FREE TECHNOLOGY ACADEMY

Open Standards

General Overview

Free Technology Academy – http://freetechnology.org
Module 1: **Introduction to Free Software and Open Standards**

Table of Contents

Introduction 6
Open Standard definition....7
About standards....7
General introduction to the OS definition discussion....8
Different definitions of “Open Standards” different definitions....9
Definitions given by Governmental bodies definitions....9
The Danish definition of the Danish Government....9
Definition given by the European Union definition....10
Practical problems....10
Definitions given by Aacademics and theorists definitions....12
Bruce Peren's definition....12
Ken Krechmer's definition....15
Definitions given by Open Standards Organiszations definitions....17
ITU-T's definition....17
Internet Engineering Task Force's definition....18
World Wide Web Consortium's definition....18
Digital Standards Organiszation' definition....19
Open Source Initiative's definition....24
The more suitable Open Standards definition....25
Organizsations and standardiszation process....25
Involved organiszations....26
Standard-Setting Organizationorganisations (SSO)....26
The International Organiszation for Standardization (ISO)....27
The International Telecommunication Union (ITU)....27
Open Standards Organiszations....28
The Internet Engineering Task Force....29
The World Wide Web Consortium....29
The Organizsation for the Advancement of Structured Information Standards....30
The Free Standards Group....30
The Institute of Electrical and Electronics Engineers....31
Standardizsation processes....32
De ijure Standards....32
De facto Standards....33
Industry-driven Standards....34
Open Standards analysis....35
Benefits 35
OS and Innovation....36
Open Standards and Information Technologies....37
Open Sstandards and Free Software....38
Open Standards in relation with software patents....40
Software pPatents....40
Policies on pPatents....41
ISO....41
IETF....42
W3C....42

OASIS....43
Summary of pPatent pPolicies of sStandards Organizationorganisations....44
RAND lLicensing tTerms and Free Software lLicenses....44
Why RAND is not appropriate for Open Standards?....45
Patent offerings to the Free Software community....48
Open File Formats....48
Open Standards and file formats....48
Open vs. proprietary formats....49
Open Fformat....49
Proprietary fFormat....50
1. Taking the risk that the recipient may not be able to access the data....51
2. Taking the risk of transmitting confidential information....51
3. Contributing to virus propagation and exposing oneself to the risk of contamination....52
4. Propping up existing monopolies in the domain of electronic communication....52
Lists of open standard formats....53
Office Aapplications Ffile Fformats....53
Microsoft Office Fformats....54
OpenOffice.org and StarOffice Formats....55
Adobe's Portable Document Format....56
PostScript....57
LaTeX....57
Rich Text Format....57
Text documents....58
ASCII....58
HTML/XHTML....58
Graphics/Image File Formats....59
GIF....59
PNG....60
MNG....60
XPM....60
TIFF....61
JPEG JFIF....61
SVG....61
Video file formats....62
Video Containers....62
AVI....62
ASF....63
MOV....63
MP4....63
Ogg....63
Video Codecs....64
MPEG Codecs....64
MPEG-1 Part 2....64
MPEG-2 Part 2....64
MPEG-4 Part 10 (H.264/AVC)....65
Sorensen....65

Windows Media Video....65
Theora....65
Audio file formats....66
WAV....66
FLAC....66
MP3....67
AAC....67
RealAudio....68
Ogg Vorbis....68
Archiving and compression file formats....68
7z....68
Bzip2....69
gzip....69
PAQ....69
Tar....70
ZIP....70
A case of study: OpenDocument and& Office Open XML....70
OpenDocument....71
General description....71
Brief history....72
Implementations and adoption....73
Office Open XML....74
General description....74
Standardiszation process....74
Inside EcmaCMA....74
International process....75
Complaints about irregularities in the process....75
Technical problems and inconsistencies with other standards....77
Criticism on OOXML in relation with OpenDocument....79
OpenDocument vs. OOXML....80
Comparing the codes....80
ODF XML respresentation of the example....80
OOXML XML respresentation of the example....80
Interoperability between both standards....82
Final words....83
Conclusions....84
Glossary....86
Bibliography and links....93
License of this document....100
GNU Free Documentation License....100
0. PREAMBLE....101
1. APPLICABILITY AND DEFINITIONS....101
2. VERBATIM COPYING....103
3. COPYING IN QUANTITY....103
4. MODIFICATIONS....104
5. COMBINING DOCUMENTS....106
6. COLLECTIONS OF DOCUMENTS....107

7. AGGREGATION WITH INDEPENDENT WORKS.....107
8. TRANSLATION.....107
9. TERMINATION.....108
10. FUTURE REVISIONS OF THIS LICENSE.....108
11. RELICENSING.....109
ADDENDUM: How to use this License for your documents.....109

Introduction

This document covers the main concepts around Open Standards and introduces the audience in the complexity of this field. Open standards is a strategic issue for industry, governments and users. Open standards enable industry to provide software and services which meet the customers' requirements today and in the future. In a networked ecosystem there is the need for end-to-end solutions with pieces of software from different vendors working seamlessly together. Hence, the concept of open standards is generally accepted in industry. However, in the software universe, there exists much controversy in relation with information control by private corporations and suppliers who develop and promote non standardized information formats and protocols.

To customers, on the other hand, open standards guarantee a high degree of flexibility and choice as they prevent single vendor lock-in by facilitating interoperability. These and many other issues regarding the Open Standards discussion are covered in the sequel.

This document is structured focusing in four main areas of understandings the topic:

1. Open Standards definitions: minimum requirements, points of controversy between definitions;
2. Open Standards in information storage (open file formats): main characteristics, conditions, analysis;
3. Open Standards case of studies: different cases of study of standardization process and adoption cases;
4. Open Standards catalogue list and classification.

This document is the result of studying and researching over a large amount of existent technical bibliography about standards, open specifications and many others relevant issues. Unlike those texts, this one aims to be an educative material with a wider scope of readers, bringing together different points of views and remixing other sources which address the Open Standards issues with a “plain language” writing style.

This text is aimed at providing the reader with a better understanding as to why open standards are

important and how they can complement Free Software in fostering a more open IT environmenttowards to a “free society”. As users and consumers, the readers and students should demand, from their software, conformance to open standards as far as possible. In addition to promoting interoperability and making more choices available, this will make it easier for Free Software to co-exist and take root in environments filled with proprietary software.

Open Standard definition

“The Internet is fundamentally based on the existence of open, non-proprietary standards”

Vint Cerf, “the father of the Internet”

About standards

The word "standard" has different meanings. Working within the context of the subject matter of this document, its meaning in everyday usage can be taken to refer to:

1. a level of quality or attainment, or
2. an item or a specification against which all others may be measured.

In technical usage, a standard is a framework of specifications which has been:

1. approved by a recognized organisation, or
2. is generally accepted and widely used throughout by the industry.

For the rest of this document, unless specified otherwise, when the word standard is used the technical meaning is implied.

The ISO (International organisation for Standardisation) says:

> “Standards are extremely important in modern society. They ensure that products and services are of adequate quality and that they can interoperate and work together even though they may be from different parties or entities. Ultimately, they raise levels of quality, safety, reliability, efficiency and interoperability, and provide such benefits at an economical cost.”

General introduction to the OS definition discussion

Before going deep in the Open Standard definition section, we must be aware that there is not just one and globally accepted Open Standard definition. Many groups, theorist, declarations, government organisations, companies and legislation bodies have created different definitions for the term. However, in the Free Software community, there are certain central agreed concepts which lead to common base of understanding around the concept. Those base of concepts allows us to have the general definition that will be described in the next session as the more precise and suitable definition. In the following pages, we will introduce each of the most important (and some of them very controversial) existing definitions from the industry, free software communities and non-free approaches.

As a basic definition we can say that an Open Standard is a standard that is publicly available and has various rights of use associated with it, and may also have various properties of how it was designed (*e.g.* open process).

The terms "open" and "standard" have a wide range of meanings associated with their usage. The term "open" is usually restricted to royalty-free technologies while the term "standard" is sometimes restricted to technologies approved by formalized committees that are open to participation by all interested parties and operate on a consensus basis.

In the next sections, we present the various existing definitions, the difference and controversy around them. This deep analysis allows us to reach a common and more suitable definition for the Open Standard concept.

Different definitions of “Open Standards”

As mentioned in the introduction of this chapter, there are various definitions for what should or should

not be considered an Open Standard. Many specifications which are sometimes referred to as standards are proprietary and only available under restrictive contract terms (if they can be obtained at all) from the organisation that owns the copyright on the specification. Those specifications are not considered to be fully “Open”.

The definitions of the term "open standard" used by academics, the European Union and some of its member governments or parliaments such as Denmark, France, and Spain preclude open standards requiring fees for use, as do the New Zealand and the Venezuelan governments. On the standard organisation side, the W3C ensures that its specifications can be implemented on a Royalty-Free (RF) basis.

Many definitions of the term "standard" permit patent holders to impose "reasonable and non-discriminatory" royalty fees and other licensing terms on implementers and/or users of the standard. For example, the rules for standards published by the major internationally recognized standards bodies such as the IETF, ISO, IEC, and ITU-T permit their standards to contain specifications whose implementation will require payment of patent licensing fees. Among these organisations, only the IETF and ITU-T explicitly refer to their standards as "open standards", while the others refer only to producing "standards". The IETF and ITU-T use definitions of "open standard" that allow "reasonable and non-discriminatory" patent licensing fee requirements.

In the next pages we will reference and in some cases analyse the existing definitions. Some of them were defined by government bodies, Standards Setting organisation (SSO), private corporations or individuals from the academic field.

Definitions given by Governmental bodies

The definition of the Danish Government

The Danish Government has attempted to make a definition of open standards which also is used in pan-European software development projects. This definition states that:

- An open standard is accessible to everyone free of charge (*i.e.* there is no discrimination between users, and no payment or other considerations are required as a condition of use of the standard)
- An open standard of necessity remains accessible and free of charge (*i.e.* owners renounce their

options, if indeed such exist, to limit access to the standard at a later date, for example, by committing themselves to openness during the remainder of a possible patent's life)

- An open standard is accessible free of charge and documented in all its details (*i.e.* all aspects of the standard are transparent and documented, and both access to and use of the documentation are free)

This is relatively similar to the definition of an Open Standard by the European Commission in its European Interoperability Framework. We will introduce that definition in the following section.

Definition given by the European Union

The European Union adopted the following definition in its European Interoperability Framework:

To reach interoperability in the context of the pan-European eGovernment services, guidance needs to focus on open standards.

The word "open" is here meant in the sense of fulfilling the following requirements:

- The standard is adopted and will be maintained by a not-for-profit organisation, and its ongoing development occurs on the basis of an open decision-making procedure available to all interested parties (consensus or majority decision, etc.).
- The standard has been published and the standard specification document is available either freely or at a nominal charge. It must be permissible to all to copy, distribute and use it for no fee or at a nominal fee.
- The intellectual property -*i.e.* patents possibly present- of (parts of) the standard is made irrevocably available on a royalty-free basis.
- There are no constraints on the re-use of the standard.

Practical problems

In theory, the definitions of the European Union or the Danish parliament would be sufficient to define an Open Standard. In practice things have proven to be more complicated because the situation with proprietary formats described above is immensely profitable for the vendor in control of that software.

Thus, ultimately, a proprietary vendor with a certain amount of market penetration has an economic incentive to violate the Open Standard and turn it into a de-facto proprietary one. This indeed has happened repeatedly in history. The European Commission antitrust investigation against Microsoft provides testimony to how deviating from an Open Standard (CIFS, the "Common Internet File System") allowed Microsoft to leverage its desktop monopoly into near total dominance on the workgroup server market. This has proven so profitable that Microsoft appears more inclined to pay billions in fines than to stop this practice.

Often this is also done by changing the implementation slightly in ways that are hard to pinpoint or can be debated within the limits of human interpretation, but make sure that the implementations of other vendors will not integrate flawlessly anymore. The economic incentive for this is huge for proprietary players that bypass a certain threshold in size.

Both these definitions were criticised by the vendors who profit commercially from the dependency cycles explained above, as well as organisations representing their interest. The usual argumentation for this criticism is generally oriented along the lines of patents that were granted on such a format or protocol, and for which the patent holder might choose to generate license revenues. The euphemism dujour for this is usually "Reasonable and Non-Discriminatory" (RAND) licensing.

This is but a euphemism because patents are, by their nature, limited monopolies granted by law to a single entity. This entity will always have the upper hand in any dispute, and indeed there are plenty of stories about formats and protocols that are theoretically known, but remain proprietary due to patent issues. The relation of patents and OS is further explained below, but you may also refer to the FTA module documentation on Legal Aspects of the Information Society.

That all other vendors not holding this patent are put in an equally bad position may indeed seem non-discriminatory, but it does not fundamentally change the balance of power of the situation.

In the final definition, we use most of the characteristics described in the European Union and the Danish parliament definitions.

Definitions given by academics and theorists

Bruce Peren's definition

Bruce Perens is a computer programmer and advocate in the open source community. He created the

Open Source Definition and published the first formal announcement and manifesto of open source. He co-founded the Open Source Initiative with Eric S. Raymond. He wrote a famous and widely used definition of OS that combines a clear specification with a set of principles[1]:

"An Open Standard is more than just a specification. The principles behind the standard, and the practice of offering and operating the standard, are what make the standard Open.

Principles

1.Availability

Open Standards are available for all to read and implement.

2. Maximize End-User Choice

Open Standards create a fair, competitive market for implementations of the standard. They do not lock the customer in to a particular vendor or group.

3. No Royalty

Open Standards are free for all to implement, with no royalty or fee. Certification of compliance by the standards organisation may involve a fee.

4. No Discrimination

Open Standards and the organisations that administer them do not favor one implementor over another for any reason other than the technical standards compliance of a vendor's implementation. Certification organisations must provide a path for low and zero-cost implementations to be validated, but may also provide enhanced certification services.

5. Extension or Subset

Implementations of Open Standards may be extended, or offered in subset form. However, certification organisations may decline to certify subset implementations, and may place requirements upon extensions (see Predatory Practices).

6. Predatory Practices

Open Standards may employ license terms that protect against subversion of the standard by embrace-and-extend tactics. The licenses attached to the standard may require the publication of reference information for extensions, and a license for all others to create, distribute, and sell software that is compatible with the extensions. An Open Standard may not othewise prohibit extensions.

1See: http://www.dwheeler.com/essays/opendocument-open.html

Practices

1. Availability

Open Standards are available for all to read and implement. Thus:

- The best practice is for the standards text and reference implementation to be available for free download via the Internet.
- Any software project should be able to afford a copy without undue hardship. The cost should not far exceed the cost of a college textbook.
- Licenses attached to the standards documentation must not restrict any party from implementing the standard using any form of software license.
- The best practice is for software reference platforms to be licensed in a way that is compatible with all forms of software licensing, both Free Software (Open Source) and proprietary. However, see *Predatory Practices* regarding license restrictions that may be appropriate for a software reference platform.

2. Maximize End-User Choice

Open Standards create a fair, competitive market for implementations of the standard. Thus:

- They must allow a wide range of implementations, by businesses, academia, and public projects.
- They must support a range of pricing from very expensive to zero-price.

3. No Royalty

Open Standards are free for all to *implement,* with no royalty or fee. *Certification* of compliance by the standards organisation may have a fee. Thus:

- Patents embedded in standards must be licensed royalty-free, with non-discriminatory terms.
- Certification programs should include a low or zero cost self-certification, but may include higher-cost programs with enhanced branding.

4. No Discrimination

- A standards organisation that wishes to support itself through certification branding should establish a premium track and a low-cost or zero-cost track.

Generally, the premium track will provide a certification lab outside of the vendor's facility to verify a vendor's implementation and enhanced branding: a certification mark that indicates a greater certainty of verification and financial support of the standard. The low or zero-cost track would provide self-certification by the vendor and baseline branding.

Open Standards and the organisations that administer them do not favor one implementer over another for any reason other than the technical standards compliance of a vendor's implementation. Certification organisations must provide a path for low and zero-cost implementations to be validated, but may also provide enhanced certification services. Thus:

5. Extension or Subset

Implementations of Open Standards may be extended, or offered in subset form. However, certification organisations may decline to certify subset implementations, and may place requirements upon extensions (see *Predatory Practices*).

6. Predatory Practices

Open Standards may employ license terms that protect against subversion of the standard by *embrace-and-extend* tactics. The license may require the publication of reference information and an license to create and redistribute software compatible with the extensions. It may not prohibit the implementation of extensions.

- The standards organisation may wish to apply an agreement similar to the *Sun Industry Standards Source License* to the standard documentation and its accompanying reference implementation. The Sun agreement requires publication of a reference implementation (not the actual commercial implementation) for any extensions to the standard. This makes it possible for a standards organisation to actively preserve interoperability without stifling innovation.

Ken Krechmer's definition

Ken Krechmer is a technician and consultant, involved in the International Telecommunications Union Recommendations and fellow researcher in the International Center for Standards Research, University of Colorado. Krechmer researching work on Open Standards has allowed him to contribute a large number of academic articles of great relevance in the field. One of his most important papers called "Open Standards requirements" defines a set of ten requirements for open standardization that expands the definition of open standards further to include not only economic effects resulting from an open standard's implementation and openness in the process of standards setting, but also the concept of openness in use.

The text bellow is a extract of the main content of Krechmer's definition:

> The term Open Standards may be seen from the following three perspectives:
>
> 1. The formal SSOs, as organisations representing the **standards creators**, consider a standard to be open if the creation of the standard follows the tenets of open meeting, consensus and due process.
> 2. An **implementer** of an existing standard would call a standard open when it serves the markets they wish, it is without cost to them, does not preclude further innovation (by them), does not obsolete their prior implementations, and does not favor a competitor.
> 3. The **user** of an implementation of the standard would call a standard open when multiple implementations of the standard from different sources are available, when the implementation functions in all locations needed, when the implementation is supported over the user?s expected service life and when new implementations desired by the user are backward compatible to previously purchased implementations.
>
> These are the very different views from the creators, implementers and users of standards on what is an Open Standard. Their combined, reasonable, but not simple expectations translate into ten rights that enable Open Standards:
>
> 1. Open Meeting - all may participate in the standards development process.
> 2. Consensus - all interests are discussed and agreement found, no domination.
> 3. Due Process - balloting and an appeals process may be used to find resolution.
> 4. Open IPR - IPR related to the standard is available to implementers.
> 5. One World - same standard for the same capability, world-wide.

6. Open Change - all changes are presented and agreed in a forum supporting the five rights above.

7. Open Documents - committee drafts and completed standards documents are easily available for implementation and use.

8. Open Interface - supports migration and allows proprietary advantage but standardized interfaces are not hidden or controlled.

9. Open Use - objective conformance mechanisms for implementation testing and user evaluation.

10. On-going Support - standards are supported until user interest ceases rather than when implementer interest declines (use).

Comparing these ten rights with the economic rights identified by J. West [10] suggests that economic rights require further rights to exist. Economic rights cannot be maintained without supporting the political rights basic to a fair political process: balance, consensus and due process. In order for the economic rights associated with compatibility to be available, some technical process (Open Changes) and technical functionality (Open Interfaces) are also required. In order for specific economic rights associated with Intellectual Property Rights (IPR) to be available, specific SSO procedures must be used.

Comparing these ten rights to the six principles proposed by B. Perens [11], the following equivalences can be found:

Availability is addressed by Open Documents.

Maximum end-user choice is addressed by Open Use.

No royalty is addressed under Open IPR.

No discrimination is addressed by Open Meeting, Consensus and Due Process.

Ability to create extension or subset is addressed by Open Interface.

Ability to prevent predatory practices is addressed by Open Change.

The six principles proposed by B. Perens map fully onto eight of the ten rights of Open Standards proposed. B. Perens does not directly address the desires for or against One World or the end user right of On-going Support. This is one affirmative test of the completeness of the rights of Open Standards proposed.

Definitions given by Open Standards Organisations

ITU-T's definition

The ITU-T is a standards development organisation (SDO) which is one of the three sectors of the International Telecommunications Union. It is worth remarking that ITU-T is not a public organisation but an association of communication corporations. The ITU-T has a "Telecommunication Standardisation Bureau director's Ad Hoc group on Intellectual Property Rights" which produced the following definition in March 2005, which the ITU-T as a whole has endorsed for its purposes since November 2005. Below you can read the ITU-T "Open Standards" interpretation :

> "Open standards" are standards made available to the general public and are developed (or approved) and maintained via a collaborative and consensus driven process. "Open Standards" facilitate interoperability and data exchange among different products or services and are intended for widespread adoption.
>
> Other elements of "Open Standards" include, but are not limited to:
>
> - Collaborative process –voluntary and market driven development (or approval) following a transparent consensus driven process that is reasonably open to all interested parties.
> - Reasonably balanced –ensures that the process is not dominated by any one interest group.
> - Due process –includes consideration of and response to comments by interested parties.
> - Intellectual property rights (IPRs) –IPRs essential to implement the standard to be licensed to all applicants on a worldwide, non-discriminatory basis, either (1) for free and under other reasonable terms and conditions or (2) on reasonable terms and conditions (which may include monetary compensation). Negotiations are left to the parties concerned and are performed outside the SDO.
> - Quality and level of detail –sufficient to permit the development of a variety of competing implementations of interoperable products or services. Standardised interfaces are not hidden, or controlled other than by the SDO promulgating the standard.
> - Publicly available –easily available for implementation and use, at a reasonable price. Publication of the text of a standard by others is permitted only with the prior approval of the SDO.
> - On-going support –maintained and supported over a long period of time.

The ITU-T, ITU-R, ISO, and International Electrotechnical Commission (IEC) have harmonized on a common patent policy under the banner of the World Standards Cooperation (WSC). However, the

ITU-T definition should not necessarily be considered also applicable in ITU-R, ISO and IEC contexts, since the Common Patent Policy does not make any reference to "Open Standards" but rather only to "standards".

There exist controversy around the ITU-T definition as it accepts that an “standard” that can be covered by patents.

Internet Engineering Task Force's definition

The Internet Engineering Task Force (IETF) develops and promotes Internet standards, cooperating closely with the W3C and ISO/IEC standard bodies and dealing in particular with standards of the TCP/IP and the Internet protocol suite. It is an Open Standards organisation, with no formal membership or membership requirements. All participants and managers are volunteers, though their work is usually funded by their employers or sponsors; for instance, the current chairperson is funded by VeriSign and the U.S. Government's National Security Agency.

In section 7 of its RFC 2026, the IETF classifies specifications that have been developed in a manner similar to that of the IETF itself as being "Open Standards", and lists the standards produced by ANSI (American National Standards Institute), ISO, IEEE, and ITU-T as examples. As the IETF standardization processes and IPR policies have the characteristics listed above by ITU-T, the IETF standards fulfill the ITU-T definition of "Open Standards".

However, the IETF has not adopted a specific definition of "Open Standard"; both RFC 2026 and the IETF's mission statement (RFC 3935) talks about "open process", but RFC 2026 does not define "Open Standard" except for the purpose of defining what documents the IETF standards can link to.

World Wide Web Consortium's definition

The World Wide Web Consortium (W3C) is the main international standards organisation for the World Wide Web (abbreviated WWW or W3). W3C engages in education and outreach, develops software and serves as an open forum for discussion about the Web.

As an important provider of Web technology ICT Standards, notably XML (Extensible Markup Language), Hypertext Transfer Protocol (HTTP), HTML, CSS and WAI, the World Wide Web Consortium (W3C) follows a process that promotes the development of high-quality standards.

Looking at the final result, the specification alone, up for adoption, is not enough. The participative/inclusive process leading to a particular design, and the supporting resources available with it should be accounted when we talk about Open Standards:

- Transparency (due process is public, and all technical discussions, meeting minutes, are archived and referable in decision making)
- Relevance (new standardisation is started upon due analysis of the market needs, including requirements phase, *e.g.* accessibility, multilingualism)
- Openness (anybody can participate, and everybody does: industry, individual, public, government bodies, academia, on a worldwide scale)
- Impartiality and consensus (guaranteed fairness by the process and the neutral hosting of the W3C organisation, with equal weight for each participant)
- Availability (free access to the standard text, both during development and at the final stage, translations, and clear IPR rules for implementation, allowing open source development in the case of Internet/Web technologies)
- Maintenance (ongoing process for testing, errata, revision, permanent access)

Digital Standards Organisation' definition

The Digital Standards Organisation (Digistan) “seeks to promote customer choice, vendor competition, and overall growth in the global digital economy through the understanding, development, and adoption of free and open digital standards ("open standards")”. This organisation have arrived a well defined concept they named “Free and Open Standard”: [2]

> We, the Digital Standards Organisation, explicitly take the side of "the market at large". We do not accept the definitions of "Open Standard" produced by vendor bodies, including W3C to some extent. We do not accept the attempts of some legacy vendors to stretch "Open Standard" to include RAND-licensed standards.
>
> An open standard must be aimed at creating unrestricted competition between vendors and unrestricted choice for users. Any barrier –including RAND, FRAND, and variants– to vendor competition or user choice is incompatible with the needs of the market at large.

2 See: http://www.digistan.org/open-standard:definition

Vendor capture

Looking at the standards "wars" that plague the telecoms and media industries, we see that the fight is not over technical quality or process. The war is to capture markets through the use of standards backed by patent pools. The goal of vendors who engage in such standards processes is to become part of the cartel and to kill or capture the other cartels.

Vendor capture is thus a persistent and fundamental danger to a standardisation process. Incompetence and human error are fixable by collective processes and time, as long as standards are free to improve. Vendor capture occurs at any stage in the life-cycle of a standard, from development through to long-term use:

- Vendors capture the standards development groups and exclude open participation in the development process. The standards which result become dumps of the vendor's software, which gives significant advantage to vendor.
- Vendors capture the standard specification through copyright and reduce competition from smaller competitors by making it expensive to simply acquire the text.
- Vendors capture the market by introducing patented technology into the standard, silently or explicitly. When many vendors do this, they create franchise standards ("reasonable and non-discriminatory" at the best of times) that enable cartel-like constructions.
- Vendors capture the market by silently extending Open Standards. Users who thought they were getting an Open Standard may instead be buying lock-in.
- Vendors capture the standards authorities. They do this by setting up their own (like ECMA) or lobbying existing ones (ISO) to promote and accept their advantageous standards.
- Vendors capture existing Open Standards. They do this by acquiring key patents, long after the market has committed to the standard. Often, such patents were held originally to "defend" the standard.

It is not enough for a standard to be open, it must also be resistant against all these forms of attack.

Free and Open Standard

The term "open" itself has many degrees of meaning from a high wall with a crack in the door to a field with no walls at all. We wish to secure the terminology by adding the word "free", in both its meanings:

- Freedom to use, improve upon, trust, and extend a standard over time.
- Freedom from all costs and tariffs associated with the above freedoms.

The ambition of freedom in this case is the removal of barriers, of friction, and of costs. It is also the ambition of remaining free over time, especially as the value of the market based on the standard increases.

A canonical definition

From the above analysis that the interests of established vendors diametrically oppose those of the market at large, we can make a simple canonical definition of "Free and Open Standard" as follows:

- A standard is a published specification.
- It is a a Free and Open Standard if it is immune to vendor capture at all stages in its life-cycle.

This definition deliberately expresses the threat. A less direct definition ignores the threat and lists the *primary properties* of a Free and Open Standard and its process:

- The standard is adopted and will be maintained by a not-for-profit organisation, and its ongoing development occurs on the basis of an open decision-making procedure available to all interested parties.
- The standard has been published and the standard specification document is available freely. It must be permissible to all to copy, distribute, and use it freely.
- The patents possibly present on (parts of) the standard are made irrevocably available on a royalty-free basis.
- There are no constraints on the re-use of the standard.

Measuring immunity

Immunity from vendor capture, if this is indeed the defining characteristic of a Free and Open Standard, can be measured. This is very useful because it gives us a tool to measure how "free and open" a standard is, and to offer concrete suggestions for improving any given standards process or specification.

We look at each aspect of the standardisation life-cycle, from development to wide-spread use, ignoring quality issues, which we assume an open process will detect and improve.

The development process

In which experts agree to work together to standardise an area of technology:

- What is the cost of entry for experts? If not very low, independent experts are excluded.
- Do vendors form the majority of developers? If so, the standard will be rapidly captured.

- Do all developers grant their copyrights and patents to the standard by contract? If not, copyright and patent claims may be used later to capture the standard.
- Can others freely fork the standard under a share-alike license? If not, a captured standard cannot be freed.
- Is the standard published frequently during development? If not, external reviewers are excluded and vendors can more easily capture the standard.
- Is the process transparent and open? If so, vendor capture is easier to see and correct.
- Is there a process for accepting specification translations? If not, foreign vendors are put at a disadvantage.
- Is the development process hosted by a not-for-profit organisation? If so, it is relatively safe from vendor capture.

While "forking a specification" may seem counter-productive, we consider this an essential freedom at this stage in the standard life-cycle. Sometimes, a standard will become "loaded" with features that favor one vendor and exclude others, and the best way forward is to fork the standard, strip-out those features, and create a competing standards process. The simple possibility of such competition forces a more honest process.

The implementation process

In which vendors - often also standards developers –take provisional specifications and use them to make products:

- Is the specification text clear and well-engineered? Vendors often inject complexity into specifications to create barriers to competing implementations.
- Are there reference implementations freely available? This promotes a level playing field. Copy-left implementations ensure that closed source derivatives that silently modify the standard are not possible.
- Is the standard free to implement for any purpose? If particular license conditions apply, this is easily used to exclude competitors.
- Can the specification be freely acquired and distributed? If not, cost can be used to create barriers for smaller competing vendors.
- Is there a process for contributing change proposals back into the standard? If not, it is easier for

vendor-developers to capture the standard.

- Are there outstanding patent or copyright claims on the standard? If so, vendors are unable to freely implement the standard.

The certification process

In which the standards group, or other groups, help define what is a valid implementation and help the market measure and compare implementations:

- Are there conformance tests freely available? This promotes a level playing field among vendors. We would argue that copy-left licenses are most appropriate for conformance tests.
- Is certification done by a not-for-profit body? If not, it is trivially captured through acquisition.
- Are there clear rules for building subsets and extensions? If this is part of the standard specification itself, users can demand better behaviour from vendors.
- Are trademarks used to enforce implementations? If not, it is easier to capture users through extended-subset implementations.
- Are trademarks held by a not-for-profit? If so, they are more immune from capture.

The deployment process

In which users use standards-based products to construct their own business infrastructure:

- Is there a real market of competing vendors? If not, this is a sign that the standard has been captured.
- Is the standard mature and well-certified? If not, users can be captured by vendors who sell premature, incompatible implementations.
- Is the standard free from patent claims? If not, users are trapped by using the standard in any way at all.

The authorisation process

In which an authority declares its support and recognition of the standard:

- Is the authority backed by the state, or by vendors? Vendor-backed authorities are by definition captive, and will readily promote franchise standards as "open".
- Is the authority meritocratic? That is, are its decisions taken by experts with proven track records in standards recognition? If not, the authority can be easily captured by vendors.

- Does the authority have clear, transparent rules on copyright and patent claims? If not, franchise standards can be injected, capturing users by deceit.
- Does the authority have a clear definition of "Open Standard" that represents the economic interests of the market at large? If not, vendors can more easily defeat Open Standards with franchise standards.

Open Source Initiative's definition

The Open Source Initiative (OSI) defines the requirements and criteria for open standards as follows:

The Requirement

An "Open Standard" must not prohibit conforming implementations in open source software.

The Criteria

To comply with the Open Standards Requirement (OSR), an "Open Standard" must satisfy the following criteria. If an "Open Standard" does not meet these criteria, it will be discriminating against open source developers.

1. No Intentional Secrets: The standard MUST NOT withhold any detail necessary for interoperable implementation. As flaws are inevitable, the standard MUST define a process for fixing flaws identified during implementation and interoperability testing and to incorporate said changes into a revised version or superseding version of the standard to be released under terms that do not violate the OSR.
2. Availability: The standard MUST be freely and publicly available (*e.g.*, from a stable web site) under royalty-free terms at reasonable and non-discriminatory cost.
3. Patents: All patents essential to implementation of the standard MUST:
    - be licensed under royalty-free terms for unrestricted use, or
    - be covered by a promise of non-assertion when practiced by open source software
4. No Agreements: There MUST NOT be any requirement for execution of a license agreement, NDA, grant, click-through, or any other form of paperwork to deploy conforming implementations of the standard.
5. No OSR-Incompatible Dependencies: Implementation of the standard MUST NOT require any

other technology that fails to meet the criteria of this Requirement.

The more suitable Open Standards definition

As described in the previous sections, there are various definitions of Open Standards, even more than those listed above. As a first conclusion, we can determine that a global unique understanding of the term “Open Standards” is still lacking. However there exist some minimal agreement in most of the definitions. We can affirm that the best existing definition are one provided by Ken Kretchmer and the European Interoperability Framework v1.

As a result of the previous analysis we can determine a common understanding of the concept. We understand Open Standards as follows.
An Open Standard refers to a format or protocol that is

1. subject to full public assessment and use without constraints in a manner equally available to all parties;
2. without any components or extensions that have dependencies on formats or protocols that do not meet the definition of an Open Standard themselves;
3. free from legal or technical clauses which limit its utilization by any party or in any business model;
4. managed and further developed independently of any single vendor in a process open to the equal participation of competitors and third parties;
5. available in multiple complete implementations by competing vendors, or as a complete implementation equally available to all parties.

Standards which sufficiently meet all the above criteria will be classified as “Open Standard”.
When a new format or protocol is emerging, clause 5 cannot possibly be met. So in these special cases the format or protocol can be recognized as an ``Emerging Standard´´ for a limited amount of time.

Organisations and standardisation process

In this chapter, we describe the most relevant actors of this field:

1. Standard-Setting Organisations (SSO) and
2. Open Standards Organisations

In the second section of this chapter, the standarisation process, categories and other procedures are introduced.

Involved organisations

In the last sections, we meet different definitions for the Open Standard concept. Some of them were defined by specific Open Standards organisations. In the next section of the text we will describe the most important ones. We difference between Standard-Setting Organisations (SSO) and Open Standards Organisations.

Standard-Setting organisations (SSO)

In this document, the term Standard-Setting Organisation (SSO) is taken to refer to an organisation which attempts to set standards or make recommendations which, *when widely* deployed, become *de facto* standards. There are many SSOs, national, regional as well as industry-based. A formal SSO refers to one which is recognised directly or indirectly by a governmental entity. Very often, there exists a formal SSO in a country which the government recognises as the national standards body and which has the authority to designate a specification as the national standard for the country. Thus, for example, in India, the Bureau of Indian Standards (BIS) is the national standards body; in the USA, the American National Standards Institute (ANSI) is the official body; while in the United Kingdom, it is the British Standards Institute (BSI).

While any organisation can come up with its own specification and call it its standard, to be an internationally acceptable standard, it has to be either set or adopted/adapted by an SSO which is recognised as an international standard-setting body. The three organisations having the highest international recognition are the International organisation for Standardisation (ISO), the International Electro-technical Commission (IEC) and the International Telecommunication Union (ITU).

ISO is an international standard-setting body made up mainly of representation from national standards bodies. IEC is a standards organisation that deals mainly in setting standards for electrical, electronic

and related technologies. A body that is an accredited representative to ISO or IEC is called a Standard Development Organisation (SDO); most national standards bodies are SDOs. ISO produces standards in many domains, including IT. Many of its standards are also developed jointly with IEC, in particular, the ISO/IEC Joint Technical Committee 1 (JTC 1) is active in setting standards for the IT domain.

The International Organisation for Standardization (ISO)

ISO is a non-governmental organisation for standards with its secretariat in Geneva, Switzerland. Membership of ISO is open only to national standards institutes or similar organisations most representative of standardisation in their country (one member in each country). Currently, there are over 150 members representing countries from all over the world.

ISO sets standards for a wide variety of industries ranging from agriculture to rubber and plastics and to IT. Standards approved by the ISO are agreed upon (by consensus) between national delegations representing all the economic stakeholders concerned –suppliers, users and governments. ISO standards are usually regarded as international standards.

The International Telecommunication Union (ITU)

ITU, one of the world's oldest international standards bodies, was established to standardise and regulate international radio and telecommunications. With the convergence of IT and telecommunications, ITU (specifically its Telecommunication Standardization Sector, ITU-T) is now also involved in specifying standards (or Recommendations as it calls them) which impact the ICT world.

ITU has its headquarters in Geneva, Switzerland, and it is an international organisation within the UN System where governments and the private sector coordinate global telecom networks and services. It started out as the International Telegraph Union in 1865 to facilitate the interoperability of the then-fledgling telegraphy system among countries. From there it has grown and evolved to the ITU of today, which is involved in the standardisation and regulation of international radio and telecommunications.

Membership of the ITU is open to governments as well as to private organisations involved in the telecommunications industry, *e.g.* carriers, equipment manufacturers, large telecommunication organisations, research bodies, etc.

ITU is divided into three sectors: Radio Communication (ITU-R), Telecommunication Standardisation (ITU-T), and Telecommunication Development (ITU-D). ITU-T is increasingly becoming an important international body for the development of IT standards due to the convergence of IT and Telecommunications.

ISO sets standards for a wide variety of industries ranging from agriculture to rubber and plastics and to IT. Standards approved by ISO are agreed upon (by consensus) between national delegations representing all the economic stakeholders concerned suppliers, users and governments. ISO standards are usually regarded as international standards.

Open Standards Organisations

Bodies dealing with standards are usually non-profit and may be government-appointed, industry-backed, non-government organisations or even voluntary ones. While almost all of these claim to be "open" , some are more open than others especially with respect to the free and easy accessibility and open participation criteria discussed in the Introduction. Some more active organisations that are generally perceived to be open include IETF, IEEE (Institute of Electrical and Electronics Engineers), OASIS, W3C and the Free Standards Group (FSG).

Note that this list is by no means an exhaustive listing of open standards bodies and indeed some may dispute the inclusion of one or more of these and/or the exclusion of other bodies if the accessibility and open participation criteria are applied strictly. However, in terms of important IT standardisation activities and relative "openness" to world-wide participation and access by organisationbig and small organisations, the organisations listed above do stand out.

Standards and/or recommendations from these bodies account for many of the standards being deployed or developed in the IT and Internet/Web industries. Many of these standards have also been adopted as standards by international SSOs like ISO.

As noted above, these non-formal SSOs often have liaisons, especially at the technical working group level, with formal organisations such as ISO and ITU-T. Therefore, there is awareness and knowledge of the work and activities of the respective working groups from the various organisations working in the same area.

The Internet Engineering Task Force

Internet networking standards and protocols, like TCP/IP, became *de facto* standards when the Internet was widely embraced throughout the world. IETF is charged with developing and promoting Internet standards. It is a voluntary organisation with membership open to any interested individual. The actual technical work of IETF is done by its working groups which are formed, based on topics, into several key areas. Each area is overseen by an area director and the area directors, together with the IETF Chair, form the Internet Engineering Steering Group (IESG), which is responsible for the overall operation of IETF. ETF is overseen by the Internet Architecture Board (IAB) which is, in turn, responsible to the Internet Society (ISOC).

The drafting and setting of specifications and standards by IETF is carried out considerably faster when compared to the formal SSOs. IETF working groups do the drafting work. A new set of specifications starts off as an Internet Draft which is placed in IETF's "Internet-Drafts" directory and also replicated on a number of Internet hosts. Interested parties are encouraged to comment on this, usually through the working group's mailing lists. Based on comments and feedback, the draft undergoes several rounds of modification and then moves on to become a Requests for Comments (RFC) document and is published.

The specifications in a RFC document may be implemented by the Internet community and it can become a *de facto* standard if it receives wide acceptance. An RFC specification for which significant implementation and successful operational experience have been obtained may be elevated to the Internet standard level and is assigned a number in the STD series while retaining its RFC number.

The World Wide Web Consortium

W3C is an international consortium which specialises in the development of protocols and guidelines for use on the World Wide Web. It is the leading body for specifications on Web technologies and applications. It calls its guidelines and specifications "Recommendations" which it considers as equivalent to Web standards. Many W3C Recommendations have been submitted to a formal standards body like ISO to become international standards.

W3C believes in complete interoperability for the Web to function and realise its full potential. Towards this end, it publishes open standards for Web languages and protocols. This makes it possible for Web technologies to be compatible with one another and to allow any hardware and software used

to access the Web to work together.

W3C is an independent body, membership is open to any organisation and there are several categories of membership depending on the nature of the organisation. W3C counts vendors of technology products and services, content providers, corporate users, research laboratories, standards bodies and governments among its members. Individuals who are not employees of W3C member organisations can also be involved by participating in the technical discussions in its many public mailing lists.

The Organisation for the Advancement of Structured Information Standards

OASIS is a non-profit, international consortium which drives the development, convergence, and adoption of e-business standards. Standards produced by the OASIS include those for security, Web services, conformance, business transactions, supply chain, public sector, and interoperability within and between marketplaces.

Membership of OASIS is open to both individuals and organisations all over the world. There are several types of membership and OASIS has a diverse membership base, counting users and vendors, governments and universities, trade groups and service providers among its members.

OASIS prides itself on its transparent governance and operating procedures. The members themselves set the OASIS technical agenda using a process designed to promote consensus and unite disparate efforts. Completed work is ratified by open ballot before it is published as an OASIS standard.

The Free Standards Group

Free Standards Group (FSG) is an independent, non-profit organisation dedicated to accelerating the use of Free Software by developing and promoting standards. It is supported by both commercial corporations in the IT industry as well as the Free Software development community. All standards produced by FSG are freely available and are distributed under free software licenses. Anyone can participate in and contribute to the FSG standards development by participating in the various FSG standards projects mailing lists.

The FSG is responsible for the important Linux Standard Base (LSB) standardisation activity and the Open Internationalisation (OpenI18N) initiative. Some LSB specifications have been submitted to the ISO/IEC JTC1 SC22 working group on GNU/Linux standardisation.

The Institute of Electrical and Electronics Engineers

The IEEE is a non-profit, technical, professional association of more than 360,000 individual members in over 175 countries. The IEEE Standards Association (IEEE-SA) is active in the development of technical standards in the fields of information technology, telecommunications, and energy and power. IEEE standards development is guided by the five basic principles of due process, openness, consensus, balance and right of appeal; it is open to all and not restricted to a particular type or category of participants.

The working groups which are developing the standards are open to the public and have well-publicised procedures regarding membership, voting, officers, record-keeping and other areas. They try to be as transparent as they can, agendas for meetings are distributed beforehand and the results of a group's deliberations are publicly available, usually through meeting minutes.

When a draft standard is deemed mature enough, it goes up for balloting to become an IEEE standard. The sponsor of the standard forms a balloting group by inviting people from an "invitation pool" The latter consists of IEEE-SA members or people who have paid a ballot fee and are interested in balloting some of the draft standards. Unlike the development stage where anyone can contribute comments, only members of the balloting group can vote in the ballot. The ballot sponsor has to take care that the balloting group is balanced with no domination by any one group or company.

Many IEEE standards have found international recognition and usage, *e.g.* the IEEE 802 series of LAN/ MAN networking standards like 802.3 (Ethernet) and 802.11 (Wireless Fidelity, Wi-Fi).

Standardisation processes

The setting or creation of new technical standards can basically follow several main processes: *de iure*, *de facto*, and industry-created standards.

De iure Standards

De iure standards are usually created by formal SSOs following procedures which have been established by these bodies. Based on a need, work on the creation of a new standard is proposed by one or more members of the organisation. This is called a new work item proposal. If there is enough support, work on drafting the new standard is started by a small committee or working group. The

working draft may go through several cycles of deliberation, voting and modifications by the working group members (as far as possible, a consensus among the members is usually sought) before it is released as a draft to other members of the main organisation or committee for scrutiny. At this level, it may be sent back to the working group for further changes and the cycle repeated until it is accepted as a draft standard for publication by the organisation. Once it is published, it becomes a formal standard from the organisation.

In SSOs, like ISO, the final acceptability of the draft is determined by a formal vote from the participating national bodies. After this final round of voting, the draft document is published.

The advantage of such a process as described above is that formal and accountable procedures are followed and each step in the process is accomplished through consensus as far as possible. The members of the SSO are given an opportunity to contribute during the drafting of the document. Some SSOs also allow contributions from invited subject-matter experts. The idea is that everyone interested in the standard should participate and the standards creation process be seen as neutral and transparent, not controlled by any particular group or party.

There are several disadvantages to the process involved in the creation of *de iure* standards. First of all, the entire standard drafting process can be quite long because of the structure and makeup of the formal SSOs. For example, in the case of ISO standards, there is commonly a time span of two to three years from the new work item proposal to the publication of a standard.

While the standard-setting process formally tries to be neutral and impartial to any group, in practice this may not be so. In some cases, vendors and commercial organisations will send their experts to participate and push their own agendas, *e.g.* the inclusion of the specifications of their particular technology into the standard. Also some formal SSOs, like ISO, allow participation mainly by the national standards body only, so direct participation is restricted. However, interested parties should be able to participate at the local level via their national standards body, that will then carry the so-called national viewpoints, which may or may not concur with those of the interested parties.

The publication of a de iure standard by no means guarantees its success in implementation and acceptance by the industry and users. Sometimes, a simpler and more practical standard from the industry may win over a more complex and difficult to implement standard simply because implementation is simpler and faster, which results in better acceptance in the industry. A classic example of this is the highly complex but more complete X.400 suite of messaging protocols which is

not widely used today as compared with the simpler but more easily implemented SMTP (Simple Mail Transfer Protocol) mail protocol that forms the backbone of Internet e-mail. The former was developed by the formal SSOs, ISO and ITU-T, while the latter came from the industry-driven IETF body.

Examples of internationally recognized SSOs which are active in putting out *de iure* standards are ISO, IEEE, ITU-T and ANSI. Examples of widely used de iure standards include:

1. EEE 802 –a set of standards for Local Area Networking (LAN)
2. ISO 10918 –a standard for the JPEG (Simple Mail Transfer Protocol) graphics compression and file format
3. ITU-T X.25 –a standard for packet switching networks

Not all standards are created from scratch. Very often, an entity (*e.g.* an industry forum or group) may propose that a standards body, like ISO, adopt or adapt its standard or specification as an international standard. Sometimes a *de facto* standard may also be submitted to a standards body for adopting/ adapting as an international standard.

De facto Standards

In the fast-moving IT industry, very often, some technology or product may become so popular that, as a result, it becomes generally accepted and widely used by a majority of users throughout the industry. As a result of this, a *de facto* standard is established that everybody seems to follow as though it was an authorised standard from a standards body. Examples of these are:

1. The FAT (File Allocation Table) file system from Microsoft
2. The Hayes command set for dial-up modem control
3. The Hewlett-Packard Printer Command Language (PCL)

The main advantage of a *de facto* standard is that widespread acceptance in its implementation and usage is assured. It is unlike a *de iure* standard where the standard is just debated and agreed upon by the committee of the SSO and hence industry acceptance is by no means guaranteed.

Since a *de facto* standard does not have to wait for committee debate and approval, changes and modifications are made much faster. Indeed, very often it tends to change as and when the product is upgraded or improved.

The main disadvantage of a standard set in this way is that, very often, it starts off as part of a product

implementation and as such will invariably include some technology and/or specification that is either owned or controlled by the vendor or group that produces the product. Unless that party is willing to give up control or at least share the control by allowing other stakeholders to be involved in developing and driving the *de facto* standard easily, there is a possibility of a lock-in later.

In some cases, after some time, a *de facto* standard may be submitted to a more independent standards body for adoption or adaptation whereby the proprietary control is relinquished and it may then become a real open standard. An example of this is the Network File System (NFS) that was originally introduced by Sun Microsystems as a means of allowing a user to access a file on a remote machine in a way similar to how a local file is used. Later on, with the widespread usage of NFS even on other vendors' systems, it became part of the TCP/IP application standards from the IETF.

Industry-driven Standards

These are sort of intermediate between the *de iure* standards set by formal standards bodies and product based *de facto* standards set mainly by vendors and owners of products. There is a trend nowadays in the IT industry for various consortia or groups to be formed among stakeholders in a particular segment of the industry. One of the functions of such a group may be to develop standards and/or recommendations deemed important and necessary for the progress of the sector. A good example of such a group is OASIS. OASIS is a not-for-profit, international consortium which drives the development, convergence, and adoption of e-business standards. It produces many Web services and Internet-related standards for e-business deployment, such as Universal Description, Discovery and Integration (UDDI) and OpenDocument Format for Office Applications. The W3C is another consortium which has influence in the Web industry. It develops interoperable technologies (specifications, guidelines, software, and tools) for Web usage, e.g. HTML, XML, SOAP, etc. Although it is not a formal standard-setting body, it does come out with recommendations on Web technologies and services that are followed by many developers and/or vendors.

While the industry may adopt and support many of the standards or recommendations from these industry consortia as *de facto* standards, the established ones are eventually submitted to be adopted by traditional international standards organisation like ISO to become a "legitimate" international standard. Many of these industry bodies have on-going liaisons with the technical committees of the international SSOs.

Open Standards analysis

Benefits

Numerous benefits are obtained if an organisation ensures that its technological and IT procurements and implementations follow open standards as far as possible. First and foremost, there is less chance of being locked in by a specific technology and/or vendor. Since the specifications are known and open, it is always possible to get another party to implement the same solution adhering to the standards being followed. Another major benefit is that it will be easier for systems from different parties or using different technologies to interoperate and communicate with one another. As a result, there will be improved data interchange and exchange. It will not be necessary to use the same software or software from a particular vendor to read or write data files. For example, if a multinational organisation requires that all its offices worldwide use office software applications which can read and write files using the Open Document format –an open, standardised XML-based file format from the organisation for the Advancement of Structured Information Standards (OASIS).[1] An individual office will have the flexibility of using whatever office software which is best suited for it and at the same time be able to read, write and exchange documents with other offices in the organisation. **The open formats for office document storage is a relevant topic in the Open Formats discussion. We analyse this specific topic in the Case Study section.**

Using open standards will also offer better protection of the data files created by an application against obsolescence of the application. If the data file format used is proprietary then, should the application become obsolete, the user may have a difficult time converting the data files to another format needed by a new application. However, if the data format follows an open standard and, hence, is known, either the new application will be able to use it as it is or it will be easier to convert the data so that the new application can use it.

It stands to reason that if a user demands that open standards are adhered to, there will be more choices

available as more vendors can participate to offer solutions and it may be possible to even mix and match solutions from multiple vendors to provide best-of-breed solutions as far as possible.

If Open Standards are followed, applications are easier to port from one platform to another since the technical implementation follows known guidelines and rules, and the interfaces, both internally and externally, are known. In addition to this, the skills learned from one platform or application can be used with possibly less re-training needed. This can be contrasted with the usage in applications of proprietary standards that are not openly published and where there is inadequate information publicly available about them.

The benefits obtained with respect to using data and file formats whose specifications are publicly published and widely accessible cannot be over-emphasised, especially with respect to an organisation that possesses huge amounts of data stored electronically. A national government is a good example of such an organisation. If the data formats are not known or easily available, the organisation may find it difficult to migrate or change its information systems since it can be prohibitively expensive or even impossible to convert data files.

OS and Innovation

Open Standards have proven to be an important facilitator for innovation. By providing an agreed, reliable and globally valid base of technology, Open Standards allow innovators to develop highly competitive, innovative technologies and solutions "on top" of the standard. At the same time they have got some safeguards regarding the potential for global market outreach. The most prominent example from the last decade is the World Wide Web. Having Open Standards, publicly available on royalty-free terms was the base for a wide wave of innovation which has, in fact, revolted the way we live, operate, and communicate. Open Standards have boosted innovation and growth.

There is a large use of the term "open" in relation with innovation: proof of that is the "Open Innovation" concept. It generally describes open innovation as, "a paradigm that assumes firms can and should use external ideas… and external paths to market."[3] Open standards and open innovation both refer to a process that involves sharing or exchanging technology across firm boundaries. The difference is that the objective of an Open Standard setting is to promote the adoption of a common

3 See: Chesbrough, H. (2003). Open Innovation. Cambridge, MA, Harvard Business School Press

standard, while the objective of open innovation is to profit from the commercialization of a new technology. In other words, open innovation might take place in a regime of either open or closed standards.

Open Standards and Information Technologies

In the IT industry, standards are particularly important because they allow interoperability of products, services, hardware and software from different parties. Without standards, users may be forced to use only hardware and software or services from one party or vendor. Internationally recognised standards define common interfaces and any changes or modifications in the standards are usually carried out by common agreement. For example, the Internet would not achieve its current ubiquitous presence, where it is accessible from almost any type of computer platform and device, if it did not use widely accepted technical standards in its networking infrastructure and supported services.

Open Sstandards and Free Software

Many people are confused between the terms Open standards and Free Software / Open Source thinking that they are one and the same or one cannot exist without the other. Open Standards is not the same as Free Software, which refers to software which follows certain principles in its creation, modification, usage, licensing and distribution. In particular, it should have the four fundamental freedoms described by the Free Software Foundation:

1. Freedom to run the programme, for any purpose;
2. Freedom to study how the programme works, and adapt it to your needs;
3. Freedom to redistribute copies so you can help others; and
4. Freedom to improve the programme, and release your improvements to the public.

For a further explanation of these concepts go to the main documentation of this module: “Introduction to Free Software”.

Free Software is software whereas Open Standards refer to standards –two different things altogether. The processes and issues involved in developing software and a standard are also very different. It is

entirely possible for a functionality in a non-free software (often called proprietary software) to be implemented following an open standard. Open standards are neutral with regard to software licensing or business models and so it is equally possible for an Open Standard to be implemented in proprietary software as it realized under a free software license. For example, proprietary software like the Microsoft Windows operating system can implement the TCP/IP networking protocols following the open standards from IETF and be compliant with them.

Widespread usage of standards, and especially Open Standards, is very important to Free Software. It makes it easier for Free Software to be compatible with proprietary software. It is a practical reality that Free Softwaree needs to coexist with proprietary software and that compatibility with the proprietary platforms is facilitated if standards are adhered to by all. If all software applications were to follow standards strictly, they should be able to interoperate and communicate among themselves well and data files could be read and written transparently. While both proprietary and Open Standards may allow this to happen, the latter are preferred by the Free Software community as they facilitate free access, open development and participation.

Free Software support may be difficult in cases where a proprietary specification is not publicly published but needs to be licensed. In the past, one way to work around this problem is to reverse engineer some proprietary product that implements the specification or protocol but, in recent times, more and more proprietary licenses have specifically forbidden this. In some countries, legislation has also been passed (*e.g.* the DMCA –Digital Millennium Copyright Act in the USA) which makes it illegal to reverse engineer a product if it is deemed that the process can assist in the circumvention of measures implemented to protect against illegal copying of the product. These developments have re-enforced the important role which Open Standards play in ensuring that Free Software can interoperate well with proprietary software. The emergence of Free Software and the Open Standards which it uses highlights the needs and benefits of open standards in a world where interoperability is required.

Free Software has also benefited much from Open Standards in that the current widespread usage and popularity of Free Software owes much to the Internet and the open standards that the Internet uses. While programmers (and many users who write their own programmes) have been freely exchanging programmes with source code since the early days of the computer, it was only after the Internet explosion in the 1990s that the idea and culture of Free Software became widely known and accepted by the mainstream IT industry. Free Software which implements the Open Standards and protocols

used on the Internet like TCP/IP, HTML, Simple Mail Transfer Protocol (SMTP), etc., were easily available and many people as well as organisations started to use them. From there, they became aware of Free Software and the latter grew in strength and acceptance as more and more used it and contributed towards it.

Some may argue that the freedom in Free Software for anyone to modify the software source code will allow and may even encourage the inclusion of code that does not conform to published standards. This is possible but in practice it is seldom done (*i.e.* modifying a Free Software mainstream product to make it non-compliant with an Open Standards and redistributing the modified software). Also Free Software project owners guard against this as they realise that it is to the advantage of Free Software if Open Standards are adhered to as much as possible. In fact, it is very natural for Free Software to promote the adoption of Open Standards since the ideals and development model of Free Software itself encourages availability, openness and participation by all –the very traits and characteristics of an Open Standard.

Free Software can play a useful role in popularising an Open Standard. A Free Software implementation of a standard usually results in an open and free-working reference implementation. Many benefits of Open Standards are negated if its only implementation is a closed and proprietary one. The availability of a Free Software implementation will spur quicker adoption and acceptance of the standard as everyone has easy access to the implementation of the standard and so can try and test it out. A very good example of this is the Internet HTTP standard. One reason why this service became universally accepted is that very early on there were free and open implementations of both the HTTP server (*e.g.,* National Center for Supercomputing Applications or NCSA HTTPd, Apache) and client (*e.g.,* NCSA Mosaic).

In conclusion, choosing open standards is highly strategic. Their benefits and positive impact are debated and seen at the highest decision making levels. Interoperability is a major requirement for the ICT sector as societies, governments and industry increasingly move towards global collaboration and integration. Open Standards built on the principles of openness, transparency and consensus lay the grounds for innovation and growth, for flexibility and choice, for global market success and fair competition. In other words, Open Standards is where society, government and industry align and where everyone is sure to benefit.

Open Standards in relation with software patents

A patent is a set of exclusive rights given by a government to a patent applicant in which the patent holder is granted the right to prevent others from making, using, selling, offering to sell or importing the invention for a specific period of time. Patents are usually granted for inventions which are considered to be non-trivial, new and novel. Patent grants are territorial in nature in that patents applied for and granted in one country are not automatically recognised in another country.

Software patents

Traditionally, patents are given mainly to physical inventions but in recent times many countries have begun to grant patents for non-physical items such as business methods and computer programmes (software). Software has become patentable in countries like the USA and Japan. The issues on the patentability of software and the way patent offices process software patent applications are very controversial. In countries where software patents are recognised, patents may be granted to functional aspects of software which are considered to be innovative and non-obvious. The expressive elements of code are not patentable. Instead, they are covered by copyright to which almost all the countries in the world subscribe to. While many countries still do not recognise software patents, most are re-examining this issue and trying to decide whether they should change their positions. The inclusion of software patents in IT-related specifications and standards has attracted a great deal of discussion.

Policies on patents

In the case of technical standards, it is not uncommon for patented items to be proposed to be included as part of the specifications. The standards development body has to decide on whether it should use such an item or look for an alternative. In the past, the development of standards related to software and IT has proceeded using mainly a 'reasonable and non-discriminatory (RAND)' terms policy whenever patents are included in a standard. Under RAND, the patent holder must be willing to negotiate rights to use the essential patent on reasonable and non-discriminatory terms. The intent of RAND was to prevent patent issues from hindering the adoption of a standard and to ensure that the cost of any necessary licenses needed, arising from the patent, are affordable. This has proved adequate in the past

but, in recent times, the increasing proliferation of patents granted to software-based innovations (including software patents) has led standards developing and setting bodies all over the world to clearly state their patent policies to ensure that they are adequate and will continue to support the development of highly successful and widely used standards as they have in the past. Since, in general, a standard is targeted for use by all in the world, it is vital that the terms of usage of any patent which is included in the standard are clearly specified.

Standards have been produced which include patented technologies and all the main standards bodies have policies with regard to the treatment of patents in the documents which they produce. The list below describes how each standard body handles patents issues:

ISO

ISO has published directives on the issue of patents in its standards development process. There is a strong recommendation to avoid references to patented items in ISO publications. Nevertheless, ISO recognises that, for technical reasons, sometimes this may not be possible and, in such exceptional situations, it does not object in principle to the inclusion of items covered by patent rights even if the terms of the standard are such that there are no alternative means of compliance. During the preparation of the ISO document, a basic text for the identification of patent rights is to be inserted into the draft documents in those cases where compliance with an ISO document may involve the use of a patent.

IETF

RFC 3979 is the main document dealing with the IETF's stand on patents. In general, IETF prefers technologies with no known patent claims or patents which offer royalty-free licensing. However, the IETF working groups have the discretion to adopt technology with a commitment of RAND terms, or even with no licensing commitment, if they feel that the technology is superior enough to alternatives with no such patents or licensing encumbrances.

In order for the working group and the rest of the IETF to have the information needed to make an informed decision about the use of a particular technology, a person contributing to the working group's discussions must disclose the existence of any patent claims that the individual is reasonably and personally aware of and that he/she (or his/her employer) owns or controls.

W3C

W3C has a very clear policy with regard to patent usage in its Recommendations. It seeks to issue Recommendations which can be implemented on a Royalty-Free (RF) basis. This has arisen from the experience it had with the WWW.

Many early standards (Recommendations) from W3C paid scant attention to patents. Later, as the Web became more commercial and software and business process patents increased, patent infringement issues surfaced as several patent holders, including some who had participated in the development of the standards themselves, sought license payments. As a result, W3C decided to have a clear patent policy governing the Recommendations which it develops.

The key position of W3C with regard to patents which are deemed essential to a Recommendation (it calls them "essential claims") is that they have to be available for implementation in accordance with the W3C RF License requirements. An "essential claim" refers to a patent for which there is no known alternative and, therefore, it is essential to the implementation of a normative part of a Recommendation.

The policy generally requires that a participating organisation in a W3C working group formally commits to the RF requirements for "essential claims" The participants are not required to disclose known patents as long as their participating organisation commits to licensing those patents according to RF requirements. In the event that a working group participant holding a patent does not want the patent to come under RF requirements, there is some flexibility in the policy in that it allows the participant to exclude specific patent claims from the RF commitment, provided the working goup is informed within a well-defined time limit. In this manner, a participant can still participate while specifying that strategic technology be excluded from the RF process and the working group is made aware of a potential patent conflict. As far as possible, the working group will try to resolve this conflict. However, in the event that it cannot be resolved, the matter is referred to the Patent Advisory Group (PAG) task force which will attempt to resolve the conflict. Ultimately after exhausting all other options, if the PAG does indeed recommend that an alternative to the RF licensing requirements be used, it has to go through several levels of review and consensus before W3C accepts the alternative.

W3C policy requiring commitment to the RF requirements by default is a stricter policy as compared with the RAND policy of ISO and IETF.

OASIS

OASIS has a published policy which governs the treatment of patents that are considered as "essential claims" (patents that are deemed essential for the implementation of a normative part of an OASIS standard), in the production of specifications and other works by OASIS.

Unlike the W3C, OASIS does not have a single licensing agreement for "essential claims"; instead it uses three types: "Reasonable And Non-Discriminatory (RAND)", "Royalty-Free (RF) on RAND Terms" and "RF on Limited Terms".

RAND defines a basic set of minimal terms that a patent holder is obliged to offer (such as granting a license that is worldwide, non-exclusive, perpetual, reasonable, and non-discriminatory, etc.) and leaves all other non-specified terms to negotiations between the patent holder and the implementor seeking a license.

RF on RAND Terms is the same as RAND with the exception that no fees or royalties are to be charged.

RF on Limited Terms specifies the exact royalty-free licensing terms which may be included in a patent holder's license and which must be granted upon request without further negotiations.

Summary of patent policies of standards organisations

As can be seen from the discussion in the previous section, most standards bodies do allow the inclusion of patents in their standards although patent-free ones are preferred. Their patent policies all revolve around allowing a RAND policy, either with some form of royalty payment or royalty-free or a mixture of both. This practice is based on the view that RAND licensing appropriately balances the legitimate rights of patent owners who contribute innovative technology to the standard, with the interests of implementors who wish to obtain access to essential patents on reasonable terms.

RAND licensing terms and Free Software licenses

The possibility that patents under Reasonable And Non-Discriminatory (RAND) terms can be included in standards has very relevant implications for software that is released under a Free Software license. Free Software licenses usually include terms which satisfy the following clauses of the Open Source Initiative's Open Source definition:

Free Redistribution

The license shall not restrict any party from selling or giving away the software as a component of an aggregate software distribution containing programmes from several different sources. The license shall not require a royalty or other fee for such sale.

Derived Works

The license must allow modifications and derived works, and must allow them to be distributed under the same terms as the license of the original software.

Distribution of License

The rights attached to the programme must apply to all to whom the program is redistributed without the need for execution of an additional license by those parties.

Although all Free Software licenses share these characteristics, the actual requirements and obligations imposed can vary from license to license. For example, the BSD license requires only copyright attribution and license reproduction, and redistributions of the software may be made under any other license. However, the Mozilla Public License (MPL) impose moderate obligations in that they require that specific files containing MPL code be distributed in source-code form and under the terms of the MPL. The GNU General Public License (GPL) requires that any work that includes GPL code, if distributed at all, be distributed under the terms of the GPL. It also clearly states that any patent must be licensed for everyone's free use or not licensed at all. Free Software licenses, then, do differ with regard to the nature and degree of rights and obligations described. Consequently, licenses like the BSD allow the usage of technology available under RAND terms but GPL does not allow any GPL-based public distribution to include any technology available under a RAND license that is not royalty-free. The licenses cited above are the most commonly used Free Software licenses with GPL by far being the most popular one. This has the implication that a large number of Free Software applications may be incompatible with RAND licensing. In connection with this issue, the Free Software Foundation has stated that RAND licensing discriminates against Free Software as it is generally not possible for software to be freely modified and redistributed under RAND licensing terms. We go deeper in RAND problems in the following section provided by OASIS:

Why RAND is not appropriate for Open Standards?

Companies desiring to collect revenue from patent royalties frequently make statements of the following kind, designed apparently to put their competitors and the paying public at ease: "We are willing to grant licenses for this technology to all under reasonable and non-discriminatory terms and conditions." By what means are we supposed to be comforted by such a promise? The notion that RAND ("reasonable and non-discriminatory") terms and conditions protect the public needs to be debunked. RAND puts the patent owner in control of a government-backed monopoly. RAND offers no certain protection against license fees that constitute oppressive global taxation, and it may be judged highly discriminatory. Hence one writer's characterization of RAND as an expanded acronym: "half of 'RAND' is deceptive and the other half is prejudiced."

The English word "reasonable" sounds assuring, and "non-discriminatory" rings with political correctness. But RAND licensing terms pose a potential threat to the health of an open, accessible Internet; this verdict has been reached by IP experts and by common persons having common sense, revealed in several thousand emails sent to the W3C public comment mailing list. The 2002-02 draft Patent Policy Working Group Royalty-Free Patent Policy from the W3C Patent Policy Working Group exposes the threat of RAND and seeks to implement an alternative royalty-free (RF) policy framework. The argument for RAND as "reasonable and non-discriminatory" is flawed at many levels. For example:

- A royalty fee *reasonable* for one wealthy US software company may not be *reasonable* for a software development project in a third-world country, or for public hospitals, museums, and elementary schools. A royalty fee paid directly by an end user (based upon metered usage of the Internet) may be *reasonable* for a CEO living in the California Silicon Valley, but quite *unreasonable* for a student in a poor region of Eastern Europe or Africa –who can barely afford to be connected to the Internet at all. Note that the MPEG LA proposal already establishes per-minute use fees for MPEG-4 video data.

- Reasonable defies formulation of an operational definition which can serve to guarantee to multiple parties that the fair interests of all will be served –once the details of a future licensing agreement are worked out. Thus Sun's Carl Cargill writes: "The lack of a clear, equitable, and easily understood definition for the term RAND is one of the major sticking points in ICT [Information and Communications Technologies] industry. Small companies and individuals fear gouging by big companies; large companies fear diminution of their IPR portfolios..." (Failed Evolution page 4, note #8). Clarity and non-ambiguity are not so problematic as "equitable," which depends largely upon one's perspective and situation in life, including ability to pay. The W3C's description of "Reasonable and Non-Discriminatory (RAND) License" in the 2001-08-10 Patent Policy Frequently Asked Questions

(FAQs) document does not hint at a method or metric for "reasonable."

- A price judged reasonable as a royalty fee for one single patent considered in isolation may become quite unreasonable when taken in aggregate (*e.g.*, when added to 25 or 125 other "reasonable" fees for companion patented technologies), or when embedded into the very foundation/core (and still developing) layered Internet architecture.
- As James Clark, Richard Stallman, and many W3C's readers have argued, RAND is discriminatory. In particular, it discriminates against open source software and core principles of the open source philosophy. As Jim Bell writes, of RAND in the W3C context: "RAND ignores the importance of open source. Open source is one of the most powerful, creative forces in computing today, and W3C patent policy should facilitate it rather than blocking it via RAND royalty fees."
- RAND also discriminates against anyone not economically empowered to pay the same royalty fee as can be paid by a wealthy, powerful software company (Microsoft, IBM) or media conglomerate (AOL Time Warner). A patent license agreement with a $15,000 initial fee disproportionately impacts and thus discriminates against small companies having a relatively small or insignificant market share.
- Uniform ('non-discriminatory') rates thus discriminate potentially against some of the highest values of an open and fair society, which recognizes the differential ability to pay (levels in tax structures, variable subscription rates, scaled consortium dues, medicare) and recognizes the benefit of a design policy that makes some resources both free and freely accessible to all (public libraries, parks). According this social principle, the Internet "information commons" needs to be free especially for the benefit of those unable to pay –not taxed under RAND terms antithetical to this core value.

Beyond these problems bound up in RAND's claims to being inherently "reasonable and non-discriminatory," RAND principles made operative in a standards setting create additional difficulties, *vis-à-vis* a royalty-free arrangement. Some have been identified previously. A brief publication by Jim Bell in connection with W3C operations lists these other concerns as follows:

- RAND can distort the standards selection process: a Working Group not fully informed about the details of RAND licenses (which are not required or allowed to be revealed in advance) may be "forced to choose between two approaches, one free and one expensive, without knowing which is which."
- RAND allows users to be ambushed and exploited: since RAND "allows the price for essential patent rights to be set after their use is mandated by a W3C standard, users cannot predict their

costs until too late and can find themselves at the mercy of a patent holder."

- RAND's scope of applicability [in the W3C setting] is inadequately defined because of the 'imprecise and moving boundary for Web infrastructure': "Even advocates of RAND licensing concede that Web infrastructure standards should be royalty free, and that RAND licensing at W3C should be applicable only to standards above the infrastructure level; but no one has been able to draw the dividing line separating the levels. In any case, such a dividing line would be steadily rising over time –as formerly optional capabilities become standard parts of the infrastructure... With RAND licensing, this natural evolution could create royalty-bearing infrastructure standards, a prospect universally agreed to be unacceptable."
- Managing RAND licensing terms is burdensome [to W3C and its Members]: "One of the greatest challenges to W3C is being able to move at a pace commensurate with its role in the industry and with the reasonable expectations of the broader community. Issues connected with RAND licensing at W3C undeniably have already proven themselves to be an immense source of [complexity], delay, and wasted effort."

Patent offerings to the Free Software community

To waylay the concerns of Free Software developers and users, and to reduce the fears of software patents infringement by Free Software developers, several commercial companies have recently offered all or part of their portfolio of software patents on a no cost basis to the Free Software community for use. IBM has announced that, for a start, it will allow royalty-free use of 500 of its software patents in any software that is released under an Open Source license (as recognized by the Open Source Initiative). Red Hat, a company well known for the development and commercial distribution of GNU/Linux, has offered unfettered use of its own software patents portfolio to Linux developers. Novell has said that it will use its existing patent portfolio to protect the Linux kernel and other Open Source programs included in Novell's offerings against potential third-party patent challenges. Sun Microsystems has released over 1,600 patents for use with software that is licensed under the Open Source Common Development and Distribution License (CDDL).

A Patent Commons Project has been started by the Open Source Development Labs (OSDL). This initiative is aimed at the creation of a central depository where software patents and patent pledges can be housed for the benefit of the open source development community and industry. Companies that have contributed and pledged patents to this project include Computer Associates, IBM, Nokia, Novell,

Red Hat and Sun Microsystems.

There is much controversy and debate over patents in software development and RAND in standardisation.

Open File Formats

Open Standards and file formats

As a starting point we define the “format” concept. First of all, it must be taken into account that computers store and transmit information in encoded form. These used to be very simple representations where certain numerical values stand for a certain character, for instance. And while their complexity has been increasing steadily with the power and complexity of computers, certain basic rules always apply. These codes used to store information are referred to as “formats”.

The first important rule is that any such choice of encoding is an arbitrary, and not a natural choice. The number 33 may represent the letter 'a' or 'z' depending on the convention for this standard. There is no right way of doing this, there are only possible ways. The second important rule is that once data have been encoded in a certain format, they can only be read by software which implements such format, and implements it exactly. Even slight deviations from the conventions of the format will easily cause massive data corruption. A common and mostly harmless form of this is lost or broken formatting in text processing software. In the worst case the data will be unrecoverable.

All formats and protocols are fundamentally arbitrary in nature, but must be followed precisely for the data which were stored in them to be recovered. Customers who saved their data in one format quickly can find themselves unable to choose another vendor who is not able to implement the same format, or unable to implement it well enough. If the only way to migrate is to loose years of data there is a very effective vendor lock-in which practically makes it impossible to choose software according to its

merits.

The solution of this situation is to choose those formats which are implemented as Open Standards. These formats are defined verifying some conditions which ensure that any software producer has enough information to implement them and, thus, guaranteeing the persistence of information.

Open vs. proprietary formats

In order to work with a file an application is required most most of the times that allows reading, editing and saving the data contained in the file.

Open format

We say that a file format is open if the mode of presentation of its data is transparent and/or its specification is publicly available meeting specifications described in the Open Standard definition. Open formats are ordinarily standards fixed by public authorities or international institutions whose aim is to establish norms for software interoperability. There are nevertheless cases of open formats promoted by software companies which choose to make the specification of the formats used by their products publicly available.

It should be noted that an open format can either be coded in a transparent way (readable in any text editor: this is the case of markup languages) or in a binary mode (unreadable in a text editor but thoroughly decodable once the format specifications are known).

Proprietary format

We will say that a file format is proprietary if the mode of presentation of its data is opaque and its specification is not publicly available, not qualifying as an Open Standard. Proprietary formats are developed by software companies in order to encode data produced by their applications: only the software produced by a company who owns the specification of a file format will be able to read correctly and completely the data contained in such files. Proprietary formats can be further protected through the use of patents and the owner of the patent can ask royalties for the use or implementation of the formats in third-party's software, as explained above in the section about standards and patents.

One of the contentious issues surrounding the use of proprietary formats is that of ownership of created content. If the information is stored in a way which the user's software provider tries to keep secret, the user may own the information by virtue of having created it, but they have no way to retrieve it except by using a version of the original software which produced the file. Without a standard file format or reverse engineered converters, users cannot share data with people using competing software. The fact that the user depends on a particular brand of software to retrieve the information stored in a proprietary format files increases barriers of entry for competing software and may contribute to vendor lock-in concept.

The issue of risk comes about because proprietary formats are less likely to be publicly documented and therefore less future proof. If the software firm owning right to that format stops making software which can read it, then those who had used the format in the past may lose all information in those files. This is particularly common with formats that were not widely adopted. However, even ubiquitous formats such as Microsoft Word can not be fully reverse-engineered.

The openformats.org initiative describes four well explained reasons to avoid proprietary file formats which are reproduced below:

1. *Taking the risk that the recipient may not be able to access the data*
    a. A proprietary format makes the use of a specific software compelling for having access to the file content. By exchanging files in proprietary formats you tacitly assume that all the recipients of your file possess the software needed for opening the file: any user that for technical reasons (*e.g.*, users working on a different platform) or financial reasons (users that cannot afford buying the required software) cannot run that specific software, will never be able to use the file.
    b. Now, let us assume the user possesses the application needed to open the file. Will this guarantee the complete accessibility of the file content? Unfortunately not: a strategy largely adopted by software producers consists in regularly upgrading the data formats they implement in their applications. Such strategy is meant to lock the user in to the use of a specific proprietary software. This way, the only possibility for the user to assure future accessibility to his/her own data or to guarantee perennial access to old files is to regularly buy updates of a specific software.
    c. [Semi-proprietary formats and predatory practices]. A similar strategy to lock the user in to a

specific data format consists in adopting an open format for storing software data at the beginning and then, progressively, modifying this format with proprietary extensions, which make the resulting format incompatible or unreadable with other software based on the original format. This strategy is often adopted to turn a public standard into a semi-proprietary format.

The adoption of proprietary or semi-proprietary formats is the result of corporate strategies that go against the user's needs, which should privilege accessibility, interoperability and perennial access of exchanged data.

2. Taking the risk of transmitting confidential information

A proprietary format encodes information which is not publicly visible. Only the producer of the format or the owner of the software which reads this format, which holds the key to totally decode the format, is able to access this information.

Often, at the moment of recording, the software adds some information to the file which is not accessible to the lay user, such as the user's name, the software's serial number, the type of operating system, the computer on which the user works, the folder in which the file is to be found, etc. Some of this information can sometimes be coded in a legible manner unknown to the author and this can then be accessible to everyone: the anecdote of the anonymous political manifesto sent out in MS Word format with the name of the author clearly legible in the properties of the document is probably the most famous case of unexpected consequences of using proprietary formats as exchange formats.

There are more serious consequences than failing to protect personal data, such as transmitting military information or trade secrets. It is somewhat curious to learn that people accept that the Ministry of Defense of a given country produces and shares documents with information accessible only to a private company in a foreign country.

Transmitting documents in a proprietary format means transmitting information nobody really knows about, other than the owner of the software that can read this format.

3. Contributing to virus propagation and exposing oneself to the risk of contamination

Most viruses are carried by infected files exchanged by users. Such viruses exploit the vulnerability of specific applications or security breaches of specific operative systems to execute malicious code. Virus proliferation in these cases relies on the fact that the majority of users use the very same kind of software and share data in the native (proprietary) format of such software. Most viruses are hence not only *platform-specific* but also *application-specific*: in many cases, simply switching to a different application makes a system immune against a class of viruses (see for instance the large number of MS Word-macro viruses). Using open formats –data formats which are software-independent, interoperable and accessible on different platforms– weakens the overall impact of viruses and discourages their propagation: it is extremely easier to create a virus exploiting known vulnerabilities of a single, largely used software and the lack of awareness of users, than adding malicious code within a format which can be read by a large number of applications and on different kinds of platforms.

4. Propping up existing monopolies in the domain of electronic communication

This problem may not be meaningful for the individual user, but it affects dramatically the community of users. By exchanging and publishing files in proprietary formats, you implicitly force your addressee to use the same software that you used for producing and storing your data. The message that is implicitly conveyed when exchanging a file in a proprietary format is *"Use software X or you won't be able to read this file"*.

This practice –which also appears when you exchange a file in a given format by considering *self-evident* that any other users possess the required software–has a twofold consequence:

a. On the one hand, this practice enforces and strengthens the usage of a proprietary format owned by a company as a *de facto* standard: this means making interoperability, accessibility and perennial access of your data "hostage" of the contingent policy of a software company. If the software producer decides or is forced to stop developing the software needed for interpreting a specific format, all the existing files encoded in this format will suddenly become unusable: since the format specifications are not publicly available, it will be impossible to retrieve the full content encoded in a file.
b. On the other hand, by propping up a *de facto* monopoly, this practice hinders fair competition between software producers –which is admittedly an essential condition for technological development– and weakens the initiatives for promoting open format specifications and public

standards –which are commonly regarded as minimum guarantees for free and fair competition.

Lists of open standard formats

Office applications file formats

Office Document Formats

Format	Organization	Published	Non-Proprietary	International Standard
DOC (text)	Microsoft	No	No	No
XLS (spreadsheet)	Microsoft	No	No	No
PPT (presentation)	Microsoft	No	No	No
SXW (text)	OpenOffice.org	Yes	Yes	No
SXC (spreadsheet)	OpenOffice.org	Yes	Yes	No
SXI (presentation)	OpenOffice.org	Yes	Yes	No
ODF (OpenDocument)	OASIS,ISO/IEC	Yes	Yes	Yes
PDF (text and presentation)	Adobe	Yes	No	Partial

Microsoft Office formats

Currently, the most popular office application is Microsoft Office (MS Office), a propietary software. This suite of office software comprises mainly (depending on the type of suite purchased) word processing (MS Word), spreadsheet (MS Excel) and presentation software (MS PowerPoint). Up till version 10 (MS Office 10), the used file formats were binary (*i.e.* non-plain text) in nature and not

publicly published. MS Word, MS Excel and MS PowerPoint use the binary Word Document Format (DOC), Excel Document Format (XLS) and PowerPoint Document Format (PPT) formats, respectively, and these are proprietary formats, being owned and controlled entirely by Microsoft.

The file formats for these applications are widely used due to the popularity of MS Office. Other software not from Microsoft, *e.g.* OpenOffice.org or StarOffice, are able to read and write files using these proprietary formats but the compatibility is incomplete. Competing products cannot be totally compatible with MS Office unless they are provided with the file format specifications by Microsoft.

Some MS Office applications like Word and Excel can save their data in what is known as the Rich Text Format (RTF) file format. This is a non-binary file format which has been developed by Microsoft for crossplatform document interchange. Technical documentation on RTF is published by Microsoft and as many non-Microsoft software support the RTF file format well, it is widely used for document exchange between MS Office and other office applications. However, the RTF format does not completely support the more complicated and sophisticated features found in MS Office, and complex documents may not be properly represented using the RTF format.

With MS Office 11 (MS Office 2003), the option to use a new XML-based file format for Word and Excel was made available. However, these XML-based formats have been criticised in some quarters for being incomplete and immature. They were not available for all the software applications in the suite and some major functionalities were not supported in those available. As a result, the traditional binary MS Office file formats remained in use mainly. In June 2005, Microsoft announced that MS Office 12, due in 2006, would deliver support for a new set of XML file formats called the "Microsoft Office Open XML Formats". The applications that would use these formats by default were Word, Excel and PowerPoint. The Case Study #1 describes the Office Open XML standard definition and its controversial standardisation process. rPlease refer to that section for further reading about this issue.

OpenOffice.org and StarOffice Formats

OpenOffice.org (OOo) is a full-fledged Free Software office application suite, comprising word processor, spreadsheet, presentation software, graphics editor and a database programme (available in OOo version 2 only). The original file formats used by OOo were XML-based. As there were several files associated with a single document, all the files were compressed and stored as a single Zip-

compressed file. OpenOffice.org is available on multiple platforms, *e.g.* GNU/Linux, MS-Windows, Mac OS X, etc., and offers multi-lingual support. It is compatible with all other major office suites. In particular, it is able to read and write MS Office file formats. The degree of compatibility is very good though not complete.

The OpenOffice.org file format was submitted to OASIS to form the basis for a new standard for office applications and this resulted in OASIS coming up with the OpenDocument Format for Office Applications (OpenDocument) v1.0 in May 2005. The OpenDocument Format has also been accepted as an international ISO/IEC standard (ISO/IEC 26300). There is more information about this format in the Case Study #1 where we compare OpenDocument and Microsoft's Office Open XML.

New versions of OOo as well as other office suites like KOffice and StarOffice now support OpenDocument as their native file formats. This will significantly improve the interoperability of office software and enhance document exchange. What is most important though is that all these office applications now use a standard open file format for storing their data. The OpenDocument format is not owned or controlled by a single vendor, instead it falls under the ambit of OASIS, an open standards body. Users can, thus, be assured that they will have access to their documents and data from a variety of software.

StarOffice shares the same code base as OOo but it is released under a proprietary commercial license. In addition to the core functionalities of OOo, it also comes with some proprietary and third-party modules, *e.g.* the Adabas B database and some proprietary clip art galleries and templates. StarOffice uses and supports the same file formats as OpenOffice.org.

Adobe's Portable Document Format

PDF is a file format developed by Adobe Systems, Incorporated for secure and reliable electronic document distribution and exchange. The format is able to preserve the look and integrity of the original document, regardless of the application and platform used to create it even if it contains complex combinations of text, graphics and images. As such, the PDF format is very useful as a format for multi-platform document exchange and distribution and for sharing information. However, one major drawback of PDF is that it is an end-form format*, i.e.*, it is not suitable for modifying or re-writing its contents.

The PDF format is a standard set and controlled by Adobe. It also contains several patents owned by Adobe but licensed royalty-free for use. Older versions and subsets of PDF (*e.g.* version 1.4) have been adopted as ISO standards (*e.g.* PDF/X for printing and graphics, ISO 15930, and PDF/A for long term preservation of electronic documents, ISO 19005). However, the industry mainly makes use of the published PDF specifications from Adobe rather than the ISO standards in implementations of software to use PDF. The specifications for the PDF format is publicly published by Adobe and it can be implemented without restrictions by anyone (provided that there are no objections from Adobe). As a result, a variety of software on many different platforms which can read the PDF format is available , and there is a (smaller) number of applications which can write out the contents of a document in PDF.

Due to its popularity and wide support, PDF can be considered a *de facto* standard as a file format for information exchange and sharing but since it is created, owned and controlled by Adobe Corporation, it does not meet the technical definition of an open standard. The PDF specification are actively being developed by Adobe with no means of open participation by interested parties and control of the specification always lies in the hands of Adobe. While the specs are openly available there are specific constraints in the implementation of the features in the specs. Thus, Adobe can, when it sees fit, impose specific constraints on another party attempting to make use of the specification. The recent decision by Adobe not to allow Microsoft to include as a native option in its MS Office 12 software to enable a user to save or export the contents in PDF format is a very clear example of this.

PostScript

PostScript is a page description language and programming language. PostScript started out as a proprietary standard, but was later submitted through a standardisation process.

LaTeX

LaTeX is a document markup language and document preparation system for the TeX typesetting program.

LaTeX is most widely used by mathematicians, scientists, engineers, philosophers, linguists, economists and other scholars in academia. As a primary or intermediate format, e.g., translating DocBook and other XML-based formats to PDF, LaTeX is used because of the high quality of typesetting achievable by TeX. The typesetting system offers programmable desktop publishing

features and extensive facilities for automating most aspects of typesetting and desktop publishing, including numbering and cross-referencing, tables and figures, page layout and bibliographies.

It is distributed under the terms of the LaTeX Project Public License (LPPL), LaTeX is free software.

Rich Text Format

The Rich Text Format (often abbreviated RTF) is a proprietary document file format with published specification developed by Microsoft Corporation in 1987 for Microsoft products and for cross-platform document interchange.

Most word processors are able to read and write some versions of RTF. There are several different revisions of RTF specification and portability of files will depend on what version of RTF is being used. RTF specifications are changed and published with major Microsoft Word/Microsoft Office versions.

It should not be confused with enriched text (mimetype "text/enriched" of RFC 1896) or its predecessor Rich Text (mimetype "text/richtext" of RFC 1341 and 1521) which are completely different specifications.

Text documents

ASCII

The American Standard Code for Information Interchange is a character-encoding scheme based on the ordering of the English alphabet. ASCII codes represent text in computers, communications equipment, and other devices that use text. Most modern character-encoding schemes are based on ASCII, though they support many more characters than did ASCII.

US-ASCII is the Internet Assigned Numbers Authority (IANA) preferred charset name for ASCII.

ASCII includes definitions for 128 characters: 33 are non-printing control characters (now mostly obsolete) that affect how text and space is processed; 94 are printable characters, and the space is considered an invisible graphic. The most commonly used character encoding on the World Wide Web was US-ASCII until December 2007, when it was surpassed by UTF-8.

An “ASCII text file” reffers to a plain text file information.

HTML/XHTML

HTML, which stands for HyperText Markup Language, is the predominant markup language for web pages. It's the standard ISO/IEC 15445:2000. It provides a means to create structured documents by denoting structural semantics for text such as headings, paragraphs, lists, links, quotes, and other items. It allows images and objects to be embedded and can be used to create interactive forms. It is written in the form of HTML elements consisting of "tags" surrounded by angle brackets within the web page content. It can load scripts in languages such as JavaScript which affect the behavior of HTML webpages. HTML can also be used to include Cascading Style Sheets (CSS) to define the appearance and layout of text and other material. The W3C, maintainer of both HTML and CSS standards, encourages the use of CSS over explicit presentational markup.

Graphics/Image File Formats

Many graphics file formats in use today are proprietary by nature, being derived and tied to the software used to create them. There are some formats that have gained wide acceptance as *de facto* standards and a few of these have emerged as open graphic file formats.

Graphic Formats

Format	Orgnization	Published	Non-Propreitary	International Standard
GIF	CompuServe	Yes	No	No
PNG	W3C	Yes	Yes	Yes
XPM	X.Org	Yes	Yes	No
TIFF	Adobe	Yes	No	TIFF/IT
JPEG	ISO	Yes	Yes	Yes
SVG	W3C	Yes	Yes	No

GIF

GIF is a bitmap image format which is widely used on the Internet, especially in its early days as this

format resulted in small graphic file sizes. Images stored as GIF files are generally limited to 256 colours. The GIF format makes use of the LZW compression algorithm that was patented in the USA by Unisys. After the GIF format found widespread use on the Web, Unisys asked for royalty payments for all software which uses GIF (this patent has since expired in the USA, in 2003). This led to the diminished use of GIF and also to the creation of alternatives to it, notably the PNG format.

The patents over the LZW algorithm expired in 2006. About this issue, the Free Software Foundation explains[4]:

> We were able to search the patent databases of the USA, Canada, Japan, and the European Union. The Unisys patent expired on 20 June 2003 in the USA, in Europe it expired on 18 June 2004, in Japan the patent expired on 20 June 2004 and in Canada it expired on 7 July 2004. The U.S. IBM patent expired 11 August 2006, The Software Freedom Law Center says that after 1 October 2006, there will be no significant patent claims interfering with employment of the format.

PNG

The PNG format was created as an alternative to GIF when Unisys decided to enforce its software patent on LZW data compression which was used in the then popular GIF format. The PNG format, like the ZIP format, makes use of the unpatented *Deflate* compression algorithm. PNG is an extensible file format for the lossless, portable, well-compressed storage of raster images. It offers indexed-colour, greyscale, and true colour image support, plus an optional alpha channel for transparency. It is fully streamable with a progressive display option making it useful for online graphics display in web pages. It also boasts robust features, providing both full file integrity checking and simple detection of common transmission errors. [10]

PNG is supported by all major graphics software and is now very widely used. It has become an open file format standard and it is a W3C recommendation as well as an ISO international standard (ISO/IEC 15948).

MNG

Multiple-image Network Graphics is a public graphics file format for animated images.

MNG is closely related to the PNG image format. When PNG development started in early 1995, developers decided not to incorporate support for animation, not least because this feature of GIF was

4 http://www.gnu.org/philosophy/gif.html#venuenote

seldom used at the time. However, work soon started on MNG as an animation-supporting version of PNG. Version 1.0 of the MNG specification was released on January 31, 2001.

XPM

The XPM (XPixMap) format is a *de facto* standard for creating icon pixmaps for use in GUIs based on the X Window System. It consists of an ASCII image format and a C library. The XPM format defines how to store colour images (X Pixmap) in a portable way while the associated library provides a set of functions to store and retrieve images to and from XPM format data.

TIFF

The Tagged Image File Format (TIFF) is a file format for digital images. It is a specification that is now owned by Adobe Systems, Incorporated. TIFF is widely used in image applications in the publishing industry and also supported by most image scanning and editing software. The specifications for the TIFF format is publicly published by Adobe and it can be implemented without restrictions by anyone. As a result, there is software available on many different platforms which can read and write the TIFF format. It has become a *de facto* standard graphics format for high colour depth (32-bit) graphics.

TIFF/IT, which is based on TIFF, is a specification for the exchange of digital advertisements and complete pages (*e.g.*, newspapers, magazines). This has been made an ISO standard (ISO 12639) as a media independent means for pre-press electronic data exchange.

JPEG JFIF

JPEG is a standardised image compression mechanism from the Joint Photographic Experts Group (JPEG). The file format which employs this compression is JFIF (JPEG File Interchange Format) and JPEG JFIF is what people generally mean when they refer to "JPEG" The JFIF file format was created by the Independent JPEG Group (IJG) for the transport of single JPEG-compressed images.

The JPEG compression uses a lossy mechanism for compressing colour or greyscale images. It works well on natural, real-world scenes like photographs, naturalistic artwork and similar material but it does not fare too well on lettering, simple cartoons or line drawings. The basic JPEG format is the most common format used for storing and displaying photographic images on the Web. One reason for this popularity is that the amount of compression can be adjusted to achieve the desired trade-off between file size and visual quality. The JPEG compression is now an ISO standard –ISO/IEC 10918 Parts 1-4.

There are potential patent issues with JPEG, especially with some of its optional features, namely arithmetic coding and hierarchical storage and so for this reason, these optional features are seldom used on the Web.

SVG

Unlike other file formats listed above that are meant for raster graphics, the SVG (Scalable Vector Graphics) format is meant for vector graphics, *i.e.* the use of geometrical primitives such as points, lines, curves, and polygons to represent images in computer graphics. SVG consists of an XML-based file format and a programming API for graphical applications. It is a W3C recommendation and is starting to become a popular choice for including graphics in XML documents. As an SVG document can include raster images such as JPEG and PNG, it can be used to add raster and mixed vector/raster graphics to XML documents.

The SVG format is important as it offers a way based on open standards to render graphics optimally on all types of devices. While currently the usage of SVG usage on the Web is somewhat limited, this should change in due course as more web browsers support it natively. For the mobile phone industry, it has become the basis for its graphics platform with the publication of the SVG Mobile profile targeted at resource-limited devices such as mobile handsets and PDAs.

Video file formats

In order that a multimedia experience can be enjoyed properly by all without any discrimination, it is important that there exist multi-platform and multi-software support for it. This underlies the important role that open standards play in relation to video formats and technologies.

Video data's storage involves more than just finding an efficient means to store raw data; other data like tags, menus and possible media manipulation information need to be stored too. There may also be a need to store audio data as video frequently has sound associated with it. Also, the data stream is usually not stored in its raw form, it is transformed into a form more suitable for storage or transmission. A type of file called a container is used to store the data and associated information and a codec is utilized for encoding and decoding the data stream. It is important that the format of the container file as well as the codec that is supported by it follow open standards.

Almost all video containers popular today are proprietary. This is due to the popularity of Apple's

QuickTime and Microsoft's Windows Media framework multimedia technologies. Some of these formats, through widespread usage, have emerged as de facto standards but remain proprietary formats all the same.

Video Containers

AVI

Audio Video Interleave (AVI) is a video container format from Microsoft containing both audio and video data. It is a Resource Interchange File Format (RIFF) file specification used with applications that capture, edit, and play back audio-video sequences. It enjoys widespread support and it is the most common container format for audio/video data on the PC.

ASF

Advanced Systems Format (ASF) is Microsoft's proprietary container designed for streaming. The codec is not specified in ASF but the most common ones are Windows Media Audio (WMA) and Windows Media Video (WMV). The ASF container structure is patented in the United States.

MOV

The MOV container is from Apple Computer's QuickTime multimedia architecture and technology. This video file format is openly documented and available for anyone to use royalty-free. As a result, there are several non-Apple video player software available which can play QuickTime video files. The proprietary Sorenson codec is usually used with QuickTime. The QuickTime format was used as the basis of the MPEG-4 MP4 container standard (see entry on MP4 below).

MP4

MPEG-4 Part 14 (MP4) is a container specified as part of the MPEG-4 international standard, ISO/IEC 14496-14. MP4 is designed to support streaming, editing, local playback, and interchange of content. Its design is based on the QuickTime format.

Ogg

The Ogg container uses a bitstream format to encapsulate data from one or more sources. It can handle both audio and video data and while the codecs are not specified, there are several open codecs associated with the Ogg project, including Vorbis (see above) for lossy compressed audio, FLAC for lossless compressed audio, Speex for speech and Theora for video.

The Ogg format has been published as an IETF document, RFC 3533.

Video Formats

Format	Organization	Published	Non-Proprietary	International Standard
AVI	Microsoft	Yes	No	No
ASF	Microsoft	No	No	No
MOV	Apple Computer	Yes	No	No
MP4	MPEG/ISC	Yes	Yes	Yes
Ogg	Xiph.Org	Yes	Yes	No

Video Codecs

MPEG Codecs

MPEG has developed several standards pertaining to video technology that are used by many digital video products on the market. The MPEG video codecs are specified in the following ISO standards:

1. MPEG-1 Part 2 (ISO/IEC 11172-2)
2. MPEG-2 Part 2 (ISO/IEC 13818-2)
3. MPEG-4 Part 2 (ISO/IEC 14496-2)
4. MPEG-4 Part 10 (ISO/IEC 14496-10)

The MPEG-2 and MPEG-4 standards make use of numerous patented technologies and the vendors of commercial products and services that use them are expected to pay patent licensing royalties.

MPEG-1 Part 2

The MPEG-1 standard that specifies the MP3 audio codec also specifies a video codec for non-interlaced video signals. This codec can be used for compressing video sequences, both 625-line and

525-lines, to bit rates of about 1.5 Mbit/s. It is used in the Video CD (VCD) specifications and the picture quality is comparable to that found for the VHS video cassette recorder.

MPEG-2 Part 2

The MPEG-2 standard specifies a video codec for interlaced and non-interlaced video signals. MPEG-2 video is not optimized for low bit-rates (less than 1 Mbit/s), but outperforms MPEG-1 at 3 Mbit/s and above. The MPEG-2 video codec is backward compatible with the MPEG-1 codec. MPEG-2 is widely adopted for video broadcasting (e.g., direct broadcast satellite and cable TV), filmmaking, and DVD discs. MPEG-2 has a lot of market acceptance and a very large installed base.

MPEG-4 Part 10 (H.264/AVC)

This video coding standard is the same as the ITU-T H.264 recommendation and the technology is also known as Advanced Video Coding (AVC). It contains several innovative features that allow it to compress video more efficiently than earlier MPEG codecs. It also possesses more flexibility, which allows it to accommodate applications in a wide variety of environments.

This is a new standard and it represents the current state-of-the-art in the series of MPEG video compression standards. It is rapidly gaining adoption in a wide variety of applications and digital broadcasting and TV systems. Apple Computer has integrated H.264 into Mac OS X version 10.4 (Tiger), as well as QuickTime version 7 while x264 is a FOSS free library for encoding H.264/AVC video streams. H.264 decoders for Windows, GNU/Linux and Macintosh as well as video servers and authoring tools are available from a number of vendors.

Sorensen

The Sorenson codec is a proprietary video codec from Sorenson Media and used by Apple's QuickTime.

Windows Media Video

This is a set of proprietary streaming video technologies developed by Microsoft as part of its Windows Media framework. It is the codec usually used in an AVI or ASF container and has support for digital rights management facilities. Microsoft has submitted WMV Version 9 to the Society of Motion Picture and Television Engineers (SMPTE) for approval as a standard under the name "VC-1".

Theora

This is a video codec from Xiph.org Foundation as part of the Ogg project. It is based on patented technology but it has been irrevocably given a royalty-free license to use the patents in the codec. The Theora codec is released under a Berkley Software Distribution (BSD) FOSS license and it is available freely for commercial or non-commercial use.

Video Codecs

Format	Orgnization	Published	Non-Propreitary	International Standard
MPEG-1	MPEG/ISC	Yes	Yes	Yes
MPEG-2	MPEG/ISC	Yes	Yes	Yes
MPEG-4	MPEG/ISO/ITU	Yes	Yes	Yes
Sorenson	Sorenson Vision	No	No	No
WMV	Microsoft	No	No	No
Theora	Xiph.org	Yes	Yes	No

Video Formats

Container - Codec Commonly Used	Usage	Open/Close
AVI - WMV	Wide	Close
ASF - WMV	Wide	Close
MOV - Sorenson	Wide	Close
MP4 - MPEG-1,2,4	Wide	Open
Ogg - Theora	Limited	Open

Audio file formats

There are two major groups of audio file formats:

1. those using lossless compression, e.g. like WAV, FLAC
2. those using lossy compression, e.g. MP3, Ogg Vorbis, WMA, AAC

In the lossless compression of a piece of data, nothing is lost during the compression and the original data is restored upon uncompressing. In lossy compression, some data is lost during compression and upon uncompressing the data is not identical to the original but possibly close to it. Lossy compression is used mainly in the compression of multimedia data like audio or video where the loss of some details is tolerable under certain conditions, e.g., the human eye is unable to discern the loss in certain details of an image or video.

WAV

WAVEform audio format (WAV) is a Microsoft and IBM audio file format for storing audio on PCs. It is the main format used on Microsoft Windows systems for raw audio storage. The WAV format is most commonly used with an uncompressed, lossless storage method (pulse-code modulation) resulting in comparatively large audio files. Today, the WAV audio format is no longer popular being superseded by other more efficient means of audio storage.

FLAC

Free Lossless Audio Codec (FLAC) is a popular lossless audio format with compression designed specifically for audio data streams, achieving compression rates of 30-50 percent. The format specification is publicly available and forms part of the FLAC Open Source project.[20] It is supported by a growing list of audio software and devices.

MP3

MPEG-1 audio layer 3 (MP3) is a popular lossy compression audio format. The MP3 specification was set by the Motion Pictures Experts Group (MPEG), a working group of ISO/IEC charged with the development of video and audio encoding standards. The compression scheme and format for MP3 forms part of the MPEG-1 video and audio compression standard specifications and is an ISO standard, ISO/IEC 11172-3.

MP3 is one of the most popular audio file formats in use today. Music files encoded with MP3 are particularly popular on music exchange and download sites on the Internet due, in part, to the relatively small size of such files and the wide availability of free software on PCs that allow easy creation, sharing, collecting and playing of MP3 files.

MP3 makes use of patented technology and so software and devices that support it are subject to royalty payments in those countries that recognize software patents. This has led to the creation of alternatives to MP3, e.g. Ogg Vorbis and WMA.

AAC

Advanced Audio Coding (AAC) from MPEG is a lossy data compression scheme intended for audio streams. It was designed to provide better quality at the same bit-rate than MP3, or the same quality at lower bitrates (and hence smaller file sizes). The compression scheme and format for AAC forms part

of the MPEG2 video and audio compression standard specifications and is an ISO standard, ISO/IEC 13818-7. This MPEG-2 AAC specification makes use of patents from several companies and a patent license is needed for products that make use of this standard.

The newer MPEG-4 standard also specifies an audio compression technology that incorporates MPEG-2 AAC. This is known as MPEG-4 AAC, and is an ISO standard, ISO/IEC 14496-3.

Apple's popular iTunes service and iPod products have music available in AAC and this has led to an upsurge in the popularity of AAC despite the required patent license royalty payments.

RealAudio

RealAudio is a proprietary audio format developed by RealNetworks for low bandwidth usage. It was first introduced in 1995 and it became popular especially for streaming audio, i.e., the audio is being played in real time as it is downloaded. Many radio stations use RealAudio to stream their programmes over the Internet.

Ogg Vorbis

Ogg Vorbis is a compressed audio format that is believed to be free of patents and royalty payments. The format originated from the Xiph.Org Foundation, a non-profit organization dedicated to producing free and open protocols, formats and software for multimedia.

Ogg Vorbis uses the Vorbis lossy audio compression scheme. The audio data is wrapped up in the Ogg container format, the name of Xiph.org's container format for audio, video, and meta-data - hence the name Ogg Vorbis. The Ogg Vorbis specification is in the public domain and is completely free for commercial or non-commercial use. There is growing support for the Ogg Vorbis format from software and hardware devices as well as online audio services.

Audio Formats

Format	Orgnization	Published	Non-Propreitary	International Standard
WAV	Microsoft	Yes	No	No
FLAC	Xiph.Org	Yes	Yes	No
MP3	MPEG/ISC	Yes	Yes	Yes
WMA	Microsoft	No	No	No
AAC	MPEG/ISO	Yes	Yes	Yes
RealAudio	RealNetworks	Yes	No	No
Ogg Vorbis	Xiph.org	Yes	Yes	No

Archiving and compression file formats

7z

7z is a compressed archive file format that supports several different data compression, encryption and pre-processing filters. The 7z format initially appeared as implemented by the 7-Zip archiver. The 7-Zip program is publicly available under the terms of the GNU Lesser General Public License. The LZMA SDK 4.62 was placed in the public domain in December 2008. The latest stable version of 7-Zip and LZMA SDK is version 4.65. The latest version of these is 9.13 beta.

The MIME type of 7z is application/x-7z-compressed.

The official 7z file format specification is distributed with 7-Zip's source code. The specification can be found in plain text format in the doc\ sub directory of the source code distribution.

Bzip2

bzip2 is a free and open source lossless data compression algorithm and program developed by Julian Seward. Seward made the first public release of bzip2, version 0.15, in July 1996. The compressor's stability and popularity grew over the next several years, and Seward released version 1.0 in late 2000.

gzip

gzip is a software application used for file compression, it usually means the GNU Project's implementation of compression/decompression tool using Lempel-Ziv coding (LZ77), where gzip is short for GNU zip; as the program was created a free software replacement for the compress program used in early Unix systems, intended for use by the GNU Project.

The GNU implementation of gzip was created by Jean-Loup Gailly and Mark Adler. Version 0.1 was first publicly released on October 31, 1992. Version 1.0 followed in February 1993.

OpenBSD's version of gzip is actually the compress program, to which support for the gzip format was added in OpenBSD 3.4 - the 'g' in this specific version stands for gratis.

FreeBSD, DragonFlyBSD and NetBSD use BSD-licensed implementation instead of the GNU version, which is actually a command line interface for Zlib that is intended to be compatible with GNU implementation's options. These implementations originally come from NetBSD, and supports decompression of bzip2 and Unix pack format.

PAQ

PAQ is a series of lossless data compression archivers that have evolved through collaborative development to top rankings on several benchmarks measuring compression ratio (although at the expense of speed and memory usage). Specialized versions of PAQ have won the Hutter Prize and the Calgary Challenge. PAQ is free software distributed under the GNU General Public License.

Tar

In computing, tar (derived from tape archive and commonly referred to as "tarball") is both a file format (in the form of a type of archive bitstream) and the name of a program used to handle such files. The format was created in the early days of Unix and standardized by POSIX.1-1988 and later POSIX.1-2001.

Initially developed to be written directly to sequential I/O devices for tape backup purposes, it is now commonly used to collect many files into one larger file for distribution or archiving, while preserving file system information such as user and group permissions, dates, and directory structures.

ZIP

The ZIP file format is a data compression and archive format. A ZIP file contains one or more files that have been compressed to reduce file size, or stored as-is. The ZIP file format permits a number of compression algorithms, but as of 2009[update], the Deflate method continues to be dominant.

In April 2010 ISO/IEC JTC 1 initiated a ballot to determine whether a project should be initiated to create an ISO/IEC International Standard format compatible with ZIP. The proposed project, entitled Document Packaging envisages a ZIP-compatible 'minimal compressed archive format' suitable for use

with a number of existing standards including OpenDocument, Office Open XML and EPUB.

A case of study: OpenDocument and Office Open XML

In this case study we analyse the process of standardisation in two specific file formats: OpenDocument and Office Open XML (OOXML). Both file formats are designed to represent spreadsheets, charts, presentations and word processing documents; also commonly described as "office documents". As we in the following pages, OpenDocument was initially approved as an standard in 2005 while Microsoft OOXML was approved in 2008 after a two-year process full of controversies and strong opposition from the Free Software community and many software companies such as IBM, Canonical and others.

The OpenDocument and OOXML constitute an interesting case of study which let us understand the complexity of this kind of processes which involve technical, economic and political issues. These two file formats have been in middle of a "battle" to establish an Open Standard in a field where Microsoft Office proprietary and closed file formats were historically known as the dominants.

In the next sections we will know each format standardisation process, technical characteristics and the elements of the debate around them.

OpenDocument

General description

The OpenDocument Format (ODF) is an XML-based file format for representing electronic documents such as spreadsheets, charts, presentations and word processing documents. While the specifications were originally developed by Sun Microsystems, the standard was developed by the OASIS Open Document Format for Office Applications (OpenDocument) TC –OASIS ODF TC, committee of the Organisation for the Advancement of Structured Information Standards (OASIS) consortium and based on the XML format originally created and implemented by the OpenOffice.org office suite (see OpenOffice.org XML). In addition to being an OASIS standard, it is published (in one of its version 1.0

manifestations) as an ISO/IEC international standard, ISO/IEC 26300:2006 Open Document Format for Office Applications (OpenDocument) v1.0.

The philosophy behind the format was to design a mechanism in a vendor neutral manner from the ground up using existing standards wherever possible. Although this meant that software vendors would need to tweak their individual packages more than if they continued down their original routes, the benefits of interoperability were understood by the participants to be important enough to justify this.

This is reflected in the specification in many ways, but specifically:

- XSL:FO—Formatting
- SVG—Scaleable Vector Graphics
- MathML—Mathematical formulas
- XLink—Embedded links
- SMIL—Synchronised Multimedia Integration Language
- Xforms—Forms definitions

Brief history

OpenDocument is a format developed as a vendor neutral standard. However, it did not originate from anywhere in particular but rather through a process of evolution…

- In 1999 StarDivision began work on an XML interchangeable file format for their StarOffice product.
- In August that year StarDivision was acquired by Sun Microsystems.
- In October 2000, Sun Microsystems released large amounts of the source code to the community driven OpenOffice.org project under an open license. At the same time, an XML community project was created with the goal of defining an XML specification for the interchangeable file format.
- In May 2002, OpenOffice.org version 1.0 and StarOffice version 6 were released using this XML format (SXW).
- Also in 2002, collaboration began with other office suites, notably the KOffice project, to further refine the interoperability of the format.

- In December 2002, OASIS had a conference call announcing the creation of what is now the OpenDocument standard.
- Then, from 2002 to 2004, the format was overhauled based on experiences to date and examination of what is required in an office format.
- In December 2004, a second draft of the XML file format was approved by OASIS for review.
- In February 2005, a third draft was published for public feedback—six years after commencement of the project and five years after public consultation began.
- In May 2005, the ODF format was approved as an OASIS standard.
- Soon after that, many office suites adopted the standard as a means to store the documents.
- In September 2005, ODF was submitted as ISO standard.
- In May 2006, ODF achieved ISO certification (ISO/IED 26300).
- In February 2007, OpenDocument Version 1.1 was approved by OASIS. It particularly addresses accessibility issues.
- ODF throughout continued and continues to grow in popularity and support.

Implementations and adoption

The OpenDocument format is used in free software and in proprietary software. This includes office suites (both stand-alone and web-based) and individual applications such as word-processors, spreadsheets, presentation, and data management applications. Prominent office suites supporting OpenDocument fully or partially include:

- Adobe Buzzword
- AbiWord
- Google Docs
- IBM Lotus Symphony
- KOffice
- Microsoft Office 2000, Office XP, Office 2003, Office 2007 with plugins (none of them provided by Microsoft)
- Microsoft Office 2007 Service Pack 2 (SP2)
- NeoOffice
- OpenOffice.org
- Sun Microsystems StarOffice
- SoftMaker Office
- Corel WordPerfect Office X4
- Zoho Office Suite

An objective of open formats like OpenDocument is to guarantee long-term access to data without legal or technical barriers, and some governments have come to view open formats as a public policy issue. Several governments around the world have introduced policies of partial or complete adoption. What this means varies from case to case; in some cases, it means that the OpenDocument standard has a national standard identifier; in some cases, it means that the OpenDocument standard is permitted to be used where national regulation says that non-proprietary formats must be used, and in still other cases, it means that some government body has actually decided that OpenDocument will be used in some specific context.

The NATO (North Atlantic Treaty Organisation) with its 26 members (Belgium, Bulgaria, Canada, the Czech Republic, Denmark, Estonia, France, Germany, Greece, Hungary, Iceland, Italy, Latvia, Lithuania, Luxembourg, the Netherlands, Norway, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, Turkey, the UK, and the USA) uses OpenDocument as a mandatory standard for all members. OpenDocument has been also officially approved by national standards bodies of Brazil, Croatia, Ecuador, Hungary, Italy, Malaysia, South Korea, South Africa and Sweden. Many sub-national governments have also adopted it.

Office Open XML

General description

Office Open XML (also referred to as OOXML or Open XML) is an ISO/IEC standardized ZIP-compatible file format originally developed by Microsoft for representing spreadsheets, charts, presentations and word processing documents.

Starting with Microsoft Office 2007, the Office Open XML file formats have become the default file format of Microsoft Office.

The Office Open XML file formats were standardised between December 2006 and November 2008, first by the Ecma International consortium (where they became ECMA-376), and subsequently, after an contentious standardization process, by the ISO/IEC's Joint Technical Committee 1 (where they became ISO/IEC 29500:2008). The process description is expanded in the following pages.

Standardisation process

Inside Ecma

More than a year after being asked by the European Union (trough the IDABC –European eGovernment Services) to standardise their Office 2003 XML formats, Microsoft submitted 2,000 pages of documentation for a new file format to the Ecma International consortium for it to be made into an Open Standard. Ecma formed a Technical Committee (TC45) in order to produce and maintain a "formal standard for office productivity applications which is fully compatible with the Office Open XML Formats, submitted by Microsoft". The technical committee was chaired by two Microsoft employees and included members drawn from Apple, Canon, Intel, NextPage, Novell, Pioneer, Statoil ASA, Toshiba, The United States Library of Congress, The British Library and the Gnome Foundation.

During standardisation within Ecma the specification grew to approximately 6,000 pages. It was approved as an Ecma standard (ECMA-376) on December 7, 2006.

International process

After achieving the Ecma approbation, the OOXML file format specification was ready to start the international evaluation process in order to become an international approved standard. OOXML was then fast-tracked in the Joint Technical Committee 1 (JTC 1) of ISO and IEC.

After initially failing to pass in the voting process went into a ballot resolution phase. Ecma produced a draft "Disposition of Comments" document which addresses the 1,027 distinct "NB comments" (that is, comments by National Bodies) which had been submitted in the letter ballot phase (formally know as Ballot Resolution Meeting –BRM). This document comprised 1,600 pages of commentary and proposed changes to be made into the OOXML specification.

The results of the ballot stated that "75% of the JTC 1 participating member votes cast positive and 14% of the total of national member body votes cast negative". There have been allegations that the ISO ballot process for Office Open XML was marred with voting irregularities and heavy-handed tactics by some stakeholders. Some of these complaints are detailed in the following section of the document.

After the ballot process, four JTC 1 members appealed the standardisation: the bodies of South Africa, Brazil, India and Venezuela. The appeals did not get sufficient support of the National Bodies voting on the ISO and IEC management boards, and consequently the go-ahead was given to publish ISO/IEC

DIS 29500, Information technology –Office Open XML formats, as an ISO/IEC International Standard. The new standard was finally published in November 2008.

Complaints about irregularities in the process

The ISO standardisation of Office Open XML was controversial and embittered, with IBM threatening to leave standards bodies which it said allow dominant corporations like Microsoft to wield undue influence. Microsoft was accused of co-opting the standardisation process by leaning on countries to ensure that it got enough votes at the ISO for Office Open XML to pass. The Wikipedia article which documents this process collects the majority of the irregularity complaints registered in various national bodies,. We reproduce some of them below:

- An article on Ars Technica sources Groklaw stating that at Portugal's national body TC meeting, "representatives from Microsoft attempted to argue that Sun Microsystems, the creators and supporters of the competing OpenDocument format (ODF), could not be given a seat at the conference table because there was a lack of chairs."
- In Sweden, Microsoft notified the Swedish Standards Institute (SIS) that an employee sent a memo to two of its partners, requesting them to join the SIS committee and vote in favor of Office Open XML in return for "marketing contributions." Jason Matusow, a Director in the Corporate Standards Strategy Team at Microsoft, stated that the memo was the action of an individual employee acting outside company policy, and that the memo was retracted as soon as it was discovered. SIS have since changed its voting procedure so that a member has to actually participate before he is allowed to vote.
- Sweden invalidated its vote (80% was for approval) as one company cast more than one vote, which is against SIS policy.
- In Finland, the vote was 16 Yes, 14 No with three abstentions. This meant that, according to rules Finland's vote was an abstention. In the end of the meeting the chairman, representing Finnish Standards Association SFS, issued his critical opinion as "an individual person" not as the chairman. Finnish IT journalists described that meeting raised strong differences in opinions.
- In Switzerland, SNV registered a vote of "approval with comments," and there was some criticism about a "conflict of interest" regarding the chairman of the UK 14 sub-committee, who did not allow discussion of licensing, economic and political arguments. In addition, the chairman of the relevant SNV parent committee is also the secretary general of Ecma International, which approved OOXML as a standard. Further complaints regarded "committee stuffing", which is

however allowed by present SNV rules, and non-adherence to SNV rules by the UK 14 chairman, which resulted in a re-vote with the same result.

- In India, an Open Source Initiative board member charges that Microsoft supplied a form letter to several non-profit organisations to "bombard the Indian IT Secretary and the Additional Director General of the Bureau of Indian Standards with letters supporting its OOXML proposal."
- Norway's vote was decided by Standard Norge; the mostly opposing viewpoints of the technical committee were ignored after members were unable to reach consensus. Membership in the technical committee had risen from 6-7 to 30 members; all of the pre-OOXML members argued in favour of a "no" vote. In October 2008, 13 of the 23 members, 12 of which are associated with the open-source movement, resigned after OOXML was ratified by ISO and all appeals were rejected.

The irregularities were also detected in the technical committees: the editorial group who actually produce the spec is referred to as “Ecma”, but in fact the work was mostly done by Microsoft people.

After the specification was officially accepted as an ISO standard, Red Hat and IBM claimed the ISO is losing credibility, and the Ubuntu founder's Mark Shuttleworth commented “We're not going to invest in trying to implement a standard that is poorly defined.” IBM issued a press release stating: “IBM will continue to be an active supporter of OpenDocument. We look forward to being part of the community that works to harmonize OpenDocument and OOXML for the sake of consumers, companies and governments, when OOXML control and maintenance is fully transferred to JTC1.”

About this, Richard Stallman from the Free Software Foundation said:

> “Microsoft corrupted many members of ISO in order to win approval for its phony 'open' document format, OOXML. This was so governments that keep their documents in a Microsoft-only format can pretend that they are using 'open standards.' The government of South Africa has filed an appeal against the decision, citing the irregularities in the process.”

Technical problems and inconsistencies with other standards

A huge amount of inconsistencies and problems where find in the OOXML definition. It has inconsistencies with the following ISO standards:

Standard	OOXML Ref	Description

Page Size Names (ISO 216)	3.3.1.61, 3.3.1.62	ISO 216 defines names for page sizes, but OOXML uses its own numeric codes for these sizes rather than the existing standard.
Language codes ISO 639	2.18.52	OOXML defines its own language codes that are inconsistent with the standard
Colour Names ISO/IEC 15445 (HTML)	2.18.46	Not only does OOXML sometimes redefine some colour names, but it also redefines some colours corresponding to names belonging to the standard. The colours “dark blue”, “dark cyan”, “dark gray”, “dark green”, “dark red” and “light gray” are different in OOXML to the existing standard.
Dates and times ISO 8601	4.17.4.1	OOXML represents the dates as integers from 31st December 1899 with the caveat that 1900 needs to be incorrectly considered a leap year, or 1st January 1004 depending on a configuration setting. This is enormously inconsistent with the standard which represents the date more descriptively.
Readable XML ISO 8879	Almost all	Tag names such as scrgbClr, algn, dir, dist, rPr rotWithShape and w are neither consistent nor human-readable. There are many other examples.

OOXML have also inconsistencies references to external or redundant non-standard resources which were even disapproved by ISO before:

Resource	**OOXML Ref**	**Description**
Vector graphics	14, 8.6.2	OOXML defines its own vector graphics XML—DrawingML. But the recognised standard for this, also recommended by W3C, is SVG. OOXML also includes Microsoft’s VML specification in contradiction to both SVG and DrawingML. VML was turned down as a W3C standard in 1999 in favour of SVG.
Objects	6.2.3.17 6.4.3.1	OOXML references Windows Metafiles and Enhanced Metafiles, both closed proprietary Microsoft formats.
Configuration	2.15.3	OOXML contains specific application configuration settings, most notably: autoSpaceLikeWord95, footnoteLayoutLikeWW8, lineWrapLikeWord6, mwSmallCaps, shapeLayoutLikeWW8, suppressTopSpacingWP, truncateFontHeightsLikeWP6, uiCompat97To2003, useWord2002TableStyleRules, useWord97LineBreakRules, wpJustification and wpSpaceWidth.

Percentages	2.15.1.95, 2.18.97, 5.1.12.41	The OOXML standard is inconsistent within itself, as well as with recognised methods, of the representations of percentage values, which can be expressed as a decimal integer (Magnification Settings—2.15.1.95), as a code made up of an integer being the real percentage multiplied by 500 (Table Width Units—2.18.97) and a real percentage multiplied by 1000 (Generic Percentage Unit—5.1.12.41).

Criticism on OOXML in relation with OpenDocument

The OOXML standard has been the subject of debate within the software industry. Opponents to OOXML included FFII, the OpenDocument Format Alliance, IBM, and many nations that voiced strong opposition during standardisation as explained above.

At over 6,000 pages in length, the specification is difficult to evaluate quickly. Objectors also claim that there could be user confusion regarding the two standards because of the similarity of the "Office Open XML" name to both "OpenDocument" and "OpenOffice".

Objectors also argued that an ISO standard for documents already exists and there is no need for a second standard. It is clearly pointless to have OOXML when OpenDocument already exists as a standard for office document interoperability. Where a standard exists for a purpose then it should be used for that. It is pointless to create a competing standard for its own sake. It seems illogical to create a competing standard when one already exists and is meeting industrial strength requirements. To intentionally create a standard which is known simply to duplicate the functionality of another just results in extra costs with no benefits to those who need to adopt it.

Google stated that "the OpenDocument standard, which achieves the same goal, is only 867 pages" and that "if ISO were to give OOXML with its 6546 pages the same level of review which other standards have seen, it would take 18 years (6576 days for 6546 pages) to achieve comparable levels of review to the existing OpenDocument standard (871 days for 867 pages) which achieves the same purpose and is thus a good comparison."

The OpenDocument Format Alliance UK Action Group has stated that with OpenDocument an ISO standard for Office files already exists. Further, they argue that the Office Open XML file-format is

heavily based on Microsoft's own Office applications and is thus not vendor-neutral, and that it has inconsistencies with existing ISO standards such as time and date formats and color codes.

OpenDocument vs. OOXML

Comparing the codes

To understand the different technical approach of each one of the implementations. The following example was created in the text processors where the standards are implemented (OpenOffice.org Writer and Microsoft Word). The XML code below represents the same document.

ODF XML representation of the example

```
<text:h text:style-name="P1" text:outline-level="1">
    Example document
</text:h>
<text:p text:style-name="Standard">
    This has some
    <text:span text:style-name="T1">
        bold formatting
    </text:span>
    , also some
    <text:span text:style-name="T2">
        with italics
    </text:span>
     , a
    <text:a xlink:type="simple" xlink:href="http://www.odfalliance.com">
        <text:span text:style-name="Internet_20_link">
            web link
        </text:span>
    </text:a>
     and a picture...
</text:p>
<text:p text:style-name="Standard">
    <draw:frame draw:style-name="fr1" draw:name="graphics1" text:anchor-
type="as-char"
                          svg:width="5.9929in" svg:height="5.4362in" draw:z-
index="0">
        <draw:image
xlink:href="Pictures/10000000000002DC00000298CDD44AEF.jpg"
                       xlink:type="simple" xlink:show="embed"
xlink:actuate="onLoad"/>
    </draw:frame>
</text:p>
```

OOXML XML respresentation of the example

```
<w:p>
    <w:pPr>
        <w:pStyle w:val="Heading1"/>
    </w:pPr>
    <w:r>
        <w:t>
            Example document
        </w:t>
    </w:r>
</w:p>
<w:p>
    <w:r>
        <w:t>
            This has some
        </w:t>
    </w:r>
    <w:r>
        <w:rPr>
            <w:b/>
        </w:rPr>
        <w:t>
            bold formatting
        </w:t>
    </w:r>
    <w:r>
        <w:t>
             , also some
        </w:t>
    </w:r>
    <w:r>
        <w:rPr>
            <w:i/>
        </w:rPr>
        <w:t>
            with italics
        </w:t>
    </w:r>
    <w:r>
        <w:t>
            , a
        </w:t>
    </w:r>
    <w:w:hyperlink w:rel="rId4" w:history="1">
        <w:r>
            <w:rPr>
                <w:w:rStyle w:val="Hyperlink"/>
            </w:rPr>
            <w:t>
                web link
            </w:t>
        </w:r>
    </w:w:hyperlink>
    <w:r>
        <w:t>
             and a picture...
```

```
            </w:t>
        </w:r>
    </w:p>
    <w:p>
        <w:r>
            <w:pict>
                <v:shapetype id="_x0000_t75" coordsize="21600,21600" o:spt="75"
                             o:preferrelative="t" path="m@4@5l@4@11@9@11@9@5xe"
                                          filled="f" stroked="f">
                    <v:stroke joinstyle="miter"/>
                    <v:formulas>
                        <v:f eqn="if lineDrawn pixelLineWidth 0"/>
                        <v:f eqn="sum @0 1 0"/>
                        <v:f eqn="sum 0 0 @1"/>
                        <v:f eqn="prod @2 1 2"/>
                        <v:f eqn="prod @3 21600 pixelWidth"/>
                        <v:f eqn="prod @3 21600 pixelHeight"/>
                        <v:f eqn="sum @0 0 1"/>
                        <v:f eqn="prod @6 1 2"/>
                        <v:f eqn="prod @7 21600 pixelWidth"/>
                        <v:f eqn="sum @8 21600 0"/>
                        <v:f eqn="prod @7 21600 pixelHeight"/>
                        <v:f eqn="sum @10 21600 0"/>
                    </v:formulas>
                    <v:path o:extrusionok="f" gradientshapeok="t"
o:connecttype="rect"/>
                    <o:lock v:ext="edit" aspectratio="t"/>
                </v:shapetype>
                <v:shape id="_x0000_i1025" type="#_x0000_t75"
                         style="width:431.25pt;height:391.5pt">
                    <v:imagedata w:rel="rId5" o:title="dalek"/>
                </v:shape>
            </w:pict>
        </w:r>
    </w:p>
```

Interoperability between both standards

The fundamental differences between OpenDocument and OOXML means they cannot be merged. OOXML was designed to be compatible with Microsoft Office documents, and OpenDocument was designed for document interchange; this as well as some “nitty gritty” differences makes merging them impossible.

Although OOXML and OpenDocument have too many namespace and structural differences to be merged, OpenDocument can store documents created for OOXML, but not vice-versa. OOXML therefore should not be a standard.

However, translators can be written, are being written and have been written to convert from one

format to the other; always with limitations and problems. In fact, two versions of the same document saved in separate standards should also look the same and have the same layout (as we have seen in the XML code comparison below). However, from a technical point of view layout will be determined by the rendering engine and the available hardware and software (resolution, text fonts, colors, etc.). Layout thus only indirectly depends on the standards. There could be differences in terms of the inner structure, even when two documents look alike. On the other hand, two documents with different layouts could have the same inner structure and contents. For this reason it is extremely important to be aware of the invariable aspects of those documents that must be transferred between different organisations, and to decide on a document format by taking into account the reasons for document exchange.

It may be concluded that many of the functionalities, especially those found in simpler documents, can be translated between the standards, while the translation of other functionalities can prove complex or even impossible. Frequently in individual cases it has to be decided, if the conversion of a document is completely translatable, translatable only to a limited extend or not at all. The individual cases are determined by different constraints. First and foremost translatability depends on the document itself together with its characteristics. In addition the application or tool used for the transformation has also to be considered. In this study, statements on translatability and its quality in principle are made. As the rules used for transformation are not standardised however, each application is allowed to use its own specific rules. Under certain circumstances, specific rules can neglect certain properties and make specific assumptions which could enhance translatability. In addition, the direction of the transformation has to be considered. In many cases, a document can be translated without any major loss from one format to the other. Even so, on a round-trip conversion it cannot be guaranteed that the initial document and the document resulting from the conversion will be identical.

Final words

Standards exist for interoperability, and office document format standards should not be different. The goal is that someone in country A working for company B using product C can interchange documents with someone in country D working for company E using product F without any thought as to what precisely A, B, C, D, E, F or any other letter actually is. It simply works. There is no need to worry if any single vendor would continue in the office suite business or not, as any other vendor could be used.

OpenDocument was created using existing standards with this interoperability in mind, using long

public consultation and design periods to achieve this. The benefits of this are evident when examining the resulting formats themselves.

It has been implemented by a large number of office products and the list is growing.

OOXML was designed by a single vendor, Microsoft, with no extensive public consultation or design input. It was largely designed to co-exist with their legacy formats using their own products. The design of the specification is such that might happen if their own legacy closed binary formats were simply XML-ised—that is binary encodings simply converted to arbitrary XML tags.

Edward Macnaghten, consultant for the OpenDocument Format Alliance UK Action Group and author of the impressive "ODF/OOXML technical white paper" concluded his analysis saying:

> "Upon examining the formats it is difficult to ascertain any technical reason why Microsoft Office documents cannot be saved and interchanged using ODF with one hundred percent reliability. ODF has the features that will deal with all Microsoft Office's quirks. However, OOXML in its current state cannot handle any applications *except* Microsoft Office.
>
> It is my opinion that Microsoft peculiarities in OOXML, together with the fact the specification is over 6000 pages long, would greatly hinder the ability of other parties to develop products that would completely, or near completely, read and manipulate documents in that format and to the extent that it would render it practically difficult to work with as a universal standard."

Conclusions

As we have seen along this book, Open Standards are a central piece of the Free Technologies infrastructures. We explored the different definitions to get an overview on how complex is this topic and we met the most relevant actors in the Open Standards discussions. At the end we listed and described the most used implementations and open formats that currently exists.

We can conclude now on how important standards are for the IT sector, communications convergence and the Internet. No single technology, group or vendor can provide for everything and, therefore, interoperability in a heterogeneous environment is required more than ever. It is only by strict adherence to standards and specifications that a high degree of interoperability can be achieved.

Standards that are open and non-discriminatory are preferred because there is no dependence on any single entity, all types of products can implement them and all interested parties can partake in their development.

Open standards are a precondition for technical neutrality. They enable interoperability, stimulate innovation and competition, enable platform independent access to digital information, and facilitate availability of knowledge and learning now and in the future.

Open Standards benefits doesn't end in the IT sector, public administrations should implement and promote them too. Because of its special role in society, the public sector must avoid to impose upon its communication partners the use of specific products, thus allowing these partners the freedom of

choice. As public administrations also have an obligation to maintain archives for future use, access to the information contained in documents must be ensured over long periods of time.

The use of Open Formats provides public administrations, businesses, educative institutions and citizens with a wide range of products and devices capable of reading, writing and manipulating documents while stimulating competition and innovation in the area of document handling. Preferably, these open document exchange and storage using open formats would be subject to formal standardisation via international standardisation procedures. Finally we can remark the importance of openness in relation with democracy, quoting the paper entitled “Open documents and Democracy”:

> “Free and open access to many types of government documents is crucial for democratic government, either because ensuring dependable, equal, and free access constitutes a condition of democracy, or because the provision or recordation of certain documents constitute core public duties […] political reasons for open standards, as they did not only focus on cost-efficiency or other purely economic imperatives, but were significantly concerned with promoting distinctly political values—either invoking the specific values of democratic equality of access or public responsibility that we articulated in this paper, or closely related values, such as the principle of citizen choice or government independence from proprietary control. We can conclude that in the present context, movement towards openness in technical standards by both governments and vendors is highly beneficial for citizens who care about democratic values.”

We hope this coursebook can serve as a useful guide towards a better understanding of the importance

of Open Standards in all described levels, and we hope this compilation promote even more progress and development of Free Technologies in a global scale.

Glossary

AAC	Advanced Audio Coding
ABI	Application Binary Interface
AMD	Advanced Micro Device
ANSI	American National Standards Institute
APDIP	Asia-Pacific Development Information Programme
API	Application Programme Interface
ARPANET	Advanced Research Projects Agency Network
ASCII	American Standard Code for Information Interchange
ASF	Advanced Systems Format
AVC	Advanced Video Coding
AVI	Audio Video Interleave
BIS	Bureau of Indian Standards
BSA	Business Software Alliance
BSD	Berkeley Software Distribution
BSI	British Standards Institute
ccTLD	Country Code Top-Level Domain
CD	Compact Disc
CDDL	Common Development and Distribution License
CE	Consumer Electronics

CII	Computer-Implemented Invention
COTS	Common Off The Shelf
DNS	Domain Name Service
DOC	Document
DOS	Disk Operating System
DTD	Document Type Definition
DVD	Digital Video Disc
EC	European Commission
ECMA	European Computer Manufacturers Association
EEE	Embrace, Extend and Extinguish
e-GIF	e-Government Interoperability Framework
EICTA	European Information, Communications and Consumer Technology Industry Association
EIF	European Interoperability Framework
ELF	Executable and Linking Format
EPC	European Patent Convention
EPO	European Patent Office
ETRM	Enterprise Technology Reference Model
ETSI	European Telecommunications Standards Institute
EU	European Union
FAQ	Frequently Asked Questions
FAT	File Allocation Table
FHS	Filesystem Hierarchy Standard
FLAC	Free Lossless Audio Codec
FSF	Free Software Foundation

FSG	Free Standards Group
FTA	Free Technology Academy
G2B	Government to Business
GIF	Graphics Interchange Format
GL	Graphics Library
GNOME	GNU Network Object Model Environment
GNU	Gnu's Not Unix
GPL	General Public License (GNU)
GSM	Global System for Mobile Communications
GUI	Graphics User Interface
HTML	Hypertext Markup Language
HTTP	Hypertext Transfer Protocol
HTTPd	Hypertext Transfer Protocol Daemon
I18N	Internationalization
IA	Internet Architecture
IAB	Internet Architecture Board
IANA	Internet Assigned Numbers Authority
ICT	Information and Communications Technology
ICTU	Dutch organisation for Information and Communications Technology
IDABC	Interoperable Delivery of European e-Government Services to Public Administrations, Businesses and Citizens
IEC	International Electro-Technical Commission
IEEE	Institute of Electrical and Electronics Engineers
IEEE-SA	Institute of Electrical and Electronics Engineers Standards Association
IESG	Internet Engineering Steering Group

IETF	Internet Engineering Task Force
IJG	Independent JPEG Group
IMAP	Internet Message Access Protocol
IOSN	International Open Source Network
IP	Internet Protocol
IPR	Intellectual Property Rights
IS	Information System
ISO	International organisation for Standardisation
ISOC	Internet Society
ISV	Independent Software Vendor
IT	Information Technology
ITU	International Telecommunication Union
ITU-D	International Telecommunication Union –Development Sector
ITU-R	International Telecommunication Union –Radio Communication Sector
ITU-T	International Telecommunication Union –Telecommunication Standardization Sector
JFIF	Joint Photographic Experts Group File Interchange Format
JPEG	Joint Photographic Experts Group
JTC	Joint Technical Committee
KDE	K Desktop Environment
L10N	Localization
LAN	Local Area Network
LCID	Locale Identifier
LDAP	Lightweight Directory Access Protocol
LSB	Linux Standard Base

LZW	Lempel-Ziv-Welch
M17N-li	Multi-lingualisation library
MAN	Metropolitan Area Network
MIME	Multipurpose Internet Mail Extensions
MLP	Mozilla Localisation Project
MNCC	Malaysian National Computer Confederation
MOV	Movie File Format
MP3	Motion Pictures Experts Group-1 Audio Layer 3
MP4	Motion Pictures Experts Group-4 Part 14
MPEG	Motion Pictures Experts Group
MPL	Mozilla Public License
MS	Microsoft
MUI	Multilingual User Interface
MyGIF	Malaysian Government Interoperability Framework
NCSA	National Center for Supercomputing Applications
NFS	Network File System
OASIS	Organisation for the Advancement of Structured Information Standards
ODP	OpenDocument Presentation
ODS	OpenDocument Spreadsheet
ODT	OpenDocument Text
OOo	OpenOffice.org
OpenI18N	Open Internationalisation
ORG	Organisation
OS	Open Standard

OSDL	Open Source Development Labs
OSI	Open Source Initiative
OSOSS	Open Standards and Open Source Software
PAG	Patent Advisory Group
PAS	Publicly Available Specification
PC	Personal Computer
PCL	Printer Control Language
PDA	Personal Digital Assistant
PDF	Portable Document Format
PIKOM	Association of the Computer and Multimedia Industry of Malaysia
PNG	Portable Network Graphics
POSIX	Portable Operating System Interface for UNIX
PPC	Power Personal Computer
PPT	PowerPoint
RAND	Reasonable and Non-Discriminatory
RF	Royalty-Free
RFC	Request for Comments
RIFF	Resource Interchange File Format
RPM	Red Hat Package Manager
RTF	Rich Text Format
SDO	Standard Development organisation
SGML	Standard Generalized Markup Language
SMPTE	Society of Motion Picture and Television Engineers
SMTP	Simple Mail Transfer Protocol

SOAP	Simple Object Access Protocol
SQL	Structured Query Language
SSL	Secure Sockets Layer
SSO	Standard-Setting organisation
STD	Standard
SVG	Scalable Vector Graphics
SVID	System V Interface Definition
SXC	StarOffice Calc
SXI	StarOffice Impression
SXW	StarOffice Writer
TCP	Transmission Control Protocol
TIFF	Tagged Image File Format
TV	Television
UCS	Universal Character Set
UDDI	Universal Description, Discovery, and Integration
UFO	Uniform Fee Only
UNDP	United Nations Development Programme
UNICODE	Unique, Universal and Uniform Character enCoding
USA	United States of America
UTF	Universal Character Set Transformation Format
VC	Video Codec
VCD	Video Compact Disc
VHS	Video Home System
VMS	Virtual Memory System

W3C	World Wide Web Consortium
WAV	WAVE form Audio Format
Wi-Fi	Wireless Fidelity
WMA	Windows Media Audio
WMV	Windows Media Video
WWW	World Wide Web
XLS	Excel
XML	Extensible Markup Language
XPath	Extensible Markup Language Path Language
XPM	XPixMap
XSL	Extensible Stylesheet Language
XSL-FO	Extensible Stylesheet Language Formatting Objects
XSLT	Extensible Stylesheet Language Transformations

Bibliography and links

"The Importance of Open Standards in Interoperability", Open Forum Europe, 2008 : http://www.openforumeurope.org/library/onepage-briefs/ofe-open-standards-onepage-2008.pdf

"FOSS Open Standards", Nah Soo Hoe, 2006: http://www.iosn.net/open-standards/foss-open-standards-primer/foss-openstds-withcover.pdf

"Open Standards: Principles and Practice", Bruce Perens: http://perens.com/OpenStandards/Definition.html

"Chapter 8: Open Standards and Intellectual Property Rights" in "Open Innovation: Researching a New Paradigm", Tim Simcoe: http://www.rotman.utoronto.ca/timothy.simcoe/papers/OpenStandards_IPR.pdf

"Sovereign Software: Open Standards, Free Software, and the Internet", G. Greve: http://www.intgovforum.org/Substantive_1st_IGF/SovereignSoftware.pdf

Open standard. (2010, May 24). In *Wikipedia, The Free Encyclopedia*. Retrieved 22:17, May 26, 2010, from http://en.wikipedia.org/w/index.php?title=Open_standard&oldid=363992271

"Open Standards requirements", Ken Krechmer, International Center for Standards Research - University of Colorado, 2005: http://www.csrstds.com/openstds.pdf

"Patents and Open Standards", OASIS: http://xml.coverpages.org/patents.html#RAND

"Four reasons not to use proprietary formats", Openformats.org: http://www.openformats.org/en3

Proprietary format. (2010, May 23). In *Wikipedia, The Free Encyclopedia*. Retrieved 22:22, May 26, 2010, from http://en.wikipedia.org/w/index.php?title=Proprietary_format&oldid=363740449

"Defining Free and Open Standard", Digital Standards Organization: http://www.digistan.org/text:rationale

"Open Documents and Democracy: A Political Basis for Open Document Standards", Laura DeNardis and Eric Tam, Yale Law School, 2007: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1028073

Office Open XML. (2010, May 12). In *Wikipedia, The Free Encyclopedia*. Retrieved 22:20, May 26, 2010, from http://en.wikipedia.org/w/index.php?title=Office_Open_XML&oldid=361748309

Standardization of Office Open XML. (2010, May 12). In *Wikipedia, The Free Encyclopedia*. Retrieved 22:20, May 26, 2010, from http://en.wikipedia.org/w/index.php?title=Standardization_of_Office_Open_XML&oldid=361745775

"ODF/OOXML technical white paper", Edward Macnaghten, ODF Alliance UK Action Group, 2007: ftp://officeboxsystems.com/odfa_ukag/ODFA%20UKAG%20Technical%20White%20Paper.pdf

"Document Interoperability: Open Document Format and Office Open XML", Dr. Klaus-Peter Eckert, Jan Henrik Ziesing and Ucheoma Ishionwu, 2009: http://www.fokus.fraunhofer.de/de/elan/_docs/WP-DocInterop-en2.pdf

OpenDocument. (2010, May 19). In *Wikipedia, The Free Encyclopedia*. Retrieved 22:19, May 26, 2010, from http://en.wikipedia.org/w/index.php?title=OpenDocument&oldid=362914625

OpenDocument adoption. (2010, May 17). In *Wikipedia, The Free Encyclopedia*. Retrieved 22:18, May 26, 2010, from http://en.wikipedia.org/w/index.php?title=OpenDocument_adoption&oldid=362681718

OpenDocument standardization. (2009, December 26). In *Wikipedia, The Free Encyclopedia*. Retrieved 22:18, May 26, 2010, from http://en.wikipedia.org/w/index.php?title=OpenDocument_standardization&oldid=334090768

7z. (2010, May 22). In *Wikipedia, The Free Encyclopedia*. Retrieved 22:16, May 26, 2010, from http://en.wikipedia.org/w/index.php?title=7z&oldid=363515763

Bzip2. (2010, May 26). In Wikipedia, The Free Encyclopedia. Retrieved 22:37, May 26, 2010, from

http://en.wikipedia.org/w/index.php?title=Bzip2&oldid=364299704

Gzip. (2010, May 14). In Wikipedia, The Free Encyclopedia. Retrieved 22:37, May 26, 2010, from http://en.wikipedia.org/w/index.php?title=Gzip&oldid=362063072

Tar (file format). (2010, May 26). In Wikipedia, The Free Encyclopedia. Retrieved 22:46, May 26, 2010, from http://en.wikipedia.org/w/index.php?title=Tar_(file_format)&oldid=364302909

ZIP (file format). (2010, May 18). In Wikipedia, The Free Encyclopedia. Retrieved 22:48, May 26, 2010, from http://en.wikipedia.org/w/index.php?title=ZIP_(file_format)&oldid=362747044

Internationalization and localization. (2010, May 11). In Wikipedia, The Free Encyclopedia. Retrieved 23:19, May 26, 2010, from http://en.wikipedia.org/w/index.php?title=Internationalization_and_localization&oldid=361467128

Locale. (2010, April 19). In Wikipedia, The Free Encyclopedia. Retrieved 23:19, May 26, 2010, from http://en.wikipedia.org/w/index.php?title=Locale&oldid=357009112

The Microsoft Developer Network (MSDN), "Locale Identifiers" http://msdn.microsoft.com/library/default.asp?url=/library/en-us/intl/nls_8sj7.asp

ISO/IEC 10646:2003, "Information technology - Universal Multiple-Octet Coded Character Set (UCS)" http://www.iso.ch/iso/en/CatalogueDetailPage.CatalogueDetail?CSNUMBER=39921&ICS1=35&ICS2=40&ICS3=

The Unicode Standard http://www.unicode.org/standard/standard.html

Kuhn, M., "UTF-8 and Unicode FAQ for Unix/Linux" http://www.cl.cam.ac.uk/~mgk25/unicode.html

ISO/IEC 10646:2003, "Information technology - Universal Multiple-Octet Coded Character Set (UCS)" http://www.iso.ch/iso/en/CatalogueDetailPage.CatalogueDetail?CSNUMBER=39921&ICS1=35&ICS2=40&ICS3=

RFC 3629, "UTF-8, a transformation format of ISO 10646" http://www.ietf.org/rfc/rfc3629.txt

Unicode. (2010, May 25). In Wikipedia, The Free Encyclopedia. Retrieved 23:20, May 26, 2010, from http://en.wikipedia.org/w/index.php?title=Unicode&oldid=364040338

ISO 639 Frequently Asked Questions (FAQ) http://www.loc.gov/standards/iso639-2/faq.html

RFC 3066, "Tags for the Identification of Languages" http://www.ietf.org/rfc/rfc3066.txt

RFC 4646, "Tags for Identifying Languages" http://www.ietf.org/rfc/rfc4646.txt

The Open Internationalization Initiative http://www.openi18n.org

The Free Standards Group http://www.freestandards.org

OpenI18N 1.3 Globalization Specification http://www.openi18n.org/docs/pdf/OpenI18N1.3.pdf

The Mozilla Localization Project http://www.mozilla.org/projects/l10n

The GNOME Translation Project http://developer.gnome.org/projects/gtp

KDE Internationalization http://i18n.kde.org

OpenOffice.org L10N and I18N Projects http://l10n.openoffice.org

Windows XP LIP FAQ http://www.microsoft.com/globaldev/DrIntl/faqs/winxp.mspx

Office 2003 Editions Localized Versions http://www.microsoft.com/office/editions/prodinfo/language/localized.mspx

Office 2003 XML Reference Schemas http://www.microsoft.com/Office/xml/default.mspx

CNET News, 1 June 2005, "Microsoft adding XML files to Office 12" http://news.com.com/Microsoft+adding+XML+files+to+Office+12/2100-7344_3-5728536.html?tag=st.ref.goo

The OpenOffice.org Project http://www.openoffice.org

Adobe Inc., "What is Adobe PDF?" http://www.adobe.com/products/acrobat/adobepdf.html

Portable Document Format. (2010, May 22). In Wikipedia, The Free Encyclopedia. Retrieved 23:21, May 26, 2010, from http://en.wikipedia.org/w/index.php?title=Portable_Document_Format&oldid=363605488

Raster graphics. (2010, May 25). In Wikipedia, The Free Encyclopedia. Retrieved 23:21, May 26, 2010, from http://en.wikipedia.org/w/index.php?title=Raster_graphics&oldid=364079065

Graphics Interchange Format Version 89a http://www.w3.org/Graphics/GIF/spec-gif89a.txt

Portable Network Graphics (PNG) Recommendation http://www.w3.org/TR/PNG/

The XPM Format and Library http://koala.ilog.fr/lehors/xpm.html

Adobe Inc., "TIFF Specifications" http://partners.adobe.com/public/developer/tiff/index.html

The JPEG Homepage http://www.jpeg.org/jpeg/index.html

JPEG JFIF http://www.w3.org/Graphics/JPEG/

JPEG image compression FAQ, part 1 http://www.faqs.org/faqs/jpeg-faq/part1/

JPEG JFIF http://www.w3.org/Graphics/JPEG/

Wikipedia (the free-content encyclopedia) entry on "Vector graphics" http://en.wikipedia.org/wiki/Vector_graphics

Scalable Vector Graphics (SVG) http://www.w3.org/Graphics/SVG/

WAV. (2010, May 25). In Wikipedia, The Free Encyclopedia. Retrieved 23:22, May 26, 2010, from

http://en.wikipedia.org/w/index.php?title=WAV&oldid=364090557

The FLAC Project Page http://flac.sourceforge.net

The Xiph.Org Foundation http://www.xiph.org

Ogg Vorbis General FAQ http://www.vorbis.com/faq.psp

Vorbis Wiki at Xiph.org http://wiki.xiph.org/index.php/Vorbis

Microsoft Developer Network, "AVI RIFF File Reference"

http://msdn.microsoft.com/archive/default.asp?url=/archive/en-us/dx81_c/directx_cpp/htm/aviriifffilereference.asp

Overview of the MPEG-4 Standard http://www.chiariglione.org/mpeg/standards/mpeg-4/mpeg-4.htm

The Ogg Encapsulation Format Version 0 http://www.faqs.org/rfcs/rfc3533.html

Xiph.org Wiki, "Projects/Formats" http://wiki.xiph.org/index.php/Main_Page

H.264. (2005, November 2). In Wikipedia, The Free Encyclopedia. Retrieved 23:22, May 26, 2010, from http://en.wikipedia.org/w/index.php?title=H.264&oldid=27134487

License of this document

GNU Free Documentation License

Version 1.3, 3 November 2008

0. PREAMBLE

The purpose of this License is to make a manual, textbook, or other functional and useful document "free" in the sense of freedom: to assure everyone the effective freedom to copy and redistribute it, with or without modifying it, either commercially or noncommercially. Secondarily, this License preserves for the author and publisher a way to get credit for their work, while not being considered responsible for modifications made by others.

This License is a kind of "copyleft", which means that derivative works of the document must themselves be free in the same sense. It complements the GNU General Public License, which is a copyleft license designed for free software.

We have designed this License in order to use it for manuals for free software, because free software needs free documentation: a free program should come with manuals providing the same freedoms that the software does. But this License is not limited to software manuals; it can be used for any textual work, regardless of subject matter or whether it is published as a printed book. We recommend this License principally for works whose purpose is instruction or reference.

1. APPLICABILITY AND DEFINITIONS

This License applies to any manual or other work, in any medium, that contains a notice placed by the copyright holder saying it can be distributed under the terms of this License. Such a notice grants a world-wide, royalty-free license, unlimited in duration, to use that work under the conditions stated herein. The "Document", below, refers to any such manual or work. Any member of the public is a licensee, and is addressed as "you". You accept the license if you copy, modify or distribute the work in

a way requiring permission under copyright law.

A "Modified Version" of the Document means any work containing the Document or a portion of it, either copied verbatim, or with modifications and/or translated into another language.

A "Secondary Section" is a named appendix or a front-matter section of the Document that deals exclusively with the relationship of the publishers or authors of the Document to the Document's overall subject (or to related matters) and contains nothing that could fall directly within that overall subject. (Thus, if the Document is in part a textbook of mathematics, a Secondary Section may not explain any mathematics.) The relationship could be a matter of historical connection with the subject or with related matters, or of legal, commercial, philosophical, ethical or political position regarding them.

The "Invariant Sections" are certain Secondary Sections whose titles are designated, as being those of Invariant Sections, in the notice that says that the Document is released under this License. If a section does not fit the above definition of Secondary then it is not allowed to be designated as Invariant. The Document may contain zero Invariant Sections. If the Document does not identify any Invariant Sections then there are none.

The "Cover Texts" are certain short passages of text that are listed, as Front-Cover Texts or Back-Cover Texts, in the notice that says that the Document is released under this License. A Front-Cover Text may be at most 5 words, and a Back-Cover Text may be at most 25 words.

A "Transparent" copy of the Document means a machine-readable copy, represented in a format whose specification is available to the general public, that is suitable for revising the document straightforwardly with generic text editors or (for images composed of pixels) generic paint programs or (for drawings) some widely available drawing editor, and that is suitable for input to text formatters or for automatic translation to a variety of formats suitable for input to text formatters. A copy made in an otherwise Transparent file format whose markup, or absence of markup, has been arranged to thwart or discourage subsequent modification by readers is not Transparent. An image format is not Transparent if used for any substantial amount of text. A copy that is not "Transparent" is called "Opaque".

Examples of suitable formats for Transparent copies include plain ASCII without markup, Texinfo input format, LaTeX input format, SGML or XML using a publicly available DTD, and standard-conforming simple HTML, PostScript or PDF designed for human modification. Examples of

transparent image formats include PNG, XCF and JPG. Opaque formats include proprietary formats that can be read and edited only by proprietary word processors, SGML or XML for which the DTD and/or processing tools are not generally available, and the machine-generated HTML, PostScript or PDF produced by some word processors for output purposes only.

The "Title Page" means, for a printed book, the title page itself, plus such following pages as are needed to hold, legibly, the material this License requires to appear in the title page. For works in formats which do not have any title page as such, "Title Page" means the text near the most prominent appearance of the work's title, preceding the beginning of the body of the text.

The "publisher" means any person or entity that distributes copies of the Document to the public.

A section "Entitled XYZ" means a named subunit of the Document whose title either is precisely XYZ or contains XYZ in parentheses following text that translates XYZ in another language. (Here XYZ stands for a specific section name mentioned below, such as "Acknowledgements", "Dedications", "Endorsements", or "History".) To "Preserve the Title" of such a section when you modify the Document means that it remains a section "Entitled XYZ" according to this definition.

The Document may include Warranty Disclaimers next to the notice which states that this License applies to the Document. These Warranty Disclaimers are considered to be included by reference in this License, but only as regards disclaiming warranties: any other implication that these Warranty Disclaimers may have is void and has no effect on the meaning of this License.

2. VERBATIM COPYING

You may copy and distribute the Document in any medium, either commercially or noncommercially, provided that this License, the copyright notices, and the license notice saying this License applies to the Document are reproduced in all copies, and that you add no other conditions whatsoever to those of this License. You may not use technical measures to obstruct or control the reading or further copying of the copies you make or distribute. However, you may accept compensation in exchange for copies. If you distribute a large enough number of copies you must also follow the conditions in section 3.

You may also lend copies, under the same conditions stated above, and you may publicly display copies.

3. COPYING IN QUANTITY

If you publish printed copies (or copies in media that commonly have printed covers) of the Document, numbering more than 100, and the Document's license notice requires Cover Texts, you must enclose the copies in covers that carry, clearly and legibly, all these Cover Texts: Front-Cover Texts on the front cover, and Back-Cover Texts on the back cover. Both covers must also clearly and legibly identify you as the publisher of these copies. The front cover must present the full title with all words of the title equally prominent and visible. You may add other material on the covers in addition. Copying with changes limited to the covers, as long as they preserve the title of the Document and satisfy these conditions, can be treated as verbatim copying in other respects.

If the required texts for either cover are too voluminous to fit legibly, you should put the first ones listed (as many as fit reasonably) on the actual cover, and continue the rest onto adjacent pages.

If you publish or distribute Opaque copies of the Document numbering more than 100, you must either include a machine-readable Transparent copy along with each Opaque copy, or state in or with each Opaque copy a computer-network location from which the general network-using public has access to download using public-standard network protocols a complete Transparent copy of the Document, free of added material. If you use the latter option, you must take reasonably prudent steps, when you begin distribution of Opaque copies in quantity, to ensure that this Transparent copy will remain thus accessible at the stated location until at least one year after the last time you distribute an Opaque copy (directly or through your agents or retailers) of that edition to the public.

It is requested, but not required, that you contact the authors of the Document well before redistributing any large number of copies, to give them a chance to provide you with an updated version of the Document.

4. MODIFICATIONS

You may copy and distribute a Modified Version of the Document under the conditions of sections 2 and 3 above, provided that you release the Modified Version under precisely this License, with the Modified Version filling the role of the Document, thus licensing distribution and modification of the Modified Version to whoever possesses a copy of it. In addition, you must do these things in the Modified Version:

- A. Use in the Title Page (and on the covers, if any) a title distinct from that of the Document,

and from those of previous versions (which should, if there were any, be listed in the History section of the Document). You may use the same title as a previous version if the original publisher of that version gives permission.

- B. List on the Title Page, as authors, one or more persons or entities responsible for authorship of the modifications in the Modified Version, together with at least five of the principal authors of the Document (all of its principal authors, if it has fewer than five), unless they release you from this requirement.
- C. State on the Title page the name of the publisher of the Modified Version, as the publisher.
- D. Preserve all the copyright notices of the Document.
- E. Add an appropriate copyright notice for your modifications adjacent to the other copyright notices.
- F. Include, immediately after the copyright notices, a license notice giving the public permission to use the Modified Version under the terms of this License, in the form shown in the Addendum below.
- G. Preserve in that license notice the full lists of Invariant Sections and required Cover Texts given in the Document's license notice.
- H. Include an unaltered copy of this License.
- I. Preserve the section Entitled "History", Preserve its Title, and add to it an item stating at least the title, year, new authors, and publisher of the Modified Version as given on the Title Page. If there is no section Entitled "History" in the Document, create one stating the title, year, authors, and publisher of the Document as given on its Title Page, then add an item describing the Modified Version as stated in the previous sentence.
- J. Preserve the network location, if any, given in the Document for public access to a Transparent copy of the Document, and likewise the network locations given in the Document for previous versions it was based on. These may be placed in the "History" section. You may omit a network location for a work that was published at least four years before the Document itself, or if the original publisher of the version it refers to gives permission.
- K. For any section Entitled "Acknowledgements" or "Dedications", Preserve the Title of the section, and preserve in the section all the substance and tone of each of the contributor acknowledgements and/or dedications given therein.
- L. Preserve all the Invariant Sections of the Document, unaltered in their text and in their titles.

Section numbers or the equivalent are not considered part of the section titles.

- M. Delete any section Entitled "Endorsements". Such a section may not be included in the Modified Version.
- N. Do not retitle any existing section to be Entitled "Endorsements" or to conflict in title with any Invariant Section.
- O. Preserve any Warranty Disclaimers.

If the Modified Version includes new front-matter sections or appendices that qualify as Secondary Sections and contain no material copied from the Document, you may at your option designate some or all of these sections as invariant. To do this, add their titles to the list of Invariant Sections in the Modified Version's license notice. These titles must be distinct from any other section titles.

You may add a section Entitled "Endorsements", provided it contains nothing but endorsements of your Modified Version by various parties—for example, statements of peer review or that the text has been approved by an organisation as the authoritative definition of a standard.

You may add a passage of up to five words as a Front-Cover Text, and a passage of up to 25 words as a Back-Cover Text, to the end of the list of Cover Texts in the Modified Version. Only one passage of Front-Cover Text and one of Back-Cover Text may be added by (or through arrangements made by) any one entity. If the Document already includes a cover text for the same cover, previously added by you or by arrangement made by the same entity you are acting on behalf of, you may not add another; but you may replace the old one, on explicit permission from the previous publisher that added the old one.

The author(s) and publisher(s) of the Document do not by this License give permission to use their names for publicity for or to assert or imply endorsement of any Modified Version.

5. COMBINING DOCUMENTS

You may combine the Document with other documents released under this License, under the terms defined in section 4 above for modified versions, provided that you include in the combination all of the Invariant Sections of all of the original documents, unmodified, and list them all as Invariant Sections of your combined work in its license notice, and that you preserve all their Warranty Disclaimers.

The combined work need only contain one copy of this License, and multiple identical Invariant

Sections may be replaced with a single copy. If there are multiple Invariant Sections with the same name but different contents, make the title of each such section unique by adding at the end of it, in parentheses, the name of the original author or publisher of that section if known, or else a unique number. Make the same adjustment to the section titles in the list of Invariant Sections in the license notice of the combined work.

In the combination, you must combine any sections Entitled "History" in the various original documents, forming one section Entitled "History"; likewise combine any sections Entitled "Acknowledgements", and any sections Entitled "Dedications". You must delete all sections Entitled "Endorsements".

6. COLLECTIONS OF DOCUMENTS

You may make a collection consisting of the Document and other documents released under this License, and replace the individual copies of this License in the various documents with a single copy that is included in the collection, provided that you follow the rules of this License for verbatim copying of each of the documents in all other respects.

You may extract a single document from such a collection, and distribute it individually under this License, provided you insert a copy of this License into the extracted document, and follow this License in all other respects regarding verbatim copying of that document.

7. AGGREGATION WITH INDEPENDENT WORKS

A compilation of the Document or its derivatives with other separate and independent documents or works, in or on a volume of a storage or distribution medium, is called an "aggregate" if the copyright resulting from the compilation is not used to limit the legal rights of the compilation's users beyond what the individual works permit. When the Document is included in an aggregate, this License does not apply to the other works in the aggregate which are not themselves derivative works of the Document.

If the Cover Text requirement of section 3 is applicable to these copies of the Document, then if the Document is less than one half of the entire aggregate, the Document's Cover Texts may be placed on covers that bracket the Document within the aggregate, or the electronic equivalent of covers if the Document is in electronic form. Otherwise they must appear on printed covers that bracket the whole

aggregate.

8. TRANSLATION

Translation is considered a kind of modification, so you may distribute translations of the Document under the terms of section 4. Replacing Invariant Sections with translations requires special permission from their copyright holders, but you may include translations of some or all Invariant Sections in addition to the original versions of these Invariant Sections. You may include a translation of this License, and all the license notices in the Document, and any Warranty Disclaimers, provided that you also include the original English version of this License and the original versions of those notices and disclaimers. In case of a disagreement between the translation and the original version of this License or a notice or disclaimer, the original version will prevail.

If a section in the Document is Entitled "Acknowledgements", "Dedications", or "History", the requirement (section 4) to Preserve its Title (section 1) will typically require changing the actual title.

9. TERMINATION

You may not copy, modify, sublicense, or distribute the Document except as expressly provided under this License. Any attempt otherwise to copy, modify, sublicense, or distribute it is void, and will automatically terminate your rights under this License.

However, if you cease all violation of this License, then your license from a particular copyright holder is reinstated (a) provisionally, unless and until the copyright holder explicitly and finally terminates your license, and (b) permanently, if the copyright holder fails to notify you of the violation by some reasonable means prior to 60 days after the cessation.

Moreover, your license from a particular copyright holder is reinstated permanently if the copyright holder notifies you of the violation by some reasonable means, this is the first time you have received notice of violation of this License (for any work) from that copyright holder, and you cure the violation prior to 30 days after your receipt of the notice.

Termination of your rights under this section does not terminate the licenses of parties who have received copies or rights from you under this License. If your rights have been terminated and not permanently reinstated, receipt of a copy of some or all of the same material does not give you any rights to use it.

10. FUTURE REVISIONS OF THIS LICENSE

The Free Software Foundation may publish new, revised versions of the GNU Free Documentation License from time to time. Such new versions will be similar in spirit to the present version, but may differ in detail to address new problems or concerns. See http://www.gnu.org/copyleft/.

Each version of the License is given a distinguishing version number. If the Document specifies that a particular numbered version of this License "or any later version" applies to it, you have the option of following the terms and conditions either of that specified version or of any later version that has been published (not as a draft) by the Free Software Foundation. If the Document does not specify a version number of this License, you may choose any version ever published (not as a draft) by the Free Software Foundation. If the Document specifies that a proxy can decide which future versions of this License can be used, that proxy's public statement of acceptance of a version permanently authorizes you to choose that version for the Document.

11. RELICENSING

"Massive Multiauthor Collaboration Site" (or "MMC Site") means any World Wide Web server that publishes copyrightable works and also provides prominent facilities for anybody to edit those works. A public wiki that anybody can edit is an example of such a server. A "Massive Multiauthor Collaboration" (or "MMC") contained in the site means any set of copyrightable works thus published on the MMC site.

"CC-BY-SA" means the Creative Commons Attribution-Share Alike 3.0 license published by Creative Commons Corporation, a not-for-profit corporation with a principal place of business in San Francisco, California, as well as future copyleft versions of that license published by that same organisation.

"Incorporate" means to publish or republish a Document, in whole or in part, as part of another Document.

An MMC is "eligible for relicensing" if it is licensed under this License, and if all works that were first published under this License somewhere other than this MMC, and subsequently incorporated in whole or in part into the MMC, (1) had no cover texts or invariant sections, and (2) were thus incorporated prior to November 1, 2008.

The operator of an MMC Site may republish an MMC contained in the site under CC-BY-SA on the same site at any time before August 1, 2009, provided the MMC is eligible for relicensing.

ADDENDUM: How to use this License for your documents

To use this License in a document you have written, include a copy of the License in the document and put the following copyright and license notices just after the title page:

```
Copyright (C)  YEAR  YOUR NAME.
Permission is granted to copy, distribute and/or modify this document
under the terms of the GNU Free Documentation License, Version 1.3
or any later version published by the Free Software Foundation;
with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts.
A copy of the license is included in the section entitled "GNU
Free Documentation License".
```

If you have Invariant Sections, Front-Cover Texts and Back-Cover Texts, replace the "with … Texts." line with this:

```
with the Invariant Sections being LIST THEIR TITLES, with the
Front-Cover Texts being LIST, and with the Back-Cover Texts being LIST.
```

If you have Invariant Sections without Cover Texts, or some other combination of the three, merge those two alternatives to suit the situation.

If your document contains nontrivial examples of program code, we recommend releasing these examples in parallel under your choice of free software license, such as the GNU General Public License, to permit their use in fr