

# Building AI: Our Shared Enterprise

Steven Minton  
InferLink Corp.

# Why are you here?







# What this Talk is About

- Open Science: The importance of collaboration
- How we share
- Why we share
- How we support sharing
- Can we do a better job?

# Myth of the Individual

---

How did “the Web” come about?

# Myth of the Individual

How did “the Web” come about?

**Peter Deutsch**

**Alan Emtage**

Robert Cailliau

Tim Berners-Lee

Mark McCahill

**Robert Kahn**

**Vinton Cerf**

Bob Alberti

Daniel Torrey

Farhad Ankelsaria

**MARC ANDREESSEN**

Paul Lindner

Michael Mauldin

**ERIC BINA**

**Rhett Jones**

**Brewster Kahle**

**Matthew Gray**

**Jon Postel**

Steven Foster

**Paul Mockapetris**

Fred Barrie

# Myth of the Individual

How did “the Web” come about?

Mark McCahill

Robert Cailliau

Tim Berners-Lee

Bob Alberti

Daniel Torrey

**Robert Kahn**

Vinton Cerf

Farhad Ankelsaria

**MARC ANDREESSEN**

Paul Lindner

**ERIC BINA**

Brewster Kahle

Matthew Gray

Jon Postel

Steven Foster



# Myth of the Individual

How did “the Web” come about?

Tim Berners-Lee

Robert Kahn

Vinton Cerf

MARC ANDREESSEN

ERIC BINA

Brewster Kahle

Jon Postel

# Myth of the Individual

How did “the Web” come about?

Tim Berners-Lee

**Robert Kahn**

**MARC ANDREESSEN**

# Myth of the Individual

---

How did “the Web” come about?

Tim Berners-Lee

# What is “Open Science”

- Open science = scientific knowledge of all kinds should be openly shared, with no barriers to dissemination and reuse.
- *“If I have seen further, it is by standing on the shoulders of giants”* -- Isaac Newton

# Open Science

- Education
- Publications
- Software
- Data
- Crowd-sourced science
- ...

# Open Science

- Education
- Publications
- Software
- Data
- Crowd-sourced science
- ...

# Open Access Publications

- Journals:

*J*ournal of *A*rtificial *I*ntelligence *R*esearch  
AN INTERNATIONAL ELECTRONIC AND PRINT JOURNAL

**JMLR**



**PLOS**

PUBLIC LIBRARY  
of SCIENCE

# Open Access Publications

- Journals:

*J*ournal of *A*rtificial *I*ntelligence *R*esearch  
AN INTERNATIONAL ELECTRONIC AND PRINT JOURNAL

**JMLR**



**PLOS**

PUBLIC LIBRARY  
of SCIENCE

- Preprints/archives:

**arXiv.org**

**CoRR**  
*Computing Research Repository*



# JAIR: An Open Access Journal

- Organized in 1992, began publishing in 1993
  - FTP → Gopher → Web/HTML
  - High impact factor
  - Thanks to:
    - Jaime Carbonell, Tom Dietterich, Jon Doyle, Oren Etzioni, Ken Forbus, Peter Friedland, Matt Ginsberg, Fausto Guinchiglia, Rich Korf, Pat Langley, Tom Mitchell, Mike Morgan, Paul Rosenbloom, Bart Selman, Peter Turney, Dan Weld, Mike Wellman
    - Exec Eds: Mike Wellman, Martha Pollack, Moshe Tennenholtz, Toby Walsh, Adnan Darwiche, Shlomo Zilberstein, and Craig Boutilier
    - Assoc Editors, Editorial and Advisory Board members, Authors, Reviewers
- Lesson: People are happy to contribute!

# Myth: Need to be radically new

- We didn't change *too much*
- Contributions:
  - Freely available
  - Fast response
  - Online appendices with code/data
- ...But still clearly a technical journal:
  - Traditional reviewing
  - Articles use standard format conventions
  - Hardcopy version

# Open Science

- Education
- Publications
- **Software**
- Data
- Crowd-sourced science
- ...

# Open Source Software

- Commercially-competitive systems



- AI Software Toolkits



- Many individual AI projects!

# GATE

- “A full-lifecycle open source solution for text processing”
- 15+ years, >5M funding, many users (both corporate and research)
- *A community of* developers, users, scientists,...

# Myth of the Rational Human

- Why do people contribute to open source projects?
  - "...to an economist, the behavior of individual programmers and commercial companies engaged in open source processes is startling."  
(Lerner & Tirolì, 2002)

# Motivations for Contributing

Motives to participate	Related literature
Software use value	[20, 25, 35, 49, 55, 60, 63, 65]
Status and recognition	[25, 28, 35, 47, 49, 55, 63]
Learning	[25, 26, 28, 61, 63, 65]
Personal enjoyment	[25, 28, 49, 55, 65]
Reciprocity	[25, 35, 55]
Getting paid	[25, 49]
Sense of ownership and control	[28, 39, 61]
Career advancement	[25, 35, 55, 65]
Free software ideology	[2, 9, 28, 29, 57]
Social identity	[2, 25]

(Fang & Neufeld, 2009)

## ■ Intrinsic vs. Extrinsic motivations

(Lakhani & Wolf, 2005)

# Practical Lessons

- People need to see the value
- They need to see how to contribute
- They need to see they are not being taken advantage of



# Open Science

- Education
- Publications
- Software
- Data
- Crowd-sourced science
- ...

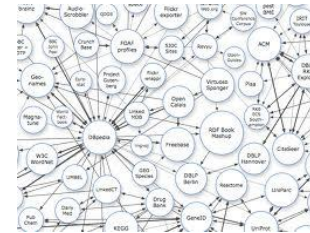


# Open Data

- Webscale Projects

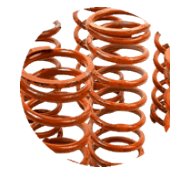


WIKIPEDIA  
*The Free Encyclopedia*



*Linked Open Data*

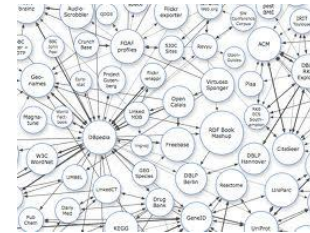
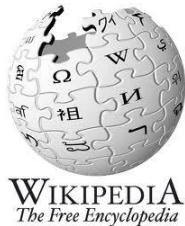
- “Free” databases from for-profit companies:



*OpenCyc*

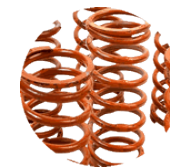
# Open Data

- Webscale Projects



*Linked Open Data*

- “Free” databases from for-profit companies:



*OpenCyc*

- Critical AI Infrastructure



# UCI's Machine Learning Repository

- Started in 1987 , rapidly became almost obligatory to use UCI data for experiments
- Founded by David Aha while *a grad student*
- Maintainers rotate over time
- Provides clear guidance on “how to cite”
- As of 2009, over 9000 citations to repository

# The Ultimatum Game

- An economics experiment



# The Ultimatum Game

- An economics experiment
- Fairness matters!



# The Ultimatum Game

- An economics experiment
- Fairness matters!
- ...uniquely human?





# Open Science

- Education
- Publications
- Software
- Data
- Crowd-sourced science
- ...

# Crowd-Sourced Science

- Classification:



# Crowd-Sourced Science

- Classification:



- Games and Competitions:



# Crowd-Sourced Science

- Classification:



- Games and Competitions:



- Knowledge/ontology creation

*Learner2*



# Crowd-Sourced Science

- Classification:



- Games and Competitions:



- Knowledge/ontology creation

*Learner2*



- Scientific collaboration



# The Polymath Project

- Tim Gower's blog, 2009
  - "Is massively collaborative mathematics possible?"
- Polymath<sub>1</sub>:
  - a new combinatorial proof of the density Hales–Jewett theorem
  - 3 months, 40 contributors
  - Result = "a simple, beautiful, combinatorial proof"
- Papers published under pseudonym "D.H.J Polymath"

# Open Science

- Education
- Publications
- Software
- Data
- Crowd-sourced science
- ....

# What Else Can We Do?

---

- Directory of Open Source AI



# What Else Can We Do?

- Directory of Open Source AI
- Crowd-sourced books & courses

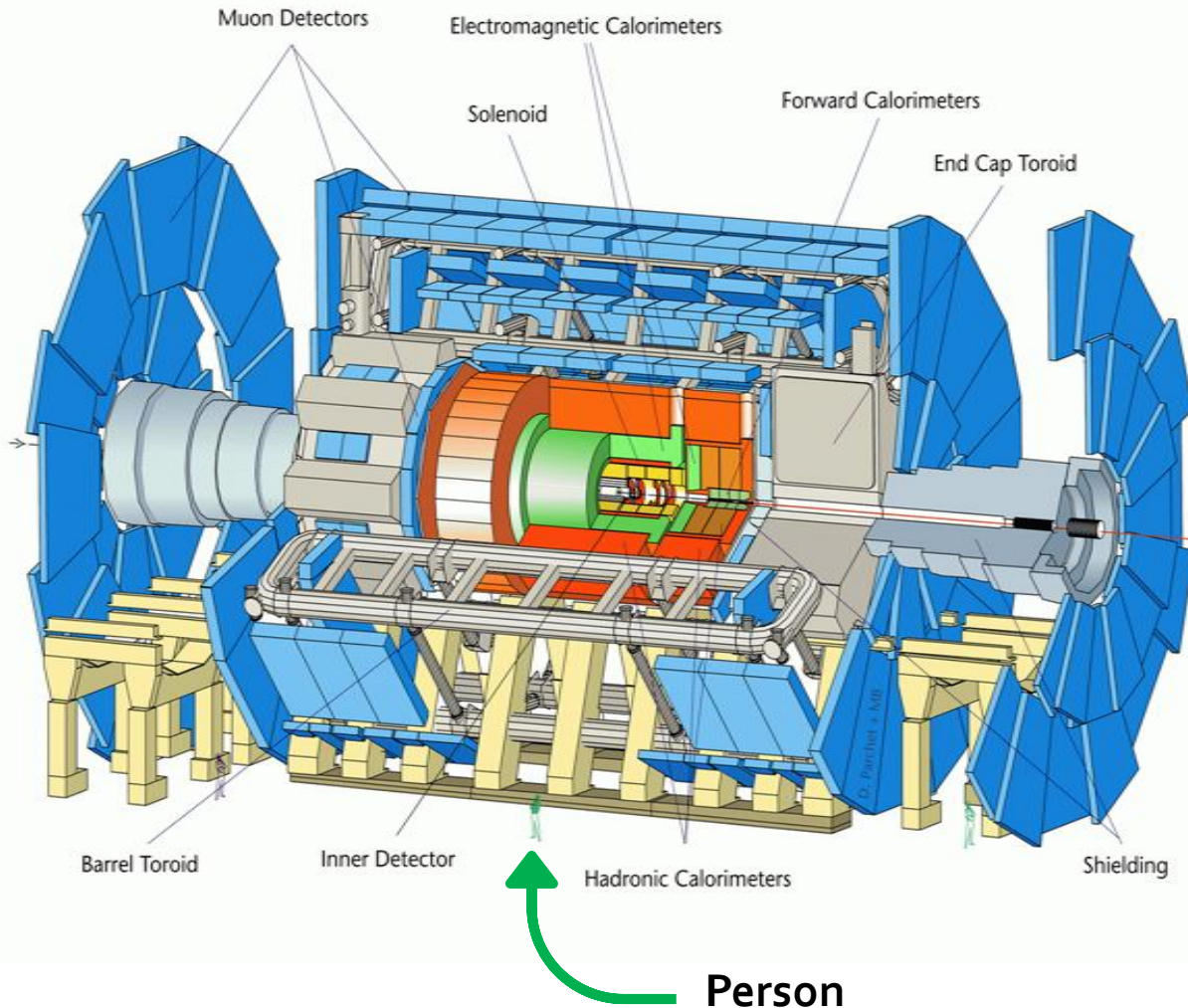
# What Else Can We Do?

- Directory of Open Source AI
- Crowd-sourced books & courses
- Virtual research groups

# What Else Can We Do?

- Directory of Open Source AI
- Crowd-sourced books & courses
- Virtual research groups
- Big AI

# Like the ATLAS experiment?



# What Else Can We Do?

- Directory of Open Source AI
- Crowd-sourced books & courses
- Virtual research groups
- Big AI

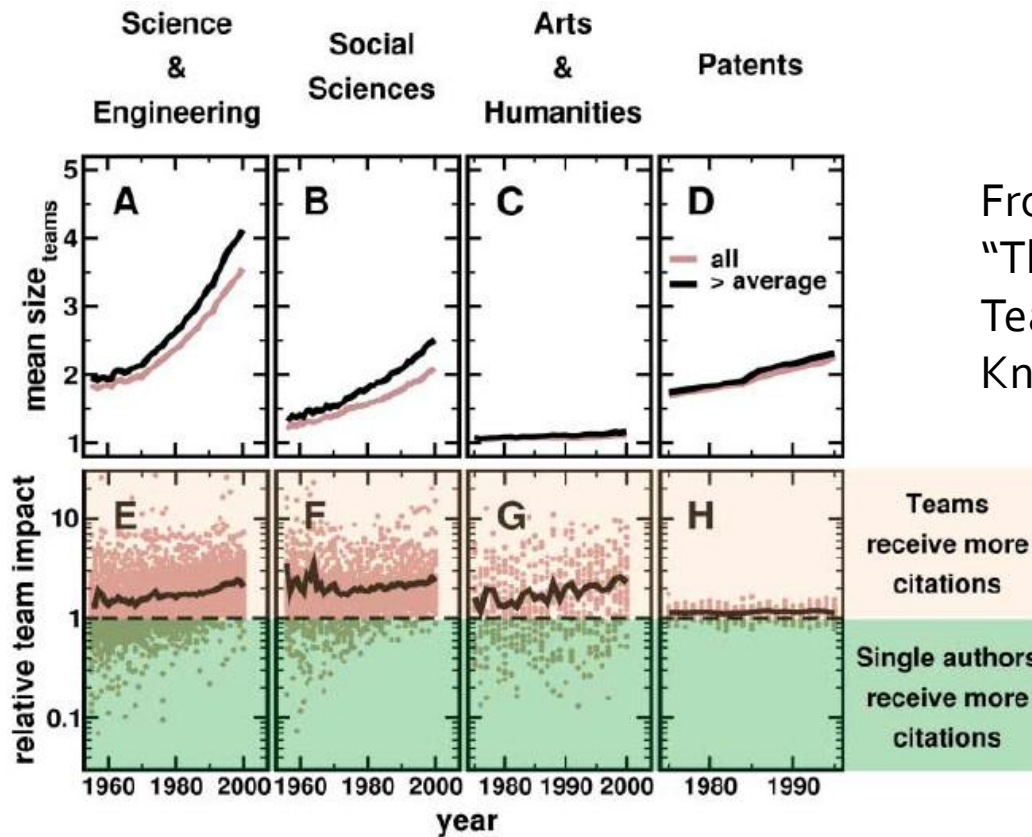
# Conclusion

- Sharing our work is of vital importance to AI
- How you can help:
  - Funders:
    - Support open source, open data
  - Senior faculty members:
    - Reward impactful contributions
  - Students and researchers:
    - Send papers to open access publications
    - Open source your code, data
    - Start your own open science project!

# Nonprofit $\neq$ no salary

- Business models:
  - Zero budget nonprofits
  - Donor-supported nonprofits
  - User-supported nonprofits
  - Social enterprises (for profit)

# Teamwork Leads to Higher Impact



From Wuchty, Jones & Uzzi, "The Increasing Dominance of Teams in Production of Knowledge", *Science*, 2007

Fig. 2. The relative impact of teams. (A to D) Mean team size comparing all papers and patents with those that received more citations than average in the relevant subfield. (E to H) The RTI, which is the mean number of citations received by team-authored work divided by the mean number of citations received by solo-authored work. A ratio of 1 indicates that team- and solo-authored work have equivalent impact on average. Each point represents the RTI for a given subfield and year, whereas the black lines present the arithmetic average in a given year.



# JAIR Submissions (1999- Present)

