National Cancer Institute

The National Cancer Informatics Program Open Development Initiative

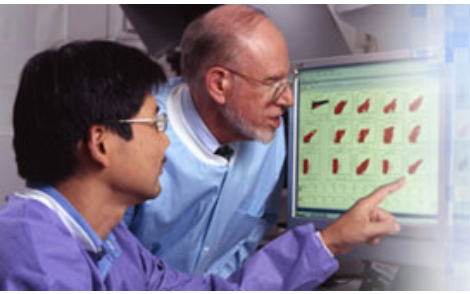
Juli Klemm, Ph.D.

Bioinformatics Open Source Conference

July 19, 2013

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES
National Institutes of Health

NCIP is a cross-NCI program to support biomedical informatics in cancer research



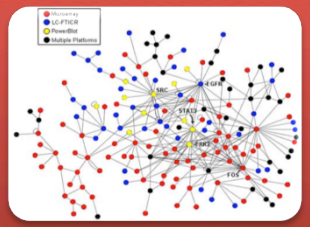
National Cancer Informatics Program

A cross-NCI program supporting biomedical informatics in cancer research

- Foster community development of informatics capabilities
- Provide access to well-described data collections
- Maintain an interoperability infrastructure and standards
- Reinforce collaborative relationships among researchers
- Train biomedical investigators to use informatics capabilities

http://cbiit.nci.nih.gov/ncip

Why is the NCI interested in open source?



Biology has become a data intensive science

- Rapidly evolving, unique needs of cancer researchers
- Big data

Sustainability

- Budget realities
- Innovation

Open science movement

- Open source, open data, open access
- Open government

The NCIP Open Development Initiative

- ***Support*** the rapid pace of innovation in cancer research
- ***Enable*** the creation of better tools by crowdsourcing innovations
- ***Empower*** the research community to govern project priorities and roadmaps
- ***Insulate*** software development from the ebbs and flows of the federal funding cycle
- ***Engage*** a broader community of talented developers to contribute to a critical public health concern

Progress to date

2011

- RFI: Models of open development for biomedical research
- Research...

2012

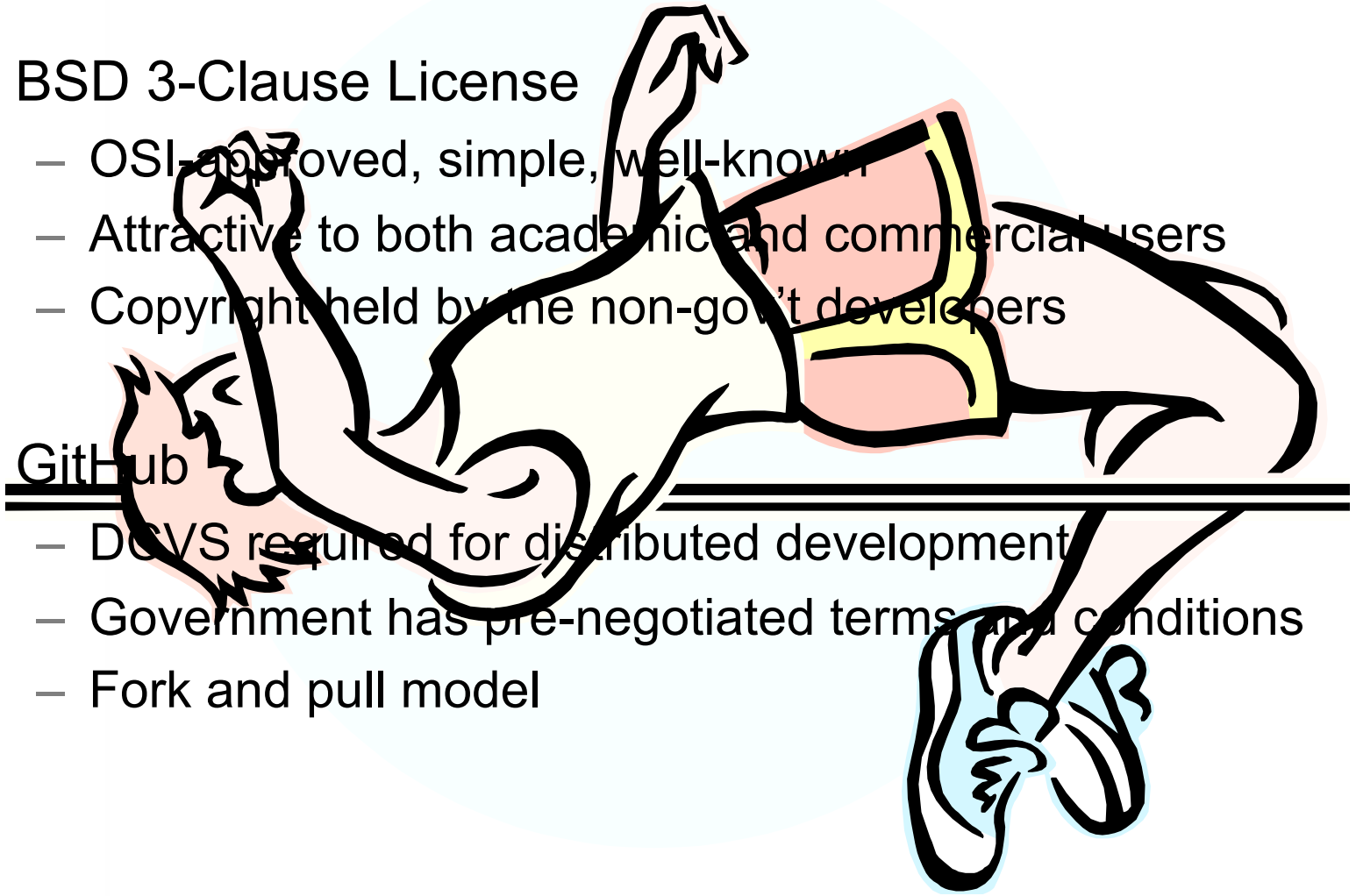
- Open source think tank
- Decision to move to BSD 3-Clause license
- NCIP GitHub Channel established https://github.com/ncip

2013

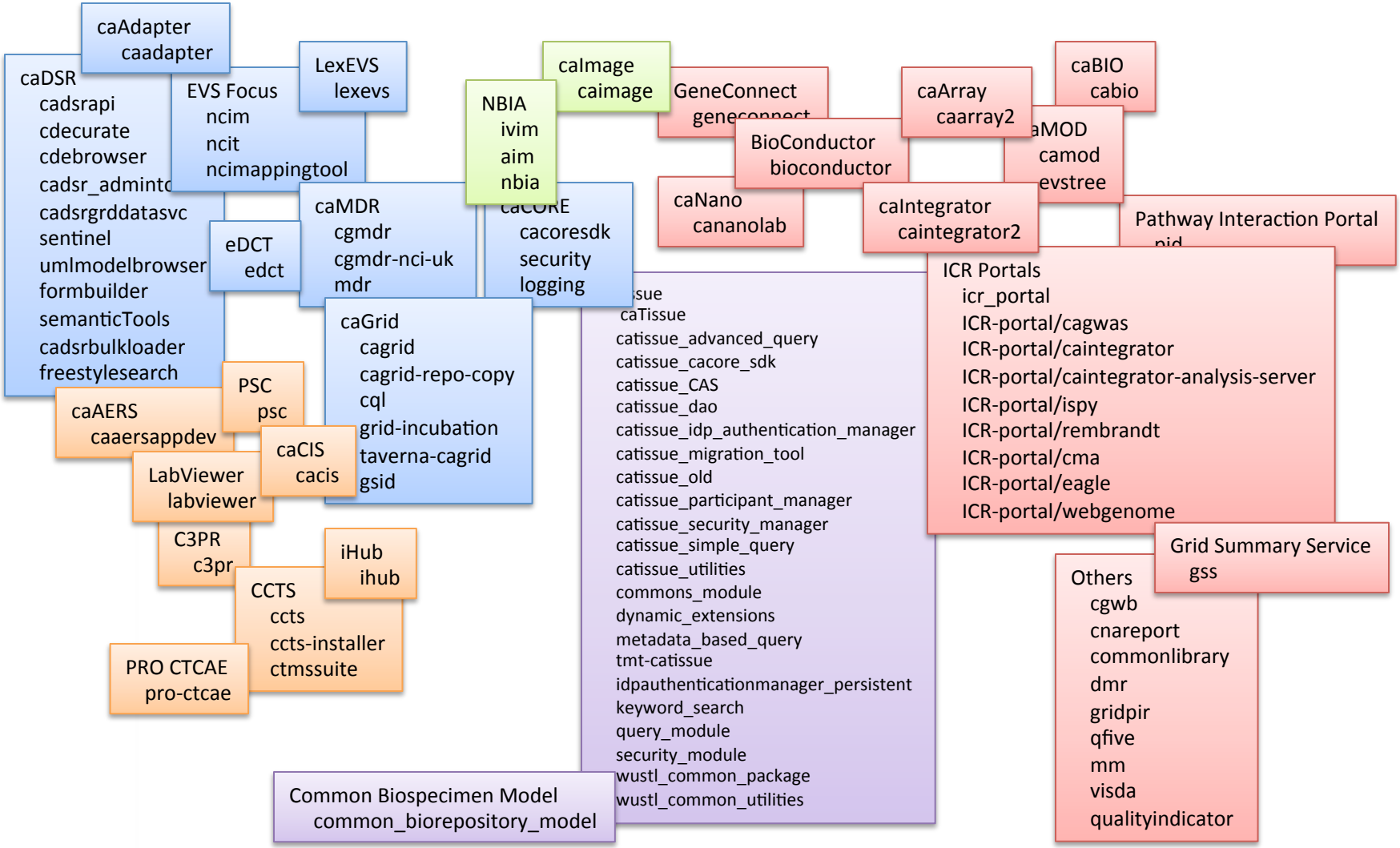
- Begin code and license migration
- Today: 112 repositories representing 47 projects migrated

Lowering the barriers: New license and new repository

- BSD 3-Clause License
  - OSI-approved, simple, well-known
  - Attractive to both academic and commercial users
  - Copyright held by the non-gov't developers
- GitHub
  - DCVS required for distributed development
  - Government has pre-negotiated terms and conditions
  - Fork and pull model



Migration Projects



Migration resources and lessons learned

- Communication, process, SOPs, templates

  https://github.com/NCIP/ncip.github.com/wiki/_pages

  - How to Migrate From SVN to GitHub
  - Templates
    - README, NOTICE, LICENSE, CONTRIBUTING
  - Access control methodology
  - Learning resources, best practices

- Repository clean-up
  - Size, content, structure
- Security considerations for building from a remote repository
  - Local mirroring

Next steps: Building Community

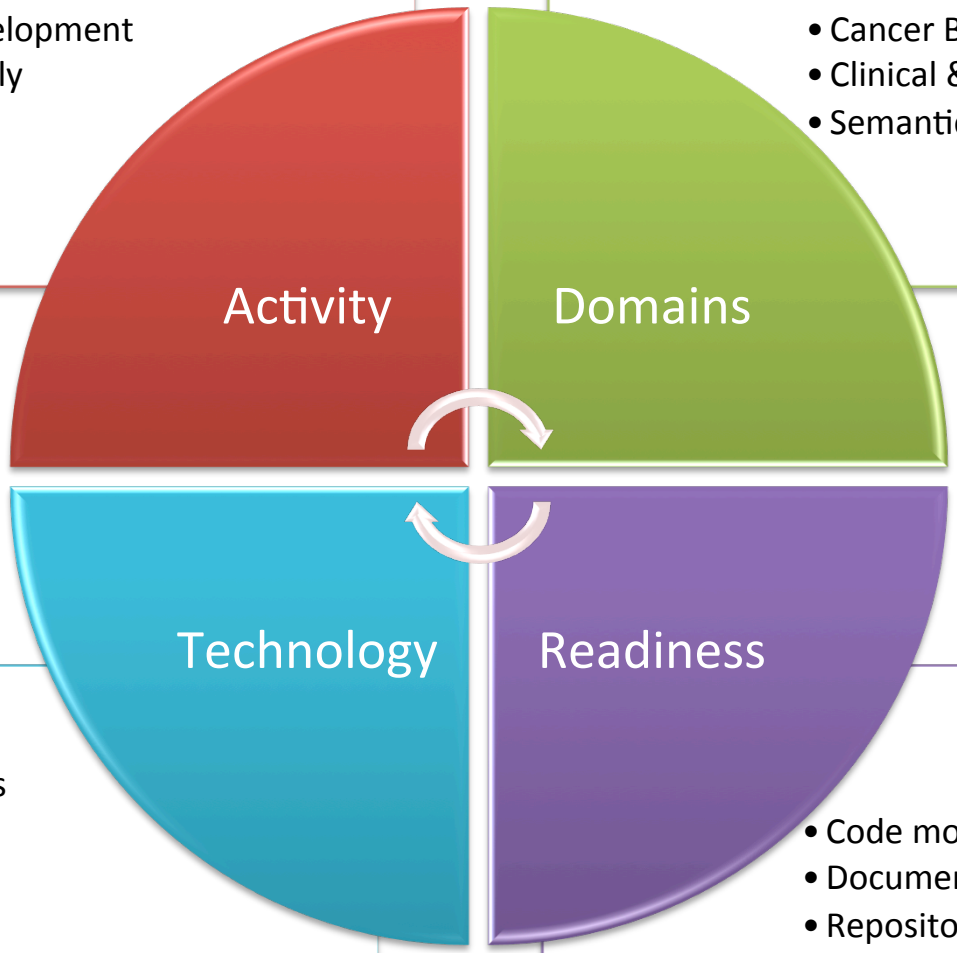
"May all of your problems be technical." – Jim Gray

Photo: Chris Ware/Keystone Features/Getty Images

Diversity is a big challenge

- Some community development
- Gov't development only
- Dormant

- Cancer Biology & Genomics
- Clinical & Translational Research
- Semantics & Interoperability



- Database applications
- Toolkits
- Data models
- Analytical scripts
- …

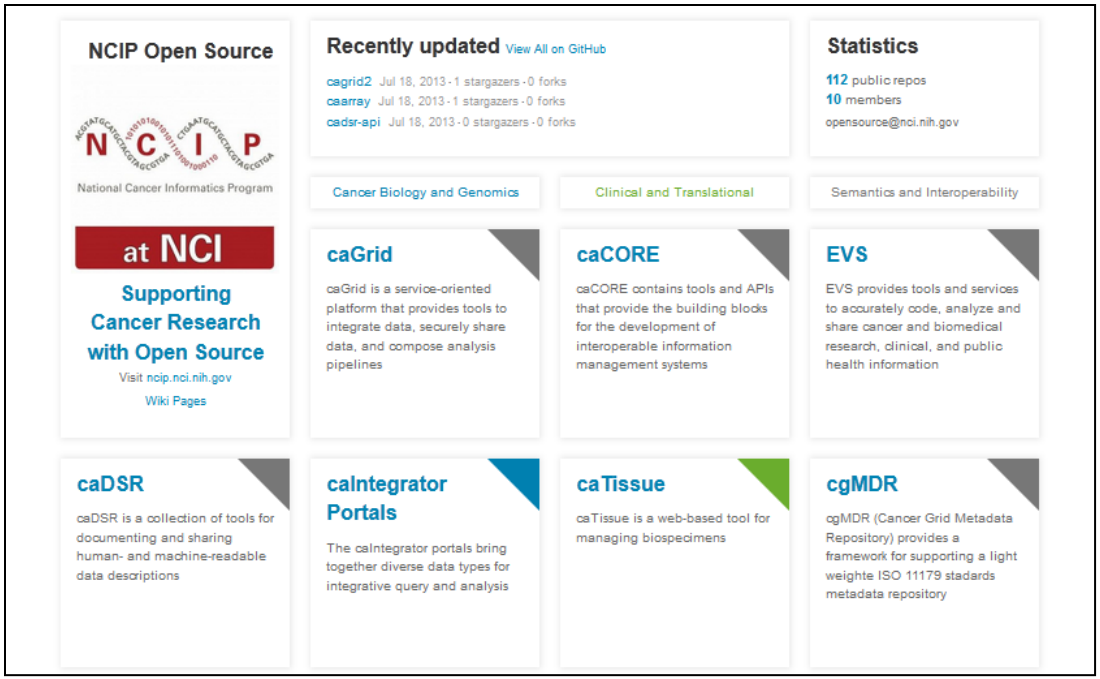
- Code modularity
- Documentation
- Repository structure

Early steps at building communities

Project Name	Initiated by	Current "Owner"	Activities
caLIMS	Government	Community	Community-led management committee, low activity
caTissue	Government	Government	Two "Code Jamborees" have been held
XIP/AVT	Community	Community	"Hack-fest" recently held – off and running

Join us!

- Fork the code
- File an issue
- Fix a typo
- Watch a repo
- Become a member



New! NCIP Channel landing page: http://ncip.github.io

Follow us on Twitter @NCI_NCIP
Connect with us on LinkedIn
NCIP Blog: http://ncip.nci.nih.gov/blog

Acknowledgments

NCI CBIIT

- Robert Shirley
- Anthony Kerlavage
- George Komatsoulis

SAIC-Frederick, Inc.

- Sreenath Nampally
- Larry Brem

Kitware

- Brad King
- Luis Ibanez

Juli Klemm
@jklemm
juli.klemm@nih.gov