

Data Sharing + Open, Rigorous & Reproducible Science


Ivo D. Dinov

Statistics Online Computational Resource
Health Behavior & Biological Sciences
Computational Medicine & Bioinformatics

University of Michigan

<https://SOCR.umich.edu>


Slides Online: "SOCR News"



1



Outline

- ❑ Pillars of Open-Science
- ❑ Rationale (Pros & Cons)
- ❑ Big Data Sharing
- ❑ *DataSifter: Statistical obfuscation*
- ❑ Case-studies
 - ❑ ALS Study
 - ❑ Population Census-like Neuroscience (UKBB)
 - ❑ Spacekime Analytics



2

Pillars of Open Data Science (HS650 / Bioinfo501)


3

Sources: Characteristics of Big Biomed Data

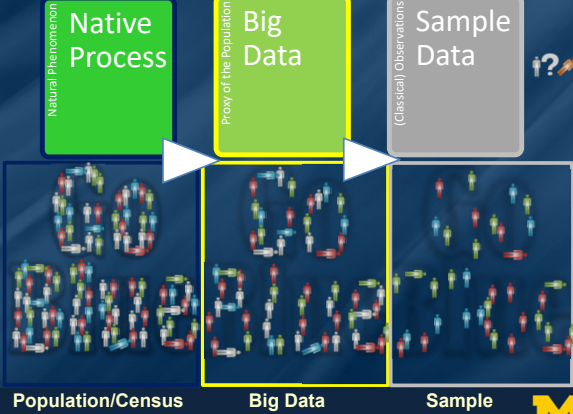
IBM Big Data 4V's: Volume, Variety, Velocity & Veracity

Big Bio Data Dimensions	Tools	
Size	Harvesting and management of vast amounts of data	Example: analyzing observational data of 1,000's Parkinson's disease patients based on 10,000's signature biomarkers derived from multi-source imaging, genetics, clinical, physiologic, phenomics and demographic data elements
Complexity	Wranglers for dealing with heterogeneous data	
Incongruency	Tools for data harmonization and aggregation	
Multi-source	Transfer and joint modeling of disparate elements	Software developments, student training, service platforms and methodological advances associated with the Big Data Discovery Science all present existing opportunities for learners, educators, researchers, practitioners and policy makers
Multi-scale	Macro to meso to micro scale observations	
Time	Techniques accounting for longitudinal patterns in the data	
Incomplete	Reliable management of missing data	


Dinov (2016) *GigaScience* | Dinov (2023) Springer



4



Population/Census Unobservable | Big Data Harmonize/Aggregate Problems | Sample Limited process view




5

From 23 ... to ... 2²³

- ❑ Data Science: 1798 vs. 2020
- ❑ In the 18th century, Henry Cavendish used just 23 observations to answer a fundamental question – “*What is the Mass of the Earth?*” He estimated very accurately the mean density of the Earth/H₂O (5.483±0.1904 g/cm³)
- ❑ In the 21st century to achieve the same scientific impact, matching the reliability and the precision of the Cavendish's 18th century prediction, requires a monumental community effort using massive and complex information perhaps on the order of 2²³ bytes
- ❑ Data & Information Science ≅ Scalability & Compression (per Gerald Friedland/Berkeley): 23 → 2²³≅10M

Cavendish (1798) *Philosophical Transactions of the Royal Society of London* | Dinov (2016) *ISM*



6

Big Data	Information	Knowledge	Action
Raw Observations	Processed Data	Maps, Models	Actionable Decisions
Data Aggregation	Data Fusion	Causal Inference	Treatment Regimens
Data Scrubbing	Summary Stats	Networks, Analytics	Forecasts, Predictions
Semantic-Mapping	Derived Biomarkers	Linkages, Associations	Healthcare Outcomes

Dinov (2023) Springer

7

Why is FAIR Data Sharing Important?

- Optimum resource utilization (low cost, high efficiency / policy, security, processing complexity)
- Democratization of the scientific discovery process
- Enhanced inference (e.g., coverage of rare events, increase of stat power)
- Increase of Kryder's Law (Data volume) \gg Moore's Law (Compute power)
- Exponential decay of data-value
- Incentivizes innovation, transdisciplinary collaborations, and knowledge dissemination
- ...

FAIR = Findable + Accessible + Interoperable + Reusable

8

Infrastructure: Cloud Ecosystem

2013

<https://soar.umich.edu/docs/BD2K/BigDataResource.html>

9

Infrastructure: Cloud Ecosystem

2020

100 STARTUPS USING ARTIFICIAL INTELLIGENCE TO TRANSFORM INDUSTRIES

<https://cmte.ieee.org/futuredirections/2020/01/02/care-for-a-little-boost-to-your-intelligence-try-ai-ai/>

10

Infrastructure: Cloud Ecosystem

2023

AI 100

<https://www.cbinsights.com/research/artificial-intelligence-top-startups-2023/>

11

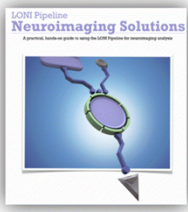
Google Cloud

DEVELOPER'S CHEAT SHEET

... a single ecosystem ...

12

Infrastructure: Cranium/Pipeline



<http://Pipeline.loni.usc.edu>

Dinov, et al. (2013) Brain Imaging and Behavior | Dinov, et al. (2014) Front. NeuroInfo.

Tools

13

Scholarly Research: OA Pubs/Sharing

- OA Pubs
 - https://en.wikipedia.org/wiki/Open_access
 - <https://arxiv.org> | <https://www.biorxiv.org>
 - Blogs (e.g., <https://TerryTao.wordpress.com>)
- Cloud Services
 - Computing (e.g., Azure, Google, AWS)
 - Storage
 - ICT (information and communication technologies)
- SW
 - <https://GitHub.com> (e.g., <https://github.com/SOCR>)
 - <http://Cran.r-project.org> | Jupyter.org | Rmarkdown.rstudio.com
 - E.g., <https://DSPA.predictive.space>
- Licensing
 - <https://www.gnu.org/licenses>
 - https://socr.umich.edu/html/SOCR_CitingLicense.html

Pubs

14

Findings: Open Science Career Assessment Matrix

Open Science activities	Metrics: Possible evaluation criteria
RESEARCH OUTPUT	
Research activity	Pushing forward the boundaries of open science as a research topic
Publications	Publishing in open access journals Self-archiving in open access repositories
Datasets and research results	Using the FAIR data principles Adopting quality standards in open data management and open datasets Making use of open data from other researchers
Open source	Using open source software and other open tools
Funding	Developing new software and tools that are open to other users Securing funding for open science activities
RESEARCH PROCESS	
Stakeholder engagement/citizen science	Actively engaging society and research users in the research process Sharing provisional research results with stakeholders through open platforms (e.g. Arxiv, Figshare, OverLeaf)
Collaboration and Interdisciplinarity	Involving stakeholders in peer review processes Widening participation in research through open collaborative projects Engaging in team science through diverse cross-disciplinary teams
Research integrity	Being aware of the ethical and legal issues relating to data sharing, confidentiality, attribution and environmental impact of open science activities Fully recognizing the contribution of others in research projects, including collaborators, co-authors, citizens open data providers
Risk management	Taking account of the risks involved in open science

Declaration on Research Assessment (DORA) | <https://sfidora.org/resource/metrics-toolkit/>

15

Findings: Open Science Career Assessment Matrix

SERVICE & LEADERSHIP	
Leadership	Developing a vision and strategy on how to integrate OS practices in the normal research practice Driving policy and practice in open science Being a role model in practicing open science
Academic standing	Developing an international or national profile for open science activities Contributing as editor or advisor for open science journals or bodies
Peer review	Contributing to open peer review processes Examining or assessing open research
Networking	Participating in national and international networks relating to open science
RESEARCH IMPACT	
Communication and Dissemination	Participating in public engagement activities Sharing research results through non-academic dissemination channels Translating research into a language suitable for public understanding
IP (patents, licenses)	Knowledge on the legal and ethical issues relating to IPR Transferring IP to the wider economy
Societal impact	Evidence of use of research by societal groups Recognition from societal groups or for societal activities. h-index, h10-index, sharing-index, other quant metrics of impact
Knowledge exchange	Engaging in open innovation with partners beyond academia
TEACHING & SUPERVISION	
Teaching	Training other researchers in open science principles and methods Developing curricula and programs in open science methods, including open science data management
Mentoring	Raising awareness and understanding in open science in undergraduate and masters' programs
Supervision	Mentoring and encouraging others in developing their open science capabilities Supporting early stage researchers to adopt an open science approach
PROFESSIONAL EXPERIENCE	
Continuing professional development	Investing in own professional development to build open science capabilities
Project management	Successfully delivering open science projects involving diverse research teams
Personal qualities	Demonstrating the personal qualities to engage society and research users with open science Showing the flexibility and perseverance to respond to the challenges of conducting open science

Declaration on Research Assessment (DORA) | <https://sfidora.org/2023/>

16

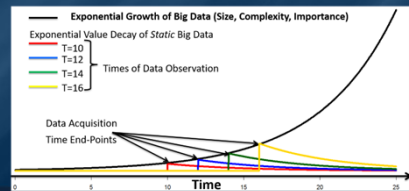
Rationale for Open Science (Cons)

- Journals impact factor (compared to pay-per-view journals, OA are newer)
- *Predatory* science (dubious quality, profit-centric, spam camouflage)
- Discovery is easy, but validity/utility of the science or tools may be difficult to evaluate *en masse*
- Extra work may be required by all scholars to sift through and identify appropriate materials
- Ambiguity of usage-rights/copyrights/licenses
- Democratization and socialization of science may suffer from some of the same downsides as social-networks
- Is science *competitive* or *collaborative*? Is it a *zero-sum* enterprise?

17

Rationale for Open Science (Pros)

- We are always stronger together
- Long-term sustainability prefers openness, inclusivity & diversity
- Optimized investments, career advancement, impact & cost-efficiency
- Expeditious discovery, innovation, productization & higher impact
- Rapid devaluation of data-hoarding, clandestine science, knowledge obfuscation
- ...



<https://www.aas.org/news/big-data-blog-part-v-interview-dr-ivo-dinov>

18

Rationale for Open Science: Kryder vs. Moore

- Moore's law = the expectation that our computational capabilities, specifically the number of transistors on integrated circuits, doubles approximately every 18-24 months.
- Kryder's law = the volume of data, in terms of disk storage capacity, is doubling every 14-18 months.
- Kryder >> Moore: Although both laws yield exponential growth, data volume is increasing at a faster pace. Thus, there are clear interests and needs for significant private, public and government engagement in opening, managing, processing, interrogating and interpreting the information content of Big Data.

19 <https://www.aaas.org/news/big-data-blog-part-v-interview-dr-ivo-dinov>

19

Reliable, Effective & Secure Data Sharing

- Why is data-sharing difficult? monopoly, preservation of *status-quo*, competitive edge, personally identifiable information, IP protection, security (on multiple levels), red tape, ...
- FAIR (Findable, Accessible, Interoperable & Reusable) Data are powerful
- Current Data Sharing Landscape? Differential Privacy, fully-homomorphic encryption, statistical obfuscation (DataSifter), ...

20

20

DataSifter

- DataSifter is an iterative statistical computing approach that provides the data-governors controlled manipulation of the trade-off between sensitive information obfuscation and preservation of the joint distribution.
- The DataSifter is designed to satisfy data requests from pilot study investigators focused on specific target populations.
- Iteratively, the DataSifter stochastically identifies candidate entries, cases as well as features, and subsequently selects, nullifies, and imputes the chosen elements. This statistical-obfuscation process relies heavily on nonparametric multivariate imputation to preserve the information content of the complex data.

21 <https://DataSifter.org> US patent #16/051,881 Marino, et al., JSCS (2019)

21

DataSifter

- A detailed description and `dataSifter()` R method implementation are available on our GitHub repository (<https://github.com/SOCR/DataSifter>).
- Data-sifting different data archives requires customized parameter management. Five specific parameters mediate the balance between protection of sensitive information and signal energy preservation.

Obfuscation level	k_0	k_1	k_2	k_3	k_4
None	0	0	0	0	0
Small	0	0.05	1	0.1	0.01
Medium	1	0.25	2	0.6	0.05
Large	1	0.4	5	0.8	0.2
Indep	Output synthetic data with independent features				

22 <https://DataSifter.org> US patent #16/051,881 Marino, et al., JSCS (2019)

22

DataSifter

23 <https://DataSifter.org> US patent #16/051,881 Marino, et al., JSCS (2019)

23

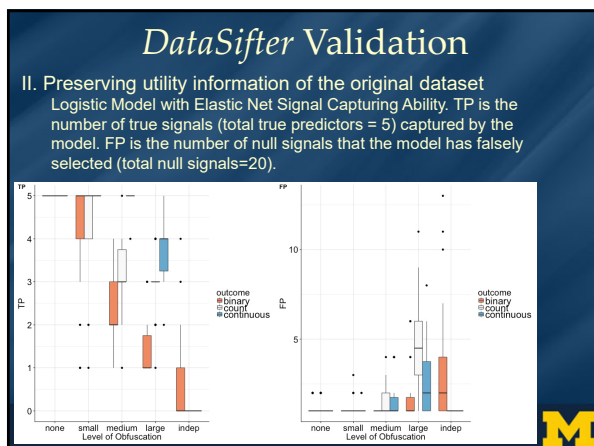
DataSifter Validation

I. Protection of sensitive information (privacy)

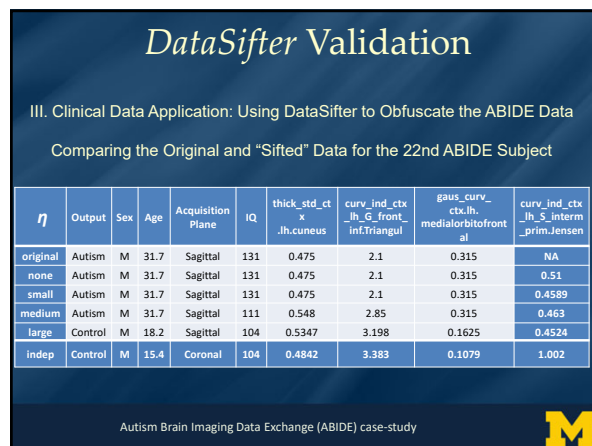
PIFV under Different Privacy Levels. Binary outcome refers to the first experiment; Count refers to the second experiment; Continuous refers to the third experiment. Each box represents 30 different "sifted" data or 30,000 "sifted" cases.

24

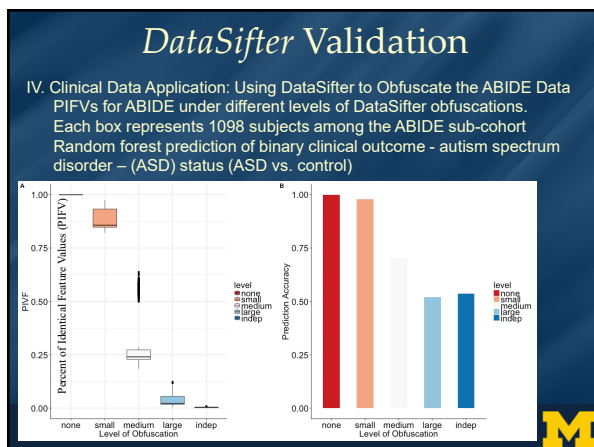
24



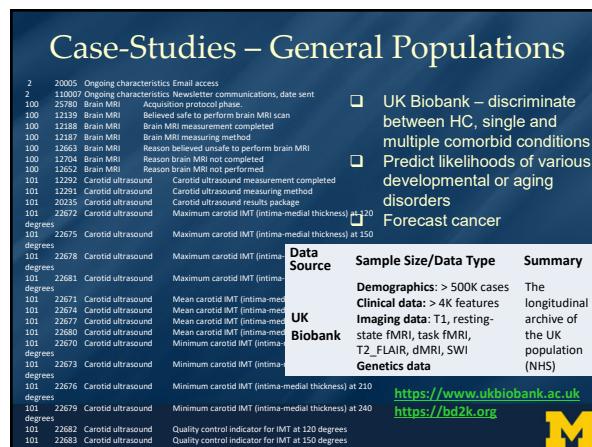
25



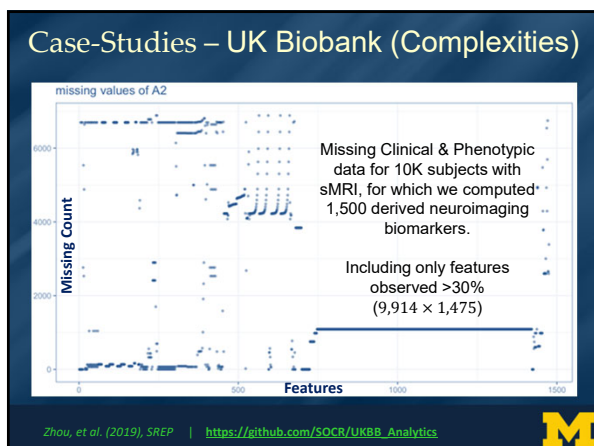
26



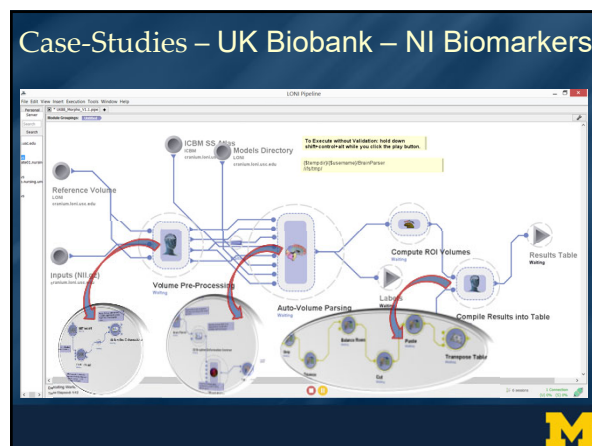
27



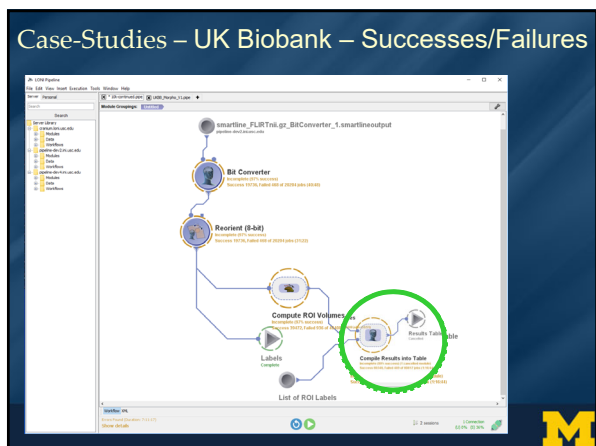
28



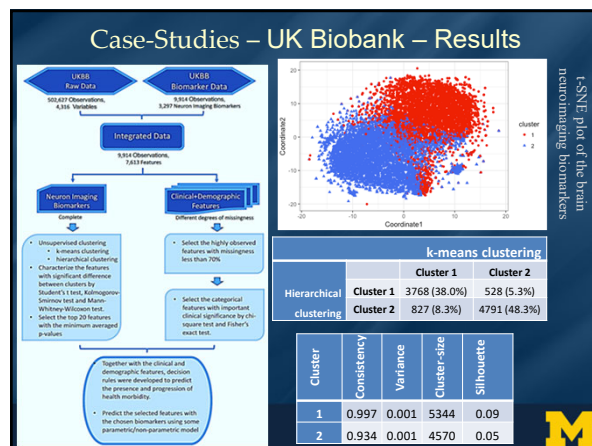
29



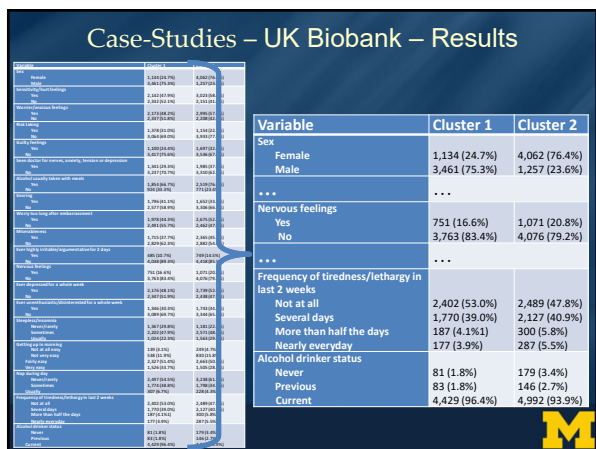
30



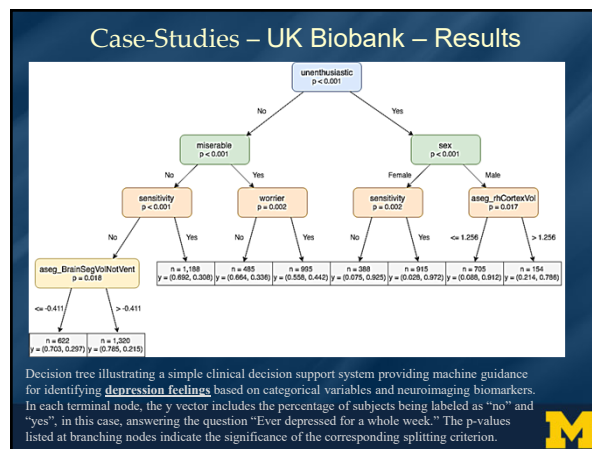
31



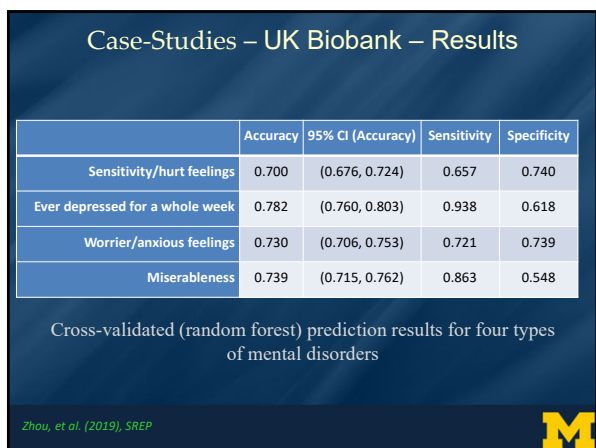
32



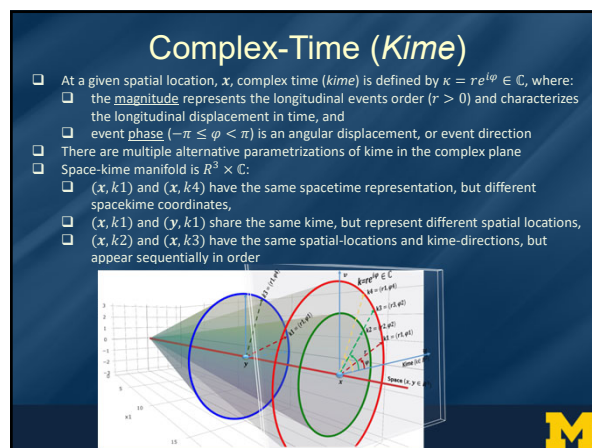
33



34



35



36

Kime Parameterizations

Conjugate Pairs $\{z, \bar{z} \in \mathbb{C}\}$

$$z = (x + iy)/2$$

$$\bar{z} = -(x - iy)/2$$

$$z + \bar{z} = x + iy$$

$$z - \bar{z} = x - iy$$

Polar $\{(r, \varphi) \in \mathbb{R}^+ \times [-\pi, \pi]\}$

$$z = r e^{i\varphi}$$

$$\bar{z} = r e^{-i\varphi}$$

$$r = \sqrt{z\bar{z}} = \sqrt{|z|^2}$$

$$\varphi = \arccos\left(\frac{z + \bar{z}}{2z\bar{z}}\right)$$

Cartesian $\{(x, y) \in \mathbb{R}^2\}$

$$x = r \cos\varphi$$

$$y = r \sin\varphi$$

$$r = \sqrt{x^2 + y^2}$$

$$\varphi = \text{atan2}(y, x)$$

37

The Importance of Kime-Magnitude (*time*) and Kime-Phase (*direction*)

Fourier Analysis
(real part of the Forward Fourier Transform)

Square Image Shape

2D Image 1 (square)
Re(FT(square))
Magnitude FT(square)
Phase FT(square)

Disk Image Shape

2D Image 2 (disk)
Re(FT(disk))
Magnitude FT(disk)
Phase FT(disk)

Fourier Synthesis
(real part of the Inverse Fourier Transform)

Square Image Shape

IFT(FT(square)) = square

Disk Image Shape

IFT using square-magnitude & disc-phase

Square Image Shape

IFT using disc-magnitude & square-phase

Disk Image Shape

IFT using disc-magnitude & disc-phase

38

Longitudinal Data Analytics

- **Neuroimaging:**
 - **4D fMRI:** time-series, represents measurements of hydrogen atom densities over a 3D lattice of spatial locations ($1 \leq x, y, z \leq 64$ pixels), about 3×3 millimeters² resolution. Data is recorded longitudinally over time ($1 \leq t \leq 180$) in increments of about 3 seconds, then post-processed
 - **State-of-the-art Approaches:** Time-series modeling or Network analysis
 - **Spacekime Analytics:** 5D fMRI kime-series, represent the hydrogen atom densities over the same 3D lattice of spatial locations, longitudinally over the 2D kime space, $\kappa = r e^{i\varphi} \in \mathbb{C}$
 - **Differences:** Spacekime analytics estimate and utilize the kime-phases

4D Spacetime 4D/5D Reconstructions 5D Spacekime

Dinov & Velev (2021)

39

Spacekime Calculus

- Kime **Wirtinger derivative** (first order kime-derivative at $k = (r, \varphi)$):
In Cartesian coordinates:

$$f'(z) = \frac{\partial f(z)}{\partial z} = \frac{1}{2} \left(\frac{\partial f}{\partial x} - i \frac{\partial f}{\partial y} \right)$$
 and $f'(z) = \frac{\partial f(z)}{\partial \bar{z}} = \frac{1}{2} \left(\frac{\partial f}{\partial x} + i \frac{\partial f}{\partial y} \right)$.
 In Conjugate-pair basis: $df = \partial f + \bar{\partial} f = \frac{\partial f}{\partial z} dz + \frac{\partial f}{\partial \bar{z}} d\bar{z}$.
 In Polar kime coordinates:

$$f'(k) = \frac{\partial f(k)}{\partial k} = \frac{1}{2} \left(\cos\varphi \frac{\partial f}{\partial r} - \frac{1}{r} \sin\varphi \frac{\partial f}{\partial \varphi} - i \left(\sin\varphi \frac{\partial f}{\partial r} + \frac{1}{r} \cos\varphi \frac{\partial f}{\partial \varphi} \right) \right) = \frac{e^{-i\varphi}}{2} \left(\frac{\partial f}{\partial r} - \frac{i}{r} \frac{\partial f}{\partial \varphi} \right)$$

$$f'(k) = \frac{\partial f(k)}{\partial \bar{k}} = \frac{1}{2} \left(\cos\varphi \frac{\partial f}{\partial r} + \frac{1}{r} \sin\varphi \frac{\partial f}{\partial \varphi} + i \left(\sin\varphi \frac{\partial f}{\partial r} + \frac{1}{r} \cos\varphi \frac{\partial f}{\partial \varphi} \right) \right) = \frac{e^{i\varphi}}{2} \left(\frac{\partial f}{\partial r} + \frac{i}{r} \frac{\partial f}{\partial \varphi} \right)$$
- Kime **Wirtinger integration:**
 Path-integral: $\lim_{|z_{i+1} - z_i| \rightarrow 0} \sum_{i=1}^{n-1} (f(z_{i+1})(z_{i+1} - z_i)) \cong \int_{z_0}^{z_n} f(z) dz$.
 Definite area integral: for $\Omega \subseteq \mathbb{C}$, $\int_{\Omega} f(z) dz d\bar{z}$.
 Indefinite integral: $\int f(z) dz d\bar{z}$, $df = \frac{\partial f}{\partial z} dz + \frac{\partial f}{\partial \bar{z}} d\bar{z}$.
 The Laplacian in terms of conjugate pair coordinates is $\Delta f = d^2 f = 4 \frac{\partial^2 f}{\partial z \partial \bar{z}} = 4 \frac{\partial^2 f}{\partial z^2 \partial \bar{z}^2}$.

Dinov & Velev (2021)

40

Quantum Mechanics, AI & Data Science

Mathematical-Physics	Data Science
A particle is a small localized object that permits observations and characterization of its physical or chemical properties	An object is something that exists by itself, actually or potentially, concretely or abstractly, physically or incorporeal (e.g., person, subject, etc.)
An observable is a dynamic variable about particles that can be measured	A feature is a dynamic variable or an attribute about an object that can be measured
Particle state is an observable particle characteristic (e.g., position, momentum)	Datum is an observed quantitative or qualitative value, an instantiation of a feature
Particle system is a collection of independent particles and observable characteristics, in a closed system	Problem , aka Data System, is a collection of independent objects and features, without necessarily being associated with a priori hypotheses
Wave-function	Inference-function
Reference-Frame transforms (e.g., Lorentz)	Data transformations (e.g., wrangling, log-transform)
State of a system is an observed measurement of all particles - wavefunction	Dataset (data) is an observed instance of a set of datum elements about the problem system, $\mathcal{O} = \{X, Y\}$
A particle system is computable if (1) the entire system is logical, consistent, complete and (2) the unknown internal states of the system don't influence the computation (wavefunction, intervals, probabilities, etc.)	Computable data object is a very special representation of a dataset which allows direct application of computational processing, modeling, analytics, or inference based on the observed dataset
...	...

Dinov & Velev (2021)

41

Quantum Mechanics, AI & Data Science

Math-Physics	Data Science
Wavefunction	Inference function - describing a solution to a specific data analytic system (a problem). For example,
Wave equ problem: $\left(\frac{\partial^2}{\partial x^2} - \frac{1}{v^2} \frac{\partial^2}{\partial t^2} \right) \psi(x, t) = 0$	<ul style="list-style-type: none"> • A linear (GLM) model represents a solution of a prediction inference problem, $Y = X\beta$, where the inference function quantifies the effects of all independent features (X) on the dependent outcome (Y), data: $\mathcal{O} = \{X, Y\}$: $\psi(\mathcal{O}) = \psi(X, Y) \Rightarrow \beta = \beta^{OLS} = (X^T X)^{-1} X^T Y$ • A non-parametric, non-linear alternative inference is SVM classification. If $\psi_x \in H$ is the lifting function $\psi: \mathbb{R}^d \rightarrow \mathbb{R}^d$ ($\psi: x \in \mathbb{R}^d \rightarrow \tilde{x} = \psi_x \in H$) where $\eta \ll d$, the kernel $\psi(x, y) = \langle \tilde{x}, \tilde{y} \rangle$: $\mathcal{O} \times \mathcal{O} \rightarrow \mathbb{R}$ transforms non-linear to linear separation, the observed data $\mathcal{O}_i = \{x_i, y_i\} \in \mathbb{R}^d$ are lifted to $\psi_{\mathcal{O}_i} \in H$. Then, the SVM prediction operator is the weighted sum of the kernel functions at $\psi_{\mathcal{O}_i}$, where β' is a solution to the SVM regularized optimization: $\langle \psi_{\mathcal{O}_1} \beta' \rangle_H = \sum_{i=1}^n p_i' \langle \psi_{\mathcal{O}_1} \psi_{\mathcal{O}_i} \rangle_H$
Complex Solution: $\psi(x, t) = A e^{i(kx - \omega t)}$ where $\frac{ \omega }{ k } = v$,	The linear coefficients, p_i' , are the dual weights that are multiplied by the label corresponding to each training instance, (x_i) .
represents a traveling wave	Inference always depends on the (input) data; however, it does not have 1-1 and onto bijective correspondence with the data, since the inference function quantifies predictions in a probabilistic sense.

GLM/SVM: <http://DSPA.predictive.space> | Dinov, Springer (2018)

42

Spacekime Analytics

- Let's assume that we have:
 - (1) Kime extension of Time, and
 - (2) Parallels between wavefunctions ↔ inference functions
- Often, we can't directly observe (record) data natively in 5D spacekime.
- Yet, we can measure quite accurately the kime-magnitudes (r) as event orders, "times".
- To reconstruct the 2D spatial structure of kime, borrow tricks used by crystallographers¹ to resolve the structure of atomic particles by only observing the magnitudes of the diffraction pattern in k-space. This approach heavily relies on (1) prior information about the kime directional orientation (that may be obtained from using similar datasets and phase-aggregator analytical strategies), or (2) experimental reproducibility by repeated confirmations of the data analytic results using longitudinal datasets.

Spacekime →
Spacekime Transforms

(1) Phase-estimation
(2) Phase-modeling
(3) Laplace Transform

5D Spacekime
3D Space R^3
(x_0, x_1, x_2)
Observed or
Computed

2D Kime $\cong R^2$
(x_3, x_4)
Computed

Data Science Analytics

FT

IFT

IFT

5D k-space
3D Space R^3
(f_0, f_1, f_2)
Observed or
Computed

K2 Kaluza-Klein $\cong R^2$
(time (t), phase (ϕ))
observed directly estimated

Experimental Science

¹ Rodriguez, Ivanova, Nature 2015

43

Spacekime Analytics: fMRI Example

- 3D isosurface Reconstruction of (2D space, 1D time) fMRI signal

4D spacetime: Reconstruction using trivial phase-angle; kime=time=(magnitude, 0)

5D Spacekime: Reconstruction using correct kime=(magnitude, phase)

3D pseudo-spacetime reconstruction:

$$f = \hat{h} \left(\underset{\text{space}}{x_1, x_2}, \underset{\text{time}}{t} \right)$$

44

Spacekime Analytics: Kime-series = Surfaces (not curves)

In the 5D spacekime manifold, time-series curves extend to kime-series, i.e., surfaces parameterized by kime-magnitude (t) and the kime-phase (ϕ).

Kime-phase aggregating operators that can be used to transform standard time-series curves to spacekime kime-surfaces, which can be modeled, interpreted, and predicted using advanced spacekime analytics.

Intensity

t time =
k magnitude

φ kime-phase

45

Bayesian Inference Representation

- We can formulate spacekime inference as a Bayesian parameter estimation problem:

$$\begin{aligned} \frac{p(\gamma|X, \phi')}{\text{posterior distribution}} &= \frac{p(\gamma, X, \phi')}{p(X, \phi')} = \frac{p(X|\gamma, \phi') \times p(\gamma, \phi')}{p(X, \phi')} = \frac{p(X|\gamma, \phi') \times p(\gamma, \phi')}{p(X|\phi') \times p(\phi')} \\ &= \frac{p(X|\gamma, \phi')}{p(X|\phi')} \times \frac{p(\gamma, \phi')}{p(\phi')} = \frac{p(X|\gamma, \phi') \times p(\gamma|\phi')}{\text{observed evidence}} \propto \frac{p(X|\gamma, \phi') \times p(\gamma|\phi')}{\text{likelihood} \quad \text{prior}} \end{aligned}$$

- In Bayesian terms, the posterior probability distribution of the unknown parameter γ is proportional to the product of the likelihood and the prior.
- In probability terms, the posterior = likelihood times prior, divided by the observed evidence, in this case, a single spacetime data point, x_{i_0} .

46

Spacekime Analytics using fMRI

- Complex-valued *finger tapping* fMRI (64x 64y 40z 160t)

fMRI time-series forecasting

Temporal Dynamics of a Voxel in Somatosensory Motor Area

On-Off fMRI time-series to Kimesurface

47

What's Next?

- Lots of "open problems" in data-science, e.g., fundamentals of data representation & analytics
- The SOCR team is developing:
 - Compressive Big Data Analytics (CBDA) technique – an ensemble learning meta-algorithm
 - DS Time-Complexity and Inferential-Uncertainty
- Need lots of community, institutional, state, federal, and philanthropic support to advance open data science methods, enhance the computing infrastructure, train/support students/fellows, and tackle the *Kryder Law* >> *Moore Law* trend
- **Web:** www.socr.umich.edu
- **Git:** <https://github.com/SOCR>
- **Projects:** www.socr.umich.edu/html/SOCR_Research.html
- **Apps:** <https://socr.umich.edu/HTML5/>

49

Acknowledgments

Slides Online:
"SOCR News"

Funding

NIH: UL1TR002240, R01CA233487, R01MH121079, R01MH126137, T32GM141746
NSF: 1916425, 1734853, 1636840, 1416953, 0716055, 1023115

Collaborators

- **SOCR:** Milen Velez, Yueyang Shen, Daxuan Deng, Zijing Li, Yongkai Qiu, Zhe Yin, Yufei Yang, Yuxin Wang, Rongqian Zhang, Yuyao Liu, Yupeng Zhang, Yunjie Guo, Simeone Marino
- **UMich MIDAS/MCAIM Centers:** Lydia Bieri, Kayvan Najarian, Chris Monk, Issam El Naqa, HV Jagadish, Brian Athey, Magdalena Ivanova



<https://SOCR.umich.edu>