

Center for Large Scale Data Systems San Diego Supercomputer Center

Member Prospectus Research Portfolio 2018-2019

James Short and Chaitan Baru
Co-Directors
September 2018
Version 1.0

Website: www.clds.ino

What is Covered in this Brief?

- **Overview of the Center for Large Scale Data Systems Research (CLDS) at SDSC, the host of Data West**
- **Membership Information**
- **Research and Education Portfolio 2018-2019**

History

- **Founded in 2012, the mission of CLDS is to help senior executives meet the challenges of leading dynamic, data-intensive organizations in tomorrow's analytics and information-driven markets**
- **Through rigorous, field-based research, informing CLDS events and education, the Center brings together senior business and technology management to address common themes of data discovery, technology adoption and organizational change**
- **More than forty organizations participate in CLDS events, education and research**

Recent Mentions of CLDS

May 6, 2017 issue of the Economist, lead article “Data is giving rise to a new economy.” Cites CLDS research on data value

<https://www.economist.com/news/briefing/21721634-how-it-shaping-up-data-giving-rise-new-economy>

Sloan Management Review. Spring 2017 article “What’s Your Data Worth?” (J. Short & S. Todd). Listed in SMR’s annual compendium of most read articles of the year

<https://sloanreview.mit.edu/article/whats-your-data-worth/>

MIT CDOIQ Symposium Interview with J. Short & S. Todd on Data Value, theCUBE

<https://www.youtube.com/watch?v=yiReE3STasE>

“Reflections on DataWest” by Paul Barth, CEO Podium Data

<https://www.podiumdata.com/blog/reflections-datawest>

Thumbnail Sketch: BlockLAB

SDSC's new blockchain research lab

Key Questions

What are the principal technology challenges in blockchains (BC), distributed ledgers, smart contracts, hybrid BC-database architectures, and related functionalities? What are the implications of BC for existing platform architectures and data governance policies? How is the blockchain marketplace evolving?

Core Concepts

Technical and economic research. On the technical side, scaling, sizing, performance, evaluation of competing DLT approaches, platform architectures. On the economic side, non-financial value exchange, trust, smart and secure smart contracts, regulation, market evolution and development

Current Activities and Outputs

Inaugural BlockLAB Sponsors Board and research workshop, Sept 25-26

Webpage: <https://www.clds.info/blocklab.html>

Thumbnail Sketch: The Business Value of Data

Key Questions

What are the current, state of the art approaches and tools in place to define the business value of data? What new tools / approaches are needed? Once data value is determined, what do we do with it? What are the data risks? Who has responsibility to define the value of data in a company?

Core Concepts

Top-down versus bottom up (heuristic) approaches to valuation. The set of business activities defining the need for valuation, classification of activities, instrumenting systems and software to provide the required information for data valuation

Current Activities and Outputs

Three related projects: (1) visualizing data relationships (heuristics approach); (2) data value business modelling (data value policies and valuation events, including monetization); and (3) valuation use cases

Thumbnail Sketch: I3 Smart City Consortium

New USC / City of LA Initiative

The Intelligent Internet of Things Integrator (I3) consortium is being formed at the USC Viterbi School of Engineering to design, develop, test and deploy IOT systems to benefit City of LA residents and industry

Core Concepts

To build community based IOT networks, constructed from groups of citizens, companies, and other entities that agree to work together on a voluntary basis to create consolidated data streams from multiple IoT devices in one common stream. To work out the engineering needed to do this, and to study the economic and behavioral impacts

Current Activities

Relationship building and initial strategy and design discussions are underway. A series of pilot projects will be kicked off in the Spring. Interested parties should contact Jerry Power at gerardpo@marshall.usc.edu and/or Jim Short at jshort@sdsc.edu

Announcement: <https://viterbischool.usc.edu/news/2017/11/smart-city-consortium-formed-usc/>

People

Center Directors



Chaitan Baru, Distinguished Scientist, SDSC

Computer scientist, has published in numerous CS areas, 15 years at SDSC, was at IBM and Univ. of Michigan prior to joining SDSC. On leave 2017-2018 at the National Science Foundation where he is Senior Advisor for Data Science in the Computer and Information Science & Engineering Directorate.



James Short, Lead Scientist, CLDS

Information scientist with principal degrees in communications and management information systems from MIT. Prior faculty appointments at the London Business School, Boston University and the MIT Sloan School. Ten years at UCSD and SDSC, was on the faculty at the MIT Sloan School prior to joining UC San Diego.

People

Principal Researchers & Affiliated Faculty



Lynda Applegate, Professor, Harvard Business School

Lynda M. Applegate is the Sarofim-Rock Professor of Business Administration at HBS and Chair of the HBS Executive Education Programs for Business Owners & Entrepreneurs, which includes the Owner President Management (OPM) program, Launching New Ventures and Private Equity/Venture Capital, and the Global Immersions for Business Owners and Entrepreneurs programs.



Mel Horwitch, Dean (ret), Central European Univ Business School

Mel Horwitch has served as Dean of the Central European University Business School, Budapest Hungary, Professor of Technology Management and Director of the Institute for Technology and Enterprise at the Polytechnic Institute of NYU, and founding Dean of Management at Theseus Institute in Sophia Antipolis, France. Horwitch's research focuses on global innovation and entrepreneurship.



Gabriel Silva, Professor, Bioengineering; Director, CENI, UCSD

Gabe Silva's research interests are in neuroscience and neural computing and engineering. He studies the neurobiology of neural signaling at cellular and cellular network scales in order to learn about the mechanisms that underlie neural computation and information processing in the brain. He achieves this by integrating advanced mathematics with experimental neurobiology. Silva is the founding director of the Center for Engineered Natural Intelligence (CENI).

People

Principal Researchers & Affiliated Faculty



Christine Kirkpatrick, Division Director, SDSC

Christine holds a Masters degree in Architecture-based Enterprise Systems Engineering from UCSD. Kirkpatrick is Division Director for IT Systems & Services at SDSC, deputy director of the NSF sponsored Western Region Big Data Innovation Hub, and the first Executive Director of the National Data Service (NDS), a U.S. based data research consortium.



James Meng, Deputy Assistant Secretary of the Navy (ret)

James Meng entered the U.S. Senior Executive Service in 1998 and retired on June 27th, 2015 as Deputy Assistant Secretary of the Navy, Business Enterprise Solutions, Office of the Secretary of the Navy. Dr. Meng led development of common data standards, system architectures and integration of business intelligence for the Department of the Navy. He established the Navy's large-scale system of systems strategies toward a boundary-less enterprise. Dr. Meng's awards include the Superior Civilian Service Medal, 2015, and the Navy Senior Executive Service Performance Awards in 2000 and 2015.



Jerry Power, Executive Director, CTM, USC Marshall School

Jerry Power is Executive Director of The Institute for Communications Technology Management (CTM) at University of Southern California's Marshall School of Business. CTM is a central player in USC's I3 Smart City Consortium, working with the City of Los Angeles in citizen directed, IOT deployment. Previously, Jerry worked for Alcatel-Lucent, Siemens and Motorola, holding positions in Marketing, Corporate Strategy, Business Development, Product Management and Product Development.

People

Industry Advisors



Kirk Borne, Principal Data Scientist, Booz Allen Hamilton

Kirk Borne is the Principal Data Scientist at Booz Allen Hamilton. He previously spent 12 years as Professor of Astrophysics and Computational Science at George Mason University where he taught and advised students in the Data Science degree programs. Before that, he worked for 18 years supporting NASA projects in various roles, including Data Archive Project Scientist for the Hubble Space Telescope. He is an active contributor on social media, where he has been named consistently the top worldwide influencer in big data and data science since 2013.



Steve Orrin, Chief Technologist, Intel Federal

Steve Orrin is Chief Technologist, Intel's Federal Division & Intel Federal LLC. Steve is responsible for Cyber Security and Cloud Strategy, Solution Architecture and Engagements. Steve is a recognized expert on enterprise security and was named one of InfoWorld's Top 25 CTO's of 2004. In 2009, Steve was a Fellow at the Center for Advanced Defense Studies. Steve is active in several professional security associations, a co-Founder of the Web Application Security Consortium, and Vice-Chair of the NSITC/IDESG Security Committee.



Hus Tigli, Founder and CEO, Xaxar Inc

Hus Tigli is the Founder and Chairman of Xaxar, established to bring innovative solutions to data centers and Cloud operations. Hus has over 25 years of experience in top leadership of technology businesses — 10 years in large international company settings at Raychem, responsible for \$900 million in revenues, and the last 15 years as the CEO or founder of start-ups in mixed signal integrated circuit design, biochips and photonic switches. Hus holds BS and MS degrees in engineering from Columbia University and an MBA from Harvard.



Membership

All information current as of September 2018

Member Benefits and Terms

Benefits

- Two half day research workshops on agreed topics annually
- Quarterly webinars
- Pre-release access to all working papers, presentations and publications
- Two (2) free seats at SDSC's flagship, annual Data West conference in December (*value of \$1500*)

Sponsorship Term Options

Option 1

The term of membership is one year, renewable annually, currently at the rate of \$6,000

Option 2

The term of sponsorship is two years, currently at the rate of \$10,000

Research Sponsor

Benefits

Research sponsors receive a tailored set of benefits based on project discussions with CLDS principals resulting in an agreed upon statement of work. Projects can be individual or pooled with other sponsors. Research sponsors receive all the benefits of Consortium Members.

Sponsorship Options

Option 1

Entry level, proof of concept / exploratory research typically starts at \$25,000 per year depending on the extent and depth of agreed-upon outputs. Research in this category is usually conducted by a single PI and does not require significant access to SDSC computational or programming resources. This category is most appropriate for exploratory research where little prior work exists, and questions and appropriate methodologies cannot be precisely defined.

Option 2

Entry level, exploratory research conducted by two or more PIs and making use of SDSC computational and programming resources typically start at \$50,000 per year, depending on the extent and depth of outputs agreed. This category of work is most appropriate after exploratory research has been concluded, and the design of experiments, primary data collection, and other methodologies are used to study more precisely defined questions.

Laboratory Sponsor

Sponsors in this category have usually conducted exploratory research and are now ready to pursue a deeper engagement, often multi-year and often with additional sponsors in a pooled sponsorship arrangement. There is no fixed formula to work conducted in this category, and sponsor pools can include public organizations and non-profits such as foundations. Laboratory sponsors are typically vendor / services organizations that plan to share research findings with others, including customers.

CLDS is spinning up a new laboratory, BlockLAB, in September 2018.

The Business Value of Data

Tools, Practices and Policies for Defining and Measuring the Value of Data

BlockLAB

New SDSC blockchain R&D laboratory

I3 Smart City (joint with USC)

Citizen directed, IOT deployment in Los Angeles

How Much Information?

An ongoing series of studies reporting on the volume and growth rate of data generated today, covering consumer media and enterprise data

Projects 2018-2019

All information current as of September 2018

Data Value

Composite group of projects researching data valuation

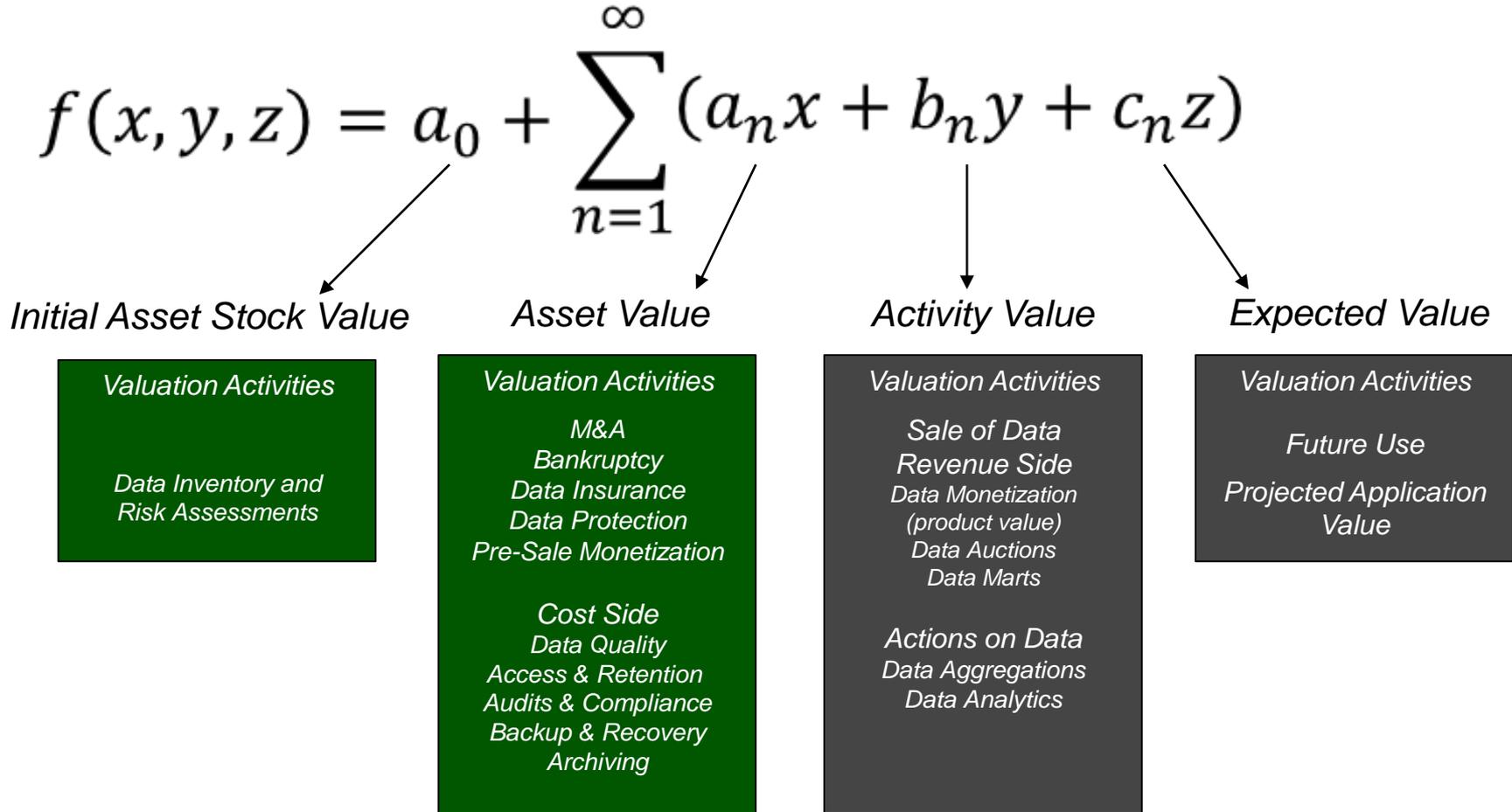
- Projects combine both heuristic approaches, focused on visualizing and quantifying data relationships *at the level of the data*, and top-down approaches, focused on the business exchange of data in data markets, data valuation in mergers and acquisitions, and securing and pricing data insurance based on the value of the data being insured

Three Projects

- Exploratory Proof of Concept Project: Blockchain and Distributed Value (with Steve Todd, Dell Technologies)
- I3 Smart City: Citizen directed, IOT applications deployment in Los Angeles (with Jerry Power, USC CTM)
- Report series on How Much Information? How much data / information is there? How is it created? Where is it? What happens to it?

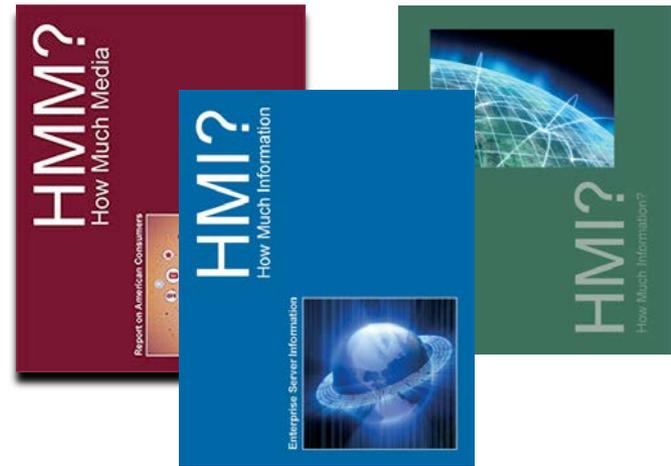
Example Valuation Model

Data Value is the sum of principal value components summed over the number of valuation events in an organization. Plus an initial data asset stock value.



How Much Information?

HMI Report Series



Goal

- To produce a rigorous set of baseline studies of data growth
 - Three reports completed (2009, 2011, 2013)
 - Five-year update reports planned for 2018 (consumer, enterprise)

Key Findings from HMM 2015

- **In 2009**, Americans consumed concurrent media for approx 1.19 trillion hours, an average of 11.4 hours per person per day. Bytes presented totaled over 11 zettabytes, or 106 gigabytes a day for the average media consumer (uncompressed), and 35 gigabytes a day compressed.
- **By 2012**, consumption had increased to 1.46 trillion hours, an average of 13.5 hours per person per day, and bytes presented totaled over 27 zettabytes, or 252 gigabytes a day for the average media consumer uncompressed, and approximately 85 gigabytes compressed.
- **By 2015**, hours consumed increased to 1.64 trillion hours, an average of over 15 hours per person per day.



Data West

**SDSC's Annual
Data Technology Conference
December 5-6, 2018**

www.datawest.org



For More Information

Contact Co-Director James Short
jshort@sdsc.edu