International Open Source Network
An Initiative of the UNDP's Asia-Pacific Development Information Programme supported by International Development Research Centre.

Free/Open Source Software Open Standards

Nah Soo Hoe

V1.1

Draft for feedback only. Not for circulation or quotation.
All feedback to be received by 5th September 2005

Draft for feedback only. Not for circulation or quotation.
All feedback to be received by 5th September 2005

IOSN - Free/Open Source Software: Open Standards

Published by
the United Nations Development Programme's
Asia-Pacific Development Information Programme (UNDP-APDIP)
Kuala Lumpur, Malaysia

Web: http://www.apdip.net/
Email: info@apdip.net

ISBN:

Table of Contents

PREFACE

INTRODUCTION

What are Standards and Why they are Important

The word “standard” has several different meanings. Working within the context of the subject matter of this document, its meaning in everyday usage [1-3] can be taken to refer to:

- a level of quality or attainment or
- an item or a specification against which all others may be measured

In technical usage, a standard [4] is a framework of specifications that has been:

- approved by a recognised organisation or
- 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 above is implied.

Standards are extremely important in modern society. They ensure that products and services are of sufficient 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 interchangeability - as well as in providing such benefits at an economical cost [5].

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 to the standards is usually done by common agreement. For example, the Internet will not achieve its current ubiquitous presence where it is accessible from almost any type of computer platforms and devices if it did not use widely accepted technical standards in its networking infrastructure and supported services.

Open Standards

Having defined what standards mean in general and technical usage let us turn our attention to the main focus of this primer - open standards. There are many opinions for what constitute open standards and different people have differing views on what this should mean or signify.

Wikipedia [7] defines open standards as publicly available specifications for achieving a specific task. Anyone is allowed to use the standard and so there is increased compatibility between components and among products from different vendors. The Open Standards Policy of the State of Massachusetts, USA [8] defines it as specifications for systems that are publicly available and are developed by an open community and affirmed by a standards body. On the other hand the European Commission's European Interoperability Framework [9] defines open standards as specifications having the minimal characteristics of:

- being adopted and maintained by a not-for-profit organisation
- its development using an open decision-making procedure available to all interested parties
- the published standard being available either freely or at a nominal charge and it must be permissible for all to copy, distribute and use it for no fee or at a nominal fee
- any patents present to be made irrevocably available on a royalty-free basis
- no constraints on the re-use of the standard

Last but not least, well known Open Source personality, Bruce Perens argues that an open standard is more than just a specification [10] and that the principles behind the standard and the practice of offering and operating the standard are what make the standard open. He proposes that open standards should follow the principles of availability, maximise end-user choice, no royalty, no discrimination, extension of subset and predatory practices and that certain practices be followed to ensure that these principles are adhered to.

From the above, we find that while there may be numerous detailed definitions and meanings given to open standards, in general, they all satisfy the following characteristics:

- easy accessibility for all to read and use
- developed by a process which is open and relatively easy for anyone to participate in
- no specific technology tie-in
- no control or tie-in by any specific group or vendor

Examples of open IT standards are:

- the TCP/IP suite of networking protocols from IETF
- the HTTP service protocol from W3C and ISO
- the Unicode coding standard from the Unicode Consortium and ISO
- the POSIX portable operating system interface from the Open Group, IEEE and ISO

Many organisations as well as governments are starting to emphasise that their IT usage follow or adhere to open standards as far as possible as they now realise that by implementing open standards they can have more flexibility in their choice of technology, vendor and solutions. In an increasingly complex IT environment, no single technology or vendor can offer good solutions in everything and so the ability to mix-and-match and to interoperate is of critical importance. Information is exchanged and stored electronically as never before and it is only by following open standards in the exchange and storage/retrieval of the data underlying the information that an organisation can be assured of access to the information both now and later when the technology or vendor may be long gone.

It should be noted here that wide usage of a standard does not necessarily mean that it is open. Numerous examples are found in the IT industry (e.g. the Portable Document Format (PDF) from Adobe Inc., the Powerpoint presentation file format from Microsoft), where some technology or file/data format associated with a popular product is very widely used, so much so that it becomes a de-facto standard, i.e. a standard established through widespread usage and acceptance in the industry. However, because this is very often based on a technology by a specific party (vendor or close group) and very often is under the control of this party it does not qualify as an open standard. There are potential pitfalls in adopting this as a standard to follow as there is no open mechanism for the users to participate in the development of the standard and there is no guarantee that later on the party in control will not try to lock-in the users into its product or technology. In some cases, the owner of the product/technology may agree to submit it to an internationally recognised standards setting body and in so doing it may then become an open standard.

Open Standards and FOSS

Many people are confused between the terms open standards and free and open source software, thinking that they are one and the same or one cannot exist without the other. (To be consistent with other publications by IOSN, the term "free and open source software" (FOSS) will be used in this document to refer to open-source software and/or free software, unless otherwise stated.) Open standards is not the same as free and open source software . In general, FOSS refers to software which follows certain principles in its creation, modification, usage and distribution [11]. In particular it should have the four fundamental freedoms:

- freedom to run the program, for any purpose
- freedom to study how the program works, and adapt it to your needs
- freedom to redistribute copies so you can help others
- freedom to improve the program, and release your improvements to the public

The excellent "FOSS General Introduction Primer" [12] from IOSN may be referred for more background and details on FOSS.

FOSS is software whereas open standards refer to standards and they are two different things altogether. The processes and issues involved in developing a software and a standard are also very different. It is entirely possible for a functionality in a non-FOSS software to be implemented following an open standard. For example, a proprietary software like the Microsoft Windows operating system can still implement the TCP/IP networking protocols following the open standards from IETF and be compliant with them.

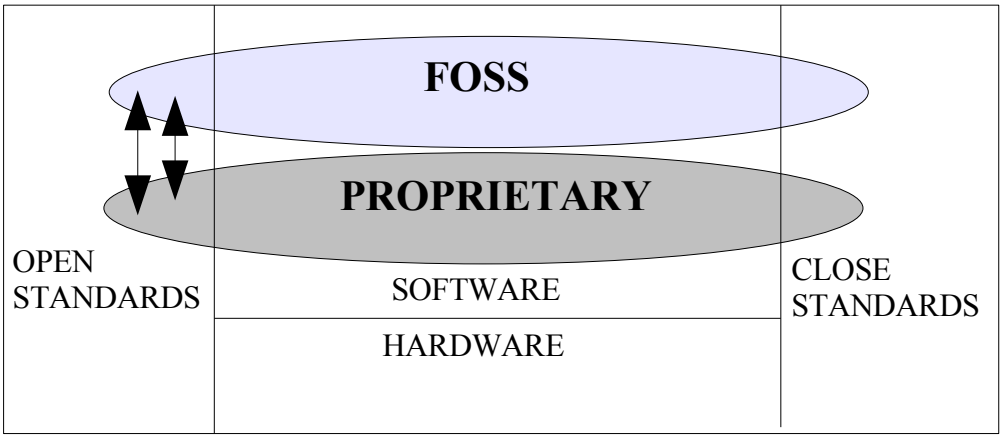

Figure: **Open standards and FOSS**
IT standards apply to hardware and software. Open as well as close standards can be implemented by both FOSS and non-FOSS. Adhering to open standards promotes interoperability between FOSS and proprietary software.

Widespread usage of open standards is very important to FOSS. It makes it easier for FOSS to be compatible with proprietary software. It is a practical reality that FOSS needs to co-exist with proprietary software and compatibility with the proprietary platforms is facilitated if the latter follows open standards. If all software were to follow open standards strictly they should be able to interoperate and communicate among themselves well and data files can be read and written transparently.

Some may argue that the freedom in FOSS for anyone to modify the software will allow and may even encourage the inclusion of code which does not conform to published standards. This is possible but in practice mainstream FOSS projects owners guard against this as they realise that it is to the advantage of FOSS if open standards are adhered to as much as possible. In fact FOSS promotes open standards as its ideals and development model encourages availability, openness and participation by all - the very traits and characteristics of an open standard.

FOSS is Useful for Popularising Open Standards
FOSS can play a useful role in popularising an open standard. A FOSS implementation of a standard usually results in an open and free working reference implementation. A lot of the benefits of open standards are negated, if the only implementation of it is a closed and proprietary one. The availability of a FOSS implementation will spur quicker adoption and acceptance of the standard as everyone has easy access to the implementation of the standard and so they 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

FOSS is Useful for Popularising Open Standards
open implementations of both the HTTP server (e.g. NCSA HTTPd, Apache) and client (e.g. NCSA Mosaic).

Adhering to open standards may be even more important than using FOSS in some cases. In situations or environments where the user is uncomfortable with FOSS or where FOSS solutions are not strong, the specification of software and technologies compliant with open standards will enable the user to have more choices and not be tied-in to any particular technology or vendor. The theoretical seamless mix and match promise of open standards will enable the user to have best-of-breed solutions irrespective of FOSS or not.

IMPORTANCE AND BENEFITS OF OPEN STANDARDS

The benefits of using open standards have been alluded to in the Introduction. Here we shall delve into more details on the importance and benefits of open standards.

Benefits of using Open Standards

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 and so 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 inter operate 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 that can read and write files using the OpenDocument format (an open standardised XML-based file format from OASIS) [13], an individual office will have the flexibility of using whatever office software best suited for it and at the same time be able to read, write and exchange documents with other offices in the organisation.

Sticking to 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 becomes obsolete, the user may have a tough 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 it will not be necessary to convert the data just so that the new application can read it.

It stands to reason that if a user demands that open standards are adhered to, there will be more choices available to it as more vendors can participate to offer it solutions and it may be able to even mix and match solutions from multiple vendors to enable it to have the best-of-breed solution 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 utilised with less or no re-training needed.

The benefits obtained with respect to sticking to data and file formats whose specifications are publicly published cannot be over-emphasised especially with respect to an organisation which 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, then the organisation may be held to ransom by a particular vendor in using its proprietary software or systems since it can be prohibitively expensive or even not possible to convert the data files.

National Considerations

From the national view point the usage of open standards by a government is even more important. In this information age, a government will need to use IT solutions to ensure that it has adequate and reliable information to enable it to govern the country effectively. It is vital that these IT implementations are compliant with standards that are open. In doing so, the government can ensure that its information systems will not be locked in later by any single technology or product. It is also in the interests of national security that open standards are followed to guard against the possibility of over-reliance on foreign technologies/products. Imagine the implications to a sovereign nation if the electronic records of its citizens are kept in databases which can be accessed readily only by proprietary software from a foreign vendor or the documents of the government are kept in a format which belongs to a vendor and thus has total control over its accessibility both now and in the future.

E-Government Projects Specify Open Standards
Many countries have started on e-Government projects or initiatives and most of these have come out with policies stating that IT standards and specifications which are open are to be followed as far as possible. Countries which have such policies include Norway, Denmark, the United Kingdom, the Netherlands, France, Brazil, New Zealand, Malaysia. The European Union's European Interoperability Framework, a framework to facilitate the interoperability of its member countries e-Government services, recommends the use of open standards for maximum interoperability. In addition, more and more public sector agencies all over the world have adopted or are considering adopting policies that require open standards.

Another important national benefit is that open standards will make it easier and in some cases the only way possible means for local companies to participate as a major player in supplying services and solutions to the government. This is because the local companies usually lack the strength and resources of multinationals and may be strong only in certain areas or solutions. The government can leverage on open standards to mix and match solutions from different suppliers in order to give the local suppliers a chance.

It is a reality in the IT world that the main language used and supported by all mainstream software is English and because of this it is sometimes difficult to produce electronic documents in another language. The availability of an open character

coding standard, Unicode [14], designed to support the worldwide interchange, processing, and display of the written texts of diverse languages of the world makes it feasible for the translation and localisation of software and electronic office documents for nations or cultures whose language is not English.

Particular Benefits of Open Standards

Open standards are particularly beneficial to some IT activities or services. Some of these are examined in greater detail here.

File Formats

Modern information systems generate data (lots of it in many cases!) which have to be stored in some form of electronic file formats for efficient storage, retrieval and exchange. If their specifications are not publicly known then only software and systems from the owner of these proprietary formats can readily access them. Also, the exchange of information is essential to the functioning of modern society. This exchange will be severely hampered if non-open file formats are utilised as products from one vendor may not be able to retrieve, access or store the information from the products of another vendor properly.

In some cases while the format may be known, it is nevertheless the property of a particular party and this party has control over the way the format evolves or is used. In such cases, the users can have very little say or control over the format and also there is the possibility that the owner may not publish the format specifications at a later stage for a new version. So while compatible systems can be created which can access the files now, there is no guarantee of this when a newer version comes out. In addition, there have been cases where when a proprietary format becomes popular and is widely used by the industry, the owner of the format at a later stage starts to impose restrictions like charging a fee or royalty charges (if it is patented) for using the format. The case of Microsoft attempting to charge flash drive makers and manufacturers of devices, such as digital cameras, a licensing fee for using its FAT file format [15] is a good example of this.

All this shows that it is of utmost importance that electronic file formats should follow some specifications which are accessible to all interested parties and also be developed by processes which are open and easy for any party to participate, in other words, they should be implemented using open standards. It is vital in today's information-centric society that the data from which information is derived can be be stored and exchanged following standards which are open so that no single party or even groups can hold the users to ransom.

Office Applications

One of the greatest successes of the PC revolution is to make it affordable and easy for office workers to use office applications for their work. Nowadays all modern offices are equipped with PCs which can perform at least some kind of basic word and spreadsheet processing as well as prepare presentation materials and documents. Currently office applications software deployment is dominated by one office suite - Microsoft Office (MS-Office). This is definitely a useful piece of office application software but some people do find that it is not well suited for their office and/or personal needs, either it is overly powerful or it may not possess certain functionalities that they need. Logically, these people then should use an alternative software more suited for them. While there exist several useful office applications software which can be considered as alternatives, in practice, most of these people opt not to use them because of the lack of full compatibility with the documents created by MS-Office. While some of these competing products claim compatibility with the MS-Office file formats, they are not 100% compatible since these formats belong to Microsoft and only if it licenses them to a competing product can it be truly compatible. This effectively results in the complete lock-in of a user into using MS-Office.

The example above illustrates that open and standardised file formats are needed to give the users the flexibility and freedom to choose and use products from different vendors and to prevent being locked in to a specific product and/or vendor. The recently published OpenDocument standard [13] from OASIS [16] for office applications offers this. Currently the office suite of applications from OpenOffice.org as well as StarOffice support this open format, while Microsoft does not. Instead, it has come up with its own XML-based file formats for its office suite. Again while the MS-Office XML schemas are publicly published and licensed for use royalty-free, they are owned by a single vendor (Microsoft) and hence are subject to the abuse discussed previously for non-open formats.

Internet Services and Applications

The Internet is perhaps the best showcase of how when technologies are implemented using mainly open standards there is almost universal accessibility, acceptance and benefits. Most of the networking infrastructure of the Internet is implemented based on open standards drawn up by the Internet Engineering Task Force (IETF) [17]. In addition, many of the services and applications running now as well as being planned for the future are being implemented following open standards and recommendations from several bodies notably, IETF, W3C [18] and OASIS. As a result, today, it is possible for one to access the major services offered on the Internet using a multitude of environments ranging from commodity PCs, hand held PDAs and mobile devices to proprietary set-top black boxes and TV sets. Without this adherence to open standards, the Internet will not be the ubiquitous entity that it is today.

STANDARD SETTING AND OPEN STANDARDS

This section will look into the standard setting processes and the more important standards bodies in the IT field and how they relate to the setting and adoption of open standards.

Standard Setting Organisations

In this document, the term Standard Setting Organisation (SSO) is taken to refer to an organisation that attempts to set standards or make recommendations which when widely deployed become de-facto standards. The are many SSOs, national, regional as well as industry-based. A formal SSO refers to one that is recognised directly or indirectly by a government entity [19]. Very often there will exist 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 is the national standards body, in USA, the American National Standards Institute (ANSI) is the official body while in the United Kingdom it is the British Standards Institution (BSI).

While any organisation can come up with its specifications and call it its standard, for this to be accepted as an international 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 ISO [20], IEC [21] and ITU [22].

The International Organisation for Standardisation (ISO) is an international standard setting body made up mainly of representation from national standards bodies. The International Electrotechnical Commission (IEC) is a standards organisation which 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 of the 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 Standardisation (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 153 members representing nations 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 ISO are agreed upon

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

The International Telecommunication Union (ITU)
The ITU has its headquarters in Geneva, Switzerland and it is an international organisation within the United Nations 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. The ITU is divided into three sectors: Radiocommunication (ITU-R), Telecommunication Standardization (ITU-T), and Telecommunication Development (ITU-D). The ITU-T is increasingly becoming an important international body for the development of IT standards due to the convergence of IT and Telecommunications.

Standard Setting Processes

The setting or creation of new technical standards can basically follow several main processes depending on the way that they are created: de-jure, de-facto and industry-created standards.

De-jure Standards

De-jure standards are normally created by formal SSOs following formal procedures which have been established by these bodies. Based on some 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 for all who are interested in the standard to participate and the standards creation process be seen as neutral and transparent and not be controlled by any particular group or party.

There are several disadvantages to the process involved in the creation of de-jure standards. First of all because of the structure and makeup of the formal SSOs the entire standard drafting process can be quite long. For example in the case of ISO standards, there is commonly a time span of 2-3 years from the new work item proposal to the publication of a standard.

While the standard setting process formally tries to be neutral and not be partial 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 and so not all interested parties can participate directly. However they should be able to participate at the local level via their national standards body which will then carry the so-called national view points, which may or may not concur with those of the interested parties.

The publication of a de-jure 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 out over a more complex and difficult to implement standard simply because implementation is simpler and faster and so there is better acceptance in the industry. A classic example of this is the highly complex but more complete X.400 suite of messaging protocols which are not widely used today as compared with the simpler but more easily implemented SMTP mail protocol which forms the backbone for Internet email. 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 recognised SSOs which are active in putting out de-jure standards are ISO, IEEE, ITU-T and ANSI. Examples of widely used de-jure standards include:

- IEEE 802 – a set of standards for Local Area Networking
- ISO 10918 – a standard for the JPEG graphics compression and file format
- 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 a specification to a standards body, like ISO, to adopt or adapt its standard 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 the majority of users throughout the industry. As a result of this a de-facto standard is established so that everybody seems to follow it like an authorised standard from a standards body. Examples of these are:

- the FAT file system from Microsoft
- the Adobe portable document format (PDF)
- the Hayes command set for dial-up modem control
- the Hewlett-Packard Printer Control Language (PCL)

The main advantage of a de-facto standard is that widespread acceptance in its implementation and usage is assured since this is already so unlike a de-jure 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 which is either owned or controlled by the vendor or group which produces the product. So 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, the possibility of a lock-in later is there.

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) which 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-jure standards set by formal standards bodies and the 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 and 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 - the Organisation for the Advancement of Structured Information Standards. OASIS is a not-for-profit, international consortium that 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 UDDI and OpenDocument Format for Office Applications. The World Wide Web Consortium or W3C is another consortium which has a lot of influence in the Web industry. It develops inter operable 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 which 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 eventually are submitted to be adopted by the 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 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 of the more active organisations which are generally perceived to be open include IETF, IEEE, OASIS, W3C and the Free

Standards Group. 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 organisations big and small, the organisations listed above do stand out.

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

As noted earlier these non-formal SSOs often have liaisons, especially at the technical working group level, with the formal ones like ISO and ITU-T and so 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 (IETF)

The Internet networking standards and protocols, like TCP/IP, became de-facto standards when the Internet was widely embraced throughout the world. The IETF is charged with developing and promoting Internet standards. It is an open, voluntary organisation with membership open to any interested individual. The actual technical work of the IETF is done by its working groups, which are grouped, 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 the IETF [23]. The IETF is overseen by the Internet Architecture Board (IAB), which is in turn responsible to the Internet Society (ISOC) [24].

The drafting and setting of specifications and standards by the IETF is carried out considerably faster when compared to the formal SSOs. The IETF working groups are the ones which do the drafting work. A new set of specifications starts off as an Internet Draft which is placed in the 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 the comments and feedback the draft undergoes several rounds of modification and after that it moves on to become an RFC (Requests for Comments) document and is published. The specifications in an RFC document may be implemented by the Internet community and it can become a de-facto standard if there is wide acceptance of it. An RFC specification for which significant implementation and successful operational experience have been obtained may be elevated to the Internet Standard level [25] and is assigned a number in the STD series while retaining its RFC number [26].

The World Wide Web Consortium (W3C)

The World Wide Web Consortium (W3C) [18] is an international consortium that 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 of the W3C Recommendations have been submitted to a formal standards body like ISO to become international standards.

The W3C believes in full interoperability for the Web to function and realise its full potential. Towards this end it publishes open standards for Web languages and protocols, making 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.

The 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. The W3C counts among its members, vendors of technology products and services, content providers, corporate users, research laboratories, standards bodies, and governments. 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 Organization for the Advancement of Structured Information Standards (OASIS)

The Organization for the Advancement of Structured Information Standards (OASIS) is a non-profit, international consortium that drives the development, convergence, and adoption of e-business standards [16]. Standards produced by 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 among its members users and vendors, governments and universities, trade groups and service providers.

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 (FSG)

The Free Standards Group [27] an independent, non-profit organisation dedicated to accelerating the use of free and open-source software by developing and promoting standards. It is supported by both commercial corporations in the IT industry as well as the FOSS development community. All standards produced by the Free Standards Group are available for free and are distributed under open source licenses. Anyone can participate 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. The LSB standardisation is tied in with the ISO/IEC JTC1 SC22 working group on Linux standardisation.

The Institute of Electrical and Electronics Engineers (IEEE)

The IEEE is a non-profit, technical professional association of more than 360,000 individual members in approximately 175 countries. The IEEE Standards Association (IEEE-SA) [28] 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 that 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 (Wi-Fi).

SOME IMPORTANT OPEN STANDARDS

This section will discuss some of the more important open standards that are either currently already available or actively being developed. The standards listed here are by no means exhaustive but they do represent some of the most widely used ones in the industry today.

Internet Networking and Applications/Services

The Internet is what it is today mainly because of the almost universal accessibility of the applications and services offered over it and its seamless connectivity. This is a direct result of the widespread use of open standards in the implementation of the Internet, both historically and currently. The standards mainly responsible for the Internet infrastructure and for the popular World Wide Web and Internet email services are highlighted here.

Transmission Control Protocol/Internet Protocol (TCP/IP)

The TCP/IP suite of networking standards provide the foundation for the network infrastructure of the Internet. All of the major services and applications on the Internet ride on top of TCP/IP. These protocols were originally developed by the pioneers of the Internet, the engineers and scientists from the universities, research institutions and companies who collaborated on the US Department of Defence's APRANET project. This evolved to become the Internet as we know it today and TCP/IP became a de-facto standard. It is now an IETF Standard and IETF is charged with its continued development.

TCP/IP is a two-layer packet-switching specification in which data to be communicated between two end-points on a network is first broken up into smaller data packets which are then individually routed through the network from the source to the destination points. The higher layer, Transmission Control Protocol (TCP) [29], manages the disassembling of the data into smaller packets at the source and the reassembling at the destination point upon receipt of the data packets. The lower layer, Internet Protocol (IP) [30], handles the addressing and routing of each packet so that it gets to the correct destination.

TCP/IP just provides the transport mechanism for sending data across the Internet or an IP network. In order for this to be useful, a service or application has to be specified and implemented. Again the IETF is mainly responsible in overseeing and setting the specifications for most of these services. The widespread implementation and acceptance of these specifications coupled with open standards bodies like the IETF and the W3C make the Internet the best showcase for open standards at work. Some of these standards are listed below.

Hypertext Transfer Protocol (HTTP)

HTTP is perhaps the most widely used Internet service protocol. It is the primary method used to access the WWW. Web content, in the form of HTML pages and possibly also other multimedia formats, is transfered from a web server to a user's web accessing agent using the HTTP protocol. HTTP was developed by W3C in co-operation with the IETF working groups. The standard most widely deployed and supported on the Web today is HTTP version 1.1 or HTTP/1.1 [31].

The HTTP protocol is a request-response protocol using a client-server model in which an HTTP client, e.g. a web browser, initiates a request by establishing a TCP connection to the server computer which will respond to the request commands sent by the client. The commands to support as well as the behaviour of both client and server is spelt out in the HTTP specification.

It is through this universal acceptance of the HTTP protocol standard that the web has become the ubiquitous information dissemination and exchange medium that it is now. One major factor in its wide acceptance by all the stake holders and players on the Internet is its open standard status.

Hypertext Markup Language (HTML)

While HTTP defines how the contents of a web page can be transmitted between a web server and a client, HTML is an open standard specifying the structure and presentation of the content. A document composed with HTML consists of the contents intermingled with symbols and tags which tell the software needed to interpret and display the HTML document the structure and presentation of the content. The HTML specification is now being maintained by the W3C. It has undergone several revisions and the most current specification is HTML 4.01 [32]. HTML is also available as an ISO standard [33], which is a subset of HTML 4.

In its simplest form an HTML document consists of the text of the document as well as tags which specify the markup needed to be performed on it. For example, in the sample below:

```
<h3>My Work Experience</h3>
<img src="mypic.png">
<p>
<b>Work Experience</b><br><br>
 1990 - 1995 System engineer<br>
 1995 - 2005 Network manager<br>
</p>
```

the tags <h3> and </h3> specify that the text enclosed within them is to be rendered as third level headings while the tag <img src="mypic.png"> specifies the display of a graphics file. The tags , specify that the text enclosed within them should be rendered as bold and the tag
 signifies a line break.

An HTML user agent software is needed to render a document made up of HTML and the most common agent is a web browser. If the W3C HTML specifications are adhered to, an HTML document can be displayed properly by any user agent (which conforms to the specifications) and this can form the basis for a standard format for textual document information exchange. One major limitation of using HTML to display a document though, is that page breaks are not easily represented or controlled.

The use of HTML in email has gained popularity as it enables one to impose some simple formating on the text as well as embed graphics and multimedia content into the message. However, it is generally considered not good practice by security-conscious users to utilise HTML in mail messages as some of the popular HTML-enabled email software have been known to possess vulnerabilities. This makes them open to potential exploitation by a rogue HTML email message which may result in the compromise of a user's system.

Email Protocols

Internet email has become almost as important as the telephone service. Every time we send off an email we assume that the mail will be relayed correctly by the mail server to its destination. When we send attachments or incorporate some non-textual content into our email we just assume that the attachment will be incorporated correctly and when the recipient gets it she will be able to get it back into its original form. All this works seamlessly irrespective of the hardware and software deployed because Internet email makes use of several important open standards in its mail transmission as well as in the encoding of email messages.

Simple Mail Transfer Protocol (SMTP)

The Simple Mail Transfer Protocol (SMTP) [34] enables the transport and routing of email from the sender to the recipient using their email addresses. This standard is client-server based whereby the SMTP client (usually the user's email software or mail user agent) will initiate a TCP connection to the SMTP server (the mail relay host). Communications between the server and client is done using the SMTP protocol. This is a simple text-based protocol where essentially the client informs the server the email addresses of the sender and the recipient(s). After that if all goes well and the server allows it (based on its mail relay policy), the client will transmit the mail message to the server. The server will then attempt to deliver it to the computer housing the

recipient's mailbox or if necessary forward the email to another server for delivery to the recipient's mailbox.

The SMTP protocol started out supporting only 7-bit ASCII text in the messages, effectively limiting it to the transmission of English-based text. Non-English language texts which make use of more than the 7-bit ASCII character set as well as binary file attachments have to be encoded by the email user agent software before transmission. The message format of this text-based mail is specified by another IETF standard, RFC 2822 [35]. The SMTP standard has been extended to support 8-bit text [36], permitting the transmission without encoding of text messages in more languages.

Multipurpose Internet Mail Extensions (MIME)

As Internet email became more and more popular, users find it a convenient, economical and efficient way to send information to one another. Soon they tried to send other types of content e.g. audio, video, images, software programs, besides text messages via email. However, since the original Internet email specifications were meant primarily for English-based text messages, some new set of specifications had to be drawn up to allow interoperability and seamless transmission of multipurpose content. This resulted in the IETF producing the Multipurpose Internet Mail Extensions (MIME) standard [37].

MIME is an extension of the basic text-based Internet mail standard. It defines mechanisms for sending other kinds of information in email. These include non-English text using character sets beyond ASCII and binary file content such as multimedia files and computer software. To support these as well as to retain backward compatibility with the simple ASCII-based mail format, a set of email headers for specifying additional attributes of a message, e.g. content type, and a set of transfer encodings which can be used to represent 8-bit data using characters from the 7-bit ASCII set are defined. The encoding of non-ASCII characters in mail message headers is also catered for in MIME allowing the usage of non-English characters in them. The MIME standard specifies a means to register new content types and transfer encodings making it flexible for supporting new multimedia types in the future.

MIME is also an important standard for the Web as the HTTP protocol makes use of mail-like MIME formating rules and syntax for its data formating.

The Extensible Markup Language (XML)

The Extensible Markup Language (XML) is a Recommendation [38] from W3C that specifies a meta markup language (a meta language is a language for describing other languages) for the creation of other markup languages for use on the WWW. HTML is a single predefined markup language and hence possesses severe limitations to

describe and represent all sorts of data for dissemination, exchange and interaction. XML, being a markup specification language, is capable of being used to design markups for describing many different kinds of data for storage, transmission, or processing by a program [39]. It describes the data but it does not tell you what you should do with the data. One should note that XML and HTML were designed with different goals in that XML was designed to store, carry, and exchange data whereas HTML was designed to display data and to focus on how data looks [40]. XML was created for deployment on the Web by using a subset of an existing, widely used international standard for text document markup - the Standard Generalized Markup Language (SGML) [41].

Due to its design goals, XML is well-suited for data transfer and exchange and as a format for document storage and processing. This and the fact that it is under the charge of an open specifications/standards body, W3C, has resulted in XML being used as the base for specifying many other data formats and exchange protocols. According to the community-based XML portal, XML.ORG [42], it is now viewed as the standard way for information exchange in environments that do not share common platforms. Special purpose languages and standards developed using XML for specific environments or activities are announced almost daily and several hundred have been adopted since XML 1.0 was released in February 1998. In particular the e-government and e-commerce segments are very active in developing and implementing XML-based specifications.

A simple XML document is shown below:

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<?xml-stylesheet type="text/xsl" href="bookcollection.xsl"?>

<BookCollection>
     <Book>
          <Title>Chronicles: Volume One</Title>
          <Author>Bob Dylan</Author>
          <Publisher>Simon and Schuster</Publisher>
          <Year>2004</Year>
     </Book>
     <Book>
          <Title>Harry Potter and the Goblet of Fire</Title>
          <Author>J.K. Rowling</Author>
          <Publisher>Bloomsbury Publishing</Publisher>
          <Year>2000</Year>
     </Book>
</BookCollection>
```

Note that while XML uses syntax tags to identify various types of data in a document file these tags are not predefined. So the document creator has to define and describe them using what is called an XML schema and associate the document with the schema. To create the schema an XML schema language is used, e.g. Document Type Definition (DTD), XML Schema and RELAX NG. The purpose of the schema is to define the legal building blocks of the XML document, i.e. the elements, data attributes, tags etc. which can appear in the document. DTD has limitations with respect to its extensibility and lack of support of several useful features e.g. data types and namespaces. XML Schema which is also another W3C Recommendation is more suitable for use in many practical web applications.

While the schema may define the legal components of the XML document it does not carry information about how to display the data. So in order for the data in an XML document to be displayed properly by say, a web browser, a display style has to be specified. The Extensible Stylesheet Language (XSL) is used to perform this. Styling is both about transforming and formatting information and the W3C specifications separate these processes. In addition, the components in an XML document have to be navigated to extract and process them. Hence the XSL Recommendation from W3C consists of three parts:

- XSL Transformations (XSLT): a language for transforming XML documents
- XSL Formatting Objects (XSL-FO): a language for formatting XML documents
- XML Path Language (XPath): a language for navigating in XML documents

An example of an XSLT transformation of the XML example document above to a web browser displayable HTML output is:

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<xsl:stylesheet version="1.0"
xmlns:xsl="http://www.w3.org/1999/XSL/Transform">
<xsl:template match="/">

  <html>
  <body>
    <h3>Book Collection</h3>
    <table>
    <tr bgcolor="#ff0000">
      <th align="center">Title</th>
      <th align="center">Author</th>
      <th align="center">Publisher</th>
      <th align="center">Year</th>
    </tr>
    <xsl:for-each select="BookCollection/Book">
    <tr>
```

```
      <td><xsl:value-of select="Title"/></td>
      <td><xsl:value-of select="Author"/></td>
      <td><xsl:value-of select="Publisher"/></td>
      <td><xsl:value-of select="Year"/></td>
    </tr>
   </xsl:for-each>
    </table>
  </body>
  </html>

</xsl:template>
</xsl:stylesheet>
```

Computer Graphics and Multimedia

In the old days of computers, the display was predominantly text-based and any graphics displayed was at best line-graphics implemented using special line drawing character sets. Computer terminals which can display full-fledged graphics were expensive and used only for special purposes or applications. Today with the proliferation of inexpensive personal computers which have the power to process and display graphics and multimedia, even the user interface is now graphics-based. One of the main attractions of the Web is its widespread support and usage of graphic images and multimedia to make the contents interesting and lively. It is important that open standards are followed as much as possible in graphics and multimedia data storage, processing and retrieval to enable diverse devices and computing platforms to offer the same degree of Web experience.

Portable Network Graphics (PNG)

In the early days of the Web when Internet links and connections were relatively slow, many of the simple images and animations displayed in web pages made use of a graphics format called GIF (Graphics Interchange Format) [43] as this format resulted in small graphic file sizes. The GIF format included the use of the LZW compression algorithm which was patented in the USA by Unisys and the latter eventually decided to ask for royalty payments for all software which utilises GIF. This led to the creation of the patent-free Portable Network Graphics (PNG) format [44] to replace GIF for use as a single-image web format. The PNG format later became a W3C Recommendation as well as an ISO international standard (ISO/IEC 15948).

PNG is an extensible file format for the lossless, portable, well-compressed storage of raster images. Indexed-color, greyscale, and truecolor images are supported, 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 [44].

The X Window System

The graphics user interface (GUI) that is now common on desktop computers uses a graphical window metaphor as the basic user interface. This window system GUI enables different programs to run simultaneously in their own individual windows and these windows can be opened, closed and resized. The windowing systems found on platforms like Microsoft Windows and Mac OS X are proprietary ones. On the other hand, Unix and Unix-like operating systems (e.g. Linux, FreeBSD) make use of an open window system - the X Window System.

The X Window System, or X, is an open windowing system standard led by the X.Org Foundation [45]. X provides a framework for the display and management of graphical information and on top of this a GUI may be built. X uses a client-server model. The X client is usually the application which sends graphical output for display on the X server. The X server interacts with the user using primarily the keyboard and mouse as input devices and the input is transmitted to the client to act upon. The X client and server may be running on the same machine or they may be on different physical devices connected together over a network. The intrinsic client-server property of X constitutes the main difference between it and other well known window systems like Microsoft Windows, which simply displays graphical applications local to the device on which the application is running on.

Being an open standard, besides Unix and Unix-like systems, X has been implemented on a variety of hardware and operating systems, including the various generations of Macintoshes, PCs running MS-DOS and Microsoft Windows as well as OpenVMS from Hewlett Packard (formerly Digital Equipment) etc.

Ogg Vorbis

Ogg Vorbis is a general-purpose compressed audio format for storing and playing digital music. It is comparable in quality to other formats such as the popular MP3; however unlike MP3, it is an open format and it claims to be free from patents. The format originated from the Xiph.Org Foundation [46], a non-profit organisation dedicated to producing free and open protocols, formats and software for multimedia.

Vorbis is the name of the audio compression scheme and this is contained in Ogg, the name of Xiph.org's container format for audio, video, and metadata - hence the name Ogg Vorbis. Vorbis is a lossy codec, i.e. it uses a compression algorithm which discards data in order to increase the compression possible. The Ogg Vorbis specification is in the public domain and is completely free for commercial or noncommercial use [47].

Software and hardware devices which support Ogg Vorbis are steadily increasing in number and may be found on the Vorbis wiki at Xiph.Org [48].

Office Documents

Office applications is one of the most widely used applications for personal computers in a modern day office. These applications include wordprocessing, spreadsheets and presentation software. There is available on the market several office applications e.g. Microsoft Office, WordPerfect Office, OpenOffice.org and Applixware Office, and each of these invariably in the past use different formats for storing their files. As a result, it was difficult to convert from one file format to another and for one application to read/write a file created by another application. It is thus a real step forward in terms

of office interoperability and productivity when, OASIS announced that it is recommending the Open Document Format for Office Applications (OpenDocument) as a standard for file format for use in office applications.

OpenDocument

OpenDocument [13] is a file format developed by OASIS for storing office documents created by applications such as spreadsheets, word processors, charts and presentations. It makes use of a royalty-free, open and vendor-independent XML-based format. The format is based on the file format of OpenOffice.org which was submitted to OASIS to form the basis for the standard. OpenDocument provides a single XML schema for text processing, spreadsheet, presentation, drawing, charting, and math documents.

Office software which have announced that they will support the OpenDocument format as their primary/native format include the office suites of OpenOffice.org, StarOffice and KOffice.

Open Standards Usage

The table below summarises the usage and penetration levels of the open standards described above in their respective domains. As can be seen, open standards are widely deployed on the Internet and in running Internet-related/derived services and applications while for the graphics, multimedia and office applications areas they are still very limited in acceptance. The limited penetration in these domains is a result of the fact that they are dominated by proprietary products like those from Apple and Microsoft which make use of their own proprietary formats and specifications (see the next section on “Comparison of File Formats”). The incentive for these vendors to support open standards or at least make their specifications more open is not strong due to their dominant positions.

Important Open Standards

Domain	Standard	Org. Resp.	Usage
Networking	TCP/IP	IETF	Universal
WWW	HTTP	W3C, IETF	Universal
Web content	HTML	W3C	Universal
Email	SMTP	IETF	Universal
Email, WWW	MIME	IETF	Universal
Doc exchange	XML	W3C	Universal
Graphics	PNG	W3C	Wide
Window system	X Window	X.Org	Limited
Audio	Ogg Vorbis	Xiph.Org	Limited

Important Open Standards			
Office documents	OpenDocument	OASIS	Limited

COMPARISON OF FILE FORMATS

This section will list, compare and discuss the degree of openness and/or lack of openness of several popular file formats. These include file formats for the following application areas:

- office applications
- graphics
- audio
- video

Office Applications File Formats

Microsoft Office Formats

The most popular office application currently is Microsoft Office (MS-Office). 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 file formats used were binary (i.e. non-plain text) in nature and not publicly published. MS-Word, MS-Excel and MS-PowerPoint use the binary DOC, XLS and PPT formats respectively and these formats are proprietary ones, being owned and controlled entirely by Microsoft. So while the file formats for these applications are widely used due to the dominant role MS-Office plays in the office applications software field, other software not from Microsoft are not able to read and/or write files created with MS-Office properly. Some of these non-Microsoft office suites, e.g. OpenOffice.org or StarOffice do have very good compatibility in reading and/or writing using these proprietary formats but the compatibility is not complete. Competing products cannot be totally compatible with MS-Office unless they are provided with the file format specifications by Microsoft.

Some of the 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 that has been developed by Microsoft for cross-platform document interchange. Technical documentation on RTF is published by Microsoft and so many non-Microsoft software do support the RTF file format well and it is widely used for document exchange between MS-Office and other office applications. However, the RTF format does not support completely the more complicated and sophisticated features found in MS-Office, and so complex documents may not be properly represented using the RTF format.

With MS-Office 11 (MS-Office 2003) an option was made available to use a new file format for Word and Excel that is XML-based [49]. However, these XML-based formats have been criticised for being incomplete and immature. They were not available for all the software 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, will deliver support for a new set of XML file formats called the "Microsoft Office Open XML Formats" [50]. The applications that will use these formats by default are Word, Excel and PowerPoint.

Office XML Open Format is also being published by Microsoft on a royalty-free basis to the industry. While this potentially will make it possible and easier for third party products to be compatible with MS-Office, the file format will still be owned and controlled by Microsoft and hence is not open.

OpenOffice.org and StarOffice Formats

OpenOffice.org (OOo) [51] is a full-fledged Open Source office application suite, comprising word processor, spreadsheet, presentation software, graphics editor and a database program (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 multi-platforms, e.g. 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 Open Document Format for Office Applications (OpenDocument) v1.0 in May 2005. New versions of OOo as well as other office suites like KOffice and StarOffice will support OpenDocument as their native file formats. This will improve significantly 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. Unlike the Microsoft Office Open XML Format, the OpenDocument format is not owned or controlled by a single proprietary vendor, instead it falls under the ambit of OASIS, an open standards body, and so users can be assured that they will have access to their documents and data from a variety of software.

StarOffice shares the same codebase 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)

The Portable Document Format (PDF) is a file format developed by Adobe Systems, Incorporated [52] 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 specifications for the PDF format is publicly published by Adobe [53] and it can be implemented without restrictions by anyone. As a result, there is available a variety of software on many different platforms which can read the PDF format and a (smaller) number of applications which can write out the contents of a document in PDF [54]. 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 falls short of an open standard.

Office Document Formats

Format Std.	**Org. Resp.**	**Published**	**Non-Prop.**	**Int.**
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
ODT (text)	OASIS	Yes	Yes	Yes
ODS (spreadsheet)	OASIS	Yes	Yes	Yes
ODP (presentation)	OASIS	Yes	Yes	Yes
PDF (text and presentation)	Adobe	Yes	No	No

Graphics/Image File Formats

A picture is worth a thousand words, as the saying goes. So it is not surprising that with the advent of powerful desktop systems which are able to display high resolution graphics, images are being utilised more and more to convey information. Modern computer systems use what is known as raster graphics to display an image on the video screen. A raster graphics image, digital image, or bitmap, is a data file or structure representing a generally rectangular grid of pixels, or points of color, on a computer display monitor [55]. Each point or pixel on the screen is represented by a value denoting its colour and this bitmap is stored in memory. Using this bitmap the entire screen is repainted 30 or more times per second by the video device resulting in the human eye seeing the image being displayed on the screen. There are many ways to create and store this raster graphics image file and so if we are to be able to exchange and share useful graphical information there is a need to have a format that is supported on multiple platforms and by various graphics software.

Many of the graphics file formats in use today are proprietary by nature, being derived and tied to the software used to create them [56]. There are some formats which have gained wide acceptance as de-facto standards and a few of these have emerged as open graphic file formats.

GIF

GIF (Graphics Interchange Format) is a bitmap image format [43] that is widely used on the World Wide Web, 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 which 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 utilises 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 patent-free Portable Network Graphics (PNG) format [44].

PNG

The PNG (Portable Network Graphics) 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. PNG is an extensible file format for the lossless, portable, well-compressed storage of raster images. It offers indexed-color, grayscale, and truecolor 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 [44].

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).

XPM

The XPM (XPixMap) format [57] is a de-facto standard for creating icon pixmaps for use in GUIs based on the X Window System [45]. It consists of an ASCII image format and a C library. The XPM format defines how to store color 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 TIFF (Tagged Image File Format) 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 [58] is publicly published by Adobe and it can be implemented without restrictions by anyone. As a result, there is available software on many different platforms which can read and write the TIFF format. It has become a de-facto standard graphics format for high color 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 prepress electronic data exchange.

JPEG JFIF

JPEG is a standardised image compression mechanism from the Joint Photographic Experts Group (JPEG) [59]. 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 [60].

The JPEG compression uses a lossy mechanism for compressing color 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 [61]. 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 [60].

SVG

Unlike the other file formats listed above which 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 [62]. SVG consists of an XML-based file format and a programming API for graphical applications. It is a W3C Recommendation [63] 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.

Graphics Formats				
Format Std.	**Org. Resp.**	**Published**	**Non-Prop.**	**Int.**
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

Audio File Formats

There are two major groups of audio file formats [64]:

- those using lossless compression e.g. like WAV, FLAC
- 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 so the original data is restored upon uncompressing it. In lossy compression some data is lost during compression and so upon uncompressing it the data is not identical to the original one 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

WAV (WAVEform audio format) 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 not popular anymore being superseded by other more efficient means of audio storage [65].

FLAC

FLAC (Free Lossless Audio Codec) is a popular lossless audio format with the compression designed specifically for audio data streams, achieving compression rates of 30–50%. The format specification is publicly available and forms part of the FLAC Open Source project [66]. It is supported by a growing list of audio software and devices.

MP3

MP3 (MPEG-1 audio layer 3) is a popular lossy compression audio format [67]. 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 which allow easy creation, sharing, collecting and playing of MP3 files.

While MP3 is an international standard, it suffers from the fact that it makes use of patented technology and so software and devices which support it are subject to royalty payments in those countries which recognise software patents. This has led to the creation of alternatives to MP3 e.g. Ogg Vorbis and WMA.

WMA

WMA (Windows Media Audio) is a lossy compression audio file format developed by Microsoft. It is a proprietary format but is widely used and supported due to the popularity of the MS-Windows platform.

AAC

AAC (Advanced Audio Coding) from MPEG is a lossy data compression scheme intended for audio streams. It was designed to provide better quality at the same bitrate 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 MPEG-2 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 which make use of this standard.

The newer MPEG-4 standard also specifies an audio compression technology which 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 [68].

RealAudio

RealAudio is a proprietary audio format developed by RealNetworks. It is especially designed to conform to low bandwidths, and it can be used as a streaming audio format, i.e. being played in real time as it's downloaded. Many radio stations use RealAudio to stream their programmes over the Internet [69].

Ogg Vorbis

Ogg Vorbis is a compressed audio format which is believed to be free of patents and royalty payments. The format originated from the Xiph.Org Foundation [46], a non-profit organisation 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 metadata - hence the name Ogg Vorbis.

The Ogg Vorbis specification is in the public domain and is completely free for commercial or noncommercial use [47].

There is growing support for the Ogg Vorbis format from software and hardware devices [48] as well as online audio services.

Audio Formats

Format	Org. Resp.	Published	Non-Prop.	Int. Std.
WAV	Microsoft	Yes	No	No
FLAC	Xiph.Org	Yes	Yes	No
MP3	MPEG/ISO	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

Video 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.

For video data, the storage of it involves more than just finding an efficient means to store the raw data. Apart from the raw data stream itself, 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 frequently a video 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 utilised for the encoding and decoding of 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 of the video containers popular today are proprietary [70]. 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 they remain proprietary formats all the same.

Video Containers

AVI

AVI (Audio Video Interleave) is a video container format by Microsoft containing both audio and video data. It is a RIFF [71] file specification used with applications that capture, edit, and play back audio-video sequences [72]. It enjoys widespread support and it is the most common container format for audio/video data on the PC.

ASF

ASF (Advanced Systems Format) 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 [73].

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 [74].

The QuickTime format was used as the basis of the MPEG-4 MP4 container standard (see entry on MP4 below).

MP4

MP4 (MPEG-4 Part 14) 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 [75].

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 [76], there are several open codecs associated with the Ogg project: Vorbis and FLAC for audio and Theora for video [77].

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

Video Containers				
Container	**Org. Resp.**	**Published**	**Non-Prop.**	**Int. Std.**
AVI	Microsoft	Yes	No	No
ASF	Microsoft	No	No	No
MOV	Apple Computer	Yes	No	No
MP4	MPEG/ISO	Yes	Yes	Yes
Ogg	Xiph.Org	Yes	Yes	No

Video Codecs

MPEG Codecs

The Motion Pictures Experts Group (MPEG) has developed several standards pertaining to video technology and they are used by many of the digital video products on the market. The MPEG video codecs are specified in the following ISO standards:

- MPEG-1 Part 2 (ISO/IEC 11172-2)
- MPEG-2 Part 2 (ISO/IEC 13818-2)
- MPEG-4 Part 2 (ISO/IEC 14496-2)
- 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 which make use of them are expected to pay patent licensing royalties.

MPEG-1 Part 2

The MPEG-1 standard which 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 disks. 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 which allow it to compress video much more efficiently than earlier MPEG codecs and 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, Linux and Macintosh as well as video servers and authoring tools are available from a number of vendors [78].

Sorensen

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

Windows Media Video (WMV)

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" [79].

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 BSD-style FOSS license which means that it is available freely for commercial or non-commercial use.

Video Codecs

Container	Org. Resp.	Published	Non-Prop.	Int. Std.
MPEG-1	MPEG/ISO	Yes	Yes	Yes
MPEG-2	MPEG/ISO	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

To establish whether a video format is open or close, we have to look at the status of the specifications for both the container and the codec. The table above lists down the popular containers as well as the codecs which are most commonly associated with them.

STANDARDS AND INTERNATIONALISATION/LOCALISATION OF SOFTWARE

Internationalisation and Localisation of Software

The internationalisation of a product e.g. software, is not the same as its localisation although they may address many similar issues. Internationalisation refers to the process whereby a product is made or adapted so that it can be used internationally (i.e. in different countries or regions all over the world with different cultures and conventions) without redesign. On the other hand, localisation addresses how a product may be tailored for a specific country, region or culture by making it linguistically and culturally appropriate. Internationalisation is often referred to using the abbreviation "I18N" or "I18n", where the number 18 refers to the number of letters omitted and similarly for localisation the abbreviation "I10n" is used.

It is important that application software which is meant for deployment in many different countries with different cultures and languages be designed with internationalisation in mind to be able to accommodate possibly different ways of expressing an item of information or peculiarities of a different language. Some of the issues which internationalisation needs to grapple with include [80]:

- Date and time formats
- Currency format
- Language peculiarities (e.g. alphabets, numerals and left-to-right script vs. right-to-left)
- Language character coding sets for textual display
- Names and titles
- Sorting of names and text
- Identification numbers e.g. social security and passport numbers
- Telephone numbers, addresses and international postal codes
- Weights and measures

While the cultural and linguistic demands may change from country to country, the core program dealing with the functionalities of a software do not change and so it is common practice to separate text and other environment-dependent data from the program code itself. This makes it easier to support internationalisation as only changes need to be made to the environment-dependent resources and minimal code changes need to be made.

The better internationalised an application is, the easier it is to localise it. This is because a well-internationalised application will have built-in support to cater for the items which are needed for localisation. These may include [80]:

- Language translation

- Hardware support for certain languages e.g. input devices and methods
- Local customs
- Local content
- Aesthetics
- Cultural values and social context.

The major work of localisation is in translating the user interface and documentation but it involves more than just translating the language used. It also needs to cater to other relevant changes such as the usage of appropriate cultural and social values, symbols peculiar to the language, display of numbers, dates, currency, appropriate input methods etc.

In software internationalisation and localisation, a set of parameters, termed a locale, is used to define the user's language, country and any special variant preferences that the user wants to see in the user interface [81]. A locale identifier usually contains at least a language and a region/country identifier. Depending on the operating platform/system used, locale identifiers can be defined in several ways. Most systems utilise the 2 and 3-letter language codes defined by ISO 639-1 and 639-2 respectively for the language identifier and the 2-letter country codes from ISO 3166-1 for the country identifier. However, MS-Windows uses a numeric Locale Identifier (LCID) which specifies the language and sort identifier [82].

Standards Important to I18n and L10n

In this section we shall look at some of the important standards which are used in i18n and l10n.

Unicode and ISO/IEC 10646

Proper rendering and display as well as practical input methods for multilingual text on a computer system are essential if efforts to make software available in multilanguages is to be successful. Standards are needed for the character code tables and character encoding methods. The character code tables assign integer numbers to characters while character encoding is a method by which characters or their respective integer values can be represented as a sequence of bytes for use by the software.

The international standards ISO/IEC 10646 [83] and the Unicode Standard (Unicode) [14] describe and define the Universal Character Set (UCS) which is a superset of all other character set standards. It guarantees round-trip compatibility to other character sets. This means simply that no information is lost in the conversion of any text string to UCS and then back to its original encoding [84].

The Unicode Standard Version 4.0 and ISO/IEC 10646:2003 make use of the same character set tables and character encoding methods, but the Unicode Standard additionally provides details of character properties, processing algorithms, and definitions that are useful to implementers [83].

ISO/IEC 10646 and Unicode define several encoding forms, UTF-8 (UCS Transformation format 8), UCS-2, UTF-16, UCS-4 and UTF-32. In an encoding form, each character is represented as one or more encoding units and apart from UTF-8, all other encoding forms have an encoding unit larger than one octet (an 8 bit byte), making them hard to use in many current applications and protocols that assume 8 or even 7 bit characters [85]. UTF-8 uses all bits of an octet for its encoding and it preserves the full US-ASCII range, the latter being encoded in one octet having the normal US-ASCII value. This is important and very useful since it is backwardly compatible with the large existing volume of software which predominantly uses US-ASCII encoding. UTF-8 encodes UCS characters as a varying number of octets, where the number of octets, and the value of each, depend on the integer value assigned to the character in the Unicode character code table.

Unicode has become the dominant encoding scheme in software internationalisation and usage in multilingual environments. Many other standards such as XML have adopted Unicode as the underlying scheme to represent text. Modern operating environments like those under Linux, Mac OS X and MS-Windows XP have support for Unicode [86].

ISO 639

The international standard, ISO 639-1, provides a two-letter code identifier (alpha-2) for the representation of names of languages while ISO 639-2 provides a three-letter identifier (alpha-3) for the languages [131]. Locale language identifiers make use of the ISO 3166 country codes to identify the language to use.

ISO 639-1 was devised mainly for use in terminology. It provides identifiers for those languages which are responsible for a major proportion of the world's literature and which also possess specialised vocabulary and terminology.

ISO 639-2 tries to provide a representation to the world's languages, for use in bibliography as well as terminology, but it is not as restrictive in scope as ISO 639-1. It was devised to include languages that are most frequently represented in the total body of the world's literature, regardless of whether specialised terminologies exist in those languages or not. The three-letter code for ISO 639-2 means that it can accommodate more languages. So, while it limits coverage of individual languages to those for which at least modest bodies of literature have been developed, other languages are still accommodated, by means of identifiers for collections of languages, such as language families [131]. Under ISO 639-2, some languages (22)

have different codes for bibliography and terminology (see Table below on sample ISO 639 codes).

Sample ISO 639-1 and 639-2 Language Codes		
639-2*	**639-1**	**Language Name**
apa		Apache languages
ara	ar	Arabic
bur/mya	my	Burmese
chi/zho	zh	Chinese
dut/nld	nl	Dutch; Flemish
eng	en	English
hin	hi	Hindi
kar		Karen
kin	rw	Kinyarwanda
tlh		Klingon; tlhIngan-Hol
may/msa	ms	Malay
nep	ne	Nepali
swa	sw	Swahili
tam	ta	Tamil
tha	th	Thai
ton	to	Tonga (Tonga Islands)
* For the 639-2 codes, where two codes are provided the bibliographic code is given first and the terminology code is given second.		

ISO 3166-1

ISO 3166-1 provides two (alpha-2) and three-character (alpha-3) codes for representing names of countries [132]. It thus provides a table of country codes just as ISO 639 provides a table of language codes. However, these two standards were developed independently, and there was no attempt to use the same code for a language as that for the country in which it is spoken, and codes from each list should be used independently. Locale country identifiers make use of the ISO 3166 codes to identify the country or region location.

The ISO 3166-1 alpha-2 code is probably best known in its usage for the country code top-level domain (ccTLD) of the Internet Domain Name Service (DNS) system. However there are several ccTLDs in use which are not part of the ISO 3166-1 two-

letter codes, e.g. “uk” for the United Kingdom (the corresponding ISO 3166-1 alpha-2 code is “gb”).

Sample ISO 3166-1 Alpha-2 Country Codes	
ISO 3166-1 (Alpha-2)	**Country/Region**
CA	Canada
DE	Germany
GB	United Kingdom
KE	Kenya
NG	Nigeria
TH	Thailand
TN	Tunisia
VE	Venezuela

RFC 3066

The IETF's RFC 3066 [87] describes a language tag for use in cases where it is desired to indicate the language used in an information object, how to register values for use in this language tag, and a construct for matching such language tags. RFC 3066 specifies use of a 2-character language code from ISO 639-1 when it exists and when a language does not have a 2-character code assigned, the 3-character code is used.

The RFC also specifies the use of optional subtags (e.g. a country code from ISO 3166) and how to register a dialect or variant information with IANA when there is no available ISO 639 code.

Internationalisation and Localisation Software Initiatives

In the past, the language supported in software was very much dependent on where the authors were from. So many common off the shelf (COTS) software was written mainly for the English language due to the dominance of countries like the USA in this area. In recent times, with the emergence of the Internet and globalisation, this predominantly single language-centric support for popular software is changing. There is growing awareness among software developers and authors that many of the software can be and will be deployed world-wide and it is important to be able to adapt the software to the local environment. As a result, there is much better support for internationalisation and localisation on modern software platforms.

For a commercial proprietary software, experience has shown that any localisation effort has to be considered in the light of economical viability and/or other benefits which the effort may bring to the vendor. This means that in many cases, versions of popular commercial proprietary software are not available for languages or cultures where commercial returns are not justified. On the other hand, FOSS can be freely modified and redistributed, and so it just needs some interested party to take the initiative to localise a software that is released as FOSS. This has resulted in many of the popular FOSS being localised (e.g. the Mozilla.org family of products, GNOME, KDE, OpenOffice.org) and made available in many languages including some rather obscure ones.

The Open Internationalisation Initiative

The Open Internationalisation (OpenI18N) Initiative [88] is a key initiative under the Free Standards Group [27]. This initiative has several active projects under it. One of them is the OpenI18N Specification which is concerned with the specification for the interfaces and functionalities that must be supported by Linux-like operating systems to run internationalised application software as well as recommendations for such operating systems to facilitate the development of internationalised application software [89]. Other active projects include:

- Linux Internationalisation Locale Name Guideline
- Common XML Locale Repository
- Internet Intranet Input Method Framework
- OpenI18N Certification Test Suite
- Multilingualisation library (m17n-lib)

All the standards, publications and documentation from the OpenI18N Initiative are freely available.

Some FOSS I18n and L10n Initiatives

Most of the FOSS i18n and/or l10n projects are community-driven. Almost all the major FOSS have good support and tools for i18n and l10n. Local users of the software are encouraged to contribute to the l10n projects.

Mozilla Family

The Mozilla Localisation Project MLP [90], relies mainly on the FOSS community to make the products from the Mozilla Foundation available for different world cultures and languages. The project is focused towards software localisation making use of the underlying internationalisation support available in the products.

The software localisation projects under MLP include:

- Mozilla (aka project Seamonkey) with over 100 languages registered
- Mozilla Firefox with over 30 languages registered
- Mozilla Thunderbird with over 50 languages registered

GNOME

The aim of the GNOME Translation Project [91] is to translate GNOME applications and documentation to every language in existence! This community-based effort currently boasts of translation projects covering well over 100 languages.

KDE

The popular KDE software also has wide support for its internationalisation and localisation initiatives [92]. There are good guides and documentation available and again community-driven projects for localisation are well supported and received. As a result KDE is currently available in over 100 languages.

OpenOffice.org

OOo has a framework and tools for both i18n and l10n [93]. OOo is now available in over 70 languages covering all the major languages and cultures of the world and also some minor ones too.

Microsoft Software

The newer versions of software from Microsoft e.g. MS-Windows XP, MS-Office 2003 have good internationalisation support and are also available in many localised native versions.

MS-Windows XP

Localised versions of MS-Windows XP is available in 24 languages and the Mulitilingual User Interface (MUI) Pack offers more localised user interface languages. The MUI Pack is a set of language specific resource files that can be added to the English version of MS-Windows. All-in-all Microsoft claims that the total number of languages supported in MS-Windows XP is in excess of 140 languages [94].

MS-Office

Localised versions of MS-Office 2003 are available in over 35 languages [95]. In addition, the MS-Office MUI offers support for other languages for which a localised version is not available.

INTELLECTUAL PROPERTY RIGHTS IN STANDARDS

For the case of technical standards, it is not uncommon that sometimes a technology that is best suited to base a specification on is proprietary and is encumbered by some intellectual property protection scheme e.g. patents. So the standards development body has to decide on whether it should use an item that has intellectual property rights protection or look for an alternative which may be inferior. In the past, the development of standards related to software and IT have proceeded using mainly a reasonable and non-discriminatory terms policy (RAND) whenever intellectual property (IPR) e.g. patents, copyrights etc. [96], are included in a standard. Under RAND the IPR holder must be willing to negotiate rights to use the item on reasonable and non-discriminatory terms. The intent of RAND was to prevent intellectual property rights issues from hindering the adoption of a standard and to ensure that the cost of any licenses, arising from the IPR, needed are affordable. This has proved adequate in the past but in recent times with the proliferation of patents granted to software-based innovations (software patents), commercial organisations have began to become more aggressive in pursuing and protecting their IPR in IT especially with regard to software patents. This has led to the concern that the inclusion of IPR in standards under RAND may not be enough and standards developing and setting bodies all over the world are re-looking at their IPR policies. Since in general a standard is targeted for use by all in the world, it is vital that the terms of usage of any intellectual property that is included in the standard are clearly specified.

Software Patents

Of particular importance to the IPR issue, is the inclusion of software patents in IT-related specifications and standards that are meant for all to implement. The issues on the patentability of software and the way patents offices process software patent applications are very controversial. Software has become patentable in some countries e.g. USA, Japan, Singapore. In these countries software 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 of these countries are re-examining this issue and trying to decide whether they should change their positions.

The software patents argument has been discussed extensively [97,98,99,100]. A summary of its pros and cons (taken mainly from a note published by the author [101]) is given below.

The proponents of software patents say that as long as a country recognises the patent system as a means to reward individual entities for technical innovation and invention it should apply equally to the functionalities found in software. With this, as with a physical hardware invention or component, the inventor is appropriately

rewarded by being granted a set of exclusive rights for a limited amount of time. Only with this can the country appropriately reward and in doing so encourages technical innovation and inventions. So in the case of software, if its functionalities are patentable, the software owner or developer can protect his software innovations against competitors and others who may want to steal his ideas and software functionalities. The patent system if applied to software will offer protection for the lone inventor or small-time developer against his ideas being usurped by larger developers and corporations. They also argue that since the patents office allow only real inventions and innovations to be patentable there should be no fear of abuse. So when viewed in this manner, software patents should be good for innovation, especially for small software companies and developers since they can come up with good ideas, patent them and then collect royalties from their inclusion in other software products.

The detractors of software patents say that it is almost impossible to create software without building on the previous ideas and work of others. For any single piece of useful software, there can literally be thousands or more of such ideas or functionalities and it will be almost impossible to ensure that no patents are violated. The small and independent developer will face a huge challenge if he wants to avoid infringing a patent as he will need to expand considerable amounts of time and financial resources to perform the check. Software creation and innovation will suffer and a strict software patents regime only benefits big corporations with a huge arsenal of registered patents since it is only mainly these organisations that has the capability to cross-license patents with other corporations. At the national level, less developed software economies will suffer as, collectively, their indigenous companies will probably possess only a relatively small number of patents to benefit from cross-licensing. The software development entry-level is raised and the cost of software will rise due to patent royalty payments leading to increased cost of doing business and providing services across the board.

Another reason software patents may probably do more harm than good is that the entry level for software inventions or innovation is very low, unlike physical inventions or pharmaceutical research. With the current state of computer power and Internet access, almost anyone with a personal computer can tinker with software creation and development and in doing so come up with all sorts of ideas for functionalities in software and file for patent protection. In theory, the patents office is supposed to scrutinise patent applications and reject non-innovative and obvious or trivial ideas. It is also supposed to reject a patent if the idea or function is already in use in some other software. However, in practice all over the world we find that contrary to this, software patents are granted too easily. The patents office is usually ill-equipped to search for software functionalities already in use (“prior art”) when considering a patent application. Patents are granted for broad, general and very often trivial concepts, e.g., double-click for handheld devices, progress bar on a computer screen, etc.

It should be noted that in the above arguments against software patents there is no distinction about the type of software that will suffer. It will affect all types of software, proprietary as well as open source, commercial as well as non-commercial.

Policies on Intellectual Property Rights

Standards have been produced which include patented technologies and all the main standards bodies have policies with regard to the treatment of intellectual property rights in the documents that they produce [102].

The documents produced by these organisations are subject to certain copyright terms and possibly other IPR protection pertaining to their access and distribution.

ISO

ISO has published directives on the issue of IPR in its standards development process [103, 104].

With regard to patents, 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 then 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.

For the case of the inclusion or reference to proprietary material (including tradenames) in ISO publications, ISO takes the same stand as for patents, in that these are not to be used as far as possible but if their usage cannot be avoided then a basic text identifying this should be included in the document.

Generally ISO publications have to be purchased. However, in certain cases when the publication is jointly produced with another body, it may be made available without charge if the collaborating organisation makes its publications available for free. Some ISO member national standards bodies have an arrangement with ISO in which they can adopt without change an ISO standard as their own national standard.

IETF

RFC 3979 [105], clarifying on the position taken in RFC 2026 [106], is the main document governing the IETF's stand on IPR. In general, IETF prefers technologies with no known IPR claims or technologies with IPR claims that offer royalty-free licensing. However, the IETF working groups have the discretion to adopt technology with a commitment of fair and non-discriminatory terms, or even with no licensing commitment, if they feel that this technology is superior enough to alternatives with fewer IPR claims or free licensing to outweigh the potential cost of the licenses.

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, all those contributing to the working group's discussions must disclose the existence of any IPR which is believed to cover or may ultimately cover the technology under discussion.

IETF documents are all available without charge from the Internet from the IETF website as well as several other sites.

W3C

The W3C has a very clear policy with regard to patents usage in its Recommendations. This has arisen from the experience it had with the WWW.

Many of the 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 of the patent holders, including some who had participated in the development of the standards themselves, sought license payments. As a result of this, the W3C decided to have a clear patent policy governing the Recommendations that it develops [107].

The key position of the W3C with regard to patents that 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 Royalty-Free License (RF) requirements. An "essential claim" refers to a patent for which there is no known alternative to it and therefore it is essential to the implementation of a normative part of a Recommendation [108].

The policy 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 the RF requirements.

In the event that a patent issue cannot be resolved within the Working Group, 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.

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

W3C documents are available without charge from the W3C website.

OASIS

OASIS has a published policy which governs the treatment of intellectual property in the production of specifications and other works by OASIS [109].

OASIS also deals with patents that are considered as "essential claims" i.e. patents that are deemed essential for the implementation of a normative part of an OASIS standard.

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" [110].

RAND defines a basic set of minimal terms 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 implementer seeking a license.

Royalty-Free (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 that may be included in a patent holder's license and that must be granted upon request without further negotiations.

OASIS standards are available from their website without charge.

Patents in Open Standards and FOSS

As can be seen from the discussion in the previous section, most of the standards bodies do allow the inclusion of patents in their standards although patent-free ones are preferred. The patent policies of these organisations are summarised in the table below. These policies all revolve around either allowing a reasonable and non-discriminatory (RAND) policy with some form of royalty payment or an entirely royalty-free one or a mixture of both.

Patent Policies of Some Standards Organisation

Organisation	RAND	Royalty-free
ISO	X	
IETF	X	X
W3C		X

Patent Policies of Some Standards Organisation		
OASIS	X	X

This issue of royalty payment for using a certain standard is of particular importance to FOSS. Since most FOSS is widely available without charge it is not practical to impose any royalty payment for using it and there is no mechanism for collecting such payments. Furthermore, having patents in software will curtail the freedom associated with its distribution and modification, two fundamental freedoms inherent in FOSS. In view of this, if patents, especially software patents, are used in a standard it will be difficult for FOSS to conform to the standard unless the patents are granted irrevocably for use royalty-free. As part of its definition of what constitutes an open standard, the European Commission's European Interoperability Framework [9], includes the property of royalty-free patents being made available on an irrevocable basis. This has led the Business Software Alliance (BSA), whose key members include some of the most prominent close-source vendors like Microsoft and Adobe, to protest that a more appropriate characteristic should be one which specifies that any patent rights needed to implement the standard is to be made available on a RAND basis [130]. This suggestion if it is ever successfully included as a key characteristic of an open standard will undermine the mutually beneficial relationship which FOSS currently enjoys with open standards.

THE LINUX STANDARD BASE (LSB)

The Linux operating system consists actually of the Linux kernel itself and the rest of the system software and tools/utilities which working together make up the operating system. Most of the system software is from the GNU Project [111]. In addition, in order for an operating system to be useful to most people, it has to be made available with support for some application software. The strong community-based history and support of Linux together with the nature of the licensing of the Linux kernel and the GNU software resulted in many people taking the kernel, system software from GNU and possibly other open-source utilities/tools, add in some application software which they deem useful and put all these together to form a working package. This working package is termed a Linux distribution or distro. Consequently, the Linux operating system comes in very many distros [112]. The large number of distros available coupled with the fact that since most of the software, if not all, included in a distro are open source and hence can be customisable to suit the requirements of a particular distro, have resulted in a fair measure of binary and configuration incompatibilities among distros. Some of the incompatibility problems include different library versions, package formats and differences in directory and file layouts. It has been recognised that if Linux is to be fully embraced and supported by mainstream computing as a legitimate alternative to proprietary operating systems there is a need to cut down on these incompatibilities so that a software package with source can compile cleanly across distros and a binary version can run properly across all distros. The Linux Standard Base Project (LSB) [113] tries to do this by specifying a standard for Linux.

What is the LSB

The LSB is a project under the Free Standards Group [27]. It attempts to develop and promote a set of binary standards that will increase compatibility among Linux and other similar systems . These standards will also enable software applications to run on any conforming system.

While the main goal of the LSB project is to increase compatibility among Linux distros by specifying and promoting standards for their use, it does not limit the applicability of the specification to only the Linux environment. The LSB specification has been written so that it can be readily implemented on any Unix-like operating system, natively or as a compatibility layer. With more work it can also be implemented on other operating systems.

The LSB is a community-based project and anyone can contribute to it by participating in the various LSB mailing lists. There is considerably good support for the LSB standard among commercial software and Linux vendors like Mandrakesoft, Miracle Linux, Novell, Progeny, Red Flag, Red Hat, IBM, Oracle, Veritas, MySQL etc.

The LSB Specification

The LSB comprises a single common (generic) specification and architecture specific specifications. The complete specification for a particular platform consists of the generic specification plus one of the architecture specifications. Architectures supported currently are IA32 and IA64 (Intel 32 and 64-bit processors), PPC32 and PPC64 (IBM's 32 and 64-bit PowerPC family) , S390 (IBM's S390 processors) and S390X (IBM zSeries processors), AMD64 (Advanced Micro Devices 64-bit processors).

The LSB defines both a set of Application Program Interfaces (APIs) for source code and Application Binary Interfaces (ABIs) for compiled binaries. A conforming implementation has to support all of the ABIs in the LSB but not all of the source level APIs.

The LSB is divided into specification modules in which a specification module refers to a unique collection of one or more functions that have value for a certain group of runtime implementations. The modules currently available are LSB-Core, LSB-C++, LSB-Graphics and LSB-I18n. Both the LSB-Core and LSB-C++ have generic and architecture-specific specifications while the LSB-Graphics and the LSB-I18n have only the generic specification. The Table below summarises the currently available modules [133].

LSB Modules		
Module	**Functional Area**	**Architectures Available**
LSB-Core	ELF	Generic, Processor-specific
	LSB	Generic, Processor-specific
	Packaging	Generic, Processor-specific
LSB-CXX	LSB-C++	Generic, Processor-specific
LSB-Graphics	Graphics	Generic
LSB-I18n	OpenI18n	Generic

The latest version of the LSB as of the writing of this document is 3.0.0. LSB 2.0.1 had been submitted to ISO to become an international standard for Linux.

LSB-Core Specification

This is the Core module of the Linux Standard Base. This module provides the fundamental system interfaces, libraries, and runtime environment upon which all

conforming applications and libraries depend. It provides specifications for the following areas:

- Executable and Linking Format (ELF)
- Base libraries
- Utility libraries
- Command and utilities
- Execution environment
- System initialisation
- Users and Groups
- Package format and installation

The specifications make extensive use of existing standardised APIs and ABIs from other bodies. Some of the normative references include those from ISO POSIX, the System V Interface Definition (SVID) and the Filesystem Hierarchy Standard (FHS).

In particular, the LSB-Core specification includes many interfaces described in ISO POSIX (ISO/IEC 9945) [114] and it specifies that such interfaces should behave exactly as specified in the POSIX standard. It is also the long term plan of the LSB to converge with ISO/IEC 9945.

One of the problems plaguing the many different Linux distros has been the various formats used in software package distribution. The LSB addresses this by specifying that applications shall be packaged in the RPM packaging format as defined in the LSB, or supply an installer which is LSB conforming (for example, by invoking LSB commands and utilities). This means that while packages are encouraged to be supplied in RPM format the LSB does not mandate the use of the rpm program or database.

LSB-C++ Module

This is the C++ module of the Linux Standard Base (LSB). It supplements the core interfaces by providing system interfaces, libraries, and a runtime environment for applications built using the C++ programming language.

Normative references include the LSB-Core, IOS POSIX and the ISO/IEC 14882 C++ Language standard.

It provides specifications for the following areas:

- Low level system information
- Base libraries
- Package information

The LSB-Graphics Module

This specification defines the graphical interface found on an LSB conforming system. Normative references include the LSB-Core and graphic libraries and specifications from the The X.Org Foundation [45].

It provides specifications for the following areas:

- Graphic libraries
- OpenGL libraries
- Package information

The LSB-I18n Module

This module corresponds to the OpenI18N Global Specification [89] from the OpenI18N Project.

LSB as an ISO Standard

LSB 2.0.1 had been submitted to ISO for use as an international standard for Linux through the ISO PAS (Publicly Available Specification) process. It has been approved as a draft international standard (DIS 23360) for publication in 2005 subject to another round of balloting and some other ISO processes.

The availability of an ISO Linux standard is an important milestone symbolically, in the development of Linux. It signifies that the Linux operating environment has come of age and is now officially recognised as a full-fledged mainstream computing platform. This will give confidence to corporations and governments, some of whom have been reluctant to use Linux till now even though they are aware that it will give them many benefits as they are not confident of its long-term viability and international acceptance. An ISO Linux standard will also help the acceptance and usage of FOSS in general as many FOSS products are implemented on Linux and Linux is arguably the most well-known FOSS product.

LSB Certification

Linux distributions which conform to the LSB can be certified as such. The LSB certification scheme is run on behalf of the Free Standards Group by the Open Group [116], a vendor and technology-neutral consortium, to ensure neutrality and

confidentiality. Certification charges are kept to a minimum to encourage developers, ISVs and Linux distributions to become LSB certified.

LSB certification is currently available for the following:

- LSB Runtime Environment
- LSB Application
- LSB Internationalised Runtime Environment

Developers and vendors are granted a license to use the LSB Certified trademark in connection with a particular product once it has passed the applicable certification test suites.

GOVERNMENT/NATIONAL OPEN STANDARDS POLICIES AND INITIATIVES

Many governments all over the world have begun to take notice of FOSS and the benefits that it can bring to a nation. As FOSS development methodologies tend to promote specifications which are freely available for all to use and to participate in their development, many governments now realise the benefits of open specifications and standards and the dangers of proprietary ones. As a result, in many countries, the government has come out with policies and/or initiatives which advocate and favour open standards. Indeed some governments place a higher priority on open standards as it can bring about increased independence from specific vendors and technologies and at the same time accommodate both FOSS and proprietary software. This is especially true for most of the e-Government projects and initiatives all over the world.

Open Standards Favoured by Governments
E-Government initiatives as well as many government agencies now favour the use of open standards where possible. Some of the open standards that are frequently specified are:
Networking protocols – TCP/IP
Networking services – HTTP, SSL, SMTP, MIME, IMAP, LDAP
Document exchange – XML and XML-based specifications
Web services – UDDI, SOAP
Database – SQL
Internationalisation - UNICODE

This section will review the government policies and initiatives of several countries with respect to open standards.

The European Union

The European Union (EU) comprises many nation states with many diverse cultures, languages at varying states of technology/technical development. For it to able to function effectively, especially in the area of information exchange, the governments concerned have to establish a proper interoperability framework and standards on data interchange.

The European Commission's Interchange of Data between Administrations (IDA) programme supports the implementation of EU legislation by facilitating the exchange of information between public administrations across Europe through the use of information technology. Under the IDA Programme is the development of the European Interoperability Framework (EIF) [117]. The EIF is a framework for the e-

Government services of the member states to facilitate the interoperability of these services at pan-European level.

The EIF version 1.0 [9] recommends the use of open standards for maximum interoperability among the e-Government services. It defines the minimal characteristics for open standards as the following:

- the standard is adopted and maintained by a not-for-profit organisation
- the development of the standard occurs using an open decision-making process and does not preclude any party from it
- the standard is published and is available either free of charge or for a nominal fee
- the published standard must be available for all to copy and distribute it either free of charge or for a nominal fee
- any patents present in the standard are to be irrevocably available on a royalty-free basis

The IDA expert group on open document formats has recommended that the European Union's public sector use open formats in their electronic documents [118]. For revisable documents, XML-based formats like the Open Document format from OASIS and Microsoft's new XML-based MS-Office formats are recommended.

United Kingdom

The United Kingdom's e-Government initiative places a lot of emphasis on open technical standards to achieve seamless information flow across the public sector and to provide citizens and business with better access to government services [119]. Its e-Government Interoperability Framework (e-GIF) defines the technical policies and specifications governing information flows across government and the public sector. Complying with e-GIF at the highest level includes the use of open standards like XML as the primary means for data integration and the implementation of Internet and WWW standards [120].

Denmark

The Danish e-Government Interoperability Framework includes recommendations and status assessments for more than 450 selected standards, specifications and technologies used in its e-Government solutions [121]. In general the Framework recommends the use of open standards and centrally agreed XML schemas (which may be provided free of charge throughout the public sector) for data interchange.

As part of the Interoperability Framework, the policy on data and document exchange specifies that documents should be published in generally available formats for which free readers exist and the use of proprietary word processing formats such as MS-

Word or formats that do not have widely available readers should be avoided for publicly available documents [122].

Netherlands

The Netherlands has its OSOSS - the Programme for Open Standards and Open Source Software in government [123]. This programme encourages the use of open standards and provides information on open-source software. The Dutch ICTU, the organisation for ICT and government programme, runs OSOSS. While targeting the public sector, the results of the Programme will be available for the private sector and individuals too. The Programme provides information and advice to the public sector on open standards. It has set up a catalogue of recommended open standards [124] for use in the public sector.

Norway

The Norwegian Government has declared that proprietary formats will no longer be acceptable in communication between citizens and government [125]. As part of its "eNorge 2009 – the digital leap" masterplan for IT, all public sector bodies in Norway are to have in place a plan for the use of open-source software and open standards by the end of 2006.

Massachusetts, USA

The Commonwealth of Massachusetts, USA, has announced an IT Policy that emphasises the importance of open standards compliance for IT investments in Massachusetts [126]. The Policy states that all prospective IT investments will have to comply with open standards referenced in the current version of the Enterprise Technology Reference Model (ETRM) of the Commonwealth. It further says that existing IT systems will be reviewed for open standards compatibility and will be enhanced to achieve open standards compatibility where appropriate. In addition, open standards solutions will be selected when existing systems are retired off or need major enhancements.

New Zealand

As part of its e-Government vision, New Zealand has come up with a supporting Information Systems (IS) Policies and Standards document. The guiding principles state that the IS Policies and Standards are to be based on open standards, wherever possible [127]. New Zealand also has an e-Government Interoperability Framework (NZ e-GIF) [128] which lists the mandatory use of many open standards for compliance.

Malaysia

The Malaysian Government Interoperability Framework (MyGIF) [129] defines the minimum set of IT standards and technical specifications for use in Government Ministries, agencies and departments. These cover the areas of interconnection, data integration, information access, security and metadata. Instead of creating new standards or specifications, MyGIF adopts internationally recognised open and de-facto IT standards and specifications for all the interoperability areas mentioned.

Others

Many other countries e.g. South Africa, Vietnam, Brazil and Peru have started initiatives and/or policies to address the digital divide and to improve their government's IT implementations by leveraging on FOSS and open standards together. These countries view FOSS and open standards as going hand-in-hand. Consequently by using FOSS solutions and products, they will be implementing open standards as most of the FOSS products make use of open standards.

Summary

As can be seen from the list above, most of the world's governments are asking for the adoption of open standards and specifications as much as possible in their country's IT usage and/or e-Government projects and initiatives. This is a good move since if more and more countries are to insist on open standards, more proprietary software vendors will be forced to open up their file formats and technology specifications and adhere to open standards in their products as much as possible.

CONCLUSION

This primer has tried to explain what technical standards are and the characteristics of what may be termed as open standards in the field of information technology. Specifications which satisfy these characteristics can be viewed as open ones and those which are in widespread use and acceptance may be regarded as open standards.

Open standards is even more important in this present information age of IT and Communications convergence (ICT) and the Internet. No single technology, group or vendor can provide for everything and therefore interoperability in a heterogenous environment is required more than ever. It is only by strict adherence to standards and specifications which are open and non-discriminatory can a high degree of interoperability be achieved.

XML and related technologies are expected to play an important role in setting new standards for better interoperability and information exchange in the areas of Web applications, services, e-commerce as well as in office applications. It is crucial that these standards are steered and developed by open standards bodies. Towards this end it is very important that bodies like W3C, OASIS and IETF remain open and support non-discriminatory policies especially with regard to intellectual property rights issues.

In many environments, the demand and usage of open standards go hand-in-hand with FOSS. As pointed out in the primer, FOSS and open standards are two distinct and different domains and it is possible to have a non-FOSS product implement open standards and a FOSS software make use of a proprietary specification. Nevertheless, in practice, FOSS is generally a great implementor and supporter of open standards in software and so it is not surprising that many developing countries see them working in tandem.

The software localisation initiatives of many countries will benefit from the setting and availability of more open standards in the relevant areas. The easy and free access to open standards related to internationalisation and localisation will encourage more local people to participate in these initiatives. Again FOSS is playing a key role in providing and encouraging localisation in many important end-user application areas e.g. web browsers, office applications and the desktop environment.

More and more governments are asking for open standards now and this is a very good sign as they are the biggest buyers and consumers of IT products and software. The vendors will have to comply with open standards and open up any proprietary file formats or specifications in response to these demands. In conjuction with this , it is hoped that more and more users too will follow suit. It is the aim of this primer to help

educate and make the reader aware of the benefits of open standards and the pitfalls of proprietary ones even though the latter may be de-facto standards and widely used.

It should be the ultimate objective of users to be able to access and use applications and services using any device, platform or interface of their choice and at the same time be able to exchange information and data from these applications/services with other users without suffering any degradation in content. The only practical way all these can be achieved is through strict conformance with open standards by all.

REFERENCES

1. Wikipedia (the free-content encyclopedia) entry on "standards"
http://en.wikipedia.org/wiki/Standards

2. Merriam-Webster Online Dictionary
http://www.m-w.com/

3. Wiktionary - a wiki-based Open Content dictionary
http://en.wiktionary.org/wiki/Main_Page

4. Webopedia - the online computer dictionary for computer and Internet technology definitions
http://www.webopedia.com/TERM/s/standard.html

5. Overview of the ISO system - Why standards matter
http://www.iso.org/iso/en/aboutiso/introduction/index.html

6. Standards: Do We Really Need Them?
http://www.masternewmedia.org/2003/12/26/standards_do_we_really_need.htm

7. Wikipedia (the free-content encyclopedia) entry on "open standards"
http://en.wikipedia.org/wiki/Open_standards

8. Open Standards Policy, IT Division, Massachusetts, USA
http://www.state.ma.us/itd/openstandards.htm

9. Euporean Interoperability Framework v1.0
http://europa.eu.int/idabc/en/document/3473/5585#finalEIF

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

11. Free Software Foundation, "Free Software"
http://www.fsf.org/philosophy/free-sw.html

12. Ken Wong, "FOSS General Introduction Primer"
http://www.iosn.net/foss/foss-general-primer/

13. OASIS Standards, "OpenDocument Format for Office Applications (OpenDocument) v1.0"
http://www.oasis-open.org/specs/index.php#opendocumentv1.0

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

15. CNET News, "Microsoft FAT patent falls flat"
http://news.com.com/Microsoft+FAT+patent+falls+flat/2100-1014_3-5390138.html

16. Organization for the Advancement of Structured Information Standards (OASIS)
http://www.oasis-open.org

17. The Internet Engineering Task Force (IETF)
http://www.ietf.org

18. The World Wide Web Consortium (W3C)
http://www.w3c.org

19. Ken Krechmer, "The Meaning of Open Standards",
http://www.csrstds.com/openstds.html

20. The International Organisation for Standardisation (ISO)
http://www.iso.org/

21. The International Electrotechnical Commission (IEC)
http://www.iec.ch/

22. The International Telecommunication Union (ITU)
http://www.itu.int/

23. Overview of the IETF
http://www.ietf.org/overview.html

24. The Internet Society (ISOC)
http://www.isoc.org/

25. RFC 2026, The Internet Standards Process, Revision 3
http://www.ietf.org/rfc/rfc2026.txt

26. Official Internet Protocol Standards
http://www.rfc-editor.org/rfcxx00.html

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

28. IEEE Standards Association
http://standards.ieee.org/

29. RFC 793, Transmission Control Protocol
http://www.ietf.org/rfc/rfc793.txt

30. RFC 791, Internet Protocol
http://www.ietf.org/rfc/rfc791.txt

31. RFC 2616, Hypertext Transfer Protocol -- HTTP/1.1
http://www.ietf.org/rfc/rfc2616.txt

32. HTML 4.01 Specification
http://www.w3.org/TR/html401/

33. ISO/IEC 15445:2000
http://purl.org/NET/ISO+IEC.15445/15445.html

34. RFC 2821, Simple Mail Transfer Protocol
http://www.ietf.org/rfc/rfc2821.txt

35. RFC 2822, Internet Message Format
http://www.ietf.org/rfc/rfc2822.txt

36. RFC 1652, SMTP Service Extension for 8bit-MIMEtransport
http://www.ietf.org/rfc/rfc1652.txt

37. RFC 2045, 2046, 2047, 2048, 2049, Multipurpose Internet Mail Extensions (MIME) Parts One - Five
http://www.ietf.org/rfc/rfc2045.txt
http://www.ietf.org/rfc/rfc2046.txt
http://www.ietf.org/rfc/rfc2047.txt
http://www.ietf.org/rfc/rfc2048.txt
http://www.ietf.org/rfc/rfc2049.txt

38. Extensible Markup Language (XML) 1.0 (Third Edition)
http://www.w3.org/TR/REC-xml

39. Flynn, P (Ed.), The XML FAQ v4.1, Cork, May 2005
http://xml.silmaril.ie/

40. W3Schools, "Introduction to XML"
http://www.w3schools.com/xml/xml_whatis.asp

41. Wikipedia (the free-content encyclopedia) entry on "SGML"
http://en.wikipedia.org/wiki/SGML

42. XML.ORG
http://www.xml.org/

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

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

45. The X.Org Foundation
http://www.x.org/

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

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

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

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

50. 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

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

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

53. PDF Specificatons
http://partners.adobe.com/public/developer/pdf/index_reference.html

54. Wikipedia (the free-content encyclopedia) entry on "Portable Document Format"
http://en.wikipedia.org/wiki/Pdf

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

56. Wikipedia (the free-content encyclopedia) entry on "Graphics file format"
http://en.wikipedia.org/wiki/Graphics_file_format

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

58. TIFF Specifications
http://partners.adobe.com/public/developer/tiff/index.html

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

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

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

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

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

64. Wikipedia (the free-content encyclopedia) entry on "Audio file format"
http://en.wikipedia.org/wiki/Audio_file_format

65. Wikipedia (the free-content encyclopedia) entry on "WAV"
http://en.wikipedia.org/wiki/WAV

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

67. Wikipedia (the free-content encyclopedia) entry on "MP3"
http://en.wikipedia.org/wiki/MP3

68. Wikipedia (the free-content encyclopedia) entry on "AAC"
http://en.wikipedia.org/wiki/AAC

69. Wikipedia (the free-content encyclopedia) entry on "RealAudio"
http://en.wikipedia.org/wiki/RealAudio

70. Wikipedia (the free-content encyclopedia) entry on "Video file format"
http://en.wikipedia.org/wiki/Video_file_format

71. Wikipedia (the free-content encyclopedia) entry on "RIFF"
http://en.wikipedia.org/wiki/RIFF

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

73. Wikipedia (the free-content encyclopedia) entry on "Advanced Streaming Format"
http://en.wikipedia.org/wiki/ASF

74. Wikipedia (the free-content encyclopedia) entry on "QuickTime"
http://en.wikipedia.org/wiki/QuickTime

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

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

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

78. Wikipedia (the free-content encyclopedia) entry on "H.264/MPEG-4 AVC"
http://en.wikipedia.org/wiki/H.264

79. Wikipedia (the free-content encyclopedia) entry on "WMV"
http://en.wikipedia.org/wiki/WMV

80. Wikipedia (the free-content encyclopedia) entry on "Internationalization and localization"
http://en.wikipedia.org/wiki/Internationalization_and_localization

81. Wikipedia (the free-content encyclopedia) entry on "Locale"
http://en.wikipedia.org/wiki/Locale

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

83. 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=

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

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

86. Wikipedia (the free-content encyclopedia) entry on "Unicode"
http://en.wikipedia.org/wiki/Unicode

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

88. The Open Internationalisation Initiative
http://www.openi18n.org/

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

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

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

92. KDE Internationalisation
http://i18n.kde.org/

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

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

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

96. Wikipedia (the free-content encyclopedia) entry on "Intellectual property"
http://en.wikipedia.org/wiki/Intellectual_property

97. Wikipedia (the free-content encyclopedia) entry on "Software patent debate"
http://en.wikipedia.org/wiki/Software_patent_debate

98. No Software Patents!
http://www.nosoftwarepatents.com/

99. Richard Stallman, "Patent Absurdity"
http://www.guardian.co.uk/online/comment/story/0,12449,1510566,00.html

100. Patents for Innovation
http://www.patents4innovation.org/

101. Nah Soo Hoe, "A Brief Note on Software Patents"
http://www.mncc.com.my/articles/nsh1.html

102. Priscilla Caplan, "Patents and Open Standards"
http://www.niso.org/press/whitepapers/Patents_Caplan.pdf

103. ISOTC Portal, "Intellectual Property Rights (IPR)"
http://isotc.iso.org/livelink/livelink.exe/fetch/2000/2122/3146825/4229629/sds_ipr.htm

104. ISO/IEC Directives, Part 1, Procedures for the technical work (Ed.5)
http://isotc.iso.org/livelink/livelink.exe?func=ll&objId=4230455&objAction=browse&sort=subtype

105. RFC 3979, "Intellectual Property Rights in IETF Technology"
http://www.ietf.org/rfc/rfc3979.txt

106. RFC 2026, "The Internet Standards Process -- Revision 3"
http://www.ietf.org/rfc/rfc2026.txt

107. W3C Patent Policy
http://www.w3.org/Consortium/Patent-Policy/

108. Overview and Summary of W3C Patent Policy
http://www.w3.org/2004/02/05-patentsummary.html

109. OASIS Intellectual Property Rights (IPR) Policy
http://www.oasis-open.org/who/intellectualproperty.php

110. OASIS Intellectual Property Rights (IPR) Policy FAQ
http://www.oasis-open.org/who/ipr/ipr_faq.php

111. The GNU Operating System
http://www.gnu.org/

112. DistroWatch
http://distrowatch.com/

113. The Linux Standard Base Project
http://www.linuxbase.org/

114. ISO POSIX (2003)
http://www.unix.org/version3/

115. Filesystem Hierarchy Standard (FHS)
http://www.pathname.com/fhs/

116. The Open Group
http://www.opengroup.org

117. European Interoperability Framework for pan-European eGovernment services
http://europa.eu.int/idabc/en/document/2319/5644

118. Documentation on the Promotion of Open Document Exchange Format
http://europa.eu.int/idabc/en/document/3439/5585

119. "Schemas and Standards"
http://www.govtalk.gov.uk/interoperability/schemasstandards.asp

120. e-Government Interoperability Framework Version 6.1
http://www.govtalk.gov.uk/schemasstandards/egif_document.asp?docnum=949

121. Danish e-Government Interoperability Framework
http://standarder.oio.dk/English/

122. The Interoperability Framework: Technical Standards: Document and data exchange
http://standarder.oio.dk/English/Tekniske_standarder/Dokument-_og_dataudveksling/index.html

123. Programme for Open Standards and Open Source Software in Government
http://www.ososs.nl/index.jsp?alias=english

124. The Dutch Government Open Standards Catalogue
http://www.ososs.nl/matrix/matrix.jsp?id=10927

125.
http://www.andwest.com/blojsom/blog/tatle/agenda/2005/06/27/Norwegian_Minister_Proprietary_Standards_No_Longer_Acceptable_in_Communication_with_Government.html

126. Commonwealth of Massachusetts - Enterprise Open Standards Policy
http://www.mass.gov/Aitd/docs/policies_standards/openstandards.pdf

127. New Zealand Government Information Systems Policies and Standards
http://www.e-government.govt.nz/docs/is%2Dpolicies%2Dstandards/

128. New Zealand E-government Interoperability Framework (NZ e-GIF)
http://www.e-government.govt.nz/docs/e-gif-v-2-1/index.html

129. Standards, Policies and Guidelines - Malaysian Government Interoperability Framework (MyGIF)
http://www.mampu.gov.my/mampu/bm/program/ICT/ISPlan/ispdoc/Interoperability%20Framework.pdf

130. BSA letter to the European Commission on the European Interoperability Framework
http://www.bsa.org/eupolicy/loader.cfm?url=/commonspot/security/getfile.cfm&pageid=22753&hitboxdone=yes

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

132. Wikipedia (the free-content encyclopedia) entry on "ISO 3166"
http://en.wikipedia.org/wiki/ISO_3166

133. Getting Started With LSB 3.0
http://www.linuxbase.org/build/lsb30.html

GLOSSARY

ABOUT THE AUTHOR

Nah Soo Hoe
nsh@pop.jaring.my

ACKNOWLEDGEMENTS