

# **Ubiquitous Science: U-Science, Citizen Science, and the Zooniverse Project**

by

**Kirk Borne**

**Department of Computational & Data Sciences**

**George Mason University**

# Outline

- Themes in Informatics Research
- U-Science: Introduction
- U-Science: Examples

# General Themes in Informatics Research

- Information and knowledge processing, including natural language processing, information extraction, integration of data from heterogeneous sources or domains, event detection, feature recognition.
- Tools for analyzing and/or storing very large datasets, data supporting ongoing experiments, and other data used in scientific research.
- Knowledge representation, including vocabularies, ontologies, simulations, and virtual reality.
- Linkage of experimental and model results to benefit research.
- Innovative uses of information technology in science applications, including decision support, error reduction, outcomes analysis, and information at the point of end-use.
- Efficient management and utilization of information and data, including knowledge acquisition and management, process modeling, data mining, acquisition and dissemination, novel visual presentations, and stewardship of large-scale data repositories and archives.
- Human-machine interaction, including interface design, use and understanding of science discipline-specific information, information needs, and uses.
- High-performance computing and communications relating to scientific applications, including efficient machine-machine interfaces, transmission and storage, real-time decision support.
- Innovative uses of information technology to enhance learning, retention and understanding of science discipline-specific information.
- REFERENCE: <http://grants.nih.gov/grants/guide/pa-files/PA-06-094.html>

# General Themes in Informatics Research

- Information and knowledge processing, including natural language processing, information extraction, integration of data from heterogeneous sources or domains, event detection, feature recognition.
- Tools for analyzing and/or storing very large datasets, data supporting ongoing experiments, and other data used in scientific research.
- Knowledge representation, including vocabularies, ontologies, simulations, and virtual reality.
- Linkage of experimental and model results to benefit research.
- Innovative uses of information technology in science applications, including decision support, error reduction, outcomes analysis, and information at the point of end-use.
- Efficient management and utilization of information and data, including knowledge acquisition and management, process modeling, data mining, acquisition and dissemination, novel visual presentations, and stewardship of large-scale data repositories and archives.
- **Human-machine interaction, including interface design, use and understanding of science discipline-specific information, information needs, and uses.**
- High-performance computing and communications relating to scientific applications, including efficient machine-machine interfaces, transmission and storage, real-time decision support.
- **Innovative uses of information technology to enhance learning, retention and understanding of science discipline-specific information.**
- REFERENCE: <http://grants.nih.gov/grants/guide/pa-files/PA-06-094.html>

**Enter ....**

U-Science



# Beyond e-Science

- **e-Science** emerged ~10 yrs ago using the web protocols that were common at that time:
  - web services, XML-based information exchange, registries, distributed data access, distributed computing (Grid) = **machine-to-machine communication**
- **U-Science** is now emerging from today's web protocols:
  - social networking, ubiquitous devices, user-centric experiences, user-led activities, user-generated content, wikis, blogs, mashups, tagging, annotation, ontologies (semantic web), folksonomies, knowledge-sharing, user recommendations = **user-to-user communication**
- The emergence of Citizen Science:
  - Anybody can participate in the science discovery process
- Anyone can annotate, tag, and label scientific results:
  - scientists, students, and citizen scientists

# What is U-Science?

- User-centered science
- Ubiquitous
- Universal
- Untethered [ <http://tw.rpi.edu> ]
- “You”-centric
  
- Citizen Science
- Semantic e-Science
  
- Think ... Social Networks ...
  - Facebook, Myspace, Youtube, Blogs, Wikis
  - Del.icio.us, flickr.com, Digg, Reddit, Stumbleupon

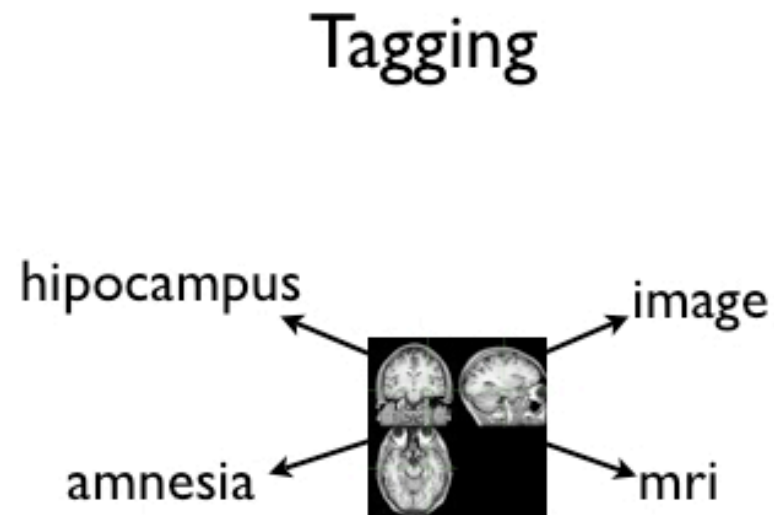
# U-Science: it is all about “U”

- Classic web is Web 1.0, and the future Semantic Web is Web 3.0.
- Between these two is the social network-oriented Web 2.0:
  - Wikis and Mashups (e.g., Yahoo Pipes, Microsoft Popfly)
  - Tagging, Annotation, Folksonomies, Microformats, Tag Clouds
  - For example: <http://www.flickr.com/> or <http://del.icio.us/>
  - **Human Computation**: The ESP game @ <http://www.espgame.org/>
    - created by [Luis Von Ahn](#) (MacArthur genius grant winner)
  - Science 2.0: [http://openwetware.org/wiki/Science\\_2.0/Brainstorming](http://openwetware.org/wiki/Science_2.0/Brainstorming)
  - **Science collaboratories**
  - **BioDAS** = Biology Distributed Annotation System: <http://biodas.org/>
  - **Heliophysics Knowledgebase**: <http://www.lmsal.com/helio-informatics/hpkb/>
  - **Wikiproteins**: <http://www.wikiprofessional.info/>
  - **Entity Describer**: <http://www.entitydescriber.org/about.html>
  - **AstroDAS**: coming soon (hopefully!)

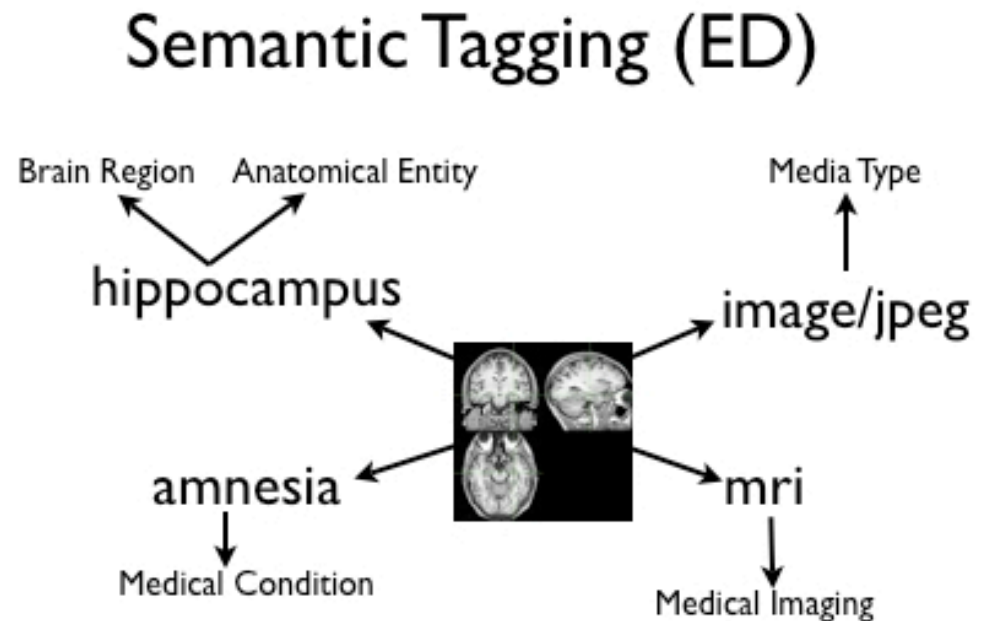
# Entity Describer – example

<http://www.entitydescriber.org/about.html>

- Before semantic tagging:



- After semantic tagging:



**It takes a human to interpret a complex image**



# Examples of U-Science

- AAVSO (Amer. Assoc. of Variable Star Observers)
- Audubon Bird Counts
- Project Budburst
- Stardust@Home
- VGI (Volunteer Geographic Information)
- CoCoRaHS (Community Collaborative Rain, Hail and Snow network)
- Galaxy Zoo (**~20 refereed pubs so far...**)
- Zooniverse (buffet of Zoos)
- U-Science (semantic science 2.0) [ref: Borne 2009]
  - includes Biodas.org, Wikiproteins, HPKB, AstroDAS
  - **Ubiquitous, User-oriented, User-led, Universal, Untethered, You-centric Science**

# The Zooniverse

<http://zooniverse.org/>

- New funded NSF CDI grant (PI: L.Fortson, Adler Planetarium; J. Wallin & K.Borne, GMU; & Chris Lintott, Oxford U; ... )
- Building a framework for new Citizen Science projects, including user-based research tools
- Science domains:
  - Astronomy (Galaxy Merger Zoo)
  - The Moon (Lunar Reconnaissance Orbiter)
  - The Sun (STEREO dual spacecraft)
  - Egyptology (the Papyri Project)
  - and more (... accepting proposals from community)

# Why ?

## ... response to the science data flood

- ***X-Informatics*** (e.g., X = Bio, Geo, Astro, ...):
  - addresses the scientific data lifecycle challenges in the era of data-intensive science and the data flood
  - defines lightweight ontologies, semantics, taxonomies, concepts, content descriptors for a science domain
  - for the purpose of organizing, accessing, searching, fusing, integrating, mining, and analyzing massive data repositories.
- ***U-Science*** (user-guided informatics-powered):
  - Human computation (e.g., tagging, labeling, classification)
    - characterized by enormous cognitive capacity and pattern recognition efficiency (carbon-based computing)
  - Semantic e-Science and Volunteer Citizen Science
  - Tagging everything, everywhere: ***Analytics in the Cloud***

# Related References

- Borne (2009): “**U-Science**”, <http://essi.gsfc.nasa.gov/pdf/Borne2.pdf>
- Borne, Jacoby, et al. (2009): “**The Revolution in Astronomy Education: Data Science for the Masses**”, <http://arxiv.org/abs/0909.3895>
- Borne (2009): “**Astroinformatics: A 21st Century Approach to Astronomy**”, <http://arxiv.org/abs/0909.3892>
- Borne (2010): “**Astroinformatics: Data-Oriented Astronomy Research and Education**”, *Journal of Earth Science Informatics*
- Dutta, Zhu, Mahule, Kargupta, Borne, Lauth, Holz, & Heyer (2009): “**TagLearner: A P2P Classifier Learning System from Collaboratively Tagged Text Documents**”, accepted paper for ICDM-2009.
- M. F. Goodchild (2007): “**Citizens as Sensors: the World of Volunteered Geography**”, *GeoJournal*, 69, pp. 211-221.
- Lintott et al. (2009): “**Galaxy Zoo: 'Hanny's Voorwerp', a quasar light echo?**”, <http://arxiv.org/abs/0906.5304>
- Raddick et al. (2009): “**Galaxy Zoo: Exploring the Motivations of Citizen Science Volunteers**”, <http://arxiv.org/abs/0909.2925>
- Raddick, Bracey, Carney, Gyuk, Borne, Wallin, & Jacoby (2009): “**Citizen Science: Status and Research Directions for the Coming Decade**”, <http://www8.nationalacademies.org/astro2010/DetailFileDisplay.aspx?id=454>