

**LSST-SCMA: New collaboration application****Received:** 21 Sep 2009 13:35:52**Keywords:** Other, Other, Other**Name:** Informatics and Statistics**Proposal for an LSST Informatics and Statistics Science Collaboration****Thomas J. Loredo** (*Cornell University, Dept. of Astronomy*), Senior Research Associate

Space Sciences Building, Cornell University, Ithaca, NY 14853-6801, USA

Email: loredo@astro.cornell.edu

Phone: 607-255-6564

**Kirk D. Borne** (*George Mason University, Dept of Computational & Data Sciences*), Associate Professor of Astrophysics and Computational Science

4400 University Drive, MS 6A2, Fairfax, VA 22030, USA

Email: kborne@gmu.edu

Phone: 703-993-8402

**G. Jogesh Babu** (*Pennsylvania State University, Dept. of Statistics*), Professor and Director of Center for Astrostatistics

326 Joab L. Thomas Building, The Pennsylvania State University, University Park, PA 16803, USA

Email: babu@stat.psu.edu

Phone: (814) 863-2837

**Eric D. Feigelson** (*Pennsylvania State University, Dept. of Astronomy & Astrophysics*), Professor

Dept. of Astronomy &amp; Astrophysics, 525 Davey Laboratory, University Park, PA 16802, USA

Email: edf@astro.psu.edu

Phone: 814-865-0162

**Alexander G. Gray** (*Georgia Institute of Technology, College of Computing*), Assistant Professor

266 Ferst Drive, Klaus Advanced Computing Building, Atlanta, GA 30332, USA

Email: agray@cc.gatech.edu

Phone: 404-933-0666

**John A. Rice** (*University of California, Berkeley*), Emeritus Professor

University of California, Berkeley, Department of Statistics # 3860, Berkeley, CA 94708, USA

Email: jariceiii@gmail.com &lt;mailto:jariceiii@gmail.com&gt;

Phone: 510 649 8168

**Joseph William Richards** (*Carnegie Mellon University, Department of Statistics*), Graduate Student

5000 Forbes Avenue, Department of Statistics # 3860, Pittsburgh, PA 15213, USA

Email: jwrichar@stat.cmu.edu &lt;mailto:jwrichar@stat.cmu.edu&gt;

Phone: 626-222-4422

**David NMI Ruppert** (*Cornell University, School of Operations Research and Information Engineering and Dept. of Statistical Science*),

Andrew Schultz Professor of Engineering

225 Rhodes Hall, 1170 Comstock Hall, Ithaca, NY 14853, USA

Email: dr24@cornell.edu &lt;mailto:dr24@cornell.edu&gt;

Phone: 607 255 9136

**Naoki Saito** (*University of California, Davis, Dept. of Mathematics*), Professor of Mathematics

One Shields Avenue, , Davis, CA 95616, USA

Email: saito@math.ucdavis.edu &lt;mailto:saito@math.ucdavis.edu&gt;

Phone: (530) 754-2121

**Chad Michael Schafer** (*Carnegie Mellon University, Dept. of Statistics*), Assistant Professor

5000 Forbes Avenue, , Pittsburgh, PA 15206, USA

Email: cschafer@stat.cmu.edu &lt;mailto:cschafer@stat.cmu.edu&gt;

Phone: 412-268-3967

**Jiayang Sun** (*Case Western Reserve University, Dept. of Statistics*), Professor of Statistics, Professor of Epidemiology & Biostatistics

Department of Statistics, 10900 Euclid Ave, Cleveland, OH 44106-7054, USA

Email: sun@case.edu &lt;mailto:jsun@case.edu&gt;

Phone: 216-368-0630

**Benjamin D. Wandelt** (*University of Illinois, Dept. of Physics*), Associate Professor

Department of Physics, University of Illinois, 1110 W. Green Street, Urbana, IL 61801, USA

Email: bwandelt@Illinois.edu &lt;mailto:bwandelt@Illinois.edu&gt;

Phone: 217 621 4820

**Larry Alan Wasserman** (*Carnegie Mellon University, Dept. of Statistics*), Professor of Statistics

Dept. of Statistics, CMU, 5000 Forbes Ave., Pittsburgh, PA 15213, USA

Email: larry@stat.cmu.edu &lt;mailto:larry@stat.cmu.edu&gt;

Phone: 412-268-8728

**Michael Joseph Way** (*NASA Goddard Institute for Space Studies*), Astronomer

NASA Goddard Institute for Space Studies, 2880 Broadway, New York, NY 10025, USA

Email: michael.j.way@nasa.gov &lt;mailto:michael.j.way@nasa.gov&gt;

Phone: 914-512-8068

**Michael Barrett Woodroffe** (*University of Michigan, Ann Arbor Statistics*), Professor Emeritus of Mathematics and Statistics

**LSST-SCMA: New collaboration application****Received:** 21 Sep 2009 13:35:52**Keywords:** Other, Other, Other**Name:** Informatics and Statistics

Statistics Department, University of Michigan, Ann Arbor, MI 48109, USA

Email: michaelw@umich.edu &lt;mailto:michaelw@umich.edu&gt;

Phone: (734) 753 3495

*James O Berger (Duke University, Department of Statistical Science)*, The Arts and Sciences Professor of Statistics

Department of Statistical Science, Duke University, Durham, NC 27708-0251, USA

Email: berger@stat.duke.edu

Phone: 919-684-4531

*Robert J Brunner (University of Illinois, Department of Astronomy)*, Associate Professor of Astronomy

1002 W. Green St., Department of Astronomy, University of Illinois, Urbana, IL 61801, USA

Email: bigdog@illinois.edu

Phone: 217-244-6099

*David F Chernoff (Cornell University, Astronomy Department)*, Professor

602 Space Sciences Bldg., Ithaca, NY 14853, USA

Email: chernoff@astro.cornell.edu

Phone: 607 255 4755

*S. George Djorgovski (Caltech, Astronomy Department)*, Professor of Astronomy

MS 249-17, Cahill Center, Caltech, Pasadena, CA 91125, USA

Email: george@astro.caltech.edu

Phone: 626 395 4415

*David A van Dyk (University of California at Irvine, Department of Statistics)*, Professor

2206 Bren Hall, Department of Statistics, University of California, Irvine, CA 92697-1250, USA

Email: dvd@ics.uci.edu

Phone: 949 824 5679

*Lee Samuel Finn (Penn State (Physics; Astronomy & Astrophysics))*, Professor

104 Davey Laboratory, Penn State University, University Park, PA 16802, USA

Email: LSFinn@PSU.Edu

Phone: 814-863-9598

*Peter Edward Freeman (Carnegie Mellon University, Department of Statistics)*, Research Associate

Department of Statistics, Carnegie Mellon University, 5000 Forbes Ave., Pittsburgh, PA 15213, USA

Email: pfreeman@cmu.edu

Phone: (412) 268-1052

*Christopher R Genovese (Carnegie Mellon University, Department of Statistics)*, Professor of Statistics

Department of Statistics, Carnegie Mellon University, 5000 Forbes Ave., Pittsburgh, PA 15213, USA

Email: genovese@cmu.edu

Phone: 412 268 7836

*Matthew James Graham (Caltech, Center for Advanced Computing Research)*, Computational Scientist

Mail Code 158-79, Caltech, Pasadena, CA 91125, USA

Email: mjpg@caltech.edu

Phone: 626 395 8030

*Jon Eric Hakkila (College of Charleston, Department of Physics and Astronomy)*, Professor

Department of Physics and Astronomy, College of Charleston, Charleston, SC 29424-0001, USA

Email: hakkilaj@cofc.edu

Phone: (843) 953-6387

*Woncheol Jang (University of Georgia, Department of Epidemiology and Biostatistics)*, Assistant Professor

129B Coverdell Center, College of Public Health, University of Georgia, Athens, GA 30602, USA

Email: jang@uga.edu

Phone: 706-583-8923

*William H. Jefferys (University of Texas at Austin and University of Vermont)*, Harlan J. Smith Centennial Professor of Astronomy (UT),

Adjunct Professor of Statistics (UVM)

253 Strong Road, Moretown, VT 05660, USA

Email: bill@astro.as.utexas.edu

Phone: 802-496-7760

*Vinay L. Kashyap (Smithsonian Astrophysical Observatory, High-Energy Astrophysics Division)*, Astrophysicist

60 Garden St., MS-83, Cambridge, MA 02138, USA

Email: vkashyap@cfa.harvard.edu

Phone: 617-495-7173

*Kevin H. Knuth (University at Albany, Departments of Physics and Informatics)*, Associate Professor

University at Albany, Physics 211, 1400 Washington Ave., Albany, NY 12222, USA

Email: kknuth@albany.edu

Phone: 518-442-4653

*Eric D. Kolaczyk (Boston University, Department of Mathematics and Statistics)*, Professor and Director, program in Statistics

111 Cimmington St., Boston, MA 02215, USA

**LSST-SCMA: New collaboration application****Received:** 21 Sep 2009 13:35:52**Keywords:** Other, Other, Other**Name:** Informatics and Statistics

Email: kolaczyk@math.bu.edu

Phone: 617-353-5208

*Ji Meng Loh (Columbia University, Department of Statistics), Associate Professor*

Dept of Statistics, Columbia U, 1255 Amsterdam Ave., New York, NY 10027, USA

Email: meng@stat.columbia.edu

Phone: 212-851-2139

*Bruce M. McCollum (Caltech, IPAC), Staff Scientist*

MS 314-6, 1200 E. California Blvd., Pasadena, CA 91125, USA

Email: mccollum@ipac.caltech.edu

Phone: 626-395-8573

*Misha (Meyer) Zalman Pesenson (Caltech, IPAC), Sr. Researcher*

MS 220-6, 1200 E. California Blvd., Pasadena, CA 91125, USA

Email: misha@ipac.caltech.edu

Phone: 626-395-8559

*Vahe Petrosian (Stanford University, Department of Physics), Professor*

Department of Physics, 382 Via Pueblo, Stanford, CA 94305-4060, USA

Email: vahep@stanford.edu

Phone: 650-723-1435

*Andrew F. Ptak (Johns Hopkins University, Department of Physics and Astronomy), Research Scientist*

3400 N. Charles St., Baltimore, MD 21218, USA

Email: ptak@jhu.edu

Phone: 410-516-4907

*Ashish Mahabal (Caltech, Astronomy Department), Research Scientist*

MS 249-17, 1200 E. California Blvd., Pasadena, CA 91125, USA

Email: aam@astro.caltech.edu

Phone: 626-395-4201

*Martin D Weinberg (University of Massachusetts, Department of Astronomy), Professor*

University of Massachusetts, Department of Astronomy, Amherst, MA 01003-9305, USA

Email: weinberg@astro.umass.edu

Phone: 413-545-3821

*Douglas J. Burke (Harvard-Smithsonian Center for Astrophysics), Astronomer*

60 Garden Street, MS-2, Cambridge, MA 02138, USA

Email: dburke@cfa.harvard.edu

Phone: 617-496-7853

*Robert Lee Wolpert (Duke University, Dept of Statistical Science), Professor*

Duke Dept of Statistical Science, 211c Old Chem Bldg, Box 90251, Durham, NC 27708-0251, USA

Email: wolpert@stat.duke.edu

Phone: 919-684-3275

*Carlo Joseph Graziani (University of Chicago, Department of Astronomy & Astrophysics), Senior Research Associate*

5640 S. Ellis Ave., Department of Astronomy &amp; Astrophysics, Chicago, IL 60637, USA

Email: carlo@oddjob.uchicago.edu

Phone: 773-702-7973

*Aneta L Siemiginowska (Harvard-Smithsonian Center for Astrophysics), Astrophysicist*

60 Garden St., MS-4, Cambridge, MA 02138, USA

Email: asiemiginowska@cfa.harvard.edu

Phone: 617-495-7243

*Joshua S. Bloom (UC Berkeley, Department of Astronomy), Associate Professor of Astronomy*

601 Campbell Hall, Berkeley, CA 94720, USA

Email: jbbloom@astro.berkeley.edu

Phone: 510-643-4621

*Merlise Aycock Clyde (Duke University, Dept of Statistical Science), Associate Professor*

223E Old Chemistry, Box 90251, Durham, NC 27708-0251, USA

Email: clyde@stat.duke.edu

Phone: 919-681-8440

*Ian Davidson (University of California at Davis, Computer Science Dept), Associate Professor*

Room 3025 Kemper Hall, Computer Science, U.C. Davis, Davis, CA 95616, USA

Email: davidson@cs.ucdavis.edu

Phone: 530-601-0385

*Tamas Budavari (The Johns Hopkins University, Bloomberg Center for Physics and Astronomy), Research Scientist*

Bloomberg Center for Physics and Astronomy, 3400 N. Charles St., Baltimore, MD 21228, USA

Email: budavari@jhu.edu

Phone: 410-516-0643

**LSST-SCMA: New collaboration application****Received:** 21 Sep 2009 13:35:52**Keywords:** Other, Other, Other**Name:** Informatics and Statistics

*Bradley Efron (Stanford University, Statistics Department)*, Professor  
 390 Serra Mall, Sequoia Hall, Stanford, CA 94305, USA  
 Email: brad@stat.stanford.edu

Phone: 650-723-2206

*Donald Q. Lamb (University of Chicago)*, Robert A. Millikan Distinguished Service Professor  
 Dept. Astronomy & Astrophysics, U. Chicago, 5640 South Ellis Avenue, Chicago, IL 60637, USA  
 Email: lamb@oddjob.uchicago.edu

Phone: 773-350-3838

*James M. Cordes (Cornell University)*, Professor  
 520 Space Sciences Building, Cornell University, Ithaca, NY 14853, USA  
 Email: jmc33@cornell.edu

Phone: 607-255-0608

*Adam M. Brazier (National Astronomy and Ionosphere Center, Cornell University)*, Astronomy Programmer  
 522 Space Sciences Building, Cornell University, Ithaca, NY 14853, USA  
 Email: abrazier@astro.cornell.edu

Phone: 607-255-0606

*Jeffrey D. Scargle (NASA Ames Research Center)*, Research astronomer  
 Mail Stop 245-3, NASA Ames Research Center, Moffett Field, CA 94035-1000, USA  
 Email: Jeffrey.D.Scargle@nasa.gov

Phone: 650-604-6330

*Christopher J. Miller (Cerro Tololo Inter-American Observatory/NOAO)*, Assistant Astronomer  
 NOAO/CTIO, 950 N. Cherry Ave. , Tucson, AZ 85719, USA  
 Email: cmiller@ctio.noao.edu

Phone: +1 56 51 205 358

**Abstract of science goals and justification (limit 200 words):**

We propose to create a new LSST Informatics and Statistics Science Collaboration (ISSC) that will pursue research and provide consultation on challenging astroinformatics and astrostatistics problems that arise in addressing a wide variety of LSST science goals. It will focus special effort on classes of problems that are cross-cutting, arising in diverse astrophysical applications. The ISSC will be interdisciplinary, with membership roughly equally divided between astronomers with significant research expertise in astroinformatics and astrostatistics, and information scientists (statisticians and computer scientists) who will bring expertise to LSST from other fields that are addressing discovery and analysis challenges with large data sets. The collaboration will provide LSST scientists access to a pool of expertise that will help astronomers find and adapt cutting-edge developments in the information sciences to the needs of LSST data analysis. Also, it will provide a venue for coordinating LSST information science research addressing cross-cutting data analysis problems. These capabilities will help the entire LSST user community to more efficiently and effectively address methodological challenges; it will also enable new science that would not be possible to pursue using conventional methods and algorithms.

## I. Science Justification (limit 4 pages)

A. Describe the science goals of the proposed collaboration and provide the context for overall significance to astronomy/physics.

The unprecedented science opportunities arising from LSST's wide-field cosmic cinematography will be accompanied by equally unprecedented data analysis challenges, due to the huge size and synoptic scope of LSST data products. While the most obvious challenges are those due to the petabyte scale of fundamental LSST databases, in fact LSST will present astronomers with new and difficult analysis problems spanning a broad range of sizes, types, and complexity, and requiring a matching breadth of methodological research to address. The challenges of science analysis---deriving knowledge and understanding from the vast LSST datasets---will require methods, algorithms, and codes that extend or transcend those in common use in astronomy today.

We propose to create a new LSST Informatics and Statistics Science Collaboration (ISSC) to help coordinate interdisciplinary interactions and pursue data analysis methodology research in collaboration with other Science Collaborations (SCs) and the broader LSST user community; we will also offer our services to the LSST Data Management team. The ISSC will be interdisciplinary, with membership split (currently 60%/40%) between astronomers and information scientists who will bring to LSST expertise gained in other communities facing large and complex data analysis challenges. By design, the ISSC includes a number of astronomer members of other SCs, to facilitate coordination of collaborative and cross-cutting activities. We have created a proposal support web site that includes a categorized team list that displays our team breakdown more clearly than is possible with the proposal personnel entry form; this site is at:

<http://inference.astro.cornell.edu/lsst/proposal09/>

To quickly see the need for the ISSC, envision a "methods cube," where the first dimension is binned by application (e.g., science problems covered by existing SCs), the second dimension is binned by problem scale (measured by the logarithm of the number of objects or database size being analyzed), and the third dimension is binned by task type (e.g., source detection, classification, multivariate density estimation, etc.). Each LSST science project corresponds to an entry in one or more cells in the methods cube. The observation motivating our effort to form the ISSC is that projecting out the application dimension produces a 2-D task/scale grid where relatively few occupied cells will have only one entry. That is, nearly every type of task will be used in multiple application areas, and many task/scale combinations will arise in multiple application areas. Each such coincidence represents an opportunity where research alliances cutting across astronomy application areas can benefit LSST science. Further, other disciplines have their own methods cubes, with numerous applications occupying the same task/scale cells as LSST applications. Each of these coincidences represents an opportunity where interdisciplinary research may benefit LSST science. The ISSC will work to identify promising areas of methodological overlap between applications and disciplines, bring existing methods to bear on these overlap problems, and build alliances pursuing research on new methods where needed. These activities promise to significantly improve the science return from LSST.

We elaborate here on this cross-cutting, interdisciplinary perspective to establish the need for the ISSC and delineate its roles in more detail. We first describe the diversity of problem scales, and then the diversity of analysis tasks that LSST scientists will face, across diverse application areas. This will motivate the main goals of the ISSC:

1. To enable new LSST science by leveraging recent developments in information sciences;
2. To focus consultation from experts in astrostatistics and astroinformatics to serve the full LSST science effort, covering advanced methods and high computational efficiency;
3. To stimulate and guide new, astronomy-focused information science research for application to LSST issues;
4. To improve cross-fertilization and reduce duplication on methodological development across the LSST community.

To establish the context for and significance of ISSC activities, we begin by surveying the range of problem scales spanned by LSST science. We distinguish three scales for the number of objects or samples ("entities") being analyzed (each possibly with multiple attributes). The three scales differ by factors of  $\sim 10^3$ , roughly delineating regimes where computational resource constraints lead to fundamental shifts in how one thinks about data analysis problems.

\* "Kiloscale" ( $10^2$  to  $10^4$  entities) - The fundamental kiloscale LSST problem is analysis of multicolor, multi-epoch photometry for a single object in the Object Catalog, where the entity is a photometric measurement (the actual samples will be obtained from associated entries in the Source Catalog). Much of this will be study of source variability in several bands which falls under the large rubric of multivariate time series analysis. A second class of kiloscale problem arises in population-level analysis of catalog data for modest-sized populations; e.g., trans-Neptunian objects (TNOs;  $\sim 20k$  expected), gamma-ray burst (GRB) hosts (potentially  $\sim 100$  per year), microlensing events, and relatively rare classes of stars, galaxies, or clusters. Kiloscale problems are generally not challenging in data processing or storage, but will benefit from a variety of statistical techniques such as outlier and change point detection, survival analysis (treatment of upper limits), periodicity searches, and statistical inference involving heteroscedastic (different for each sample) measurement errors.

\* "Megascale" ( $10^5$  to  $10^7$  entities) - Megascale problems include population-level analysis of larger populations (e.g., quasars; variable Galactic stars; low redshift galaxies); population-level re-analysis of previous survey catalogs (e.g., SDSS, FIRST) supplemented by LSST follow-up; and analysis of multicolor, multi-epoch image data for extended objects where the entity is a pixel. Megascale problems require processing datasets occupying storage of several megabytes to gigabytes, corresponding to "low-volume" queries against LSST data products. In this regime, statistical methods must be computationally efficient and advanced data visualization techniques are needed.

\* "Gigascale" ( $10^8$  to  $10^{10}$  entities) - A class of gigascale problem using Level-One LSST data products is analysis of calibrated image data for a single LSST field or a modest number of fields. A second class, using Level-Two object catalogs, includes population-level analysis of very

large samples of stars (about 20 billion total expected) or galaxies (about 10 billion total expected). A third class includes the search for rare or serendipitous objects based on queries over large parts of the Level-One source catalog. Gigascale problems require very efficient (and thus limited) processing of data streams from datasets occupying terabytes of storage, corresponding to "high-volume" queries against LSST data products. Several nightly DMS pipeline tasks also fall into this category.

Roughly speaking, kiloscale problems will raise challenges that are essentially statistical: devising methods for handling novel types of data and models that optimally extract information from LSST data and provide careful uncertainty quantification. Gigascale problems will raise challenges that are more essentially computational, in the realm of informatics: finding algorithms that make it possible to do specific, relatively straightforward tasks across enormous databases; sophisticated statistical modeling will typically not be possible. Megascale problems occupy a middle ground, where significant innovation may be required on both the statistical and computational fronts.

Many important problems have a hierarchy of scales. A key example is nightly pipeline processing: localized image processing is needed to identify sources and associate them with previously detected objects; but this apparently kiloscale problem must be executed for billions of sources per night, making the overall problem gigascale and constraining the level of statistical sophistication.

Having identified representative scales for LSST data analysis problems, we now identify representative tasks, many occurring across a range of scales.

**DISCOVERY** - These tasks address detecting and identifying astronomical objects, structures, and events:

- \* Adjusting thresholds to control the number of false detections or associations, taking into account the huge multiplicity of tests that a large survey performs. This is a fundamental DMS task, determining the behavior of nightly detection pipeline processing. Control of the False Discovery Rate (FDR) under multiple testing is a hotbed of statistics research to treat, for example, bioinformatics issues arising in the use of DNA microarray technology. The InCA group (described below) has described initial application of FDR control to astronomical source detection in modest-sized catalogs, pointing to the value of this line of research for LSST astronomy.

- \* Faint source detection in multicolor/multi-epoch data cubes. The DMS deep detection pipeline will address this task for Data Releases (DRs). It will also arise in re-analyses of Level-One data products that attempt to push the boundary of LSST to the dimmest sources, e.g., for TNO, faint star (e.g., white dwarf) and faint (high redshift) galaxy and AGN studies, particularly using "deep drilling" data. While in many cases, the faintest sources will be found in images merged from many epochs, in cases where observing conditions change or transients are present, the faintest sources may be found in localized portions of the data cubes. Finding these sources is both a statistical and computational challenge.

- \* Classification of objects within a population, including both supervised classification (assigning objects to previously-identified classes) and unsupervised classification (where the number and characteristics of classes are derived from the data), arising in the study of all sizable populations, from minor planets to distant galaxies. A wide variety of multivariate classification tools are available (see, for instance, the text 'Pattern Classification' by Duda, Hart and Stork 2002), often far more capable than the traditional color-index cuts familiar in photometric survey analysis. The options are reduced for gigascale problems due to computational limitations.

- \* Flexible and adaptive study of variability and transients in time series. The detection of short-lived transients or marginal variability in sources is a problem in statistical inference where the heteroscedastic measurement errors due to changing observing conditions will play an important role. Once variability is established, both time domain and frequency domain time series techniques are relevant. A number of algorithms for establishing autocorrelated and/or periodic variations can be applied to the time series in both a single photometric band or all bands simultaneously. Simple nonparametric measures like the partial autocorrelation function or the minimal string length measure may be appropriate for scanning for variability and periodicity in gigascale LSST catalogs.

- \* Cross-matching within and between large catalogs, with accurate accounting for astrometric uncertainties. This task arises in all application areas involving correlative analysis, and becomes particularly challenging if the target catalog has significant direction uncertainties. Current astrostatistics research efforts are addressing this task by marrying directional statistics, product partition models, and Monte Carlo methods for searching the space of likely associations and accounting for multiple comparisons. These procedures can take into account knowledge of the local densities of objects in the catalogs.

**MODELING & ANALYSIS** - These tasks involve analysis of images of known sources, or of catalogs of classified objects or events:

- \* Image analysis, including multi-frame super-resolution, and multiscale deconvolution with uncertainty quantification. When instrumental or atmospheric distortions cause blurring, likelihood-based deconvolution based on the known point spread function can sharpen the image. Large volumes of image segments can be reduced and characterized using sparse representations such as wavelets, curvelets or compressive sensing. Density estimation techniques like locally weighted regression can reconstruct image regions with reduced noise and with variance images to help adjudicate the statistical significance of features.

- \* Design and comparison of photometric redshift (photo-z) algorithms, including calibration of redshift uncertainties needed for a wide variety of extragalactic research, ranging from luminosity function estimation to dark energy studies using Type Ia supernovae (SNe Ia) and weak lensing. Relevant information science developments are in multivariate density estimation, Bayesian and neural network clustering, and nonparametric regression and density estimation. These are all areas with many relevant recent developments in machine learning and statistics. Also, statistical study can provide guidance for spectroscopic measurements to improve predictions.

- \* Flexible modeling of multivariate distributions for populations, accounting for selection effects including truncation in a single survey, censoring in a follow-up survey, and measurement error in all surveys. This class of task arises in nearly all population modeling applications; e.g., orbit/size/color distributions of minor planets; color-magnitude diagrams of stars; distance estimator calibration for galaxies (Tully-Fisher and fundamental plane

relations); luminosity functions of galaxies, AGN, and transient populations. Relevant, active research frontiers in information sciences include semiparametric and nonparametric inference with heteroscedastic measurement error, and nonparametric and parametric survival analysis.

\* Characterizing multicolor variability. A wide variety of temporal behavior will be discovered and studied by LSST. Examples include: periodic variability of minor planets (to study rotation, geometry, and composition); periodic variations of stars due to rotation and pulsation including Cepheid and RR Lyrae period-luminosity-color correlations for use as low-*z* distance indicators; characterizing the smooth multicolor light curves of SNe Ia for use as distance indicators in dark energy studies; characterizing the wide range of smooth and chaotic variability across AGN populations. Astronomers have already developed a variety of methods for periodicity searches from unevenly spaced data, although research is needed for incorporation of measurement errors. Relevant frontier information science studies include: nonparametric harmonic analysis; multivariate nonparametric regression with sparse Gaussian processes and PCA-based dimension reduction; and space-time modeling (translated to wavelength-time) with multivariate stochastic processes built over overcomplete bases (for sparse representation).

In addition to the specific information science research frontiers identified above, the general area of algorithm development for large-database proximity searches is having a profound impact on the ability to deploy increasingly sophisticated techniques on ever-larger datasets. Examples include the use of kd-trees and metric ball trees to accelerate nearest-neighbor search, k-means clustering, and Gaussian process regression. These developments play key roles in recent and current analyses of SDSS data.

All areas of LSST research will also need advanced data visualization techniques. For kiloscale problems, "Grand Tour" movies of rotating datacubes with interactive brushing and classification are useful. For megascale and gigascale problems, quantile contour maps and shaded density maps must be rapidly produced from portions of the data. Shaded parallel coordinate maps with brushing may also be powerful interactive visualization tools for multivariate data with more than three dimensions.

The list above is hardly exhaustive, neither in the tasks listed nor in the applications cited. Incomplete though it may be, it already makes clear that LSST will pose an almost dizzying variety of research-level data and science analysis challenges. It also makes clear that there are many opportunities to share expertise and research resources across applications, science collaborations, and disciplines.

In summary, nearly every frontier LSST science effort will face significant astroinformatics or astrostatistics challenges arising from the scale, dimension, and novelty of LSST data. Astronomers have faced similar challenges with other data sets in recent decades. Two new disciplines have emerged in response to the uniquely interdisciplinary nature of these challenges:

\* Astrostatistics weds the tools of statistics to the needs of astronomers. It focuses on probabilistic modeling of data and quantification of uncertainty.

\* Astroinformatics addresses the wide variety of concerns arising in managing, exploring, and analyzing extremely large data sets whose size thwarts straightforward application of standard methods (including optimal astrostatistical methods). It relies heavily on developments in the emerging fields of informatics and knowledge discovery from databases (KDD).

The emergence of these disciplines marks a growing realization within the astronomy data analysis community that the information science disciplines--primarily statistics and computer science--have significant expertise to offer astronomers. It also marks the realization among information scientists that astronomy poses compelling research-level challenges that push the boundaries of information science.

The ISSC will help focus the attention and link the resources of these new disciplines for the needs of LSST. Building this relationship will be synergistic, improving LSST science, and also fostering the growth of astroinformatics and astrostatistics, as disciplines within astronomy, and as disciplines within the information sciences.

The ISSC will be unique among current LSST SCs in that it will focus its work on methodological issues arising in the pursuit of a wide variety of astronomical science, rather than on a specific astronomical research area. By design, all of its work will overlap with that of other SCs (which will be facilitated by the fact that some of the ISSC members are already members of other LSST Science Collaboration teams); in many cases, ISSC efforts will overlap with the research interests of multiple other SCs. We ask that proposal reviewers and LSST leaders consider the unique role of the ISSC, understanding that it differs in structure and purpose from collaborations devoted to specific astronomical problems.

*B. What aspects of LSST data and/or features of operation do you need to realize your scientific goals?*

For the catalog-based research activities of our group, access to the LSST science database object catalog, source catalog, and (VO)event catalog is essential. We will make use of the 100+ science attributes per object for analysis and mining. Multi-dimensional and temporal data analyses require this.

For image-based research activities, access to the LSST images and pixel data is required, both in the set of calibrated images and in the annual data release sky template image set.

*C. Will additional data or information (not provided by LSST) be needed to realize your scientific goals?*

The ISSC itself will not directly need additional data or information not provided by LSST to realize its goals.

*D. Provide us with some information on your background, skills and experience that are relevant to the general area described above.*

Proposing members of the ISSC include a large fraction of U.S. scholars active in astrostatistics and astroinformatics today. It includes members of major cross-disciplinary collaborations including the California-Harvard Astrostatistics Collaboration (CHASC), the International Computational Astrostatistics group (InCA), the Center for Astrostatistics at Penn State, and ongoing collaborations between Cornell astronomers and Duke U. statisticians. The membership includes senior information scientists with decades of experience collaborating with scientists in many fields. Among them are members of the National Academy of Sciences, MacArthur Fellows, fellows of major societies, and distinguished professors. Others are

mid-career astronomers who have developed expertise in statistics or computer science. Yet others are young researchers who have begun their careers on the astronomy-statistics-informatics interfaces. Further details about the team appear in Section IV.

*E. Provide a quantitative estimate of the level of effort you are willing to dedicate to this work, and when you will begin.*

The level of effort devoted to ISSC activities will vary greatly across the ISSC membership, ranging from leading vigorous LSST-focused methodology research programs, to offering occasional consultation to astronomers seeking to benefit from the expertise of ISSC team members. Section VI, describing our management plan, presents more detailed information about the various levels of effort of team members.

The ISSC hopes to begin its work with a "live" team meeting at the January 2010 AAS meeting, which will be attended by many team members.

## II. Experimental Design (limit 2 pages)

*A. Provide an experimental design. State how LSST is an efficient vehicle for attaining your science goals. Provide a conceptual path to attain the science goal.*

The ISSC has no experimental design requirements of its own; it will "inherit" the requirements of the varied science projects that ISSC scientists will partner with. ISSC can play a role in assisting with optimization of experimental designs for varied LSST science tasks. Experimental design is a well-developed area of statistics with a long, fruitful history and a large literature that has been barely tapped by astronomers.

The main goals of the ISSC are presented in Section I. In pursuance of these goals, the ISSC will engage in a variety of concrete activities---both proactive and responsive---to build and support fruitful alliances between astronomers in different SCs (including the ISSC) and in the broader astronomical community pursuing related methodology research, and between astronomers and information scientists. Initial proposed ISSC activities include:

- \* Providing a centralized, prominent point of contact for addressing information science research issues relevant to LSST. This will facilitate access to expertise within the ISSC team, and provide a mechanism for astronomers and information scientists to connect with colleagues with complementary expertise. An ISSC web site will support communication activities. Mechanisms under consideration include maintaining a moderated, archived mailing list or blog for scientists to contact the ISSC, and a Wiki to provide a persistent, easily updated record of our efforts. This would include both formal internally refereed reports and informal responses to questions.

- \* Facilitating creation of interdisciplinary teams to pursue astroinformatics and astrostatistics research supporting LSST, including forming teams that will apply for future grants. The collaboration will be especially proactive in trying to assemble teams pursuing research on topics that cut across multiple other science collaborations.

- \* Informing the LSST science community about relevant recent and cutting-edge developments in informatics and statistics that can impact LSST research, by forming ad hoc working groups to write reports describing developments and providing pointers to key literature and potential collaborators. Some of these documents may become "living reviews" that are regularly maintained, if the relevant fields are particularly active. The ISSC web site will archive these reports.

- \* Organizing sessions on LSST-driven methodological developments and challenges at LSST All-Hands and American Astronomical Society meetings, as well as major gatherings in the information sciences (e.g., Joint Statistical Meetings, International Statistical Institute congresses, Neural Information Processing Systems conferences).

- \* Building and maintaining an annotated index of software implementing advanced data analysis algorithms of potential use to LSST scientists.

*B. What additional data or information (not provided by LSST) will be required to realize your scientific goals?*

Such requirements will be inherited from partnering applications; we do not anticipate ISSC-specific activities will have any nonstandard LSST data requirements.

### III. Overlap (limit 0.5 page)

*A. What aspects of your science goals overlap with the published goals of (one or more) existing LSST science collaborations? Overlap is not a disqualification: many problems require multiple approaches.*

The ISSC will exist to serve other SCs and the broader LSST user community. It will keep abreast of methodological efforts within individual SCs and may collaborate in their development. Several ISSC members are institutional members of several SCs and can facilitate such collaborations. The ISSC will also be available to advise and collaborate with the LSST Data Management group on methods for pipeline processing and computational strategies.

*B. In the event of significant overlap with other collaborations, explain the reasons why your proposed collaboration is needed (e.g. attack the same problem in different ways, overlap is a subset of the goals of the current proposal, etc.).*

The methods, skills, and literature in information sciences relevant to LSST needs are so vast that SCs committed to specific astronomical goals may often not have access to the latest or most appropriate methods. The ISSC has diverse skills, and can consult with a yet larger group of statisticians, engineers, mathematicians and computer scientists to bring the highest level of expertise to address LSST challenges.

#### IV. Resources (limit 1 page)

*A. Please state what resources you (as a collaboration collective) expect to have to conduct your science. Include institutional support, access to students/post-docs. Note that the LSST construction project cannot provide financial resources. The science collaborations are an anticipated route for developing grant proposals for designing, planning, and carrying out LSST science. Members may also apply as individuals for such grants.*

Our primary resource is the talent and experience of our diverse team. Various members are funded leaders in astrostatistics and astroinformatics research addressing problems with direct relevance to LSST science. The proposed ISSC membership includes scientists in some longstanding and productive interdisciplinary astronomy/information science research teams, including several that have trained both astronomer and statistician students and postdocs. We outline 6 of these groups here; further information is on the proposal support web site.

The International Computational Astrostatistics (InCA, <http://incagroup.org/>) group (formerly PiCA, Pittsburgh Computational Astrostatistics) is hosted at Carnegie Mellon University and University of Pittsburgh. The InCA team currently includes 28 astronomers, statisticians, and computer scientists. It began in 1999 and has repeatedly received NSF and NASA grants. InCA research has largely focused on data analysis problems in cosmology, including significant contributions to SDSS and CMB science. Drs. Freeman, Genovese, Miller, Park, Richards, Schafer and Wasserman are ISSC team members affiliated with InCA. Drs. Gray and Jong were affiliated with the InCA group before moving to faculty positions at other institutions.

The California-Harvard Astrostatistics Collaboration (CHASC, <http://www.ics.uci.edu/~dvd/astrostat.html>) is currently comprised of 17 astronomers and statisticians, mostly affiliated with the Center for Astrophysics and the statistics departments at Harvard and UC Irvine. CHASC began in 1996 with initial funding from the Chandra X-ray Center; it has received subsequent support from NASA and NSF grants in both astronomy and statistics. Their initial focus was on methodological issues in X-ray spectroscopy and imaging (including image deconvolution and quantifying systematic error). They are now addressing the analysis of optical stellar multidimensional color-magnitude diagrams. Drs. Kashyap, Meng, Park, Siemiginowska and van Dyk are ISSC members associated with CHASC.

Penn State University researchers began astrostatistical studies in the mid-1980s and formed the Center for Astrostatistics (CASt, <http://astrostatistics.psu.edu/>) in 2003. Supported by NSF and NASA grants, they have conducted a wide variety of astrostatistics research on survey-related problems. CASt also organizes several services to the astronomical community: the Statistical Challenges in Modern Astronomy research conferences held every 5 years since 1991; the Summer School in Statistics for Astronomers held annually since 2003; the StatCodes and R tutorials software sites; the R-based VStat Web service for the Virtual Observatory; and annotated bibliographies in statistics. Their Web site receives ~400,000 page-hits annually. Drs. Babu and Feigelson are ISSC members associated with CASt.

Cornell astronomers and statisticians at Duke University and Cornell pursue a variety of Bayesian astrostatistical research efforts. They treat diverse applications including time series (exoplanets, pulsars, AGN, dynamic spectra of supernovae and GRBs), and population analysis using data with selection effects and measurement error (outer solar system, GRBs, AGN, catalog cross-matching). This collaboration has funded multiple statistics post-docs and graduate students. Drs. Loredó, Chernoff, Berger, Clyde, Ruppert, and Wolpert, are ISSC members associated with the Cornell-Duke collaboration.

University of Michigan astronomers and statisticians have a long-standing collaboration studying the structure of dwarf spheroidal galaxies, requiring development of nonparametric statistical methods for modeling dark matter distributions. Dr. Woodroffe is the ISSC member from this group.

Dr. Borne, the initial ISSC Chair, is a faculty member in both astrophysics and computational & data sciences at George Mason University. His GMU department (Computational & Data Sciences) is one of very few in the nation that places diverse domain scientists (astronomers, physicists, chemists, and statisticians) within the same science department. Easy access to this group will enhance the broader impacts of our efforts (both to and from our team).

*B. Will you and your co-applicants apply (either individually or co-ordinated through the collaboration) for research grants to support your science? If yes, what will you want the grants to cover?*

Many ISSC members already have ongoing grant-funded research programs in areas overlapping LSST concerns. These members plan to continue their astroinformatics and astrostatistics research; formation of the ISSC will encourage increased focus on research problems impacting LSST science. An additional goal of the ISSC is to facilitate building interdisciplinary collaborations between SC members and ISSC members which we anticipate will pursue grant funding from information science and interdisciplinary programs.

Research funds are available from the National Science Foundation, including from crosscutting programs not traditionally tapped by astronomers. These include: Accelerating Discovery in Science and Engineering through Petascale Simulations and Analysis (PetaApps); Foundations of Data and Visual Analytics (FODAVA); Cyber-Enabled Discovery and Innovation (CDI); Computational Mathematics; Data-intensive Computing; Focused Research Groups in the Mathematical Sciences (FRG); Interactions between Mathematical Sciences and Computer Sciences (MSPA-MCS); Information and Intelligent Systems (IIS); and Statistics. Research funding is also available through the NASA Applied Information Systems Research (AISR) program and through DOE programs, such as the Mathematics for the Analysis of Petascale Data (MAPD) program.

*C. Will you have, or will you apply for, access to additional observing facilities that bear upon your science results from LSST?*

ISSC activities will not directly require non-LSST observing resources, but some proposals with ISSC participation are likely to seek such resources, initiated by collaborators in other SCs.

**V. Contributions to the Broader LSST Effort (limit 1.5 pages)**

*A. How will the creation of your proposed science collaboration add value to the LSST mission? Describe what you (as a team) bring to the project, e.g. tools that help tune LSST data acquisition strategy, data analysis methods and tools, data processing facilities, peripheral activities that enhance the value of LSST data (such as follow up observations of LSST alerts, algorithms and computational resources that can create additional data products that increase the usefulness of the LSST).*

Unlike most LSST scientific collaborations which focus on an astronomical topic, the principal purposes of the ISSC are to serve the entire LSST community through its expertise in astrostatistical and astroinformatical methodology. This entire proposal, particularly Sections I and II above, addresses how the ISSC will add value to the LSST mission.

*B. Please also describe any other ways in which you think your proposed collaboration can assist in the execution of the LSST project and/or increase its scientific productivity.*

A vibrant ISSC will facilitate both the data analysis and the science analysis of the LSST project in many ways, as described in sections I and II above.

## VI. Management Plan (limit 2 pages)

A. Please provide a management plan for how the science goals will be achieved. This is the place to justify the make up of the collaboration, so please specifically state what the roles and work loads of each of the proposed members will be.

We propose a three-tiered management structure for the collaboration. There will be very limited resources available to bring a significant number of team members together in person, so telecons, web-based interaction, and email will be the main sources of interaction within the collaboration. Our experience with similar large collaborations in the past indicates that telecons with more than a half dozen or so people are of limited value. Our three-tiered structure reflects this wisdom, trying to limit the need to have large telecons or emails broadcast to large groups.

**CORE TEAM (5 members):** The core team will initially consist of Babu, Borne, Feigelson, Gray, and Loredo. This team will be the contact point for LSST, and will have regular telecons (initially monthly) to monitor and guide the activity of the larger collaboration. Borne will serve as the initial Chair of the collaboration. The core team deliberately consists of not only astronomers with strong methodological credentials (Borne, Feigelson and Loredo), but also a statistician (Babu) and a computer scientist (Gray) representing these information science communities. The Core Team will maintain this diversity as individual members cycle off.

**ACTIVE TEAM (currently 35 members):** This will consist of scientists who plan to undertake significant astroinfo/astrostat research directly relevant to LSST. Nearly all astronomer members of the ISSC are on the active team; numerous information scientists pursuing astronomy research are also on this team. Active team members will take lead roles in producing collaboration reports, forming ad hoc working groups on various topics as the need arises. Leaders of these working groups will report to the core team telecons as necessary. Focussed telecons will connect subsets of the active team and scientists on other science collaborations to discuss specific cross-cutting issues as they arise.

**SUPPORT TEAM (currently 11 members):** Support team members will serve largely as consultants to the core and active teams, handling questions from LSST scientists fielded to them from those teams. Information scientists who are interested in LSST science but who do not expect to directly undertake significant LSST research populate this team. A few astronomers who plan to limit ISSC participation to a consulting role are support members, but we expect most astronomer members to be active team members. Although support team members will make limited time commitments to ISSC activities, they represent a broad and deep pool of expertise and networking opportunities, and are thus a critically valuable ISSC resource. As specific difficult problems arise, this team will broaden to include additional experts in the information sciences.

The current members of the ISSC are listed below, with their roles (Core/Active/Support) identified. We feel the depth and breadth of expertise in our team is the strongest argument in favor of creating the ISSC. Accordingly, we have prepared a proposal support website providing more detailed descriptions of the interests and roles of each team member (the team is too large for such information to be included here):

<http://inference.astro.cornell.edu/lsst/proposal09/team.html>

Astronomers & physicists (30):

Joshua Bloom (A), University of California, Berkeley  
 Kirk Borne (C), George Mason University  
 Robert Brunner (A), University of Illinois at Urbana-Champaign  
 Tamas Budavari (A), Johns Hopkins University  
 Douglas Burke (A), Harvard-Smithsonian Center for Astrophysics  
 David F. Chernoff (A), Cornell University  
 James M. Cordes (A), Cornell University  
 George Djorgovski (A), California Institute of Technology  
 Eric Feigelson (C), Penn State University  
 L. Samuel Finn (A), Penn State University  
 Peter Freeman (A), Carnegie Mellon University  
 Matthew Graham (A), California Institute of Technology  
 Carlo Graziani (A), University of Chicago  
 Jon Hakkila (A), College of Charleston  
 William Jefferys (S), University of Vermont and University of Texas at Austin  
 Vinay Kashyap (A), Harvard-Smithsonian Center for Astrophysics  
 Kevin Knuth (A), University at Albany  
 Donald Q. Lamb (A), University of Chicago  
 Thomas Loredo (C), Cornell University  
 Ashish Mahabal (A), California Institute of Technology  
 Bruce McCollum (A), California Institute of Technology  
 Christopher Miller (A), Cerro Tololo Inter-American Observatory  
 Misha (Meyer) Pesenson (A), California Institute of Technology  
 Vahe Petrosian (A), Stanford University  
 Andy Ptak (A), Johns Hopkins University  
 Jeffrey Scargle (A), NASA Ames Research Center  
 Aneta Siemiginowska (A), Harvard-Smithsonian Center for Astrophysics  
 Ben Wandelt (A), University of Illinois at Urbana-Champaign

Michael Way (S), NASA Ames Research Center  
Martin Weinberg (A), University of Massachusetts Amherst

Information scientists (21):

Jogesh Babu (C), Penn State University  
James Berger (S), Duke University and Statistical and Applied Mathematical Sciences Inst.  
Adam M. Brazier (S), Cornell University  
Merlise Clyde (S), Duke University  
Ian Davidson (A), University of California, Davis  
Bradley Efron (S), Stanford University  
Chris Genovese (A), Carnegie Mellon University  
Alexander Gray (C), Georgia Institute of Technology  
Woncheol Jang (S), University of Georgia  
Eric D. Kolaczyk (S), Boston University  
Ji Meng Loh (S), Columbia University  
John Rice (A), University of California, Berkeley  
Joseph Richards (A), Carnegie Mellon University  
David Ruppert (A), Cornell University  
Naoki Saito (S), University of California, Davis  
Chad Schafer (A), Carnegie Mellon University  
Jiayang Sun (S), Case Western Reserve University  
David van Dyk (A), University of California, Irvine  
Larry Wasserman (A), Carnegie Mellon University  
Robert Wolpert (A), Duke University  
Michael Woodroffe (A), University of Michigan

*B. How many additional members can you accommodate from future cycles of collaboration proposals? 14*

## VII. Data Products

*A. Will your proposed research involve generating new or "unusual" data products that can be of wider interest, or be useful to others? If so, please describe them.*

The ISSC plans on producing at least four types of unusual data products, all of which will be disseminated on the team's Web site as well as sent to appropriate groups within the LSST community.

First, we will provide formal reports on topics chosen within the SC or requested by other SCs. These will be drafted by one or more members, and circulated to the full team for improvements, and refined by the Core Team. These may be "living documents" updated as circumstances change and experience develops.

Second, we will provide informal answers to questions and discussions where individual members play equal roles. Again, the questions will be generated both within the ISSC and triggered by other members of the LSST community. Members of the Support Team will be solicited to contribute in areas of their expertise, and additional members will be sought as needed.

Third, we will provide some small codes, or scripts in advanced languages (such as IDL, R, MatLab, or Python), with sample data products. These will serve to prototype new methods, such as refined classifications of catalogued objects or characterization of the time series of individual objects.

Fourth, the ISSC will encourage members to publish their individual research efforts on innovative methods and algorithms in the spirit of "reproducible research," including the posting of free public software that can reproduce figures and other quantitative results reported in journal publications. This will facilitate further use of new algorithms, and accelerate adaptation and improvement of algorithms.

The ISSC will not provide formal data products, but will assist the Data Management team or other SC in implementing new methods they feel are valuable.

*B. Will they be made publicly available, and how?*

See VII A above.

## VIII. Education and Public Outreach

*A. Education and Public Outreach (EPO) is an important component of the LSST mission. How can your proposal contribute to LSST EPO?*

LSST has many unique features that appeal to the public and education communities. Among these will be the open data access policy for such an enormous astronomical object database and image archive. Educators, members of the public, science media specialists, and congressional science staffers will face a daunting challenge in determining what to do with this data flood. Among the solutions to this dilemma that the LSST EPO core team has already been formulating is the deployment of online Citizen Science projects, perhaps one new project each year or two, e.g., asteroid classification, light curve classification, unusual optical transient tracking, gravitational lens modeling, merging galaxy modeling, and others already being discussed. Several of these ideas are already documented in the LSST Science Book. We plan to augment and enhance those activities in several venues (formal education, informal education including planetaria and major science centers, and news media) by providing tools and algorithms for multi-variate data exploration, visualization, understanding, and interpretation. Such tools can be incorporated into the citizen science projects or into other web-based LSST data portals.

We will further contribute to the development of STEM education on the national front in the critical area of Data Sciences, using the LSST data products and LSST science themes as our foundation. Numerous national study committees (including the National Academies, National Science Board, and National Science Foundation) have issued a call to action in the area of data sciences research and education. We have documented our vision and some plans to contribute to this area in a formal position paper submitted in early 2009 to the Astro2010 Decadal Survey in Astronomy & Astrophysics. Details can be found (including specific references to the national study committees' reports) at the Astro2010 website (Borne et al. 2009):

<http://www8.nationalacademies.org/astro2010/publicview.aspx>

[http://mason.gmu.edu/~kborne/Borne\\_data\\_sciences\\_education\\_CDH\\_EPO.pdf](http://mason.gmu.edu/~kborne/Borne_data_sciences_education_CDH_EPO.pdf)

As an example, several of our institutions have already been (or soon will be) developing new courses and curricula in data sciences, some of which are dedicated specifically to astronomy (i.e., astroinformatics and astrostatistics) while others include astronomy within a broader multi-science discipline environment. Members of our collaboration team are already contributing (or will contribute) significantly to these efforts. Such undergraduate and graduate training programs address many objectives, including these: students are trained to access large distributed data repositories, to conduct meaningful scientific inquiries into the data, to mine and analyze the data, and to make data-driven scientific discoveries. These skills are necessary skills as major sky surveys have become a core research tool for a significant fraction of astronomical researchers. The LSST data repository will provide the raw material for all of these efforts.

Finally, we note that one of the members of our core team (and the initial Chair of the ISSC team) is Kirk Borne, who has been a member of the LSST EPO effort for many years (informally since 2001, and formally since 2005). We will have a seat at the table as all LSST EPO efforts are planned, discussed, decided, and implemented.

*B. Who will be responsible for overseeing this contribution?*

Kirk Borne, currently a member of the LSST EPO core team, will oversee ISSC EPO activities.