



Why free/open-source EBMT?

Mikel L. Forcada, CNGL, DCU, Dublin, Ireland

3rd Workshop on EBMT, November 12, 2009

A lesson from Moses (and Giza++, etc.)?

- “Is the success of SMT due to the fact that it is the best way to do corpus-based machine translation or is it because many SMT software packages are readily available to researchers under free/open-source licences that allow use as well as collaborative improvement?”
- “Shouldn't EBMT practitioners start to think about putting together their tools, their engines and their data and releasing them under open licences to extend their use both in academia and in industry?”

[EBMT3 Call for Papers]

Layout

- What is free/open-source software?
- What is copyleft?
- The benefits of copyleft
- Creating a platform
- Benefits for research
- Doing business with free/open-source software

What is free/open-source software? /1

Free software is also called **open-source** software:

1. anyone can use it for any purpose
2. anyone can examine it to see how it works and modify it for any new purpose
3. anyone can freely distribute it
4. anyone may release an improved version so that everyone benefits

For conditions 1 and 3 to be met, anyone should be able to access the source code, hence the name **open source**.

Definitions:

- <http://gnu.org/philosophy/free-sw.html>
- <http://opensource.org/docs/osd>

What is free/open-source software? /2

It is important to know:

- Free/open-source licences rest on copyright law and existing intellectual property law to work
- Therefore, they legally protect the rights of authors, copyright holders, and intellectual property owners under the particular terms chosen.

What is copyleft?

- Obviously a play on the word **copyright**.
- Copyleft, when added to a free/open-source licence, means that modifications have to be distributed with the same (copylefted) licence.

Copylefted and non-copylefted licenses

Free/open-source software and contents can be **copylefted** using, for instance:

- the GNU General Public License (<http://www.gnu.org/copyleft/gpl.html>),
- the Creative Commons Attribution-Sharealike licence, version 3.0

or not **copylefted** using, for instance:

- The “three-clause” or “Simplified” BSD (“Berkeley Software Distribution”) licence,
- the MIT Licence,
- the Apache Software Licence
- the Creative Commons Attribution licence, version 3.0

The benefits of copyleft

- Copyleft secures the existence of a software or knowledge **commons**.
- It **protects** the commons from **private appropriation** (incorporation into non-free/closed-source software) by third parties,
- enables **communities of programmers** (also researchers) to build shared bodies of free/open-source software and resources,
- ... by requiring that all derivative work is always distributed under the same licence

Creating a platform

[Suggested reading: Behlendorf (1999) “Open-source as a business strategy”]

- A set of free/open-source software can be organised as a free/open-source **platform**.
- By establishing standardised procedures, interfaces and formats to exploit the components of the platform.
- It attracts external users and developers and helps creating a community.
- A modular design allows for the creation of free/open-source as well as proprietary or “closed” extensions.

Free/open-source software benefits research/1

Suggested reading: Pedersen (2008) “Empiricism Is Not a Matter of Faith”

- Free/open-source software ensures radical **reproducibility** of empirical research results
- The community of researchers can debug, improve, and extend the software
- Users and developers may help documenting the software
- Lowers the bar for entry to the project for students and colleagues
- Makes it easier to retake abandoned research later
- Projects can survive their original authors
- Clear licences protect the rights of authors

Free/open-source software benefits research/2

Benefits to researchers (through improved “research metrics”)

- Extended usage of the freed/open-sourced software improves the **visibility** of the involved researchers (citations, promotion, media coverage...)
- It stimulates the involvement of actors from the whole **technology transfer** continuum (from basic researchers to developers, vendors, final users).

Doing business with open-source/free software/1

Free/open-source software induces changes in business models:

- from **licence-centered** models to **service-centered** models
- from **vendor lock-in** to **technological partnership** (if vendor goes out of business, other companies can maintain the code and provide services)

Doing business with open-source/free software/2

- Note that third parties may engage in business without permission from the original authors (limited vulnerability, but developers retaining substantial competitive advantage — deeper know-how):
- Copylefted:
 - Installing, configuring, etc. for a customer.
 - Adapting the code (a price break may be offered in exchange for the right to distribute the modifications).
 - Distributing the software and charging for the distribution cost.
- Non-copylefted:
 - All of the above, plus:
 - Creating closed-source software (back to licence-centered models)

Doing business with open-source/free software/3

A hybrid model: dual licensing

- The copyright owner can always exploit the code using a closed licence and sell licences to customers
- “Dual licensing” a common practice, with some caveats (Behlendorf 1999)

Why free/open-source EBMT? A summary

Using free/open-source licenses for EBMT platforms will benefit the field:

- Increased visibility of EBMT research
- Easier collaboration through the creation of an EBMT commons
- Reproducibility of EBMT research
- A boost to EBMT technology transfer
- Protection from private appropriation (if copylefted licences chosen)
- A wide range of business activities possible