National Cancer Institute

The National Cancer Informatics Program (NCIP) Hub

A platform for collaboration and sharing of data, tools, and standards amongst the cancer research community

Ishwar Chandramouliswaran

September 29 2014

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES

National Institutes of Health

National Institutes of Health (NIH)
Turning Discovery into Health

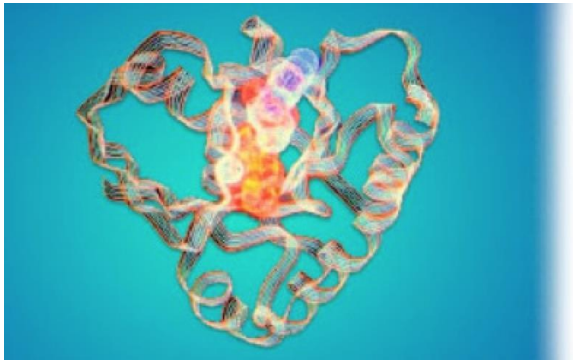
- Agency of the United States Department of Health and Human Services
- Primary agency responsible for biomedical and health-related research

- Formed in1887 - Hygienic Laboratory
- Headquarters in Bethesda, Maryland, U.S.
- Annual budget US $30.9bn (as of 2010)
- Agency Executive is Dr. Francis Collins

- 27 Institutes/Centers, each focusing on particular diseases/body systems

National Cancer Institute

- Federal Government's principal agency for cancer
- Established under the National Cancer Institute Act of 1937
- Coordinates the 'National Cancer Program':
  - Research
  - Training
  - Education



National Cancer Institute

1-800-4-CANCER

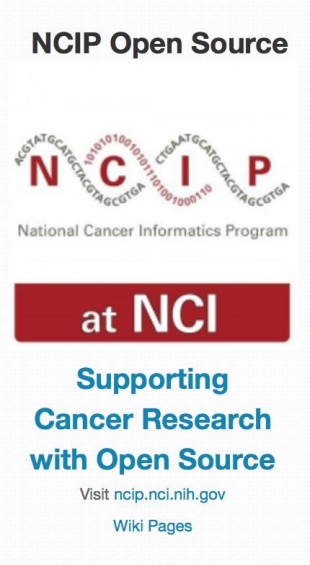
Center for Biomedical Informatics & Information Technology

- Situated in the Office of the NCI Director
- Assess and meet informatics needs of NCI Programs
- Focus is on innovative programs and technologies supporting open science to accelerate research in the cancer informatics community

National Cancer Institute

Empowering Cancer Research via Science as a Service (ScaaS)

GitHub Channel



Open Development Initiative

Cancer Cloud

Democratize Access

NCIP Hub

Foster Communities

Realize Open Science to Accelerate Research

- Foster a sustainable community of expert members to support open innovation of tools, data and standards supporting cancer research.

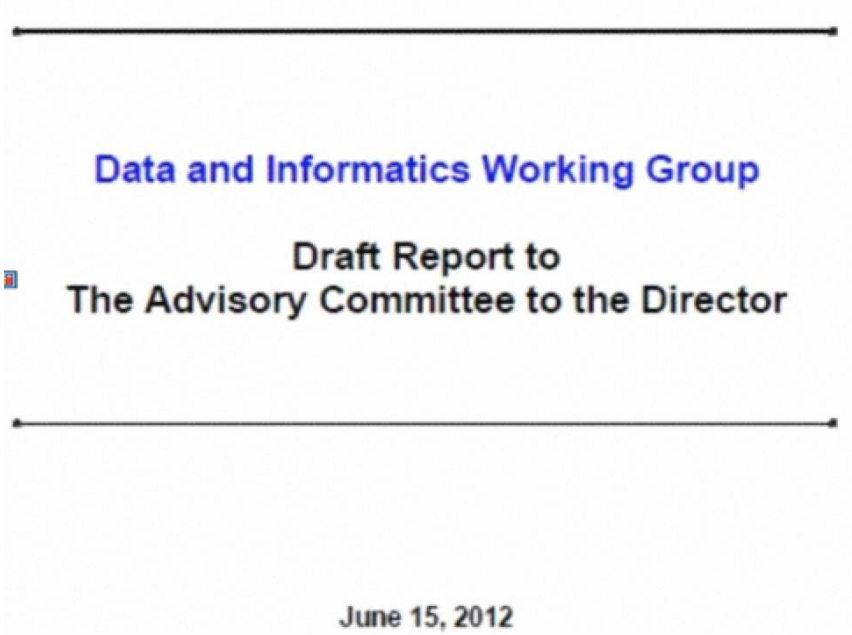
- Promote mechanisms to democratize access to data, tools, and standards within NCI and across the cancer research community.

- Enable the scientific goals of the NCI through the collaborative development and support of informatics programs.

Align with NIH Informatics Objectives

Mission and vision of the Data Informatics Working Group

http://acd.od.nih.gov/diwg.htm


Data and Informatics Working Group

Draft Report to
The Advisory Committee to the Director

June 15, 2012

NIH Informatics Needs – Problems to Solve

- Locating the digital assets – breaking the silos
- Access to the asset – lowering the barrier
- Organizing, managing and processing/sustainability
- Accelerating new methods for analyzing (data)
- Sharing/collaboration - policies, practices, etiquettes
- Training researchers to use these assets effectively

Create an Adaptive & Collaborative Environment

Recommendations of the DIWG Committee:

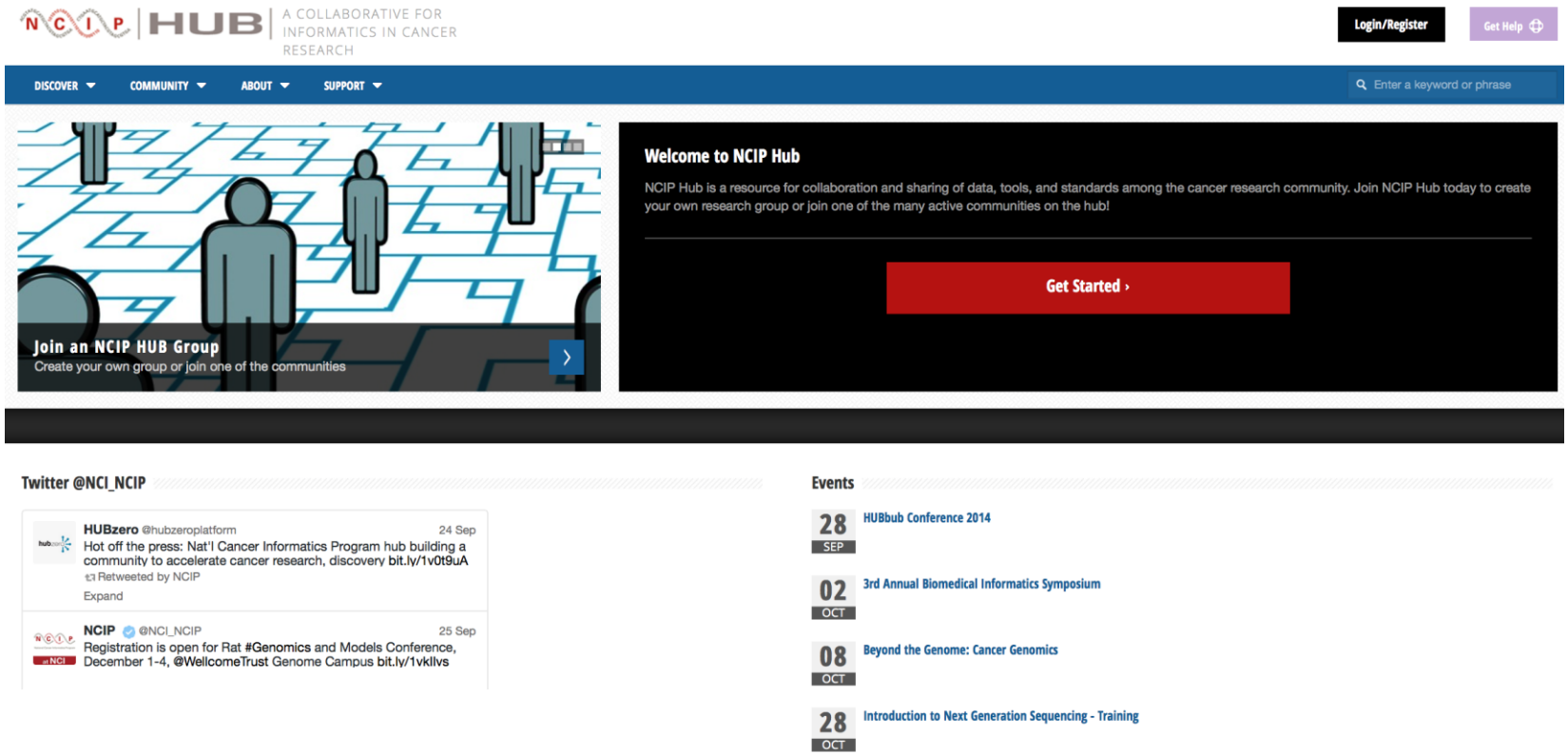
- Promote data sharing through catalogs
- Support informatics methods & applications
- Build capacity via training in quantitative sciences
- Develop NIH-wide strategic plans
- Commitment to support these recommendations

The Big Data to Knowledge Program (BD2K)

Enable the biomedical enterprise to maximize the value of biomedical data

http://bd2k.nih.gov

NCIP Hub
A Collaborative for Informatics in Cancer Research



NCIP Hub, A 360º Approach to Developing and Sustaining NCI Informatics Resources

- Serve as front-end to the cancer informatics assets
  - Access cancer related data/resources
  - Support open development projects
  - Marketplace for relevant analysis tools

- Enable sustainable communities
- Serve as a model for sharing and as an educational resource
- Enhance transparency and incentivize collaboration

VISION: NCIP Hub

Serve larger community

Serve NCI

NCIP Hub

Biologist – end user

Bionformatician – HTP user

Computational Scientist - Developer

Marketplace – *democratize resources*

- Access and explore databases
- Upload lab/clinical data generated
- Access computational tools
- Analyze data and share results
- Rate tools and provide feedback

Resources

Usage
Ranking
Feedback
Collaboration

AppStore – *open development*

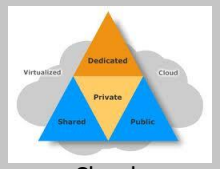
- Develop & upload computational tools
- Download other tools
- Access to data to test tools
- Crowd-sourced development
- Gather real-time requirements

Collaboratory – *sustainable communities*

- News & related blog posts
- Create or join groups for collaboration
- Explore and analyze data and tools
- Contribute other learning resources
- Publications, citations

Access data

Share data



Cloud Services

Run Analysis

Share Source Code

Community Development

Run Analysis

crowd-sourced research

contributions & learning

Community Forums

Access NCI and Other Capabilities

Why HUBzero?

- ✓ Broad set of features for scientific community
- ✓ Worked for diverse scientific domains
- ✓ Serves both individual and power users
- ✓ Provides transparent analytics
- ✓ Open source; NSF funded, sustainable research project

- ✓ Aligns with program & cancer community needs
  - Establish catalog of informatics resources
  - Content driven by the community
  - Support training needs
  - Reduce redundancies & enhance transparency
  - Reduce costs
  - Model for collaboration (internal & external)
  - Opportunities to innovate

Where are we today?

- Community site setup www.nciphub.org
- Hosted at Purdue University
- Understanding users & use at other federal agencies
- User group & webinars to teach & learn
- Lead user activities (test drives) underway
  - 500 members, 40 groups
  - 130 resources

Table 3: Miscellaneous Statistics

Domains Served	1,140
Cummulative Interactive User Sessions [10]	4,811
Cummulative Session Time	635 days
Visitors [11]	10,506
Visits [12]	38,230
New Accounts	409
Max User Logins	41 users on 2014-03-20
Web Server Hits	3,047,175

Quantitative Imaging Network: Collaborative development and algorithm evaluation

- Share (open source) tools for image analysis
- Develop & deploy statistical analysis & metrology tools
- Co-locate data & tools
- Evaluate the performance of quantitative algorithms
- Perform algorithm comparisons
- Encourage network member participation
- Make recommendations for best practices

Informatics Technology for Cancer Research: Communication & Collaboration

- Serve as platform for communication within ITCR
- Establish a central access point program generated data, tools & other resources
- Provide individual project management support
- Promote inter-project collaboration

Data (meta) Standards
Create Awareness, Discuss Gaps & Promote Adoption

STANDARDS PORTAL (PILOT)

The Standards Portal is envisioned as a potential solution for bringing together standards of interest in a unified and meaningful way.

TCGA CDE STANDARDS

TCGA version 2.6 CDEs are the standard elements in validation of clinical data, available for download at the Data Matrix. This resource presents the CDEs for search and exploration, with materials to clarify the CDEs and describe best practices for adopting them.

ADVERSE EVENTS STANDARDS

This updated version of naming and grading of adverse events in cancer clinical trials is the result of collaboration among academic, government and pharmaceutical industry clinical investigators. The effort is based on the need for reproducible, consistent and readily usable adverse events reports.

REQUEST AUTHORIZATION TO UPLOAD

Does your community want to upload preferred standards to the Standards Portal?

EXPLORE STANDARDS USING DATAVIEWS CDES, CRFS AND VOCABULARY

Use these DataViews to Explore the Standards Portal!

CDEs Complete View ⇢
CDEs Common Data Fields ⇢
CRFs Complete View ⇢
CRFs Common Data Fields ⇢
Preferred Content Vocabulary ⇢

Put DataViews on all your resource and group pages! Here is a list of all DataViews at the Standards Portal

All Standards Portal DataViews ⇢

You can filter DataViews by Group:

DISCUSSION AROUND STANDARDS

This group site supports public discussion about discovery of standards, with gateways for easy access to all standards portal groups and resources.

The Standards Portal is a community-driven dynamic web site that centralizes access to valueable and practical information around standards to facilitate adoption and prioritization.

Our collaborative environment provides a means for research communities to identify and discuss gaps and overlaps for the evolution of preferred content, with powerful support for content sharing, search and discovery.

NEW IN RESOURCES

IN THE SPOTLIGHT

Nanotechnology Working Groups/Initiatives

- Nanotechnology working group meetings
  - Collaborative paper development

- National Nanotechnology Initiative: www.nano.gov
  - Discussion boards on data readiness levels

- Introduced the nanoHUB to the nanoEHS community https://us-eu.org

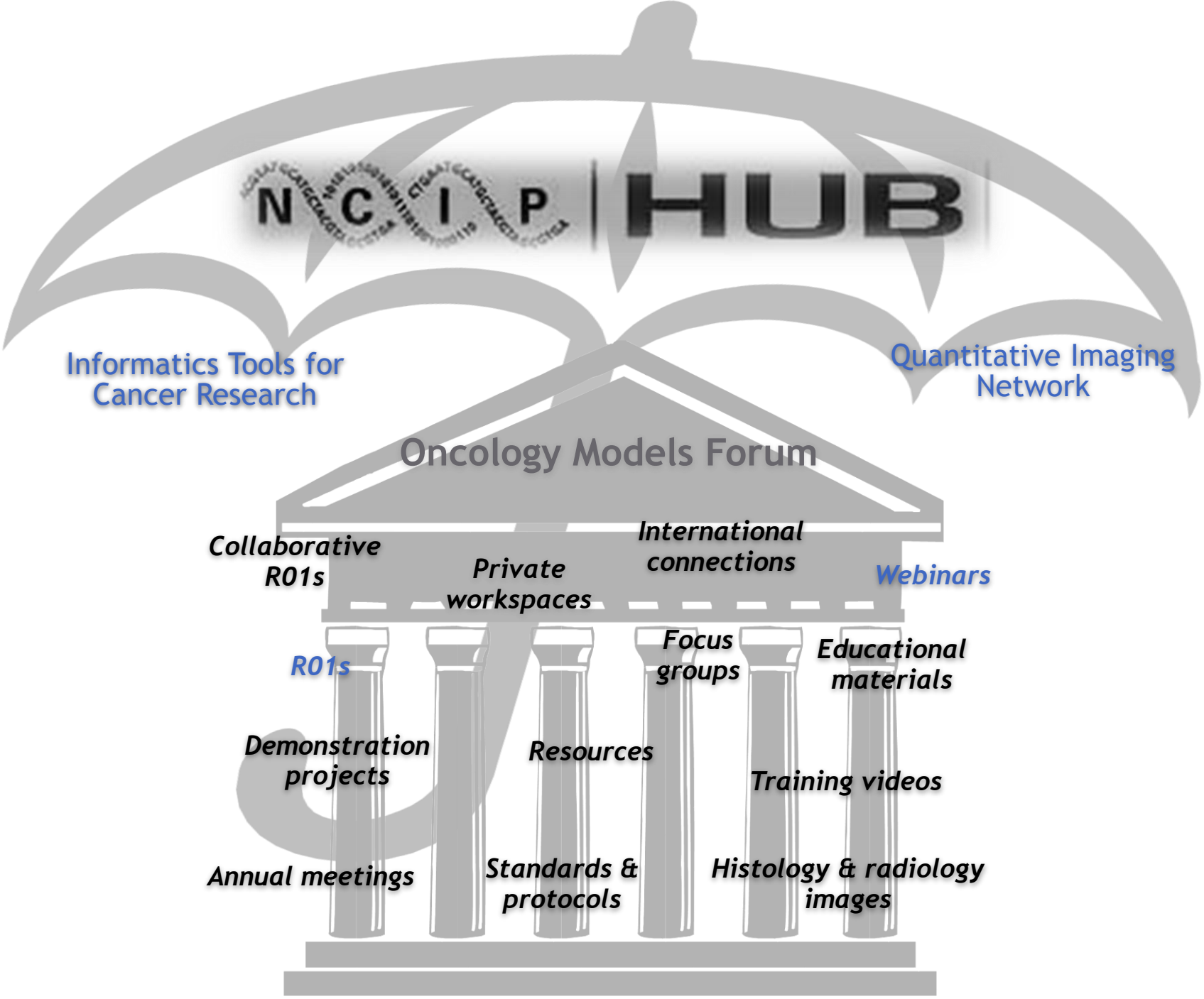
Partnerships to Advance Cancer Health Equity

- Program in Center to Reduce Cancer Health Disparity
- All about building partnerships between minority institutions and cancer centers
- Collaboration for scientific discovery and academic learning

Establish and sustain a CPACHE Community Forum in NCIP Hub

The Oncology Models Forum:
Open Interactive Platform for Translational Research

- Establish a knowledgebase supporting animal models of human cancer
- Provide an educational resource
- Enable
  - collaborations,
  - demonstrate best practices, and
  - share educational materials
  - better reproducibility of results


NCIP
HUB
Informatics Tools for Cancer Research
Quantitative Imaging Network
Oncology Models Forum
Collaborative R01s
Private workspaces
International connections
Webinars
R01s
Focus groups
Educational materials
Demonstration projects
Resources
Training videos
Annual meetings
Standards & protocols
Histology & radiology images

Challenges Ahead

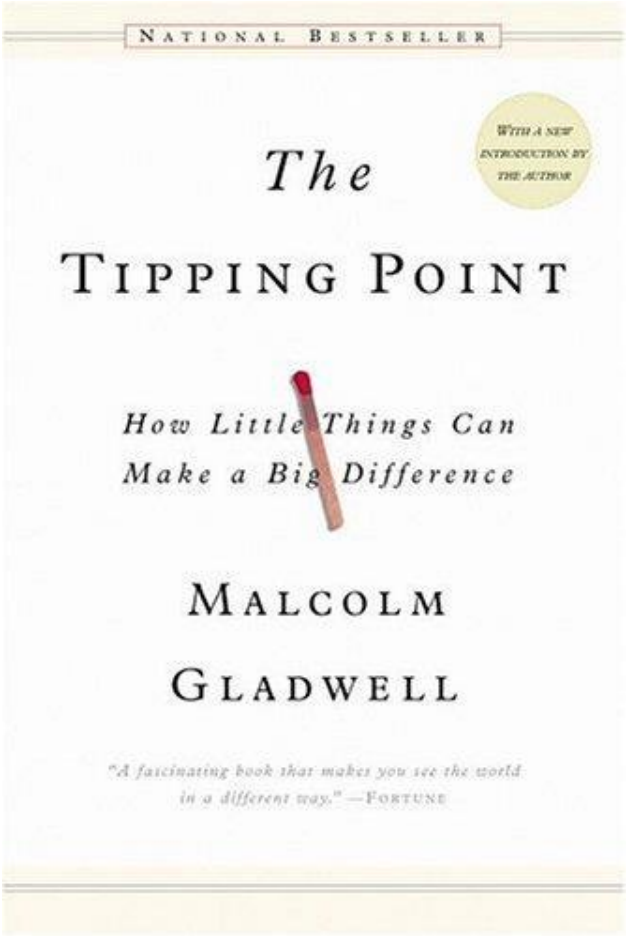
Community:

- Content to drive a vibrant community
- Identify the right partners/community champions
- Platform adoption, engage & incentivize users
- Encourage open-ness
- Co-exist with other platforms, tools

Operational:

- Acquire HUBzero admin & web development skills
- Comply with Federal Information System Requirements
- Lower the barrier for community utility

What is the NCIP Hub Tipping Point?



Ishwar Chandramouliswaran
Email :Ishwar.chandramouliswaran@nih.gov

NCIP Hub :https://nciphub.org/members/1100
LinkedIn :www.linkedin.com/in/ishwarc/
Twitter :https://twitter.com/ishwartweets

Anticipated Impact of NCIP Hub

- Greater visibility to NCI bioinformatics at lower costs
- Serve as intellectual capital to plan new studies/projects
- Bridge gap between bench-scientists & bioinformaticians
- Transformative platform for cross-disciplinary interactions
- Model for collaboration – encourage and incentivize sharing

Outreach Solution for Cancer Informatics Assets

- **Marketplace/repository**
  - Repository for resources with *DOIs*
  - Publications, seminars, reports, tools, data
- **AppStore for scientists**
  - Tools uploaded as Apps by the community
  - Accessible from any machine
  - Cloud powered for intensive computation
  - Support workflows/pipelines
  - *Toolkits* to build interfaces
- **Collaboratory for team science**
  - Groups, Calendars, wikis, blogs, collections
  - Linux Workspace - *shared*
  - Collaborative tool development (*ToolsForge*) and user support
  - Data management *(data store)*
- **Learning Management System (MOOCs)**

Current Efforts

- QIN
- ITCR Program
- PACHE – objectives, supplement tasks
- Standards Portal
- Nano-medicine initiatives
- Oncology Models Forum
- *TCRN Committee*
- *Quantitative Image Informatics (QIICR)*
- *Pathway Interaction Database Prioritization*
- *Computational Genomics Group*