Image Sharing with iDASH:
Disrupting Data Ownership to Facilitate Imaging Research

Brian E. Chapman, Ph.D.

Division of Biomedical Informatics

Department of Medicine

University of California, San Diego

"Once upon a time, there was a deep divide between scientists who did and did not have easy access to scientific contents (books, journals, lectures, data), hardware (microscopes, imaging devices, lab instruments, supercomputers) and people (professors, networks of peers, knowledgeable support personnel)."

Data as Wealth

- Proximity to data greatly enhances research leverage

- Access to data more than quality of ideas may determine career success

Individual Implications

- Data collection becomes significant part of job
- Competitive advantage to ~~hoard~~ not share data
- Consequence: Scientists often work in isolation

"Unfortunately, this isolation led some of them to completely lose perspective of the university as a place to create, propagate and disseminate meaningful scientific knowledge not only to its own members or peers, but to society as a whole."

Societal Implications

- Research progress is slower
  - Mismatch between best data and best ideas
- Research cost is higher
  - Researchers forced to duplicate data collection
- Research is more limited
  - Testable hypothesis limited to diseases for which researchers can collect sufficiently large cohorts
  - What about patients who have had a Wilms' tumor, carcinoid, and a melanoma?

Data is also a Liability

- HIPAA rules require
  - Expensive hardware/software "vault"
  - Savvy de-identification tools
  - Implication: privileged investigators
- Most researchers probably don't meet standards
- Severe financial liability for violations
- Major universities investing in costly HIPAA compliant infrastructure
  - What about smaller institutions?

Can a National Resource Help?

NIH Roadmap

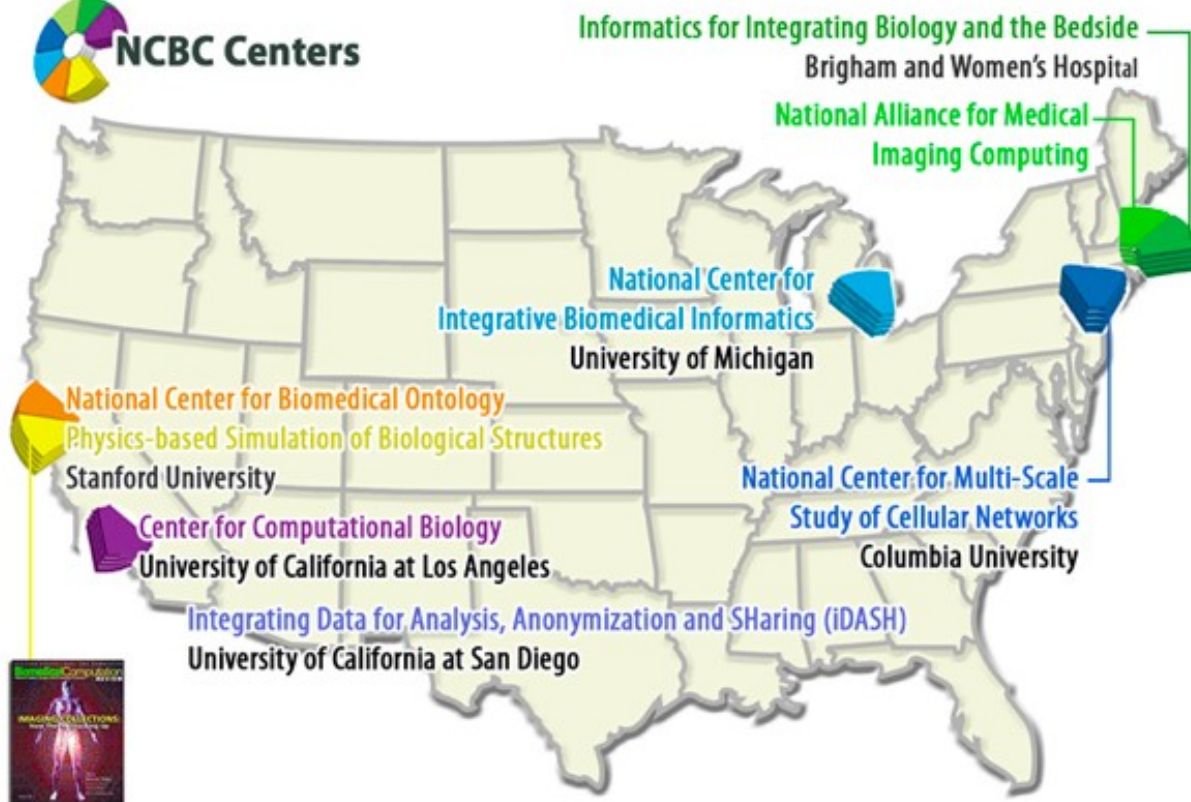# National Centers for Biomedical Computing

| Home | NCBC Summary | Calendar | All Hands Meetings | DBP Interactions | Biositemaps | Working Group Archive |

NCBC Centers

Informatics for Integrating Biology and the Bedside
Brigham and Women's Hospital

National Alliance for Medical Imaging Computing

National Center for Integrative Biomedical Informatics
University of Michigan

National Center for Biomedical Ontology
Physics-based Simulation of Biological Structures
Stanford University

National Center for Multi-Scale Study of Cellular Networks
Columbia University

Center for Computational Biology
University of California at Los Angeles

Integrating Data for Analysis, Anonymization and SHaring (iDASH)
University of California at San Diego

Sidebar icons:
- CCB
- i2b2
- NAMIC
- Simbios
- NCBO
- NCIBI
- MAGNet
- iDASH
- Biomedical Computation REVIEW

Sharing Data

- Today
  - Public repositories (mostly non-clinical)
  - Limited data use agreements
- Tomorrow
  - *Annotated* public *databases*
  - Informed consent management system
  - Certified trust network

Sharing Computational Resources

- Today
  - Computer scientists looking for data, biomedical and behavioral scientists looking for analytics
  - Duplication of pre-processing efforts
  - Massive storage and high performance computing limited to a few institutions
- Tomorrow
  - Processed de-identified, 'anonymized' data shared
  - Secure biomedical/behavioral cloud

iDASH

De-Identification, Anonymization

- Count queries do not prevent re-identification
- Removal of identifiers does not prevent re-identification
- Current "k-anonymization" methods are problematic
  - Cell swaps, cell removal, generalization

- Development of "sufficiently similar" data for research helps protect privacy
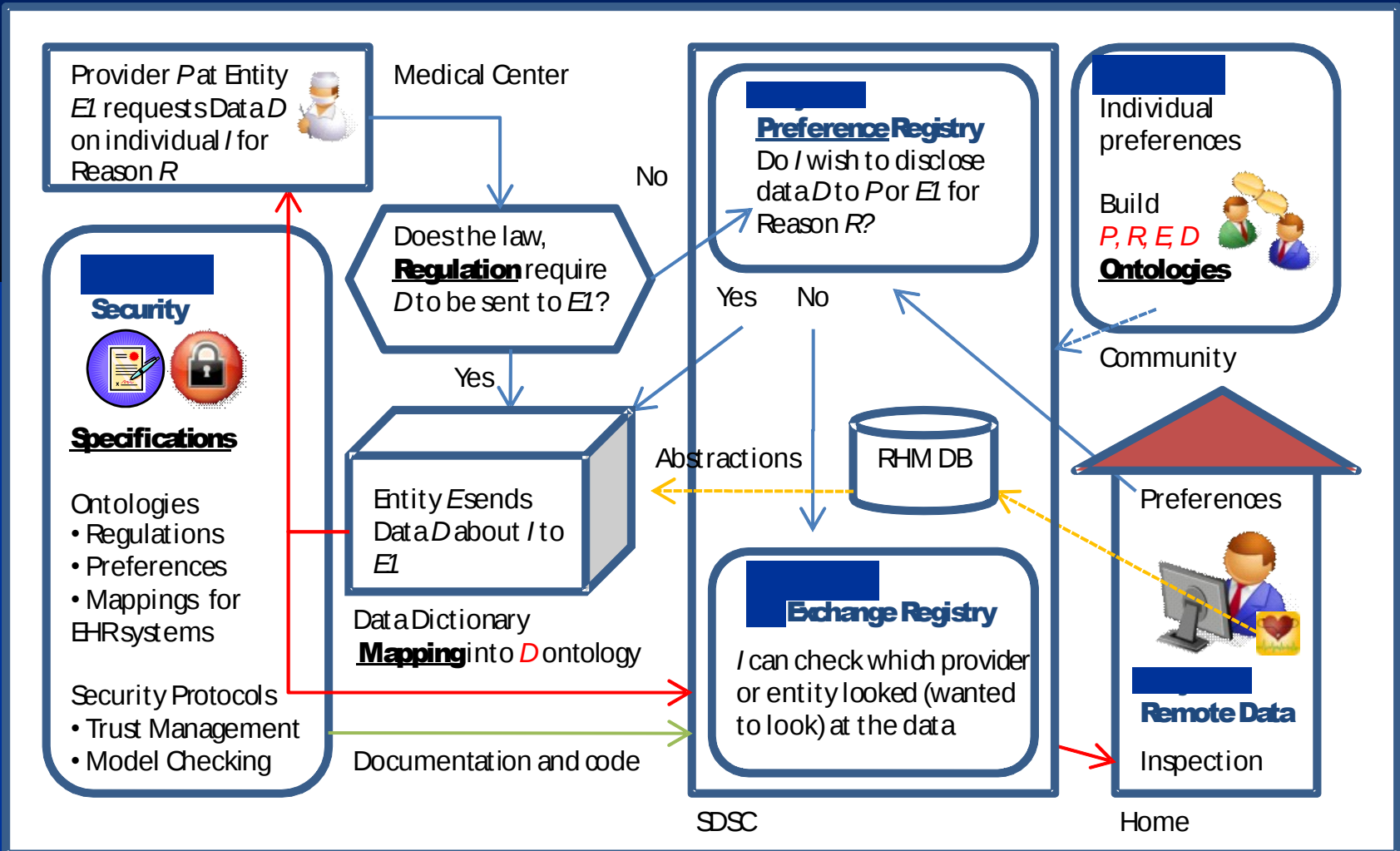- Institutional/provider de-identification is also necessary

Kawasaki Disease
(Driving Biological Project, PI Jane Burns)

- Specific Aim 1: Discover miRNAs associated with KD

- Specific Aim 2: Identify rare genetic variants that may play a functional role in disease susceptibility and outcome

- Specific Aim 3: Create a KD data warehouse and web-based data analysis system aimed at facilitating discoveries using clinical and molecular data

iDASH

Consent Management

Provider *P* at Entity *E1* requests Data *D* on individual *I* for Reason *R*

Medical Center

Does the law, **Regulation** require *D* to be sent to *E1*?

No

Preference Registry
Do *I* wish to disclose data *D* to *P* or *E1* for Reason *R?*

Individual preferences

Build
P, R, E, D
Ontologies

Security

Specifications

Ontologies
• Regulations
• Preferences
• Mappings for EHR systems

Security Protocols
• Trust Management
• Model Checking

Yes

Yes   No

Community

Abstractions

RHM DB

Preferences

Entity *E* sends Data *D* about *I* to *E1*

Data Dictionary
Mapping into *D* ontology

Exchange Registry

I can check which provider or entity looked (wanted to look) at the data

Remote Data

Inspection

Documentation and code

SDSC

Home

*Portions of this framework are being developed by a grant from the AHRQ.

Has this been tried before?

- NCI dbGaP
  - Hard to use, non-clinical data
- NCI National Biomedical Imaging Archive
  - Isolated from clinical context
- BIRN
  - Limited scope/limited success
- caBIG

iDASH Sharing

- Data

- Software
  - Open source software packages to share
    - Analysis tools
    - De-identification tools
  - iDASH ultimately shareable itself
- Computational Resources
  - Data storage, CPU cycles, memory
- Online analysis/annotation tools

iDASH Imaging Projects

- Kawasaki Disease
  - DSA, CT, MRI, US images
  - Longitudinal data
  - Coronary aneurysms
    - Risks
    - Progression
    - Phenotyping
- Pulmonary embolism
  - CTPA

iDASH Imaging Projects

- Pulmonary embolism with CTPA
    - Image querying
    - Online annotation/analysis
- Cervical Carotid MRI/MRA
- Intracranial MRA
- Pediatric brain tumors

iDASH Imaging Tools

- Graphical programming
  - Yahoo pipes, Labview, Orange
- Standard analysis background
  - ITK, R
- caBIG Annotation and Image Markup
- Privacy/Security
  - Watermarking
  - Face erosion

What About Words?

- Does the radiology report have any value?
- "Nobody cares about the report!"
  - High inter- and intra-reader variability
  - Quality concerns
- But
  - Report does obviously contain useful information
  - Initial annotation of the image

Finally for Fun
A Little Philosophy

Modern Imaging Has Transformed Medicine, But…

- Carl Jaffe: "No one in clinical [drug] trials takes radiology seriously." (CaBIG, Dec. 2005)

    1. The inability to quantitatively monitor therapy
    2. The inability to validate findings
    3. Lack of transparency and data sharing between institutions
    4. Failure to integrate clinical information in the image assessment
    5. The unreliability of site interpretations of the imaging studies

"**Monumental achievements come with monumental costs.**" *(Meet You In Hell)*

- Medicare medical imaging expenditures increased 20%/year since 1999

- Radiation exposure
    - Underappreciated risk (BMJ 2004), Medical sources = Natural sources
    - "CT scanners may account for about 1.5 to 2 percent of all cancers in the United States." (Reuters)
    - Good time to be an MR physicists

Have We Gotten Off Track?

If we step back to Pythagoras

- ca. 570 to ca. 490 BCE
- Mystic, philosopher, mathematician?
- "Pythagoras is known for the *honor* he gives to number and for removing it from the practical realm of trade and instead pointing to correspondences between the behavior of number and the
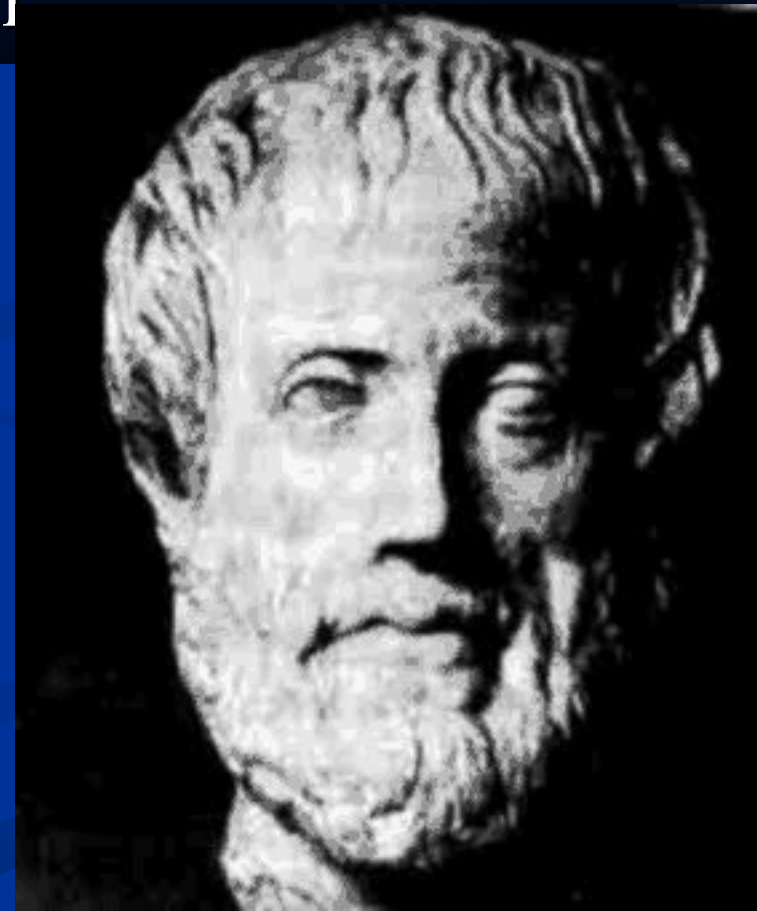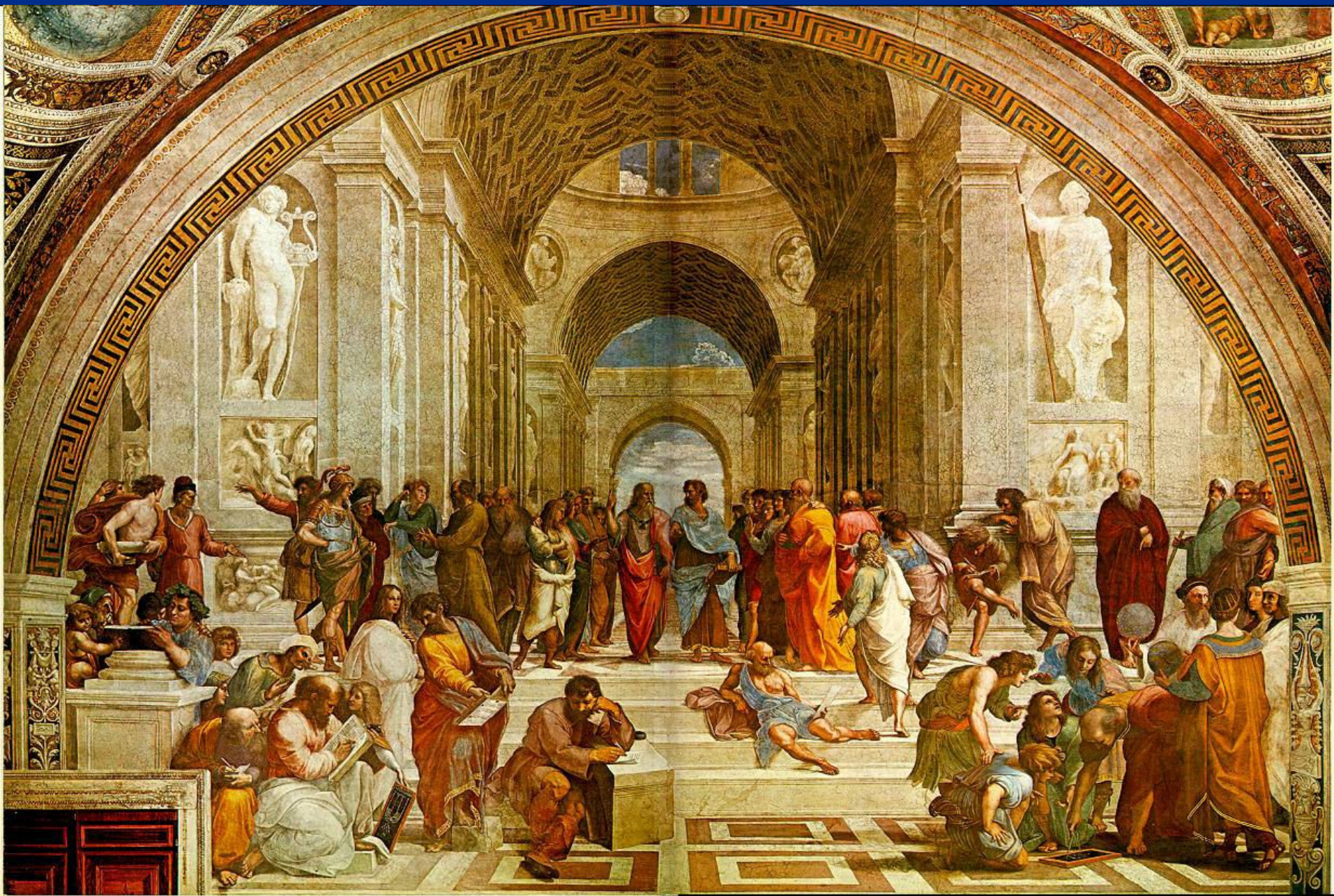
And then Forward a bit to Plato

- 429–347 BCE.
- "*let no one ignorant of geometry enter*"
- *Ontology of mathematics*
- *Mostly a promoter*
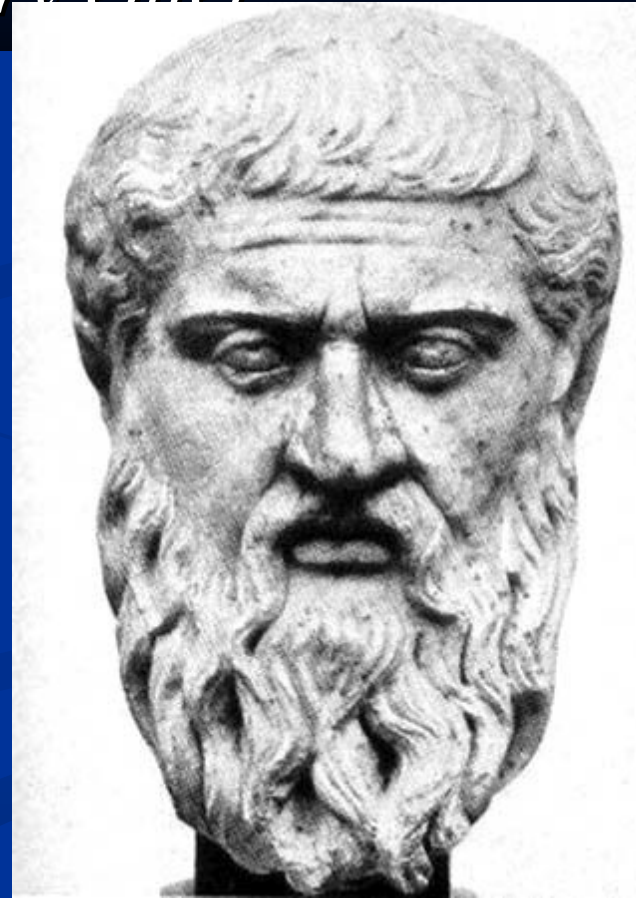
After Plato, Aristotle

- 384–322 B.C.E.
- Emphasis on logic, empirical classification
  - Collecting constitutions
  - Grounded, systematic thinker

- Plato
  - The mystic pointing up to heaven

- Aristotle
  - The realist keeping things grounded

"Plato and Pythagoras stand nearer to modern physical science than does Aristotle. The two former were mathematicians, whereas Aristotle was the son of a doctor, though of course he was not hereby ignorant of mathematics. The practical counsel to be derived from Pythagoras, is to measure, and thus to express quality in terms of numerically determined quantity. But the biological sciences, then and till our own time, have been overwhelmingly classificatory. Accordingly, Aristotle by his Logic throws the emphasis on classification

"The popularity of Aristotelian Logic retarded the advance of physical science throughout the Middle Ages. If only the schoolmen had measured instead of classifying, how much they might have learnt!

"Classification is necessary. But unless you can progress form classification to mathematics, your reasoning will not take you very far."

~Alfred North Whitehead, *Science and the Modern World*

Eratosthenes, the Model Informatician

- Born circa 285 BCE in Greek north African city

- Appointed head librarian of the Museum in Alexandria in 245 BCE
  - Literary critic
  - Historian
  - Mathematician
  - Pen-pal of Archimedes
  - Geographer

Erastosthenes' Measurements

- Eratosthenes estimated the polar circumference of the earth as 252,000 stadia (Greeks liked whole numbers)
  - Depending on definition of stadia, this measurements is between a 1% underestimation or a 16% overestimation
- Also estimated the solar distance to within 16% error
- The patron saint of imaging informatics?

Qualitative Medicine

- The impact of medical imaging is not maximized because it remains dominantly a qualitative rather than a quantitative discipline.

- "Too complex"

- Problem primarily cultural
  - "A radiologist with a ruler is a radiologist in trouble."

- Culture common to medicine

Quantitative Medicine

- Certainly many initiatives to increase quantitative imaging
- Rulers are now acknowledged as needed
- But much still to measure
  - Measure more than we know that we need
  - Economic (Bill Johnson)

Medicine and the Future

- Chuck Friedman: Creating a rapidly learning medical community
  - 17 years to 17 days
- Best practice might still be lousy practice
- ANW: "If only the [physicians] had measured instead of classifying, how much they might have learnt!"
- Hope that iDASH and myself can contribute

Thank You