

Type of data	Description	Source & Status
Synthetic populations and activity-based networks	A statistically valid representation of individuals from all countries; includes household composition, activity sequences and locations (e.g., home, work, school), social contact networks. Integrates 40+ public and commercial global datasets: Census, Landsat, OpenStreetMap, National Household Travel Survey, Multinational Time Use Study, European Social Survey, Microsoft US Building Footprints, National Center for Education Statistics.	Developed by UVA and used in numerous studies
Health fitness data	Activity data: 6M individuals; 169 countries; 791M activities tracked, 153M days of physical activity data over several years, gender, age, height, weight, and inferred location. Data on food, weight, workouts, goals for 1.7M users, demographic information for reach user, including age, gender, height, and ZIP Code/country, 320M workouts and 3B food entries over 5 years, 159 countries, 7 with >10K users.	Already available via Stanford team members
Electronic medical records and insurance claim data	Anonymized medical school health records with demographics, diagnosis, treatment, infections etc. All Payers Claims dataset, with information on medications, diagnoses and procedures. Electronic health registry covering the entire population of Denmark from 1996-2010. Covers all hospital encounters (inpatient admissions, outpatient visits, and ER visits).	Stanford, UVA, Yale, Iowa purchasing claims data
Detailed mobility	User locations from over 1B installed instances of a popular app, providing movement patterns of frequent users.	Data sharing agreement
Anti-microbial data	World's largest repository of anti-microbial resistance and use data.	Via CDDEP
Social media data	Purchased and publicly collected Tweets from different parts of the world (over 20B items from online media and 160B Tweets collected by Stanford team since 2008, >100TB total), filtered by various health related words.	Available. Purchasing additional
Prescription sales data	Weekly over the counter point of sales data of influenza-like illness (ILI) medications from Nielsen at a zip code level.	Discussion for purchase
Survey data	Detailed health behavior survey of 2200+ individuals related to influenza, <a href="http://habitlab.stanford.edu">habitlab.stanford.edu</a> , an app with 12,000+ active users, can serve as an IRB-consented data source for activities that tie to behavior change.	Led by Stanford team
Surveillance data	Influenza: from CDC, state health departments, WHO and Pan American Health Organization. Cholera in Bangladesh: International Centre for Diarrheal Disease Research, Bangladesh.	Available from Health agencies
Genomic sequence data	Data from Nextstrain for carrying out phylodynamic analysis, near real-time virological surveillance.	Some available, others under negotiation
Biomedical data	Stanford Biomedical Network Dataset Collection: 30+ datasets with relationships between biomedical entities, such as genes, drugs, diseases, tissues, functions, side-effects. Stanford Biomedical Knowledge Graph: large knowledge graph of proteins, genes, functions, diseases, and drugs with 22M+ nodes and 2.2B edges, built from publicly available datasets.	Publicly available; built by our Stanford team
Socio-political event data	Data on conflicts and related events from ACLED and the GDELT project.	Available publicly
Satellite-based data	NLDAS, GLDAS, MODIS, Dartmouth Flood Observatory: data on environmental conditions such as temperature, precipitation, flooding, that can be relevant to infectious disease spread.	Available publicly