The Open Source Paradigm and the Production of Scientific Research:

A Future Vision and Implications for Developing Countries

Charles M. Schweik

Department of Natural Resources Conservation and
Center for Public Policy and Administration
University of Massachusetts, Amherst

Introduction

- **Challenges in communication of scientific research**
  - Space limitations in paper media
  - Library resources lacking
  - Challenge of repeatability
  - Distributed and uncoordinated research efforts
- **The technology-innovation lag (Drucker, 1999)**

A Future Vision?

A “next-generation E-journal”, that combines:

1. The Web as an interactive platform for all stages of the research process;

2. Peer-review for quality control and an incentive for participation;

3. Principles of Open Source Programming as a model for global collective action

Presentation Outline

- **Design Principles of Open Source Projects**
- **Why OS may be a New Paradigm for Global Collaborative Research**
- **The Emergence of Open Content Experiments**
- **An Example: Research in Landuse/Landcover Change Modeling**
- **Conclusions**

Principles of “Open Source” Software Licensing

- Free distribution of the software
- Readable source code
- Improvements fall under the same license
- Past authors’ contributions documented
- High profile success stories: Linux Operating System, Apache Web Server – very complex software

Design Principles of Open Source Programming Projects

- Internet-based collaboration
- Volunteer programmers/testers
  - (in some cases) organizations pay employees to participate
- Modularity and Parallel development
- Peer review
- Incremental release schedules

OS Project Lifecycle

- Initiation phase
  - Individual or small group with a “critical need” or vision
  - “Kernel” with “plausible promise”:
    - Need highly skilled or (ideally) people prominent in their field
  - Modular design
  - Decision to embrace OS license
  - Communication systems (email lists, etc.)
  - Version Control Systems (CVS, Subversion)
  - Project Governance – hierarchies, rules, norms

OS Project Lifecycle: Growth, Stability or Decline?

- The goal is growth in participation
  - With more eyes, difficult problems are easier to solve
  - OS Enterprise (e.g., Linux, Apache)
    - Large network of developers
    - Regional coordinators, many languages
    - Complex systems of coordination and core staff established

Incentives to Participate

(Sources: Hars and Ou, 2002; Feller and Fitzgerald, 2002; Lakhani et al. 2002; Bonaccorsi and Rossi, 2003; Hertel, Niedner and Herrman, 2003)

Intrinsic motivations	- Enjoyment, self-esteem

Incentives to Participate

(Sources: Hars and Ou, 2002; Feller and Fitzgerald, 2002; Lakhani et al. 2002; Bonaccorsi and Rossi, 2003; Hertel, Niedner and Herrman, 2003)

Intrinsic motivations	- Enjoyment, self-esteem
Altruism or social/political motivations	- “Software should be free!”

Incentives to Participate

(Sources: Hars and Ou, 2002; Feller and Fitzgerald, 2002; Lakhani et al. 2002; Bonaccorsi and Rossi, 2003; Hertel, Niedner and Herrman, 2003)

Intrinsic motivations	- Enjoyment, self-esteem
Altruism or social/political motivations	- “Software should be free!”
To build skills	- OS as distance learning; Peer-review

Incentives to Participate

(Sources: Hars and Ou, 2002; Feller and Fitzgerald, 2002; Lakhani et al. 2002; Bonaccorsi and Rossi, 2003; Hertel, Niedner and Herrman, 2003)

Intrinsic motivations	- Enjoyment, self-esteem
Altruism or social/political motivations	- “Software should be free!”
To build skills	- OS as distance learning; Peer-review
Self-promotion	- Become known

Incentives to Participate

(Sources: Hars and Ou, 2002; Feller and Fitzgerald, 2002; Lakhani et al. 2002; Bonaccorsi and Rossi, 2003; Hertel, Niedner and Herrman, 2003)

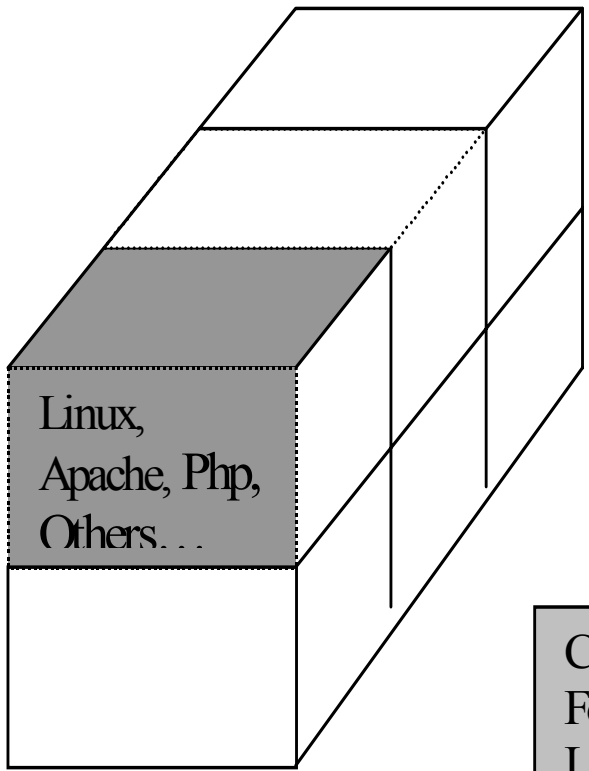
Intrinsic motivations	- Enjoyment, self-esteem
Altruism or social/political motivations	- “Software should be free!”
To build skills	- OS as distance learning; Peer-review
Self-promotion	- Become known
Personal needs	- A software gap - A complex problem (can’t do it alone)

A Problem :
Most literature to date focuses on a small number of successful “enterprise” projects.

STAGES: 1.Initiation

2.Going Open

3.Growth



OUTCOMES:

Success

Failure

Current Focus in Literature: “Enterprise Projects”

TYPE OF CASE

Why an Open Source Approach Might be a New Paradigm for Global Collaborative Research

- The entire research process and products are shared
- Incremental publishing – faster communication of new findings
- Possibly reaching a larger (global) audience? -- because of the Internet and open licensing
- Increased speed of innovations? … **why**?

The Internet as "Innovation Commons" Lessig's *The Future of Ideas* (2001)

- Example: Web exponential growth from 1994-1999? Two reasons
  - Internet end-to-end (e2e) design
  - Inclusion of the "view source" function in web browsers
  - The Web is arguably the largest and most successful distance learning program in history
- Alternatively, proprietary software blocks contributions from end-users and outside developers (Hars and Ou, 2002)

The Emergence of "Open Content" Licenses

- Extending OS software license principles to digital content
- Creativecommons.org
- 11 license variants
- Examples:
  - "By Attribution" – Others can copy, distribute, display and perform copyrighted work – and produce new derivatives from it – but only if credit is given to the original author
  - "No derivative works" – People can copy, distribute, display, and perform the work verbatim, but cannot derive new works from it.

The Emergence of Open-Content Licensing "Experiments"

- **Books ("eBooks") in the public domain**
  - Project Gutenberg (http://promo.net/pg)
- **Encyclopedias**
  - Nupedia (www.nupedia.com)
  - Wikipedia (http://www.wikipedia.com)
- **Legal arguments**
  - Openlaw (http://eon.law.harvard.edu/openlaw; legal arguments)
- **Music**
  - OpenMusic Registry (www.openmusicregistry.org)
- **Instructional material**
  - OpenCourseWare (http://web.mit.edu/ocw)
  - World lecture hall (http://www.utexas.edu/world/lecture)
  - Open Content for Education (http://www.life-open-content.org)
- **Peer-reviewed Scientific Research**
  - Public Library of Science (www.publiclibraryofscience.org)

An Example: Open-Content Research on Landuse/ Landcover Change Modeling

- A complex problem
- An issue of global interest and importance
- Connections to Global Change research program
- Many possible participants (academics, scientists, policy analysts, local/regional governments)

The Traditional Approach

- Over 20 models exist (EPA, 2000; Agrawal, 2003)
- Many quite complex
- Utilize different modeling technologies and approaches
- High transaction costs to learn/apply limits model use

An Open Content Approach: Step 1: Identify a core group of willing project participants

- Modelers
- Data Providers
- Scientists (academic, other professionals)
- Practitioners/other stakeholders

Step 2: Develop Kernel(s) with Plausible Promise

- Kernels
  - Models (open source license)
  - Theories (open content license)
  - Data (open content license)

- All kernels need to be modular to support parallel development

Step 3: Consider Incentives to Participate

Researcher (not publisher) incentives are VERY similar to OS programmers!

Intrinsic motivations	- Enjoyment, self-esteem
Altruism or social/political motivations	- Trying to solve a problem; or “Knowledge should be free!”
To build skills	- Distance learning; Peer-review
Self-promotion	- Become known
Personal needs	- A knowledge gap - A complex problem (can’t do it alone)

Step 3: Incentives continued…

- Scientists/academics may be paid to participate if in their area of research
- The challenge of (especially) junior researchers
  - Protect intellectual property for tenure
  - Why a peer-reviewed e-Journal is necessary – publishing incentive
  - Crucial – to figure out how to document intellectual property contributions
  - Can all components of the research process be treated as a form of publishing (or service)?

Step 4: Establish Systems of Governance

- Project governance
  - Team structure/responsibilities
  - Rules of operation
  - Conflict resolution mechanisms
- Select appropriate open content licenses
  - Are different licenses for models, theories and data kernels?
- System of peer-review for all three kernel types (data, models, theory and results)

Step 5: Establish Project Infrastructure

- Components of a next generation e-Journal:
    - Communication systems
    - Version control systems
        - Data - Metadata and data server
        - Models – Metadata and versioning system
        - Theory, empirical research and results
            - » Similar to today's e-journals
            - » Volumes, Issues, but incremental releases
        - Hyperlinks between kernels
    - Distance learning component
    - Special incremental issues on a particular direction
        - Last as long as there is research excitement
        - E.g., "markets" of models

Our Current Strategy

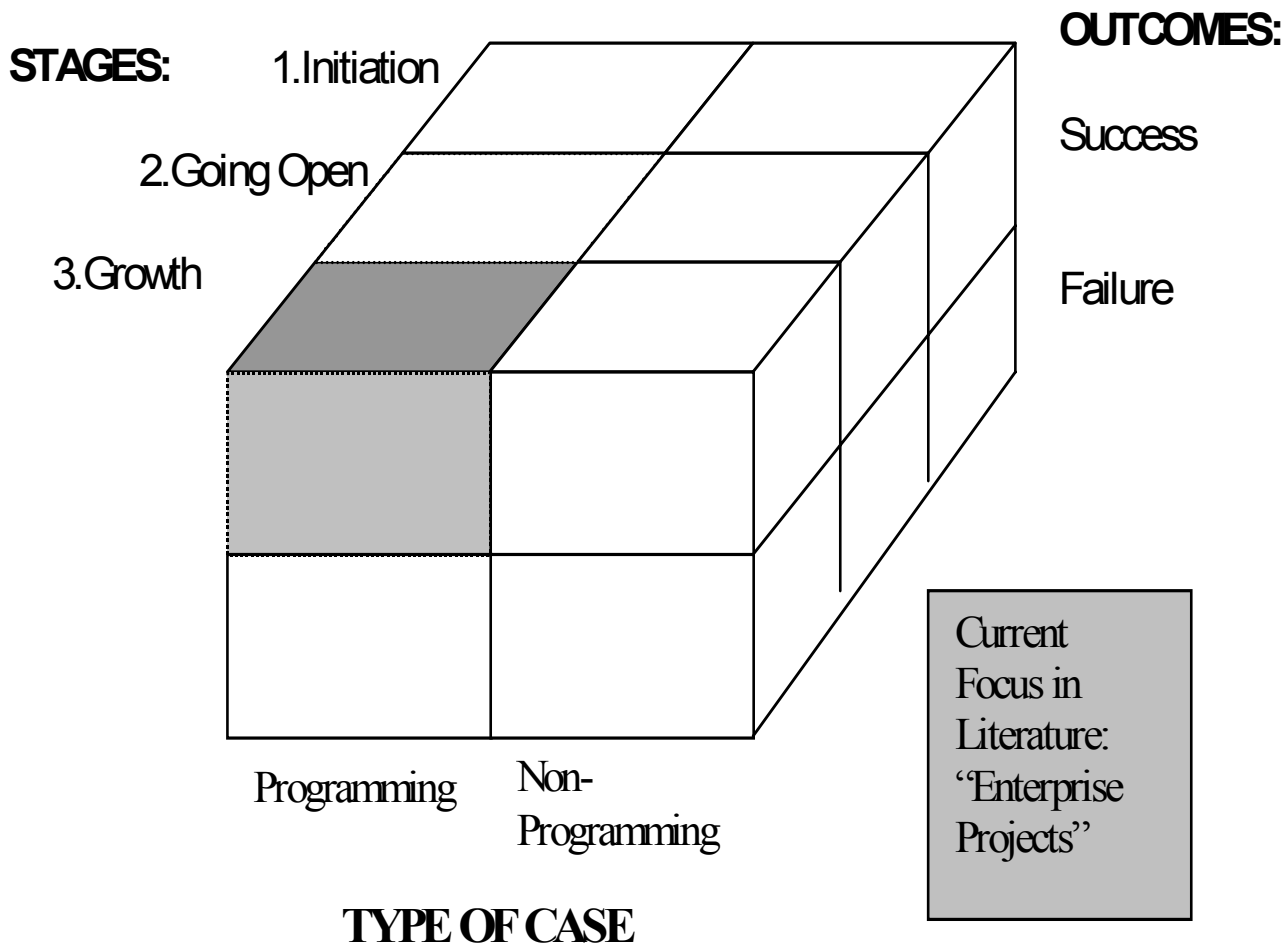
- Establish core group
  - US NSF Long Term Ecological Research Network
  - US NSF Human Dimensions of Global Change Network
  - USDA Forest Service

- Select 1 or 2 available models with plausible promise

- Establish
  - Communication systems
  - Version control systems
    - Data - Metadata and data server
      - » Open Research System
    - Models – Metadata and Versioning
      - » Concurrent Versioning System, Subversion, U of Vermont
- Look for additional funding to support effort
- Eventually move to e-Journal concept or partner with existing and relevant e-journal to publish theory and empirical findings

CONCLUSIONS

1. Need More Research on Factors that Lead to Successful or Failed Open-Content Projects



2. Need To Encourage the Development of Next Generation E-Journals

- Similar to e-Commerce and e-Government movements
- E-journals will not be cheap – but has great potential for global collaboration and innovation
- Question of who pays if no subscription
  - Pay as you can afford model?
  - Role of NGOs and Governments to support e-journal infrastructure?

3. Some e-Journal Design Issues

- Open-content derivatives and plagiarism
  - Careful historical records of people's contributions (Dr. Cowan's talk)
  - More research on OS programming needed
- Data and models as publications?
  - Systems for quantifying importance of contributions
- Design for low bandwidth (allow for quick downloads and log off)
- Careful attention to content delivery (keynote address)

4. Implications for Developing Countries

- Depends on Internet access capabilities – e.g., Nepal
- Open approach could create new opportunities for contributions and learning
  - Example – Agent-based Indiana/Nepal farmer models
  - Allows for differing interpretations (Prof. Avgerou's talk)
- LULC change modeling just one example… ask about sharing of environmental institutions

Published papers can be found at:

- *Public Administration and Management: An Interactive Journal* (2000) - www.pamij.org

- *First Monday* (Jan 2003) – www.firstmonday.org

Collaborators

J. Morgan Grove
USDA Forest Service

Tom Evans
Indiana University

Research Sponsors:

USDA Forest Service Global Change Program
USDA Northeastern Forest Research Station
USDA, Massachusetts Experiment Station
Dept of Natural Resources Conservation, UMass