projects, carefully sorted into folders. In order to participate to the selection procedure, they assemble the needed files in a new folder to further work with them and not to accidentally damage the originals on the server.

Angle Shape Studio has to make it possible for New Buildings Corp. to open all of the different file formats (especially the drawings). For this reason, in the e-mail they are sending, besides all the files in question, they also include links and instructions for the customer to be able to download and install all of the viewers needed for viewing their projects.

7.2.3.2. Selection Procedure 2nd stage

Angle Shape Studio will proceed with the planning process, according to the received instructions regarding all the aspects of the intended construction. The plan will be provided with some detailing documentation, including a rough calculation of the project and some explanatory documentation.

In this case, the 3D visualization requested from New Buildings Corp to be from at least three view points, will be delivered as 3 separate, fully rendered images. The scenes will be containing all the detailed images of the building itself and the surroundings.

The results of the above steps are then all put together in one folder, zipped and sent via e-mail to the submitter. If the e-mail wasn't of enough capacity for the project to be sent, the studio can set up an FTP account for the client to download the project material.

7.2.3.3. The Job Order

To obtain all the required permits the architects will e-mail a copy of the document meant for the permitting process. There will be multiple documents according to the permitting process. Each of these permits has then to be signed and collected for further usage in the process of requesting the building license. The recipient can print out the document and send it by mail or sign it and send it as a PDF format.

All the document exchange between the construction bureau and the contractor are performed via the data boxes, including the whole process of the gaining of the permit.

Angle Shape Studio will perform daily inspections in the field and for the scope will have to work with multiple plans and drawings. These can be saved in their original formats, although it highly depends on the device, they will be displayed on. For certain devices, they in fact would have to be exported to another file format, such as a raster image file. The on-location inspector can work with the files he brings along in the device, or can access the company server via a mobile internet connection. It is therefore easy for him to obtain all the additional information in case of necessity.

The last step for Angle Shape Studio, after the successful completion of the building process, is the handover of the documentation to the submitter. The studio is going hand all the documentation on an optical media with high durability in time and they will let the files related to the project available on their FTP server to be downloaded. They will also make a printed copy of the documentation to ensure its usability in long time periods. Finally the shipment will be performed by one of the Angle Shape Studio employees.

8. Analysis of the case study

8.1. Prevalently printed solutions

Using prevalently printed solution is the most classical approach to managing documents. It has been used since ancient times and it is the most commonly used method to date.

Paper documents have the advantage of physically existing, and it is easy for most of the people to handle them and use them. The user has also the immediate possibility to see what information the document contains. On the other hand paper can be very clumsy to administrate, or to store. In the following SWOT, there will be a description of the Angle Shape Studio using prevalently printed solutions in the job process.

8.1.1. SWOT

Strengths

- The document is always palpable.
- Easy orientation on a printed document (especially on blueprints).
- No need for other devices to work with them.
- Paper is “eternal” in the sense of archiving.
- Prints are very secure against failure by chance.
- Prints usually offer much bigger viewing area, compared to electronic devices

Weaknesses

- Prints can be damaged while working in the field.
- Difficult to store above a certain point.
- Difficult to search and administrate.
- High cost in terms of printing and material.
- Environmental stress.
- Need for significant amount of space for storage and cataloguing.
- Usually need a special employee to operate the filing cabinet.

Opportunities

- Clients are still more used to paper.
- Some of the bureaucracy is requested to be printed out.
- The possibilities of printed communication are always 100% valid.
- In some cases, prints should be handed to somebody on the construction site, with no option of returning.

Threats

- Costs of printing material can be raised due to unexpected situations on the market or taxes.
- Printed documents can be lost, stolen or compromised.
- It is mandatory to date to perform all the communication with the bureaus via data boxes (prints are useless then).
- Confidential documents have to be destroyed securely.
- In case of damage of the document repository, all the documents can be lost with no option of recovery.

8.1.2. The Costs

The calculations of the costs are based on actual offers, drawn from the internet. The quantities are based on the author's experience. Because of the broad variety of sizes, while printing different types of plans and blueprints, the unit was set to an equivalent of A3 papers. The printer is the Konica Minolta Bizhub C280 set GATE, the plotter is the Canon imagePROGRAF iPF755. The project is supposed to take approximately 2 years to complete (24 months).

Table 3 – Monthly Fixed Costs (Printing)

Item	Amount
Archive 10 m²	5 000.00 CZK
Printer depreciation (linear, 72 mo)	1 805.00 CZK
Plotter depreciation (linear, 72 mo)	1 190.00 CZK
Total	7 995 CZK

Both the printer and the plotter have an expected lifetime of 6 years (72 months). The cost of the archive corresponds to the cost of 10 m² of leased office area.

Table 4 – Variable Costs – 1st portfolio for the selection procedure

Item	Amount
300 pages A4, 1-sided, BW	176.60 CZK
100 A3 equivalent pages, 1-s, color	158.75 CZK
Additional material	85.00 CZK
Postal delivery	43.00 CZK
Total	461.75 CZK

Table 5 – Variable Costs – 2nd portfolio with the documentation for New Buildings Corp.

Item	Amount
70 A4 pages A4, 1-sided, BW	62.10 CZK
30 A4 pages A4, 1-sided, color	17.70 CZK
20 A3 equivalent pages, 1-s, color	31.75 CZK
Additional material	85.00 CZK
Postal delivery	43.00 CZK
Total	239.55 CZK

Table 6 – Variable Costs – The permits

Item	Amount
40 A4 pages, 1-sided, BW	23.55 CZK
10 A3 equivalent pages, 1-s, color	15.90 CZK
Additional material	150.00 CZK
Postal delivery	136.00 CZK
Total	325.45 CZK

Table 7 – Variable Costs – The construction process

Item	Amount
800 A4 pages, 1-sided, BW	471.11 CZK
30 A3 equivalent pages, 1-s, color	47.65 CZK
Additional material	85.00 CZK
Total	603.76 CZK

Table 8 – Variable Costs – handover of the final documentation

Item	Amount
100 A4 pages, 1-sided, BW	59.00 CZK
30 A4 pages, 1-sided, color	26.60 CZK
60 A3 equivalent pages, 1-s, color	95.25 CZK
Additional material	85.00 CZK
Postal delivery	43.00 CZK
Total	308.85 CZK

Table 9 – Cost Grand Totals

Item	Amount
Fixed costs (24 months)	191 880.00 CZK
Variable costs (for the entire project)	1939.36 CZK

8.2. Using prevalently PDF solutions

PDF is the broadly accepted file format for exchanging documents. It can be easily created using other existing documents of various sorts and distributed to the needed recipients. More than that, the PDF document can be secured to ensure authenticity and protection of confidential information. PDF documents can contain a digital signature and they can be used as valid documents for correspondence with the authorities. A signed PDF document has the same weight of a signed printed document. Following is the SWOT of Angle Shape Studio, employing the PDF technology in their workflow.

8.2.1. SWOT

Strengths

- PDF files can be digitally signed.
- Improved settings on security of the document through encryption.
- Protection against theft of data inside of the PDF document.
- The emitter of the PDF document can be in control of its existence (using appropriate tools).⁴
- Multiple types of file formats can be combined in a single PDF.
- The document can contain 3D active objects, forms that can be filled up and sent immediately, multimedia, text and images.
- PDF can contain also the original files, besides their content transformed to PDF format.
- Easy creation using both free and commercial tools.
- Indexability and searchability of the file archives.
- Included metadata in the files improve their cataloguing and allow automation of keeping a clean file archive.
- Multiple ISO standards for the PDF file format, following various types of its employment.

⁴ For instance the Adobe LiveCycle Rights Management ES2

- Everything is included in one and only file.

Weaknesses

- Possible mistake in the making process of a PDF file can lead to data unrecoverability.
- Low level of password protection can mislead about the security of the file, due to its easy crackability.
- To allow drawing full advantage from using the PDF format, the operator should be educated for the purpose.
- To make a PDF, usually there is a need of an external application or a plugin.
- To create a fully functional PDF file, with rich media content and 3D elements, the highest version of Adobe Acrobat is required (the version type Extended).
- Being an electronic document, it requires backup to ensure data safeguard.

Opportunities

- Integration of the PDF file format to many document management systems.
- Indexed by search engines on the internet.
- Format openable on the vast majority of devices.
- Easy understandability of the usage of a pdf document.
- The reader comes at no cost.
- The reader is built in various languages for various platforms.
- PDF files can be read even in non-Adobe readers.
- If backed up properly online, the file is virtually indestructible (cloud principle).

Threats

- Non-Adobe readers can misbehave on non-standardized PDFs or rich content PDFs.
- The recipient of a PDF file is willing to edit that file, with no option of getting the original file.

- The recipient does not have a PDF reader and has not the required rights to install the reader on the target device.
- The recipient has not the ability to install the reader on the target device.
- The recipient requires a printed version besides the electronic one.
- There is a competition file format that aims at the same goals, as PDF. The format is the XPS document from Microsoft.

8.2.2. The Costs

Table 10 – Monthly Fixed Costs (PDF solution)

Item	Amount
Acrobat CS4 Extended, dep. 48 mo.	460.20 CZK
Amazon S3 services	4252.00 CZK
UPC Fiber Power 100/10 Internet	1765.00 CZK
ReadyNAS Pro, dep. 120 mo.	208.00 CZK
6TB WD RE3, dep 48 mo.	370.00 CZK
Total	7055.70 CZK

While in the printing solution, there was a high percentage of a depreciation cost, in the case of the PDF solution, the fixed consist more of online backup solution and the write-off of the Acrobat software, which is supposed to last for 4 years (48 months), before getting obsolete.

As the online backup solution, it was chosen to use the Amazon S3 service. The cost for the first 50TB of data is \$0.150⁵ per GB. The studio is supposed to have about 1.5 TB of backed-up data on the service. There are also other costs for connected to the usage of the S3 service, but for the purpose of the planned usage, they are negligible.

In order to allow backups over the internet, the upstream bandwidth is required to be at least 10 Mbit/s. The service chosen for the occasion is the UPS Fiber Power 100/10 internet connection service.

⁵ At the time of the creation of this thesis, the exchange rate of the US Dollar was 18.19 CZK

The last fixed cost connected to the prevalent usage of PDF as distribution channel, is the local disk array for storage/backup. The chosen solution is the Netgear ReadyNAS Pro, with six 1 TB disks Western Digital RE3 WD1002FBYS⁶. The NAS is meant to have a lifetime of 10 years (120 months), the disks are meant to have a lifetime of 4 years (48 months), before being exchanged with bigger capacity disks. This device also provides the studio with its built-in FTP server.

Table 11 – Cost Grand Totals

Item	Amount
Fixed Costs (24 months)	169 336.80 CZK

Because of the way the PDF approach works, there is no need for additional variable costs during the whole process of fulfilling the order. The fixed costs are lower, but if there was needed a big print, the company would need to outsource the print job. The saving over the printed solution amounts to 18.75%.

8.3. Using prevalently conventional electronic solutions

The most spontaneous action when exchanging data between multiple subjects is to send them in their native file format (such as a Microsoft Word .doc or .docx document). The sender of the document takes for granted in fact, that the recipient has the appropriate program or environment to open that particular type of file. This is however not always the case, so by performing such file exchange, there can arise problems of non-compatibility, such as vastly happened between the years 2007 and 2009 with the Microsoft Office 2007 suite⁷. In the majority of cases, this exchange still happens, even at a government level, where all the emitted documents should be platform-independent. The following SWOT shows how the company embraces communication using just proprietary formats.

⁶ Heavy duty disks, meant for RAID usage and 24/7 running.

⁷ Microsoft introduced their new Open XML format, as the main format of the applications in the suite. The format is based on XML, so the files are a collection of markup tags, rather than being a binary file. This brought some compatibility issues, especially, when trying to open the new file in the older versions of the Office suite applications. In order to open the new file type in an older application version, a conversion to the older file format was needed. Many people didn't realize this situation for a while, so they were misled by the fact, that their document worked normally, but the recipient could not open it at all.

8.3.1. SWOT

Strengths

- No need of conversion, just saving the file.
- The file is editable.
- The file maintains all its original information.
- The saving of the file does not require other programs or plugins.

Weaknesses

- Possible compatibility issues.
- The file size can be much bigger, than needed.
- The file is openable only in the application of creation
- Importing one file-type into a program from a different make can lead to issues in its presentation.
- Being an electronic document, it requires backup to ensure data safeguard.

Opportunities

- The file may be opened with no loss of information in the same program it was created on the recipient's side.
- The file can be imported into other programs and worked with.
- Proprietary format files can be visualized with a free reader, usually provided by the maker of the software.
- If backed up properly online, the file is virtually indestructible (cloud principle).

Threats

- The file has only very limited (if any) possibilities of protection.
- The information in the file can be stolen or tempered with.
- No control over the "faith" of the file.

- The recipient may not have all the components to correctly visualizing the content of the file, even if having the same program of the same version as the emitter.
- The recipient has not the rights or the will to install the reader to open a proprietary file.
- The maker of the program can cease the support for a specific file format.
- The reader may not always be available.

8.3.2. The Costs

In this case, there is a prevalent use of native file formats. Therefore, there is no need to purchase special software, like Adobe Acrobat. The company will however make some final prints.

Table 12 – Monthly Fixed Costs (conventional electronic solutions)

Item	Amount
Amazon S3 services	4252.00 CZK
UPC Fiber Power 100/10 Internet	1765.00 CZK
ReadyNAS Pro, dep. 120 mo.	208.00 CZK
6TB WD RE3, dep 48 mo.	370.00 CZK
Total	6 595.00 CZK

The studio has to maintain all the data storage and backup solutions, because it relies mainly on electronic documents. There are no printer costs as well, because for the intended printing volume, the printer costs would be an unnecessary investment.

Table 13 – Variable Costs – handover of the final documentation

Item	Amount
100 A4 pages, 1-sided, BW	100.00 CZK
30 A4 pages, 1-sided, color	300.00 CZK
60 A3 equivalent pages, 1-s, color	114.00 CZK
Additional material	85.00 CZK
Postal delivery	43.00 CZK
Total	642.00 CZK

In order to print the final documentation, the company has to outsource the prints to an external printing company. In this case, the final documentation is more than 200% of the in-house printing solution expense. It does however save all the fixed costs for the printers.

Table 14 – Cost Grand Totals

Item	Amount
Fixed costs (24 months)	158 280.00 CZK
Variable costs (for the entire project)	624.36 CZK

Although the cost of the final documentation is twice as much as if it was printed in-house, the overall variable cost for the project is considerably lower, than in the first case. The missing costs for the printers also lower the fixed costs to about 82% of the in-house printing solution.

8.4. Final Outcome

This case study compared three methods used in the majority of cases for publication of documents and presenting them to the recipient. The **printing solution** turned out to be the most expensive, due to the very high costs of both printers and prints – see *Table 9 - Cost Grand Totals*.

The storage for printed documents can also be very expensive, especially in high priced office areas. The prints however can be the only way that is suitable for certain clients. For the purpose of communicating with the state authorities, there is an electronic way, that is mandatory nowadays – the data boxes.

For the **PDF solution**, there are no variable costs, as all the communication and document exchange is handled electronically – see *Table 11 - Cost Grand Totals*. The concept of the PDF format made it possible to abolish all the paper communication between subjects. Of course, the habits of many people cannot be just broken, so it still remains a goal for the future. However, all the principles stated in this case study are applicable to date and can be used without hesitation.

The PDF solution has the second biggest fixed costs, due to the need of having powerful storage and backup systems that are reliable and accident-proof. In order to take full advantage of the PDF format, it is required to have the appropriate version of Adobe Acrobat Professional.

The case study does not calculate with a device to display the electronic data in the field (such as a laptop or a tablet pc), because the document counts on the fact that the studio already has at least 1 notebook to use for the purpose. Moreover, there is a vast choice of devices suitable for the purpose of being a field reader for the electronic documents. The company can then purchase a device of choice for the on-location inspections. This would raise the fixed costs by the depreciation of such device.

The last approach mentioned was the **usage of prevalently conventional electronic solutions**. This method is used in most of the cases, when a printer is not an option. It

is the easiest and the cheapest, but if done properly, it generates costs as well – see *Table 14 - Cost Grand Totals*.

Similarly to the PDF solution, there has to be a data storage system and an online backup. Unlike the PDF however, proprietary files of different programs require proprietary readers from different makers. The reader may not be available though for some file types.

It is also required to hand a printed copy at the end of the process, because file compatibility is not ensured. This generates variable costs for the printing process that has to be undertaken by outsourcing to a printing facility.

Recommendation:

For the best results in flexibility, protection of sensitive data, speed of communication and cooperation, the option is to combine both the PDF solution and the print by outsourcing. The costs generated this way can be higher, than those generated by adopting a purely electronic solution, the versatility of such approach will however counterbalance it.

9. Conclusions

This diploma thesis presented the PDF format, its history, features, ISO standards and it demonstrated its practical usage. The PDF format is a broadly accepted file format for exchanging, archiving and protecting documents. There are many tools that are capable of creating, opening and working with PDF documents. The PDF document evolves to respond to the requests of the market. To take full advantage of the capabilities of the PDF format, including interactive 3D models, flash-based portfolios and so on, the user has to use Adobe Acrobat Pro Extended.

Using the PDF file format for exchanging and storing documents has benefits such as ease of use, increased security options and interoperability, as well as detriments such as the need of the reader, the need to purchase additional pieces of software and hardware components and their administration.

The case study focused on the comparison between three methods of handling documents: by printing, by using PDF and by using other electronic solutions. The outcome of this case study is the résumé of the properties and costs of using each of these solutions. Finally, there is a recommendation, based upon the knowledge gained during the elaboration of the case study solution.

PDF format is a great tool for moving the office and communication from paper to digital. It is conceived from the beginning to work as a paper replacement and adds several features that cannot be present even in the best of paper document system.

10. Bibliography

1. **Warnock, John.** The Camelot Project. *Planet PDF*. [Online] 1991. [Cited: 10 29, 2009.] http://www.planetpdf.com/planetpdf/pdfs/warnock_camelot.pdf.
2. **Adobe Systems, Inc.** Adobe - Executives. *Adobe.com*. [Online] Adobe Systems, Inc. [Cited: 12 24, 2009.] <http://www.adobe.com/aboutadobe/pressroom/executivebios/johnwarnock.html>.
3. **Adobe Systems, Inc.** PDF Reference, Sixth Edition, version 1.7. *Adobe.com*. [Online] 11 2006. [Cited: 1 3, 2010.] http://www.adobe.com/devnet/acrobat/pdfs/pdf_reference_1-7.pdf.
4. **Shea, Dan.** Introduction to PDF security. *Planet PDF*. [Online] Planet PDF. [Cited: 03 27, 2010.] <http://www.planetpdf.com/enterprise/article.asp?ContentID=7145&page=1>.
5. **Adobe Systems, Inc.** A primer on electronic document security White Paper. *Adobe.com*. [Online] 2007. [Cited: 03 29, 2010.] http://www.adobe.com/security/pdfs/acrobat_lifecycle_security_wp.pdf.
6. **ISO.** Document management — Portable document format — Part 1: PDF 1.7. *Adobe - PDF Developer Center: PDF reference*. [Online] 7 2008. [Cited: 11 13, 2009.] http://www.adobe.com/devnet/acrobat/pdfs/PDF32000_2008.pdf.
7. **King, James C.** ISO Ballot for PDF 1.7 Passed! *Inside PDF*. [Online] 12 4, 2007. [Cited: 12 24, 2009.] http://blogs.adobe.com/insidepdf/2007/12/iso_ballot_for_pdf_17_passed.html.
8. **Adobe Systems, Inc.** PDF/X Files and Adobe Creative Suite 3. *Adobe.com*. [Online] 11 6, 2007. [Cited: 12 27, 2009.] http://www.adobe.com/designcenter/acrobat/articles/acr8ap_pdfx_02.html#Table1.
9. **Adobe Systems, Inc.** *PDF as a Standard for Archiving White Paper*. San Jose : Adobe Systems, Inc., 2005.
10. **Adobe Systems, Inc.** Creating PDF/E-ready files using Adobe® Acrobat® to create ISO 24517-1:2008-compliant files. *Adobe.com*. [Online] 3 2008. [Cited: 12 29, 2009.] <http://www.adobe.com/enterprise/pdfs/pdf-eready-guide-ue.pdf>.
11. **Leurs, Laurens.** The PDF/X-4 file format. *Prepressure.com*. [Online] 1 24, 2008. [Cited: 12 27, 2009.] <http://www.prepressure.com/pdf/basics/pdfx-4>.

11. Appendixes

11.1. Calculation of the print cost per page

Attachment Table 1 – Cost calculation for a single-side, color, A4 paper

Item	Quantity	Price	Unitary price	Note
A4 paper 80g	500 pcs	216 CZK	0.4320 CZK	
Developer unit Cyan	1 pc	10290 CZK	0.0181 CZK	for 115000 pages
Developer unit Magenta	1 pc	10290 CZK	0.0181 CZK	for 115000 pages
Developer unit Yellow	1 pc	10290 CZK	0.0181 CZK	for 115000 pages
Toner Cyan	1 pc	2900 CZK	0.1000 CZK	26000 pages on 5% coverage
Toner Magenta	1 pc	2900 CZK	0.1000 CZK	26000 pages on 5% coverage
Toner Yellow	1 pc	2900 CZK	0.1000 CZK	26000 pages on 5% coverage
Drum	1 pc	9111 CZK	0.1012 CZK	for 90000 pages
Total per A4 page			0.8874 CZK	

Attachment Table 2 – Cost calculation for a single-side, Black&White, A4 paper

Item	Quantity	Price	Unitary price	Note
A4 paper 80g	500 pcs	216 CZK	0.4320 CZK	
Developer unit Black	1 pc	6370 CZK	0.0112 CZK	for 570000 pages
Toner	1 pc	1122 CZK	0.0774 CZK	29000 pages on 5% coverage. Usually the page is 10%.
Drum	1 pc	8200 CZK	0.0683 CZK	for 120000 pages
Total per A4 page			0.5889 CZK	

Attachment Table 3 – Cost calculation for a single-side, color, A3 paper

Item	Quantity	Price	Unitary price	Note
A3 paper 80g	500 pcs	209 CZK	0.4180 CZK	
Ink Cyan	130g	1267 CZK	0.1949 CZK	According to an independent test (HTTP://H10088.WWW1.HP.COM/GAP/DOWNLOAD/HP_DJ_T1200_V_CANON_IPF755CTR.PDF), performed by Buyers Laboratory Inc., the line art print on the Canon imagePROGRAF iPF755 consumed 107.7 g of ink to print 200 sheets of the size Arch D
Ink Magenta	130g	1267 CZK	0.1949 CZK	
Ink Yellow	130g	1267 CZK	0.1949 CZK	
Ink Black	130g	1267 CZK	0.1949 CZK	
Ink Matte Black	260g	2534 CZK	0.3898 CZK	
Total per A3 equiv. page			1.5875 CZK	