File Formats, and Free Software

Simon Bridge *

March 20, 2008

A *format* is a published specification for storing digital data, usually maintained by a standards organization.

1 Open Formats

Open formats must be free of legal restrictions on use. For example, an open format must be implementable by both proprietary and Free software, using the typical licenses used by each. In contrast to open formats, proprietary formats are controlled and defined by private interests.

The primary goal of open formats is to guarantee long-term access to data without current or future uncertainty with regard to legal rights or technical specification. Governments have increasingly shown an interest in open format issues.

1.1 And Free Software

Formats may be open or closed, free or proprietary. They can be recognised by an international standards body or not. The use of non-open (closed) formats in Free Software threatens the freedom of the software. However, propriatory formats are often used, provided they are completely free (no IP encumberance).

1.2 International Organisation for Standardization

While ISO acceptance is not needed to be an open format, many countries (including NZ) have laws concerning the adoption of ISO standards. So acceptance is often seen an a springboard to widespread adoption.

Examples of ISO standards used in NZ include the metric system of units and measures (ISO 31 et al) and many Health and Safety standards.

Closed formats have a lock-in effect which can have profound consequences, especially in government use. For example, IPONZ[1] publishes patents online in Microsoft Word (DOC) format. This means that everybody who applies for a NZ patent pays Microsoft Corporation for the priviledge.

Some people have questioned why New Zealand innovators should be forced to pay a de-facto tax to Redmond. Extend this to Inland Revenue, Justice, Health, Education ... and the issue becomes clear. This license money could be going to tax cuts, reducing crime, hospitals, schools, and public servant salaries.

[1] Intellectual Property Office of New Zealand

1.3 Extensible Markup Language (XML)

[Open, FOSS, n/a]

This is a meta-format. It allows users to create their own markup language. It is similar to ISO/IEC 24824-1 and IBMs old GML. There are a great many ISO standards based on XML, and it is not restricted to any particular media.

2 Documents

Open document formats have been the focus of some international controversy over the last year, associated with Microsoft Corporation's application to have it's Office Open XML format accepted as an open standard by the International Organisation for Standardization (ISO).

2.1 Open Document Format (ODF)

[Open, FOSS, ISO/IEC: 26300]

ODF is a file format for electronic office documents, such as spreadsheets (ods), charts (odg), presentations (odp) and word processing (odt) documents. The idea is to make it easy to share this information.

This is a Free open standard developed by Sun Microsystems and the OASIS[2] consortium. It is based on the XML format originally created and implemented by the OpenOffice.org office suite.

2.2 Portable Document Format (PDF)

[Open, Prop. ISO/IEC 32000]

PDF is a fixed-layout format used for representing two-dimensional documents in a manner independent of the application software, hardware, and operating system. It is a proprietary open standard owned by Adobe Systems.

2.3 LaTeX(TEX)

[Open, FOSS, n/a]

Pronounced "lay-tek" it is the publication standard of choice in academia, because of it's use of TEXhigh quality typography. It's license, the LPPL[3] qualifies as a Free software licence, but it is incompatible with the GNU GPL.

2.4 Plain Text

This is not a format. Technically the label "text" applies to any unformatted file intended to be human-readable. Dispite this, plain text forms the basis of every other digital standard.

[2]Organisation for the Advancement of Structured Information Standards
[3]LaTeX Project Public Licence

3 Images

Apart from Graphics Interchange Format (GIF), image formats have been relatively uncontraversial. Gif images used to be encumbered by the patent on the compression algorithm, which has recently expired. Despite this, gif images have been widely used, illegally, on the internet.

3.1 JPEG 2000 (JP2)

[Open, Prop. ISO/IEC 15444-1]

Joint Photographic Experts Group wavelets compression format for photographs. Not to be confuse with ISO 10918-1 (JPEG). JP2 and JPG files are useful for compressing photographs but are known for adding artifacts making them awkward for graphs, line-images, and scientific work.

3.2 Portable Network Graphics (PNG)

[Open, FOSS, ISO/IEC 15948]

This is a raster (bitmapped) image format that employs lossless data compression. PNG (pronounced "ping") was created to improve upon and replace the GIF format, as an image-file format not requiring a patent license. Used for small images and where artifacts are undesirable.

The only compelling reason for gif images now, is for producing animated graphics.

3.3 Scalable Vector Graphics (SVG)

[Open, Prop. n/a]

SVG is an XML specification and file format for describing two-dimensional vector graphics, both static and animated. It is an open standard created by the World Wide Web Consortium's SVG Working Group.

Vector Graphics describe a picture in terms of algorithms/objects. This is distinct from raster graphics which describe images in terms of a pixel map. SVG format is used for simple, scaleable drawings, like icons.

3.4 Bitmap

Not a file format. The term bitmap comes from the computer programming terminology, meaning just a spatially mapped array of bits. Now, along with pixmap, it commonly refers to the similar concept of a spatially mapped array of pixels.

MS Bitmap format (BMP) is an old, very simple, method of encoding pixel-map images.

4 Multi-Media

Multimedia formats are another focus of debate - particularly concerning DRM and HDCP "copy protection" methods. These "protections" have come under fire for, amongst other things, requiring implimenters to violate the laws of physics and infringing on the "fair use" laws in many countries.

4.1 Ogg (OGG)

Ogg is a free, open standard container format maintained by the Xiph.Org Foundation. The Ogg format is unrestricted by software patents and is designed to provide for efficient streaming and manipulation of high quality digital multimedia.

The name Ogg refers to the file format which can multiplex a number of separate independent free and open source codecs for audio, video, text (such as subtitles), and metadata.

Typically contains free/open format Vorbis (audio), Theora (video), and Speex (speach) files. May also contain FLAC lossless audio.

4.2 Free Lossless Audio Codec (FLAC)

FLAC is a file format for audio data compression. Being a lossless compression format, FLAC does not remove information from the audio stream, as lossy compression formats such as MP3, AAC, and Vorbis do.

4.3 Waveform Audio Format (WAV)

An IBM/MS format, it is still a commonly used, lossless, file type, suitable for retaining files of high quality, or use on a system where high fidelity sound is required and disk space is not restricted.

4.4 Motion Picture Entertainment Group (MPEG)

This is a non-free, non-open (ISO) format collection which has been agressively reverse engineered over the years. Playback, even for trivial uses, technically requires a licence fee. The mpeg decoders are encumbered by a number of different patents creating a legal nightmare for implimentation sufficient to deter the majority of for-profit GNU/Linux distributors.

5 Intellectual Property

Intellectual Property is a blanket term used to cover four, quite different, legal concepts involving ownership as it applies to intagible products such as algorithms and production methods ... and software.

These things are; Patents, Copyright, Trade Marks, and Trade Secrets.

Disclosure: I am not a lawyer. What follows is a laymans "in a nutshell" description.

5.1 Copyright

This extends legal protection to your *expression* of ideas. In terms of software, it restricts who can copy, modify, and distribute that expression; called a *work*.

The copyright holder may keep these rights exclusive, or share them with others. Typically in the form of a licence. Copyright protection extends your entire life plus 50 years, in NZ. In the USA and UK this is 70 years. In Mexico, 100.

5.2 Patents

These are designed to protect the *application* of ideas, typically in the form of a product or a process, and last for 20 years in most countries. Key aspec: right to exclude others from making, using, or selling that application.

It is possible to hold a copyright but still infringe on a patent.

Patent process is lengthy and secretive. As a result it is difficult to know if you are infringing or not.

5.3 Trademarks

These are a source identifier. Adjectives. They are always used in conjunction with the thing it applies to.

They can be perpetual. But they are prone to becoming a generic name for the product.

5.4 Trade Secrets

These have no formal protection in NZ law. Companies protect their trade secrets with non-disclosure agreements, and security. Once a trade secret is out, it's gone.

In software, the trade secret is usually the source code.

6 Licenses

By default, in New Zealand, all expressions are covered by copyright with all rights reserved to the author of the work. Patents have to be applied for. Rights are granted by licenses.

Licences exist on a continuum depending on the rights granted or taken away (see figure).

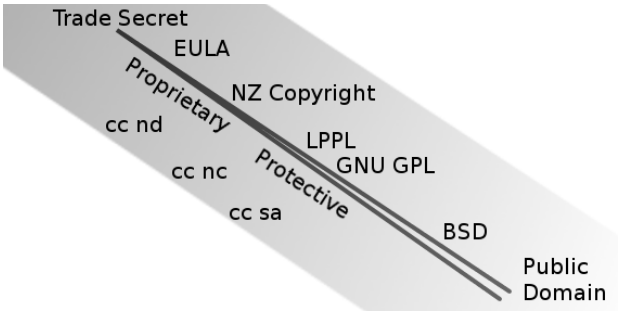

Figure 1: Licence Continuum

Most proprietary licenses take away rights you enjoy by default, and impose additional obligations, this is why you have to agree to them. They are a contract. Open Licences grant you rights, not normally available to you in law, so you don't have to agree to them. But it is important that you know about them.

Sources

- Wikipedia http://en.wikipedia.org/wiki/Open_format
- Free Software Foundation http://gnu.org
- IPONZ http://www.iponz.govt.nz
- World Wide Web Consortium http://www.w3c.org