INFO5992 Week 5 Cheatsheet

Distributed Innovation: Crowdsourcing, Open Data, FOSS

1. CROWDSOURCING

Definition (Estellés-Arolas & González-Ladrón-de-Guevara, 2012)

A type of participative online activity where an individual/organization proposes to a group of varying knowledge and number, via a flexible open call, the voluntary undertaking of a task that entails **mutual benefit**.

Key Elements:
- Open call format
- Wide network of potential contributors
- Variable complexity and modularity
- Mutual benefit (users + organization)

The 8-Step Crowdsourcing Process

1. Company has a problem
2. Company broadcasts problem online
3. Online "crowd" is asked to give solutions
4. Crowd submits solutions
5. Crowd vets solutions
6. Company rewards winning solvers
7. Company owns winning solutions
8. Company profits

Types of Crowdsourcing (Brabham, 2011)

Type	How it Works	Kinds of Problems	Examples
Knowledge Discovery & Management	Organization tasks crowd with finding and collecting information	Information gathering, organization, reporting problems	Peer-to-Patent, SeeClickFix
Broadcast Search	Organization tasks crowd with solving empirical problems	Ideation problems with empirically provable solutions, scientific problems	InnoCentive, Goldcorp Challenge
Peer-Vetted Creative Production	Organization tasks crowd with creating and selecting creative ideas	Ideation problems where solutions are matters of taste/market support, design/aesthetic problems	Threadless, Doritos Crash the Super Bowl Contest
Distributed Human Intelligence Tasking	Organization tasks crowd with analyzing large amounts of information	Large-scale data analysis where human intelligence is more efficient/effective than computer analysis	Amazon Mechanical Turk, Subvert and Profit

Why People Engage in Crowdsourcing

- Earn money
- Develop creative skills
- Network with creative professionals
- Build portfolio for future employment
- Challenge oneself
- Socialize and make friends
- Pass time when bored
- Contribute to large project of common interest
- Share with others
- Have fun

Crowdsourcing for Innovation - Another Classification

Intermediary Platforms:
- Research & Development (e.g., InnoCentive, NineSigma)
- Marketing, Design & Idea (e.g., 99designs)
- Collective Intelligence & Prediction (e.g., Kaggle)
- HR and Freelancers (e.g., TopCoder, Amazon Mechanical Turk)

Creative Co-creation:
- Examples: Threadless, CreateMyTattoo.com

Corporate Initiatives:
- Product ideas crowdsourcing (e.g., IBM InnovationJam)
- Branding and Design crowdsourcing (e.g., Fluevog)

Peer Production:
- Examples: Linux, Wikipedia

Public Crowdsourcing:
- Examples: Foldit (protein folding game), NASA Open Innovation

Crowdfunding

Crowdsourcing specifically for sourcing **funds** from a crowd.

Popular Platforms:
- GoFundMe
- Indiegogo
- Kickstarter
- Fundly
- JustGiving

Challenges in Crowdsourcing (Why Crowdsourcing Fails)

- Poor task design
- Inadequate incentives
- Lack of quality control
- Insufficient contributor engagement
- Unclear goals/instructions
- Privacy and confidentiality concerns
- Coordination problems
- Need for evaluation mechanisms

Management Strategies

- Test and filter qualified contributors
- Anonymize/obscure sensitive data
- Carefully outline processes
- Evaluate submitted answers
- Provide clear instructions
- Offer incentives
- Monitor progress and provide feedback
- Use quality control measures
- Build community support
- Establish technical standards

---

2. OPEN DATA

Definition (Open Definition)

"Open data and content can be freely used, modified, and shared by anyone for any purpose"

Open means anyone can freely access, use, modify, and share for any purpose (subject, at most, to requirements that preserve provenance and openness).

Value of Open Data

- **Economic Value:** Increased efficiency, new products/services, consumer surplus (cost savings, convenience, better products)
- **Big Data Impact:** Replacing/supporting human decision making
- **Business Opportunities:** New products and services

- **Government's Central Role:** Major provider of open data

Estimated Value: $3-5 trillion in economic value annually across seven sectors

Self-Reinforcing Cycle

Benefits increase as individuals perceive advantages and help improve accuracy/detail of information → Requires:
- Private industries and public agencies to cultivate vibrant ecosystem
- Policies to protect stakeholders
- Technologies and talent to collect/analyze data
- Vigilant, savvy providers and users

Examples

Government Open Data:
- Australia: data.gov.au
- NSW: data.nsw.gov.au
- Includes: electoral boundaries, crime data, census data, bus stop data, electricity consumption, pollution, education data

Scientific Data:
- Nature Scientific Data journal
- Computer Vision/Medical Imaging datasets (Papers With Code)
- Osteoarthritis Initiative (OAI)

Community-Built:
- OpenStreetMap
- OpenAddresses

NASA:
- EOSDIS Worldview
- Open Innovation challenges

---

3. FREE AND OPEN SOURCE SOFTWARE (FOSS)

History: Richard Stallman & GNU (1980s)

- Stallman was refused access to Xerox laser printer source code (1980)
- Convinced people should be free to modify all software
- Founded Free Software Foundation (FSF) in 1985
- Created GNU project: "GNU's Not Unix" (recursive acronym)
- Philosophy: "technical means to a social end"

Stallman's Argument:
- Software is critical for running a free society
- If companies/governments control software, it can restrict/monitor people
- Source code must be available for all software
- Without free source code, few powerful people would dominate computing

Free Software (FSF Definition)

"Free" as in "free speech" NOT "free beer"

Four Essential Freedoms:
0. Freedom to run the program for any purpose

1. Freedom to study how program works and change it (requires source code access)
2. Freedom to redistribute copies to help others
3. Freedom to distribute copies of modified versions to others (requires source code access)

Open Source Software (Eric Raymond, 1998)

Why "Open Source" term was created:
- "Free software" was ambiguous (confused with freeware/shareware)
- Intimidating to companies (ideological approach)
- Focused on **pragmatic aspects** rather than "rights"

Key Principle: "Given enough eyeballs, all bugs are shallow"

Open Source Definition (OSI)

Software must be:
- Freely redistributable
- Source code available for free or reasonable reproduction cost
- Modifications and derived works allowed under same terms
- Can protect integrity of author's source code (allow patches)
- No discrimination against people/groups
- No discrimination against fields of endeavour
- Not restricted to specific products
- Not place restrictions on other software
- Technology-neutral

Free Software vs Open Source

Aspect	Free Software	Open Source
Philosophy	Social movement	Development methodology
Focus	User rights/freedoms	Pragmatic benefits
License Range	Narrower (more restrictive)	Wider (more permissive)
Mixing with Proprietary	Less ability	More ability

When not distinguishing: Use FOSS (Free and Open Source Software) or FLOSS (Free/Libre and Open Source Software)

Examples of FOSS

OSS with Copyleft:
- Linux kernel
- MariaDB
- Eucalyptus

OSS without Copyleft:
- Apache web server
- OpenCV
- Chromium

Other Examples:
- Android, Chrome, Firefox
- Alfresco, Hadoop, Tomcat
- MySQL, OpenOffice
- Node.js, OpenStack

FOSS Licenses

Spectrum from Permissive to Restrictive:

```
Public Domain → MIT → BSD → Apache → GPLv2 → GPLv3 → AGPL
[Least restrictive]                              [Most restrictive/copyleft]
```

Public Domain

- No intellectual property rights
- Examples: Shakespeare's works, Beethoven's music, expired copyrights
- Rarely used for software

MIT License

- User can do anything with software
- Must maintain original copyright notice
- No warranty
- Very permissive

BSD Licenses (2-clause, 3-clause, 4-clause)

- Similar to MIT
- Must acknowledge use if redistributing
- 3-clause adds: can't use organization name to endorse products without permission
- Very permissive

Apache Software License

- Similar to BSD but with additional provisions
- Addresses patent rights
- Permissive

GNU General Public License (GPL)

Copyleft license - more restrictive
- Can use and modify code
- MUST release all modified code under same license
- Any code that "touches" it must also be GPL
- Two main versions: GPLv2 and GPLv3
- Ensures derivative works remain free

GPL v3 Key Requirements:
- Modified work must carry prominent notices
- Entire work must be licensed under GPL
- Must display Appropriate Legal Notices
- Compilation with other works ("aggregate") doesn't force GPL on other parts

AGPL (Affero GPL)

- Even more restrictive than GPL
- Designed for network/web services
- Must share source code even for software used over network

"Copyleft"

- Play on "copyright"

- Method for making program free AND requiring all modified/extended versions to be free
- Example: GNU Public License (GPL)
- Symbol: ɔ (reversed C in circle)

Copyright vs Copyleft

- **Copyright:** Legal concept providing exclusive rights to creators
- **Copyleft:** Licensing method encouraging unrestricted sharing, modification, and utilization

Choosing a License

Use Permissive License (BSD, Apache) if:
- Want many companies to adopt software
- Don't care if they make changes available
- Have deep expertise (they'll come back to you)
- Want maximum adoption

Use Restrictive License (GPLv3) if:
- Want to ensure companies make changes available
- Want to benefit from community improvements
- Want to keep software free/open

Dual-licensing:
- Offer software under both GPL and proprietary license
- Companies not wanting to share changes can negotiate proprietary license
- Common business model

Most Common Licenses (Black Duck data)

1. GNU GPL 2.0 (largest share)
2. MIT License (second largest)
3. Apache License 2.0 (third largest)
4. BSD variations
5. Others: LGPL, MPL, EPL, CDDL, etc.

Over 2,400 different licenses exist across 1.1M+ open source projects

Creative Commons

Standardized way to grant copyright permissions for creative/academic works

License Types (from least to most restrictive):
- CC BY (Attribution)
- CC BY-SA (Attribution-ShareAlike)
- CC BY-NC (Attribution-NonCommercial)
- CC BY-ND (Attribution-NoDerivatives)
- CC BY-NC-SA (Attribution-NonCommercial-ShareAlike)
- CC BY-NC-ND (Attribution-NonCommercial-NoDerivatives)

Importance of FOSS for Innovation

Usage Statistics:
- 99% of Fortune 500 companies use open source
- 80% of IT departments increasing open source use
- 35% of enterprise software based on open source
- 56M+ developers contributing to open source

- 140M+ projects on GitHub
- 10,000 lines added to Linux daily

Why Important:
- Accelerates innovation
- Reduces development costs
- Enables rapid prototyping
- Allows customization
- Avoids vendor lock-in
- Builds on proven technology
- Community support and improvements

Usage Areas:
- Internal IT infrastructure (Linux)
- Web services (Apache, Tomcat, JBoss)
- Software development (frameworks, libraries)
- Data science (Python, R, Jupyter)
- Machine learning (TensorFlow, PyTorch)
- Containers (Docker, Kubernetes)

Open Source Business Models

1. **Sell Support and Services**

2. Example: Canonical (Ubuntu)

3. **Sell Certified Version**

4. Example: Cloudera (Hadoop)

5. **Sell "Enterprise Edition"**

6. Example: MySQL (standard vs community edition)

7. **Dual Licensing**

8. Copyleft for free, proprietary for modifications

9. Example: Qt

10. **Other Advantages**

11. Example: Google (Android ecosystem)

12. **Freemium Model**

13. Example: Elementary OS ("pay what you want")

Company Examples

Facebook Open Source:
- Accelerates world innovation
- Users build apps more quickly
- Benefits from improvements others make
- Builds better from start (knowing it will be open)
- Attracts talent interested in challenges

Google Open Source:
- "Open source makes it possible for us to work together"
- Use, Release, Support model
- Notable projects: Android, Chrome, Angular, TensorFlow, Kubernetes

Apple:
- Swift programming language (open source)

- Success proportional to number of developers
- Attract best developers for ecosystem

Microsoft:
- Porting .NET to Linux/Mac
- 2,000+ open source projects
- Azure supports Linux workloads (20%+ of VMs)
- "Fabric and cloud anywhere" vision

Open-Source Lab Model (e.g., AMPLab at UC Berkeley)

Structure:
- Multi-year collaborative effort
- Industry sponsors (Amazon, Google, IBM, SAP, etc.)
- Develops open source software
- Forms companies from research

Examples from AMPLab:
- Software: Spark, Mesos, Tachyon, GraphX, MLBase
- Companies: Mesosphere ($122M), Databricks ($47M), Tachyon Networks ($17M)

Benefits to Companies:
- Deep knowledge of technology as developed
- Influence technology direction
- Protection from disruption
- Access to talent pool

Benefits to University:
- Clear research focus
- Additional funding
- Market interaction
- Business incubation
- Stronger innovation ecosystem

Challenges Using FOSS

1. **License Compliance**

2. Meeting obligations (notices, attribution)

3. Understanding license terms

4. **Code Contamination Risk**

5. Accidentally introducing GPL code into proprietary products

6. Legal requirement to release proprietary source

7. **Quality Assurance**

8. Ensuring adequate quality of final product

9. Unknown quality of some OSS components

10. **Security Vulnerabilities**

11. Known vulnerabilities in underlying code

12. Need for timely updates

Managing FOSS Usage

Open Source Governance:
- <50% of Global 2000 IT companies had governance program (2014)
- Only 1/3 had OSS policy

Tools for Management:
- Black Duck Software
- Palamida
- FOSSology (by HP, now open source)
- OpenHub.net
- Search over 21B lines of FOSS code

4. CASE STUDY: INNOVATIONS WITH MAPS

OpenStreetMap (OSM)

What it is:
- Initiative to create and provide **free geographic data** (street maps) to anyone
- Collaborative project to create free editable map of the world
- "Wikipedia of maps"

OpenStreetMap Foundation:
- International not-for-profit organization
- Supports (but doesn't control) OSM Project
- Encourages growth, development, distribution of free geospatial data

How it works:
- Crowd sourcing + Open source + User innovation
- 2M+ users collecting data
- Methods: manual survey, GPS, aerial imagery
- Government and out-of-copyright data
- Open data license

Innovation Aspects:
- **Open Source:** Enables innovations, new services/businesses
- **Crowd Sourcing:** Over 2M users collecting data using GPS, manual survey, free sources
- **Open Data License:** Allows anyone to use and build upon

OSM-Based Innovations

KartaView (formerly OpenStreetCam):
- Street-level imagery collection
- Powered by community
- Free to access
- Helps improve OpenStreetMap

Open Source Routing Machine (OSRM):
- High-performance routing engine
- Uses OSM data
- Open source project
- Provides turn-by-turn navigation

Pokémon Go:
- Switched from Google Maps to OSM (December 2017)
- Players edit OSM to improve gameplay

- Issues: some add spurious data (fake parks, delete schools)
- Demonstrates OSM's reach and impact

Comma AI - OpenPilot:
- Open source advanced driver assistance system
- Works on 200+ car models
- Features: automated lane centering, adaptive cruise control, lane change assist
- Open source and community supported
- 90M+ miles driven, 6000+ active users, 315+ contributors

Benefits of Open Maps

- Free alternative to proprietary maps
- Community-driven improvements
- Rapid updates in crisis situations
- Customizable for specific needs
- No licensing fees
- Supports innovation and new businesses

---

5. KEY CONCEPTS SUMMARY

Distributed Innovation Approaches

Approach	Key Feature	Example
Product Platforms	Third parties build on platform	iOS App Store, Android
Web APIs	Programmatic access to services	Google Maps API, Twitter API
Crowdsourcing	Tasks distributed to crowd	Amazon MTurk, Kaggle
Open Data	Free access to datasets	data.gov.au, OpenStreetMap
FOSS	Source code freely available	Linux, Apache, Python
User Innovation	Users create and share innovations	OpenStreetMap editing
Platform Ecosystems	Multiple sides interact	Apple ecosystem, Android ecosystem

Critical Success Factors

For Crowdsourcing:
- Clear task definition
- Appropriate incentives
- Quality control mechanisms
- Community engagement
- Proper evaluation systems

For Open Data:
- Accessible formats
- Good documentation
- Regular updates
- Privacy protection
- Clear usage terms

For FOSS:
- Active community
- Good documentation
- Clear licensing
- Responsive maintainers
- Welcoming to contributors

Impact on Innovation

Accelerates Innovation:
- Parallel development by many contributors
- Diverse perspectives and approaches
- Rapid iteration and improvement
- Lower barriers to entry

Enables New Business Models:
- Services around free software
- Freemium approaches
- Ecosystem plays
- Data-driven services

Democratizes Innovation:
- Anyone can contribute
- Geographic barriers reduced
- Access to tools and data
- Learning opportunities

6. EXAM TIPS

Key Definitions to Know

- Crowdsourcing (full definition)
- Open Data
- Free Software (4 freedoms)
- Open Source Software
- Copyleft
- GPL, MIT, BSD licenses

Be Able to:

- Classify crowdsourcing examples by type
- Explain why people engage in crowdsourcing
- Compare Free Software vs Open Source philosophies
- Choose appropriate FOSS license for scenarios
- Identify FOSS business models
- Explain benefits and challenges of each approach

Important Figures

- Richard Stallman (Free Software, GNU, FSF)
- Eric Raymond (Open Source, "Cathedral and Bazaar")
- Linus Torvalds (Linux kernel)

Key Organizations

- Free Software Foundation (FSF)
- Open Source Initiative (OSI)
- OpenStreetMap Foundation
- Apache Software Foundation
- Linux Foundation

Common Exam Questions

1. Types of crowdsourcing with examples
2. Difference between Free Software and Open Source
3. FOSS license comparison and selection
4. Benefits and challenges of distributed innovation
5. Business models for open source
6. Case study analysis (like OpenStreetMap)

---

7. IMPORTANT QUOTES

Jeff Howe (2006) - Original Crowdsourcing Definition:
"Crowdsourcing represents the act of a company or institution taking a function once performed by employees and outsourcing it to an undefined (and generally large) network of people in the form of an open call."

Richard Stallman:
"Free software is a matter of liberty, not price. To understand the concept, you should think of 'free' as in 'free speech,' not as in 'free beer.'"

Eric Raymond:
"Given enough eyeballs, all bugs are shallow." (Linus's Law)

On Open Source vs Free Software:
"Open source is a development methodology; free software is a social movement." - Richard Stallman

---

8. USEFUL LINKS

Crowdsourcing Platforms:
- Amazon Mechanical Turk: mturk.com
- Kaggle: kaggle.com
- InnoCentive: innocentive.com
- 99designs: 99designs.com

Open Data:
- Australia: data.gov.au
- NSW: data.nsw.gov.au
- World Bank Open Data Toolkit

FOSS Resources:
- Free Software Foundation: fsf.org
- Open Source Initiative: opensource.org
- GNU Project: gnu.org

- GitHub: github.com
- OpenHub: openhub.net

Maps:
- OpenStreetMap: openstreetmap.org
- OSRM: project-osrm.org

License Information:
- Choose a License: choosealicense.com
- TLDRLegal: tldrlegal.com
- Creative Commons: creativecommons.org

End of Cheatsheet