Informatics Strategic Planning Workshop Aug 27-28, 2015

Name	Email	Title	Affiliation
Ivo Abraham	abraham@pharmacy.arizona.edu	Professor	Pharmacy Practice and Science
David Allen	davida@tla.arizona.edu	Vice President	Tech Launch Arizona
Parker Antin	pba@email.arizona.edu	Associate Dean, Research	College of Agriculture and Life Sciences
Neal Armstrong	nra@email.arizona.edu	Associate VP for Research	Office for Research & Discovery
Jane Bambauer	janebambauer@email.arizona.edu	Associate Professor	Law
Jennifer Barton	barton@email.arizona.edu	Associate VP for Research	Office for Research & Discovery
Dennis Beal	dwbeal@email.arizona.edu	AVP Business Development	Office for Research & Discovery
Tom Bever	tgb@email.arizona.edu	Regents Professor	Linguistics
Sandipan Bhattacharjee	bhattacharjee@email.arizona.edu	Assistant Professor	Pharmacy Practice and Science
Ali Bilgin	bilgin@email.arizona.edu	Assistant Professor	Biomed. Eng. & Elect. & Computer Eng.
Lyle Bootman	bootman@pharmacy.arizona.edu	Dean	Pharmacy
Ron Breiger	breiger@arizona.edu	Profesor	Sociology
Shane Burgess	sburgess@cals.arizona.edu	Director	College of Agriculture and Life Sciences
Chuck Cairns	cairnsc@email.arizona.edu	Interim Dean	COM, Bio-prep. & Resp., Health Sec.
Andrew Capaldi	capaldi@email.arizona.edu	Assistant Professor	Molecular and Cellular Biology
Jane Carrington	janec@email.arizona.edu	Assistatnt Professor	Nursing
Christopher Castro	castro@atmo.arizona.edu	Associate Professor	Atmospheric Sciences
Lisa Chan	lisac@aemrc.arizona.edu	Profesor	COM, Emerg. Med., Medical Informatics
Hsinchun Chen	hchen@eller.arizona.edu	Regents Professor	Management Information Systems
Elliot Cheu	echeu@email.arizona.edu	Associate Dean	College of Science
Anne Cress	cress@email.arizona.edu	Deputy Dean, Research Affairs	College of Medicine
Hong Cui	hongcui@email.arizona.edu	Associate Professor	School of Information
Susannah Dickinson	srd@email.arizona.edu	Assistant Professor	School of Architecture
Jenn Earl	jenniferearl@email.arizona.edu	Professor	Sociology
Kimberly Andrews Espy	kespy@email.arizona.edu	Senior VP For Research	Office for Research & Discovery
Sandiway Fong	sandiway@email.arizona.edu	Associate Professor	Linguistics
Jeremy Frumkin	frumkinj@email.arizona.edu	Associate Dean	Technology Strategy
Caroline Garcia	cmgarcia@email.arizona.edu	Associate VP for Research	Office for Research & Discovery
Lynn Gerald	lgerald@email.arizona.edu	Professor	Public Health
Paulo Goes	pgoes@email.arizona.edu	Department Head	Management Information Systems
Jeff Goldberg	jgoldberg@arizona.edu	Dean	College of Engineering
Lucio Guerrero	luciog@email.arizona.edu	Director, Communications	Office for Research & Discovery
Christopher Hamilton	chamilton@email.arizona.edu	Assistatnt Professor	Lunar and Planetary Laboratory
Michael Hammer	mfh@email.arizona.edu	Research Scientist	Arizona Research Labs
John Hartman	jhh@cs.arizona.edu	Associate Professor	Computer Science
Larry Head	larry@sie.arizona.edu	Professor	Systems and Industrial Engineering
Bryan Heidorn	heidorn@email.arizona.edu	Director	School of Information
Bonnie Hurwitz	bhurwitz@email.arizona.edu	Assistant Professor	Agricultural-Biosystems Engineering
Ozan Jaquette	ozanj@email.arizona.edu	Assistant Professor	Education Policy Studies and Practice
JP Jones	jpjones@email.arizona.edu	Dean	Social and Behavioral Sciences
Stephen Kobourov	kobourov@cs.arizona.edu	Professor	Computer Science
Frank Lederman	FrankLederman@comcast.net	Facilitator	
Brian Liesveld	brianliesveld@email.arizona.edu	Prog. Mgr., Def. & Sec. Res. Inst.	Office for Research & Discovery
Wei-Hsuan Jenny Lo Ciganic	lociganic@email.arizona.edu	Assistatnt Professor	Pharmacy Practice and Science
David Lowenthal	dkl1@email.arizona.edu	Associate Department Head	Computer Science
Yves Lussier	yves@email.arizona.edu	Associate Director	BIO5 Informatics
Eric Lyons	ericlyons@email.arizona.edu	Assistant Professor	School of Plant Sciences
Fernando Martinez	fernando@arc.arizona.edu	Director	BIO5 Institute
Joanna Masel	masel@email.arizona.edu	Associate Professor	Ecology and Evolutionary Biology
Ann McGuigan	amcguigan@email.arizona.edu	Director, Research Development	Office for Research & Discovery
Cecile McKee	mckee@email.arizona.edu	Associate Dean, Research	Social and Behavioral Sciences
Usha Menon	ushamenon001@gmail.com	Associate Dean, Research	College of Nursing
Nirav Merchant	nirav@email.arizona.edu	Director	Info. Tech. Arizona Research Labs
Terry Monks	monks@pharmacy.arizona.edu	Department Head	Pharmacology and Toxicology

Ida Moore	imoore@email.arizona.edu	Professor	College of Nursing
Chris Nelson	acnelson@email.arizona.edu	Associate Dean, Research	Planning
Melynda Noble	melnoble@email.arizona.edu	Administrative Associate	Office for Research & Discovery
Lisa Ordonez	lordonez@email.arizona.edu	Vice Dean	Eller College of Management
Kim Patten	kjpatten@email.arizona.edu	Associate, Research Development	Office for Research & Discovery
Jon Pelletier	jdpellet@email.arizona.edu	Professor	Geosciences
Ryan Perkl	rperkl@email.arizona.edu	Professor	Computer Science
Larry Peterson	larrypeterson@email.arizona.edu	Assistant Professor	Planning
Phil Pinto	ppinto@as.arizona.edu	Associate Professor	Astronomy
Todd Proebsting*	proebsting@email.arizona.edu	Department Head	Computer Science
David Raichlen	raichlen@email.arizona.edu	Associate Professor	School of Anthropology
Sudha Ram	ram@eller.arizona.edu	Professor	Management Information Systems
Ken Ramos	ksramos@email.arizona.edu	Associate Vice Dean	Percision Health Sciences, Col. of Med.
Joaquin Ruiz	jruiz@email.arizona.edu	Dean	College of Science
Joellen Russell	jrussell@email.arizona.edu	Associate Professor	Geosciences
Jeffrey Schatzberg	jschatzberg@eller.arizona.edu	Dean	Eller College of Management
Carlos Scheidegger	cscheid@email.arizona.edu	Assistant Professor	Computer Science
Steve Schwartz	sschwartz@email.arizona.edu	Professor	Chemistry and Biochemistry
Tricia Serio	tserio@email.arizona.edu	Department Head	Molecular and Cellular Biology
Joan Shaver	jshaver@email.arizona.edu	Dean	College of Nursing
Yotam Shmargad	yotam@email.arizona.edu	Assistant Professor	School of Information
Richard Snodgrass	rts@cs.arizona.edu	Professor	Computer Science
Michelle Strout	mstrout@email.arizona.edu	Professor	Computer Science
Mihai Surdeanu	msurdeanu@email.arizona.edu	Associate Professor	Animal and Comparative Biomed. Sci.
Brian Ten Eyck	bteneyck@email.arizona.edu	Assist. Dean, Research Develop.	College of Engineering
Willem Van Leeuwen	leeuw@email.arizona.edu	Associate Professor	Geog. & Dev. & Natural Res. & Environ.
Donata Vercelli	donata@arc.arizona.edu	Professor	Cellular and Molecular Medicine
Joe Watkins	jwatkins@math.arizona.edu	Professor	Mathematics
Austin Yamada	akyamada@email.arizona.edu	Director, Def. & Sec. Res. Inst.	Office for Research & Discovery
Michele Zacks	mzacks@email.arizona.edu	Associate, Research Development	Office for Research & Discovery
Dennis Zaritsky	dzaritsky@as.arizona.edu	Professor	Astronomy
Dajun Zeng	zeng@email.arizona.edu	Professor	Management Information Systems
Hao Helen Zhang	haozhang@email.arizona.edu	Professor	Mathematics

Attendees names in green; other invitees in red

Informatics Strategic Planning Workshop
Flipchart Notes – Aug. 28

Physical Infrastructure:

- Secure, HIPAA-compliant, ITAR-compliant computing environment
- Building for iAnalytics / informatics / “Data7” similar to BIO5
- Protected science data nework
- Research data archive
- On-demand resources for large projects

Major Collaborations:

- Center for iAnalytics / informatics / “Data7”
- Genomics / environment / phenotypes
- Data / network privacy
- Research computing

Other Issues and Needs:

- Understaffed in analytics
- Barriers created by RCM (for ICR)
- How to maintain virtual network and share best practices
- Criteria for promotion / tenure discourage work on proposals
- Data science training
- Visibility for research computing on campus
- Roles for library and UITS
- Develop faculty for large center grants
- Faster development of small proposals
- Making full use of our influence with funding agencies
- Strategy for increasing influence in DC

Strategic Planning Workshop

Informatics

August 27-28, 2015

Frank Lederman

2

Goals for Day Two

- Breakout sessions for each program theme (2 x 80 min)
- Report-backs to entire group
  - What are the most exciting ideas?
  - Summarize key results & issues
  - What items need further work?
- Group discussion
  - Review strategic issues and common resource needs
  - Develop action plan

Breakout Sessions - Key Topics

- What makes (or could make) us unique?
- What grand challenges are addressed?
- What new strategic technical capabilities?
  - What new competitive advantage?
  - What can we demonstrate in the short-term?
  - How can capabilities be extended to other parts of the university?
- Specific opportunities for external funding
- Resources required
  - Program-specific
  - Infrastructure
- What partnerships can help us?
- Other program risks
- Strategic issues

We Need a Strategy to:

- Support the vision and/or strategy of parent organization
- Guide allocation of resources
  - What will grow and what will shrink?
- Show focus
- Inspire objectives that are *required* to achieve the vision
- Help unify the organization around a common set of goals

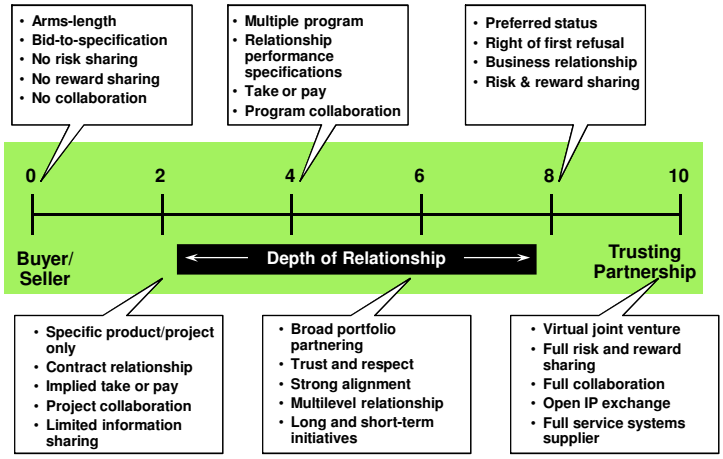
Strengths and Weaknesses

- Internal focus – emphasis on competitive capabilities
- What are our core competencies?
  - Transcend several departments or groups
  - Provide true competitive advantage / differentiation
  - Produce a barrier for others to enter
- Examples
  - Unique facilities difficult and/or expensive to duplicate
  - Unique skills, such as a critical mass of talent in one area that took years to establish
  - Local climate / ecology
  - Key partnerships / relationships
  - Weaknesses: resource limitations in required areas

What can we do?

Sources and Partners

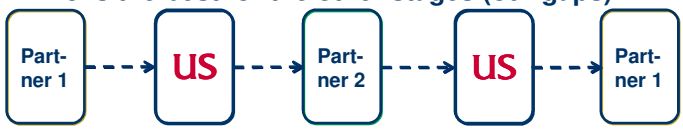
Classifying the nature / depth of a relationship



How deep of a relationship do you need?

Strategic Relationships

- **Analyze the "supply chain" of your technology**
  - For what capabilities / stages are we truly the best?
  - We can afford to be the best for only one or two stages
  - Who is the best for the other stages (our gaps)?



- **Analyze the strategy for possible partners**
  - What is *their* strategy for their gaps (why they need us)?
  - If they won't tell us, put ourselves in their shoes & guess

A mutually-strategic relationship is a competitive advantage

Opportunities and Threats

- External focus
- What are the external forces affecting your organization?
  - What changes have you seen and do you envision?
- Examples
  - Evolving model for competing for federal funding
  - New budget constraints
  - Changing demographics for students, legislature, philanthropy, …
  - Changes in partners and / or their strategies
- What are the strategic options to respond?
  - What strategic capabilities are required?
  - Which programs should grow? Which should shrink?

What should we do?

Breakout Sessions – Key Topics

- What makes (or could make) us unique?
- What grand challenges are addressed?
- What new strategic technical capabilities?
  - What new competitive advantage?
  - What can we demonstrate in the short-term?
  - How can capabilities be extended to other parts of the university?
- Specific opportunities for external funding
- Resources required
  - Program-specific
  - Infrastructure
- What partnerships can help us?
- Other program risks
- Strategic issues

Breakout Session Locations

A. Cloud Commons & Persistent Data
— Visitor Center Conf. Rm. 1 John Hartman

B. Heterogeneous Dataset Integration
— VC Conf. Rm. 2 Ivo Abraham & Yves Lussier

C. Data Analytics: Visualization, Text Mining, Statistical Analysis
— Casita 1900 Bryan Heidorn, Hong Cui

D. Necessary Infrastructure
— Sahara Room Nirav Merchant

Breakout Session A

Cloud Commons and Persistent Data

John Hartman

The premise of this breakout was that the University of Arizona should develop a "cloud commons" centered on environmental science. A "cloud commons" is a space in the cloud in which scientific data are stored and experiments performed, allowing researchers to easily collaborate and build off each other's results. One of the initial discussions focused on whether or not a cloud commons is likely to happen. It was pointed out that the university will likely retain its own computing resources, cloud or otherwise, because of security and privacy concerns, costs, and the need to train scientists on high-performance computing.

One of the challenges in developing a cloud commons occurs because of how computing resources are currently funded. Typically, computing resources appear as capital expenditures in research grants, with on-going operational costs borne by the university. This gives PIs a disincentive to use the cloud where all costs are explicit. However, just because some costs are not directly visible to a PI doesn't make them go away, and it seems likely that cloud service providers will be able to offer commodity computing and storage resources at a lower price than the university. For this reason, there was general consensus that in the near future at least some of the university's research will be done in the cloud and it would be good to plan accordingly.

One of the University of Arizona's strategic advantages in developing a cloud commons is that it has expertise in dealing with large, high-value datasets, including iPlant, iMicrobe, Tara Oceans, LSST, and the Catalina Sky Survey, to name a few. Developing procedures and processes for handling large datasets will be integral to a cloud commons. They must be curated and documented so that they are used properly. Software required for experimental computations and simulations must be packaged and included with the datasets in a way that allows others to use them. Experimental results must be annotated and attributed properly. All of this must occur throughout the data lifecycle, which may extend well beyond the initial funding that generated the dataset.

The general consensus of the group was that the University of Arizona should develop a facility for helping faculty perform research on large datasets in the cloud. This facility would provide faculty with the necessary resources and personnel to develop a cloud commons for their research programs, covering topics such as security and privacy, documentation, provenance, distribution services, data representation, discoverability, cloud-optimized algorithms, and assessment. There was discussion as to what form this facility should take. A center containing dedicated personnel was discussed, but it was noted that whatever form it takes the facility must enable cloud-commons innovation so as to attract the top talent.

Breakout Session B

Heterogeneous Data Integration (HDI)

Yves Lussier and Ivo Abraham

1. What makes us unique?

4 type of uniqueness:

1) niches of research excellence in data science over heterogeneous datasets
   a. Ontology
   b. Climate sciences air/water
   c. Environmental
   d. Security
   e. Genome/Phenotype
   f. Privacy
   g. Social sciences

2) Prestigious UA centers and institutes and prominent programs working in heterogeneous datasets

3) Unique datasets of international value.
   a. iPlant
   b. Banner UA environment + Health Plan
   c. evolving iEarth
   d. Asthma, chronic illnesses, etc.

2. What grand challenges are being addressed?

1) Several challenges beyond the theme of HDI are being addressed in a niche researcher format (e.g. feeding the world, climate, precision medicine, etc.).
2) Team science projects
   a. Research in semantic interoperability and ontology
3) Comprehensive projects (in order of opportunities)
   a. Genome/environment/phenotype
   b. Preserving privacy while integrating data

   A rising tide lifts all boats (innovation & technical support)
   c. Making structured data more accessible
   d. Make unstructured data in structured data
   e. Portability of data solutions

4. What are specific opportunities for external funding?

NSF, NIH, DOD, DOE, Philantropy, for profit and non-profit corporations/organizations

5. **What resources are required (program-specific and/or infrastructure)?**

- Data scientist faculty and technical staff, service group, resourcing key pilots in team science.
- HIPAA compliant computing environment
- Security of UITS system

5. **What partnerships would help us?**
   Stanford + Buffalo + Oregon (ontologies); Chinese Earth & Soil Institute + National Oceanic & Atmospheric Admin, Vanderbilt +UCSF (genome/phenome), UCSD (Privacy computations), Defense & government support corporations.

7. **Other program risks**
   - Universities above-mentioned competitors,
   - Lack of legislative support in lobbying for (e.g. earmarks)

Breakout Session C

Data Analytics: Visualization, Text Mining, Statistical Analysis

Brian Heidorn and Hong Cui

What is analytics? **Analytics should lead to actionable discoveries**

1. What makes us unique?

iPlant, Bio5
Statistics of GIDP with 30 faculty, law, statistics, etc.--open the doors to collaborate.
Systems Biology area is open
RCGC: Big Data taxonomy, need methods to classify people
#1: Eller dept; Optics, LPL,

Large Data sets!

Astronomy: eg. LSST
- LPL
- Astronomy Corporations
- Optics

Medicine & Health: Universities others are 5 years ahead
- Others do genomic data already funded
- Area of aging, fall detection, large aging population, eye care group
- Biodetection, senior citizen

Location/Environment strength and weakness: Dry, aging, poor… population studies; first and third world
Genomics
Earth Science
Security: Cyber security
Space Situational Awareness

Space situational awareness, astronomy
Statistics
Simulations in astronomy
Visualization
Transportation: connected vehicles
Social movements online, online protests
We have no statistics department… GIDP is a strength

2. What grand challenges are being addressed?

Is the data good or bad, is the science easy or hard… silver
Some tools might be better web sites and maybe a database?

Is there life on other planets
What are the origins of life?
Precision Medicine:
Climate change in marginal environments
Social and political science: How to get people to fix the climate?
- How to get people to eat well?

Nirav: cyberinfrastructure
- Domain science + statistics + engineer

3. What new capabilities will be developed?

No one knows what other people are doing?
Need to form methods to collaborate
Campus-wide seminar on campus: top 20 list of campus stuff
Leverage strengths to other areas. LPL
Underpowered in statistics
Few courses to teach analytics
Statistics projects - to NSF fellows Joe
How can we help researchers adapt their techniques to new areas of research. e.g. recommender system for application of analytic techniques?
Need physical space as well as online collaboration and information exchange
Need corporate relationships to generate real problems in the real world

What places do communication well?
U of Washington has University-wide lectures
Yale provides lunch matchmaking
Person assigned to do matchmaking
Co-taught classes across fields

Need to educate other fields about the settled science and breaking science

What longer-term competitive advantage will be created?

Add a couple of people who can work
Fire detection from satellites
Tenure track faculty's concern about interdisciplinary work may threaten their tenure case.
Center in Analytics -focused or broaden? theoretical vs. applied vs. both? initiated top-down or bottom-up?
Need secure data store (w/ security and computation)
Research Computing Committee may be tasked
Knowledge management system for data analytics (a social network of ….)
Collaboration with industrial partners
RDS assoc deans of research, help connect across colleges.
Cluster hires: add non-tenure track

NRT: Industry partners,
Need to establish statistics departments: now departments need their own statistics courses
Concept for the center for data science: high dimensional data science
GIDP in statistics is a plus if it helps people get together
Virtual institute
There are 20 data science centers in the country as examples
Bring together statisticians, CS and engineering sometimes and domain scientists
Need to identify from strengths above
Need people to make new stats then new algorithm
Space situational awareness
Need education outreach

What could be an early demonstration of the new capabilities?

shortage of 1.5 data scientists: graduate

Talking to kids
Need statistics taught to the experts in the domains.

How can these capabilities be extended to other parts of the university?

Many centers have built a building. VPR building
May have service units not on tenure track.
Science UITS, data scientists. Not part of college but university resource
Career path for PhDs who will not be faculty.
Institute for support tenure track people succeed. (Rotate through)
Allow young faculty to double or triple support
People can change tracks from teaching to instrumentation

Interactive working groups: virtual

Center for space awareness: people responsible to make it happen
Space Situational Awareness
Transportation situational awareness
Need bridge funds to keep centers together

4. What are specific opportunities for external funding?

Space Situational Awareness
LSST: security playground Homeland Security
Internships for graduate student support
DARPA
DOT IDIQ: iPlant for transportation
NIH largest funder in life science. UA lower than peers
Analytics in K-12 education
Institutes exist in other places. Intel has supported
Berkeley had data science institute
Moore Foundation Data Driven Discovery
Unique about Tucson: Raytheon, border region
Align org charts in colleges to include Assoc Deans of research.

Breakout Session D

Infrastructure

Nirav Merchant

Institutional Strengths/Unique

- Presence of large projects
  - Astronomy
  - Optics
  - iPlant
  - Geosciences
  - Asthma/ARC
  - Expert domain scientists in many discipline that have “private” data sets
  - Collaborative CISE community

Strategic Needs

- A institutional unit/umbrella organization that focuses on large scale analytics, informatics etc. (Bio5 for informatics ?)
- Houses technical and scientific staff, faculty that are well versed working with “real world problems”
- Has a service center
- Benefits multiple UA groups (Astronomy, Engineering etc)
- Clearing house for best practice (eHarmony, SciVal)

Strategic need

- Data Science Training
  - Special topics workshop
  - Workforce development
  - Intro Statistics
  - Professional Masters in Computational Tools
- Secure/Isolated (network) infrastructure platform for ITAR, Industry, Medical/Protected information.
- Campus Science DMZ equivalent for “protected science data”
- Long term archive for research data

Strategic Needs (for large projects)

- Integrated Project Assessment (College of Ed.)
- Integrating STEM center
- Provide “on demand” resources for special projects need (temporary large space etc.)
- Elevate the visibility/presence of “Research Computing” at higher level for campus
- “inoculate” the group for “big data” training
- Role of library and UITS (more domain awareness)
- Sharing of best practices/collaboration via “monthly party”

Opportunities

- Many large scale NSF funded in earth sciences
- Standard DOE, NIH etc
- Space Sciences
- Industry collaborations for large data analysis (establish domain specific consortia at UA)
- UA cross discipline Data Management (Data Commons) capabilities for other universities as fee for service
- International collaborations for establishing similar cyberinfrastructure (Chile, Middle east)

Opportunities

- 100% engagement goals, provide cohesive training and learning opportunities (mentored), embedded indviduals
- Similar to Academic advising/Stat has a good example

Barriers to Productivity

- Institutional systems
  - Making it efficient to develop new proposals (large)
  - Preparing faculty to apply/manage large center grants
  - RDS services are helping, more visibility for it
  - Quick turn around on smaller budget proposal (ez pass)