Data Resources, Analytical Tools & Cloud Services for Clinical Decision Support

Ivo D. Dinov

Statistics Online Computational Resource
Health Behavior & Biological Sciences
Computational Medicine & Bioinformatics
Michigan Institute for Data Science

University of Michigan

https://SOCR.umich.edu

Slides Online:
"SOCR News"

SCHOOL OF NURSING
STATISTICS ONLINE COMPUTATIONAL RESOURCE (SOCR)
UNIVERSITY OF MICHIGAN

Pillars of Open-Science

- Data Resources
- Analytical Tools
- Cloud Services
- Clinical Decision Support Systems

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
Complexity	Wranglers for dealing with heterogeneous data
Incongruency	Tools for data harmonization and aggregation
Multi-source	Transfer and joint modeling of disparate elements
Multi-scale	Macro to meso to micro scale observations
Time	Techniques accounting for longitudinal patterns in the data
Incomplete	Reliable management of missing data

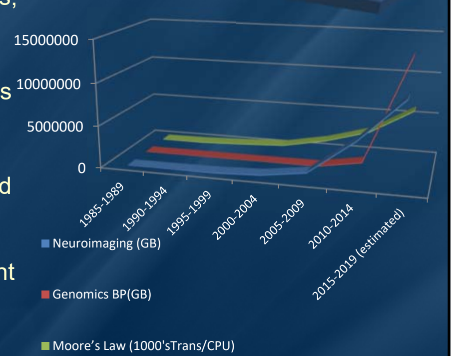
Example: analyzing observational data of 1,000's Parkinson's disease patients based on 10,000's signature biomarkers derived from multi-source imaging, genetics, clinical, physiologic, phenomics and demographic data elements

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

Dinov (2016) GigaScience Dinov (2018) Springer

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.



Dinov (2016) SMSI | https://www.aaas.org/news/big-data-blog-part-v-interview-dr-ivo-dinov

Data Sources

- UKBB https://www.ukbiobank.ac.uk
- MIMIC-III https://mimic.physionet.org
- SOCR Data Archives https://wiki.socr.umich.edu/index.php/SOCR_Data
- NIH Databases https://eresources.nlm.nih.gov/nlm_eresources

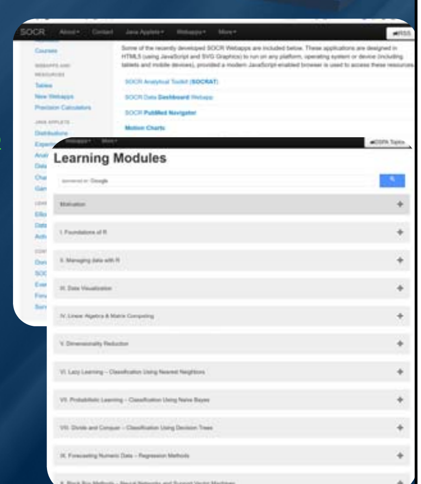
Data Source	Sample Size/Data Type	Summary
UK Biobank	**Demographics**: > 500K cases **Clinical data:** > 4K features **Imaging data**: T1, resting-state fMRI, task fMRI, T2_FLAIR, dMRI, SWI **Genetics data** **SARS-CoV-2 virus tests**	The longitudinal archive of the UK population (NHS)
MIMIC-III	ADMISSIONS, DIAGNOSES, ICUSTAYS,MICROBIOLOGY, PRESCRIPTIONS, PROCEDURES_ICD,SERVICES	ICU Data for over 40K patients
NIH Databases	100's of open-access DBs	

FAIR = Findable + Accessible + Interoperable + Reusable

- ❑ Data Resources
- ❑ **Analytical Tools**
- ❑ Cloud Services
- ❑ Clinical Decision Support Systems

Analytical Tools

- ❑ SOCR Webapps
  (HTML-based) https://socr.umich.edu/HTML5
- ❑ Data Science and Predictive Analytics (DSPA)
  (R-based) https://dspa.predictive.space



- ❑ Data Resources
- ❑ Analytical Tools
- ❑ **Cloud Services**
- ❑ Clinical Decision Support Systems

Cloud Services

Cloud Services

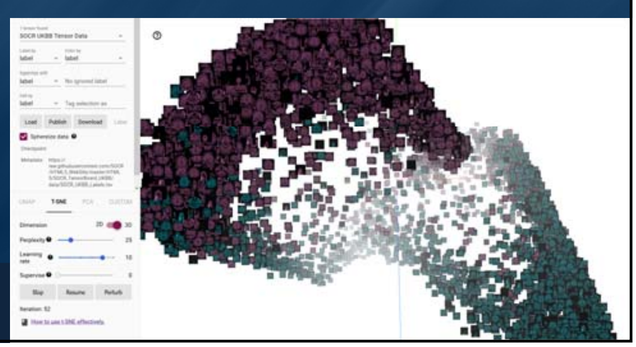
- ❑ Google Cloud & BigQuery
  https://cloud.google.com/bigquery
  Example: https://myumi.ch/QAoMj
- ❑ Big Data Cloud Services Ecosystem
  https://socr.umich.edu/docs/BD2K/BigDataResourceome.html

1 Connections to SQL Database
  1.1 Set a connection
  1.2 Managing and closing existing connections
2 Basic Functions in RODBC
3 Querying with SQL
4 Fetching Results
5 Important SQL clauses
  5.1 Basic SELECT
  5.2 SELECT from multiple tables
6 Connecting to Google `BigQuery`
  6.1 Navigating Google `BigQuery`
  6.2 Billing in Google `BigQuery`
  6.3 Case Study I: Open Library in `BigQuery`
  6.4 Case Study II: MIMICIII Intro
  6.5 Case Study III: MIMICIII & Acute HF
7 Additional Applications
  7.1 Cross-Data-Source Projects
  7.2 JOINs, EXCEPT, INTERSECT
  7.3 Database Management with SQL
  7.4 Querying Automation

- Data Resources
- Analytical Tools
- Cloud Services
- **Clinical Decision Support Systems**

Clinical Decision Support

- Pressure Injury Prediction Model (PIPM) https://myumi.ch/O49zG



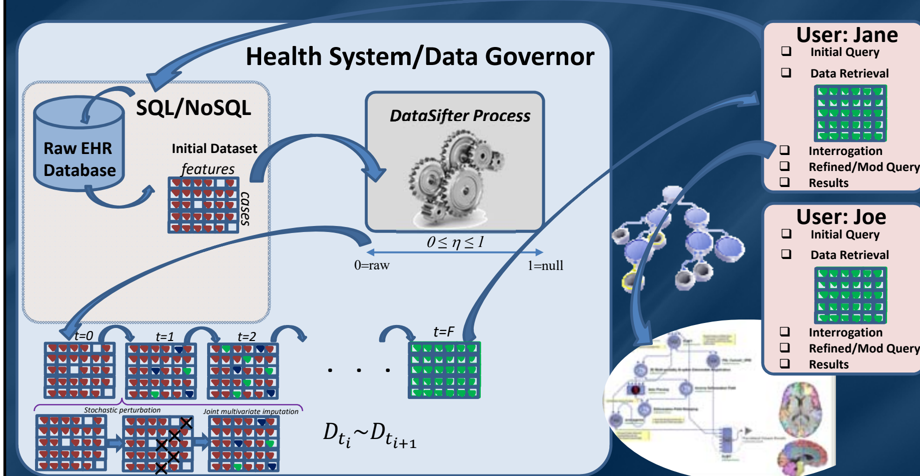
- SOCR TensorBoard (10K*200 tensor)

https://socr.umich.edu/HTML5/SOCR_TensorBoard_UKBB

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), …

DataSifter: Reliable, Effective & Secure Data Sharing

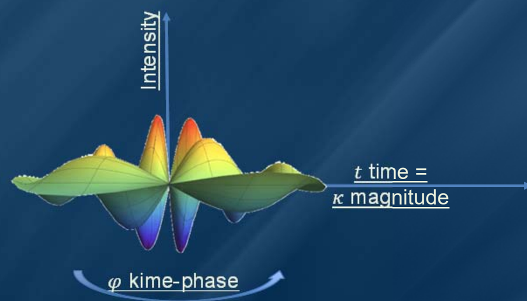


http://DataSifter.org US patent #16/051,881 Marino, *et al.*, *JSCS* (2019)

Spacekime Analytics: Longitudinal Time-series ➔ Kime Surfaces

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.



Dinov & Velev (2021) De Gruyter | https://spacekime.org

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 $Kryder\ Law \gg Moore\ Law$ trend
- **Web**: https://SOCR.umich.edu
- **Git**: https://github.com/SOCR
- **Projects**: https://socr.umich.edu/html/SOCR_Research.html
- **Apps**: https://socr.umich.edu/HTML5/

Acknowledgments

Slides Online: "SOCR News"

Funding

NIH: P20 NR015331, U54 EB020406, P50 NS091856, P30 DK089503, UL1TR002240, R01CA233487

NSF: 1916425, 1734853, 1636840, 1416953, 0716055, 1023115

Collaborators

- SOCR: **Milen Velev, Daxuan Deng, Yueyang Shen, Zijing Li, Yongkai Qiu, Zhe Yin, Yufei Yang**, **Yuxin Wang**, Alexandr Kalinin, Selvam Palanimalai, Syed HusainJuana Sanchez, Dennis Pearl, Kyle Siegrist, Rob Gould, Nicolas Christou, Hanbo Sun, Tuo Wang, Yi Wang, Lu Wei, Lu Wang, Simeone Marino
- **UMich MIDAS/MNORC/AD/PD Centers**: Chuck Burant, Kayvan Najarian, Stephen Goutman, Stephen Strobbe, Hiroko Dodge, Chris Monk, Issam El Naqa, HV Jagadish, Brian Athey

https://SOCR.umich.edu