# Jupyter for Health: Exploring Patient Generated Health Data, Standards, and Tools

**Speakers**

Simona Carini, Bill Chen, Jessilyn Dunn

**Contributors**

Simon Johnson, Ryan Lovett, Krish Patel, Fernando Pérez, JP Pollak, Min Ragan-Kelley, Travis Sauer, Kate Alexandra Scherz, Ida Sim, Ben Tseytlin, Maryam Vareth

# Speaker Intros




## Jessilyn Dunn, PhD

*Assistant Professor, Department of Biomedical Engineering, Department of Biostatistics & Bioinformatics, Duke University; Faculty Member of the Duke Clinical Research Institute*

*jessilyn.dunn@duke.edu*





## Simona Carini, MA

*Programmer / Analyst*
*University of California, San Francisco*
*Open mHealth*

*simona.carini@ucsf.edu*





## Bill Chen

*PhD Student, Department of Biomedical Engineering, Duke University*

*bill.chen@duke.edu*

# Agenda

🟢 **9:00 – 9:20 | Introduction to the JupyterHealth Project**

- Motivation and mission statement
- Introduction to JupyterHealth

📊 **9:20 – 10:00 | Hands-on Data Exploration**

- Setup your account for JupyterHealth Exchange
- Overview of data streams used in today's tutorial
- Introduce the CGMSandbox
- **Code along:** How to retrieve health data from the JupyterHealth Exchange
- **Code along:** Visualizing wearable health data using built-in overlays

⌚ **10:00 – 10:30 | Wearables and Patient Generated Data**

- Continuous Glucose Monitoring (CGM) technology
- What is a digital biomarker?
- Challenges in digital biomarker discovery

☕ **10:30 – 10:50 | Break**

🧬 **10:50 – 11:30 | Code along: Extracting Digital Biomarkers**

- Extracting CGM-based biomarkers
- Using overlays to help us interpret measures
- Using sleep and diet data to help us extract better biomarkers

{ } **11:30 – 11:45 | Data Standards: Open mHealth, IEEE 1752**

📦 **11:45 – 12:30 | Wrap-Up, Discussion, Resources**

- Recap of the session
- **Discussion**: Questions, feedback, share your experience
- Repositories, documentation, and other resources
- How to stay connected with the project

jupyterhealth

*A subproject of JupyterHub, powered by Project Jupyter*

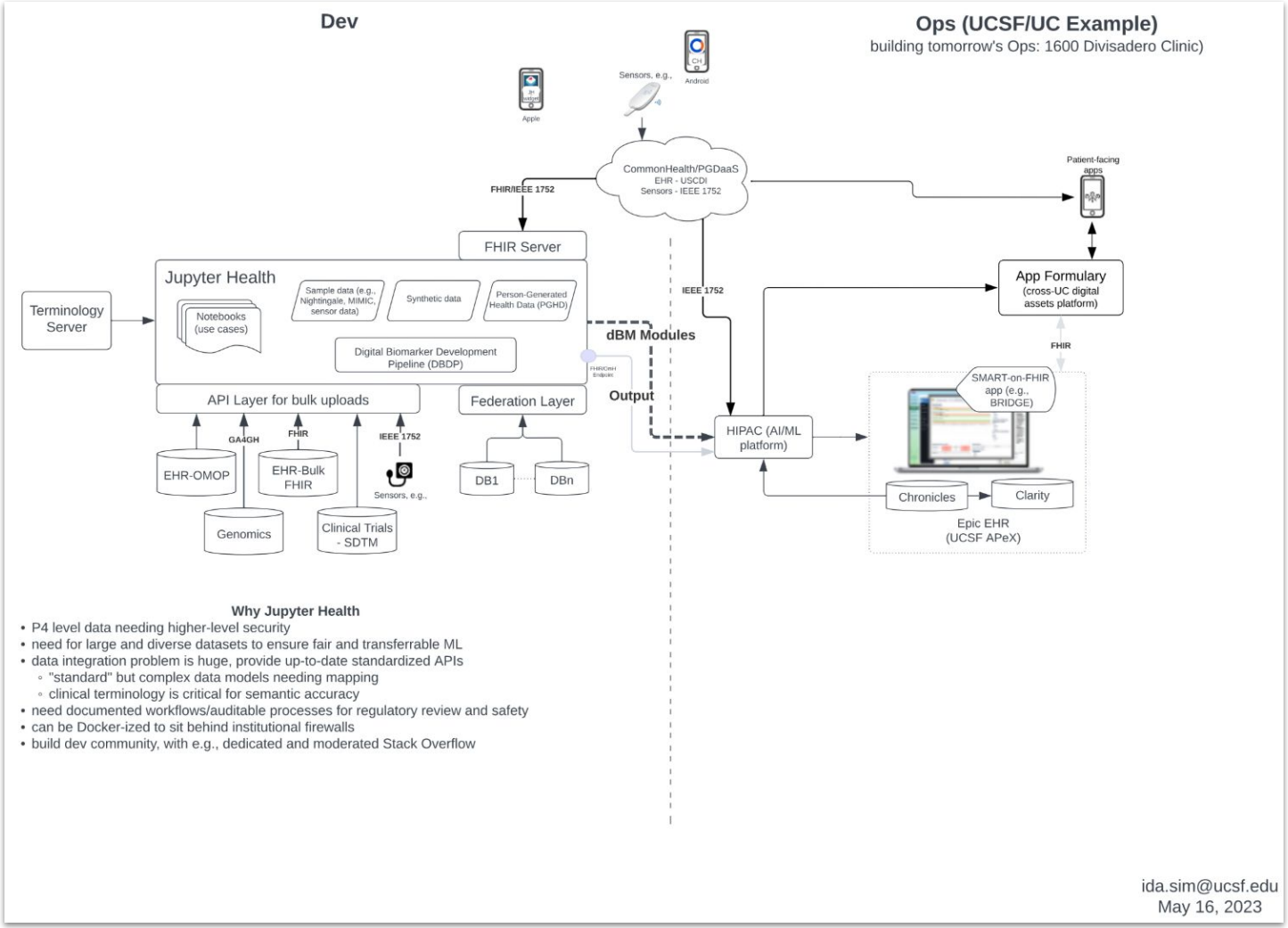



jupyterhealth.org

# 2023

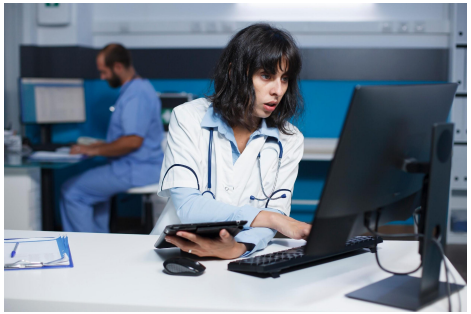
Ida's 1-pager.

It had all the key elements!



# The Problem

***Fragmented, Reactive Healthcare Systems***



- Healthcare systems operate in silos, limiting real-time data access for clinicians and patients.
- Inefficient workflows shift focus away from care, wasting time and resources.
- Despite advancements like wearables, care delivery remains episodic, not proactive.
- Integration challenges restrict the potential of life-saving algorithms and remote monitoring.

# The Opportunity

***A New Era of Connected, Data-Driven Healthcare***



- The rise of wearables, IoT, and AI creates a pivotal moment for healthcare transformation.
- Real-time data integration can enable personalized care and earlier interventions.
- A unified platform is essential to securely connect diverse data, bridging clinical care and research.

# jupyterhealth enables:





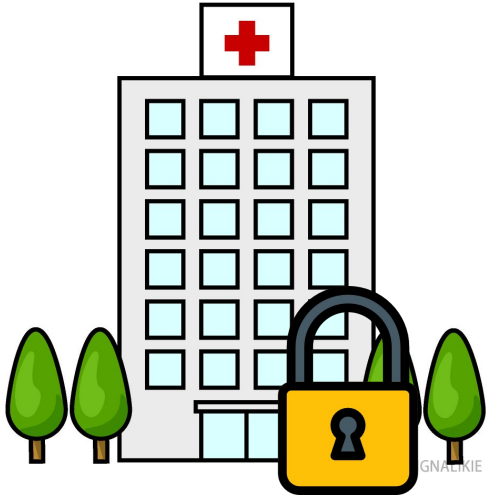
JupyterHealth Exchange





JupyterHub







1000 flowers blooming

# JupyterHealth

***open-source modular platform*** with privacy-preserving real-time data acquisition and reusable analytic pipelines to drive precision health care



Many individualized bouquets that mix and match over time

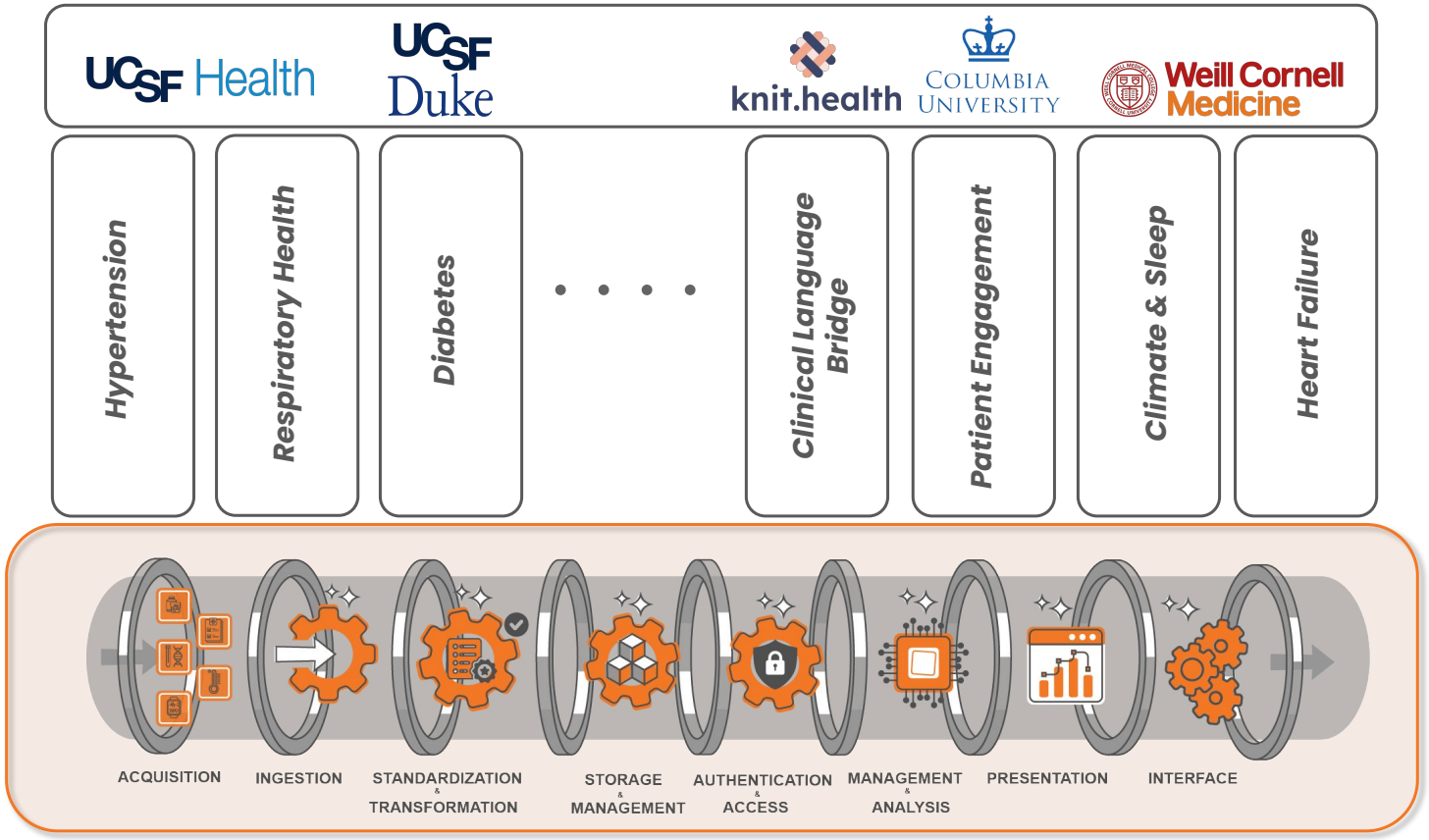
**Current State**: *Healthcare today is like a garden of 1000 blooming flowers—piecemeal, fragmented, with tools and platforms existing in isolation.*

**The Vision**: *Precision bouquets: modular, tailored arrangements of tools like glucometers, mobility and sleep trackers, nutrition advice based on your microbiome, customized to meet each person's unique and evolving needs.*

**The Solution:** ***JupyterHealth*** *enables florists-at-scale, providing an* ***open-source modular platform*** *to arrange and adapt precision bouquets, seamlessly connecting* ***wearables, IoT devices, mobile apps, and Electronic Health Records (EHRs)*** *into a* ***unified, secure, and AI-powered ecosystem***

# Ecosystem Overview:

*Horizontal Foundation and Vertical Solutions*

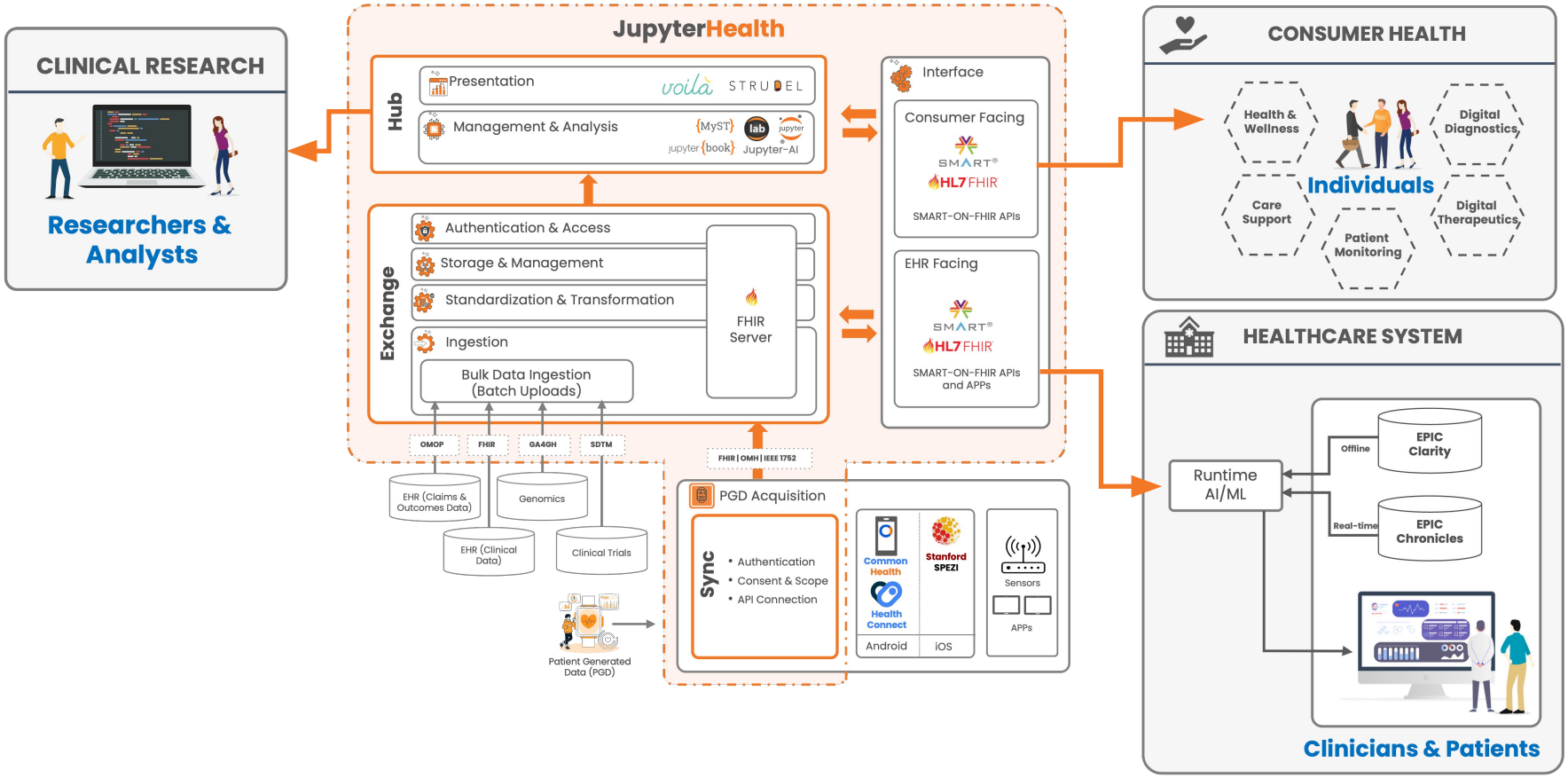



**EARLY ADOPTERS**

**Vertical applications** provide targeted solutions for specific use cases...

**JupyterHealth** combines data interoperability with the analytics power of Project Jupyter to create a scalable, interactive foundation for healthcare transformation.

# JupyterHealth Architecture & Data Flow



# JupyterHealth MVP Key Components

- **Data acquisition**: Bring data into the JupyterHealth ecosystem. Direct or facilitated. For now, focused on Patient Generated Data.
- **Data management**: JupyterHealth Exchange securely stores patient generated data. Provides scoped, authenticated access via standards-based FHIR API.
- **Data analysis**: A special JupyterHub provides access data from JupyterHealth Exchange. Incorporate data from any other source. Use the analytical and visualization packages of your choice.
- **EHR integration**: Display finished analyses and visualizations for participating patients in the EHR, supported by SMART on FHIR.

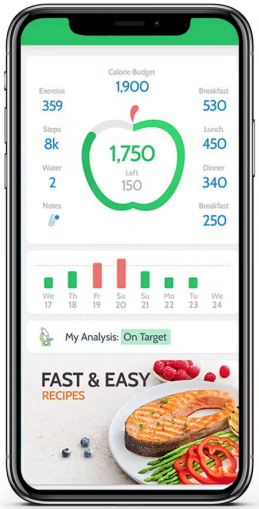
# Data Acquisition: Approach Overview

- **Focus on Patient Generated Data**: JupyterHealth currently supports IEEE 1752, Open mHealth and FHIR data from wearables, sensors, apps, and other external data sources.
- ***Out of the box* support for some devices and apps:** Use the CommonHealth app to collect data through Google Health Connect or thousands of EHR endpoints. Use Stanford's Spezi toolkit to build mobile apps that can contribute data.
- **Consent and management**: JupyterHealth supports patient invitation links with scoped data requests. Tools like CommonHealth, Health Connect, HealthKit, etc, support further scope refinement by users.
- **Open and Modular**: We provide documentation, templates, and APIs that make it easy to swap in or write new tools for acquiring data.

*Thousands of devices, apps, and services used by hundreds of millions of people can contribute data based on this model*

Types of data that we will be looking at today.

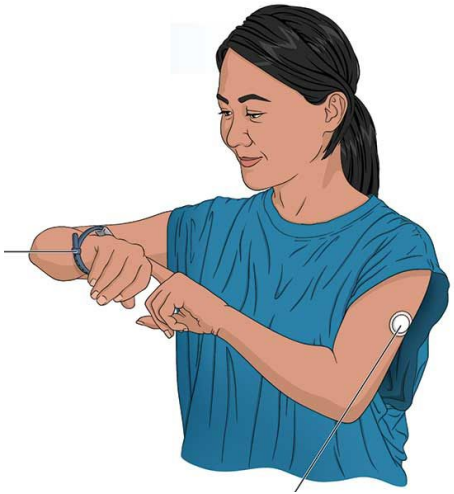
## Digital Food Logs





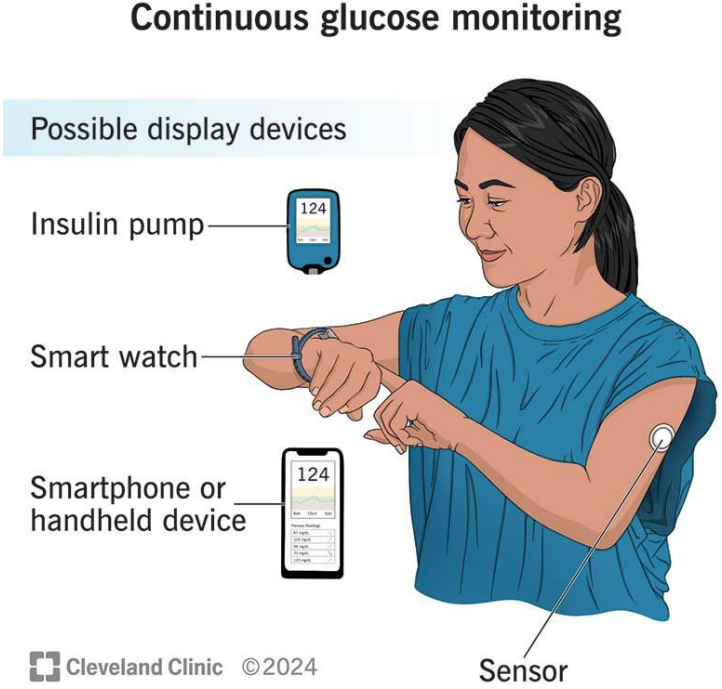


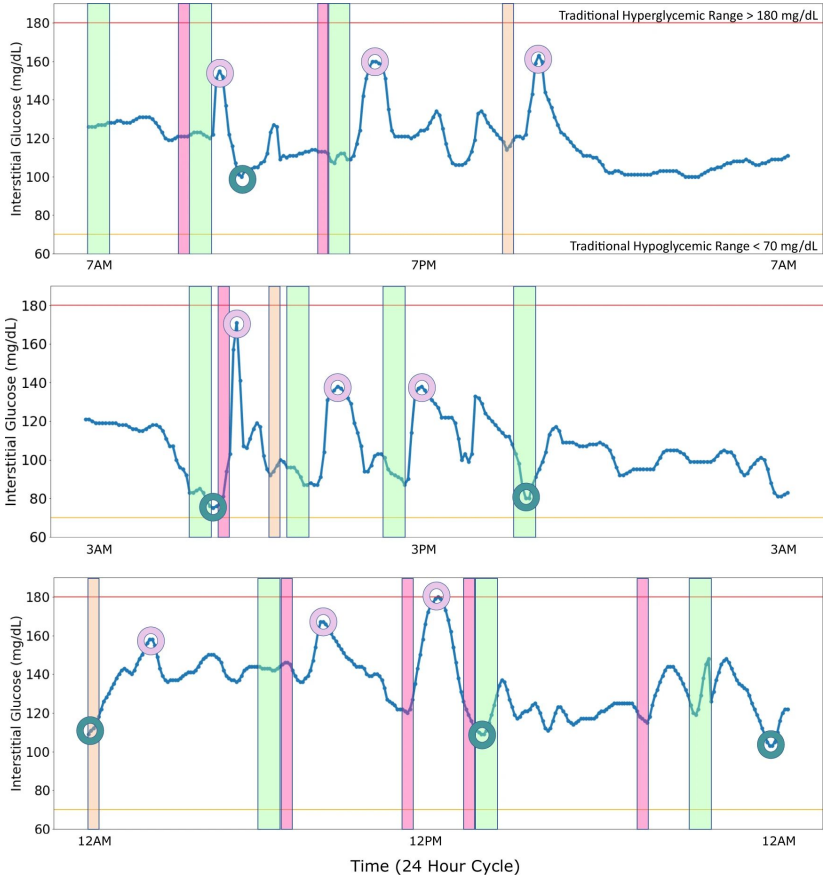
## Smartwatch



## CGM

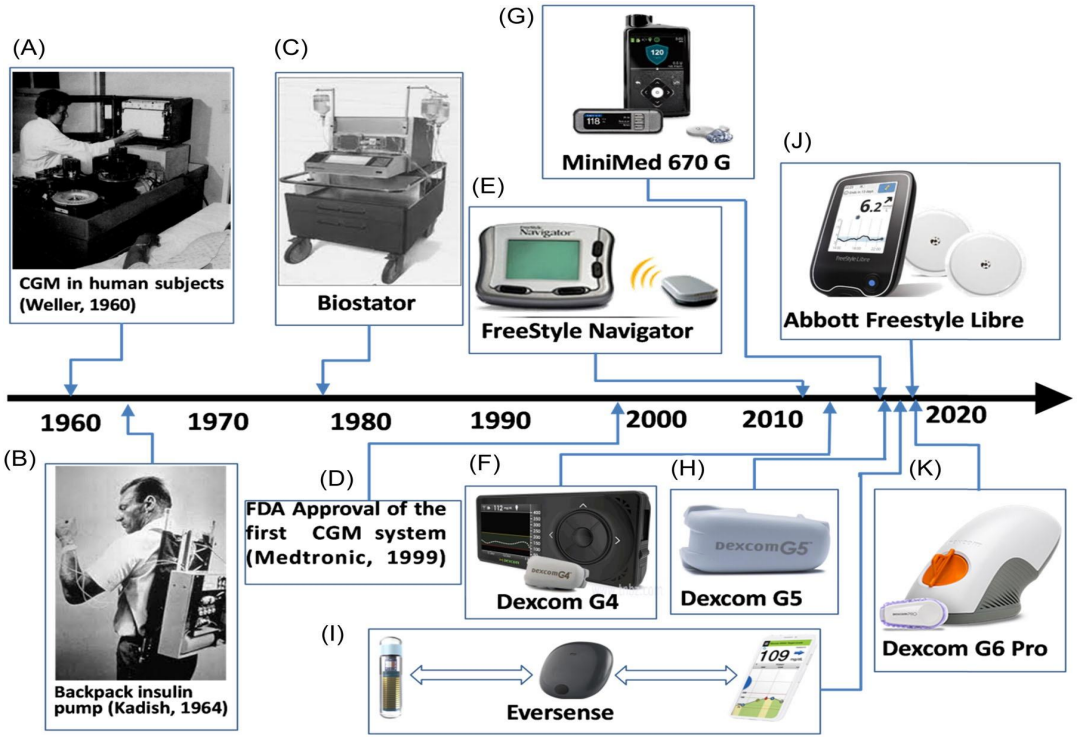
# Diabetes monitoring



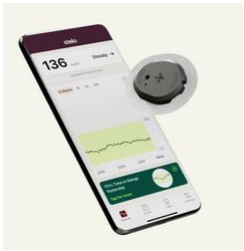

Bent, B., Cho, P.J., Henriquez, M. et al. Engineering digital biomarkers of interstitial glucose from noninvasive smartwatches. npj Digit. Med. 4, 89 (2021). https://doi.org/10.1038/s41746-021-00465-w

# CGM Innovation Over the Years



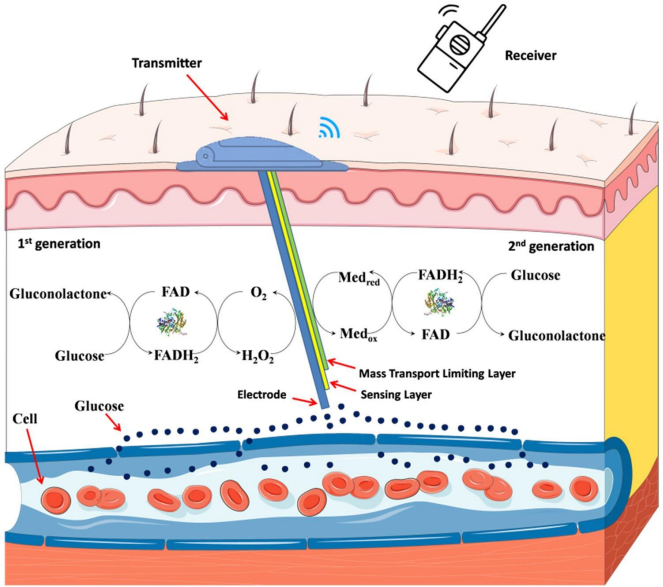

August 2024 - Dexcom Stelo
September 2024 - Abbott Lingo



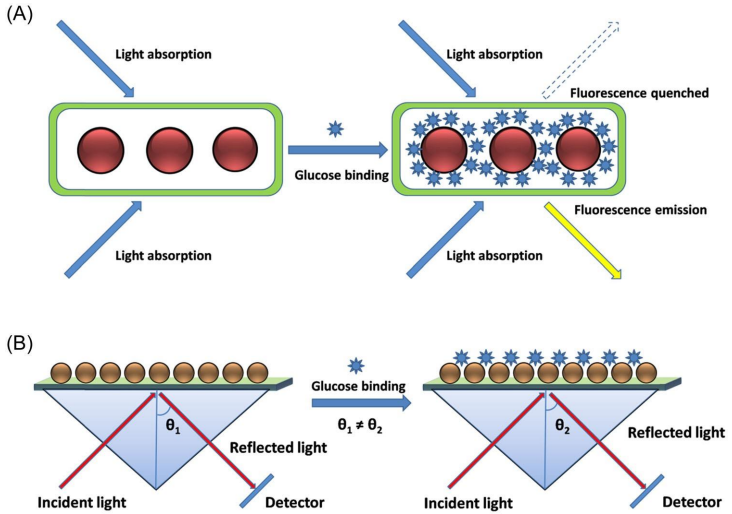


Jin, Xiaofeng & Cai, Andrew & Xu, Tailin & Zhang, Xueji. (2023). Artificial intelligence biosensors for continuous glucose monitoring. Interdisciplinary Materials. 2. 10.1002/idm2.12069.

# Electrochemical

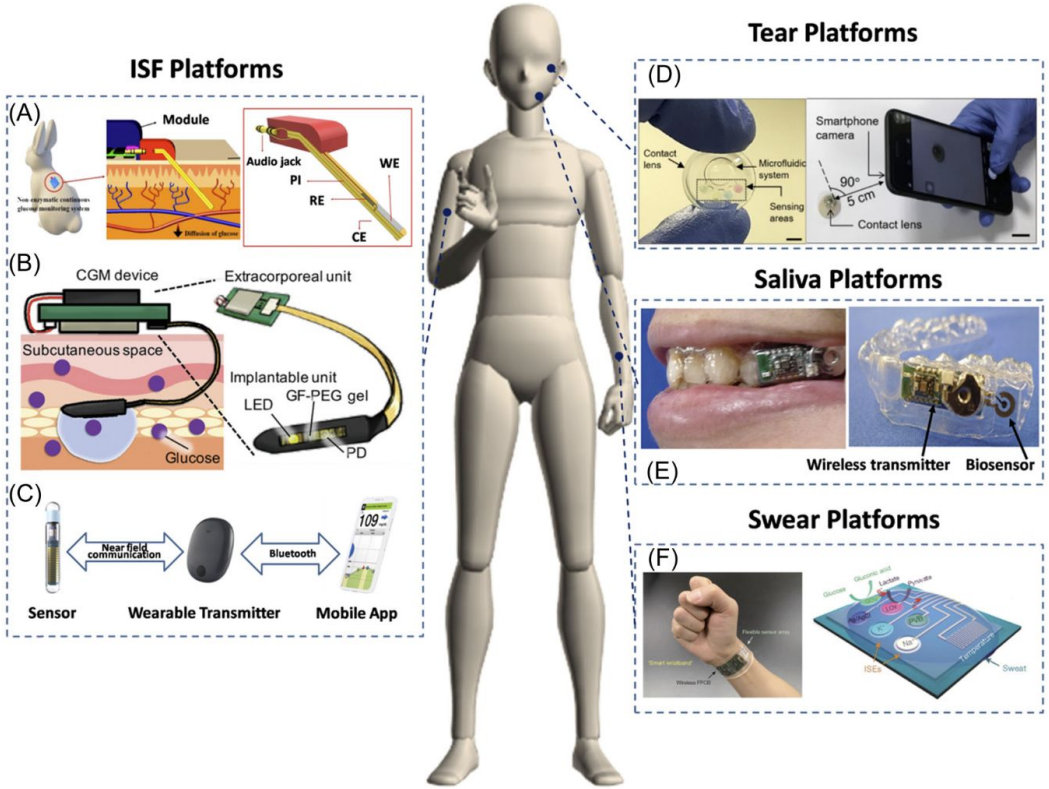



# Optical




Jin, Xiaofeng & Cai, Andrew & Xu, Tailin & Zhang, Xueji. (2023). Artificial intelligence biosensors for continuous glucose monitoring. Interdisciplinary Materials. 2. 10.1002/idm2.12069.

# Future: Completely non-invasive glucose sensing



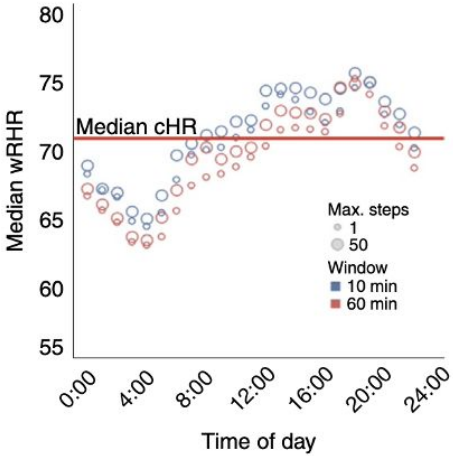

Jin, Xiaofeng & Cai, Andrew & Xu, Tailin & Zhang, Xueji. (2023). Artificial intelligence biosensors for continuous glucose monitoring. Interdisciplinary Materials. 2. 10.1002/idm2.12069.

# Why wearables?

Continuous biosignal measurement can now be seamlessly integrated into daily life

Wearables capture more consistent biosignals than in-clinic measurements

- Biosignals change over the course of the day
- Normative values are insufficient– no one is average







Dunn, Kidzinski et al. *Nature Medicine*, 2021
https://news.mit.edu/2020/sensors-monitor-vital-signs-0423

# Wearables Data Deluge

- Wearables generate massive amounts of data
- How can we use that data to make actionable insights?
- Solution: Digital Biomarkers

# Digital data becomes a Digital Biomarker when a relationship to a health outcome is established

**Target Outcomes** $= f($ **Measurements** $)$

- **Clinical measurements**
- **Survey data**
- **Illness events**

- **Heart rate**
- **Skin Temperature**
- **Skin Conductance**
- **Accelerometry**
- **Location**



The Digital Biomarker Discovery Pipeline (DBDP) provides open source ***data***, ***code***, ***algorithms***, and ***educational resources*** to make discovering digital biomarkers more accessible and establish best practices for the field.




DBDP.org

Bent et al. *JCTS*, 2020.

# Current Challenges In Digital Biomarker Discovery

- Device/Data Validity & Accuracy
- Interpretability  & FAIR principles
- The “Data Deluge”
- Regulatory oversight of tools
- “ELSI”, Privacy, Security

# Extracting Digital Biomarkers

## Extracting CGM-only Biomarkers

| Category | Measure | Definition |
|---|---|---|
| Mean-glucose measures | Time in Range (TIR) (70-180 mg/dL) | Percent of time where glucose measures are in range 70-180 mg/dL. |
| | Glucose Management Indicator (GMI) | Linear transformation of the mean glucose value, meant to estimate HbA1c. $\text{GMI} = (.02392 * mean(G)) + 3.31$ |
| High-glucose level measures | Mean Amplitude of Glycemic Excursions (MAGE) | Average amplitude of peak-to-trough excursions identified via moving-average crossings, keeping only excursions ≥ SD×multiplier. $\text{MAGE} = mean\{\text{peak–trough amplitudes} \geq k \cdot \text{SD}\}$ |
| | High Blood Glucose Index (HBGI) | Measures of the frequency and extent of high blood glucose readings using Kovatchev transform. $f = 1.509(ln(\text{G})1.084 - 5.381)$, $r = 10f^2$, $\text{HBGI} = mean(r \mid f > 0)$ |
| Variability measures | Coefficient of Variation (CV) | Coefficient of variation of all glucose values. $\text{CV} = 100 * sd(\text{G})/mean(\text{G})$ |
| Low-glucose level measures | Low Blood Glucose Index (LBGI) | Measures of the frequency and extent of low blood glucose readings using Kovatchev transform. $f = 1.509(ln(\text{G})1.084 - 5.381)$, $r = 10f^2$, $\text{LBGI} = mean(r \mid f < 0)$ |

# A software package to help you with extracting CGM-based biomarkers

cgmQuantify

**TABLE 1.**
COMPARISON TO OTHER OPEN-SOURCE CGM SOFTWARE

| CGM Software | Number of metrics that are clinically validated | Functions for plots? | Python | R | Access |
|---|---|---|---|---|---|
| cgmquantify | 28 | Yes | Yes | Yes | PyPi CRAN DBDP MD2K |
| EasyGV[8] | 12 | No | No | No | Need to request access |
| cgmanalysis [9] | 21 | Yes | No | Yes | CRAN |
| CGManalyzer [10] | 7 | Yes | No | Yes | CRAN |

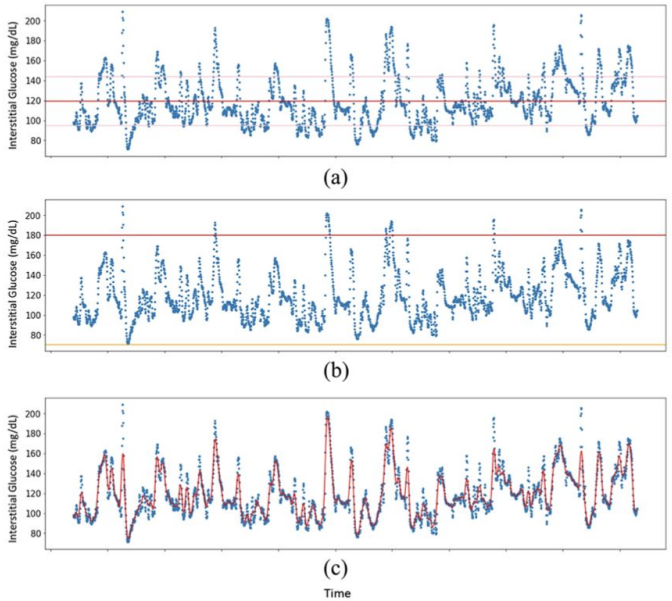



**FIGURE 1.** Visualizing longitudinal CGM data with the cgmquantify Python package. Shown are a visualization with indicators of the mean interstitial glucose level (red) and 1 SD from the mean (pink) (a), a visualization with indicators of hyperglycemic (>180 mg/dL glucose, red) and hypoglycemic (<70 mg/dL glucose, yellow) (b), and a plot with LOWESS smoothing of the glucose data (c).

Bent, Brinnae, Maria Henriquez, and Jessilyn P. Dunn. "Cgmquantify: Python and R software packages for comprehensive analysis of interstitial glucose and glycemic variability from continuous glucose monitor data." IEEE Open Journal of Engineering in Medicine and Biology 2 (2021): 263-266.

# Data Standards

- Large amounts of data from (one or more) sensors and apps: data deluge
- Rapidly changing landscape: New sensors, apps, person-reported outcomes (PROs) & continually changing software, hardware, algorithms
- Challenges to align, harmonize, combine data from different sources for different purposes, including integration into frontline workflow (EHR-based)

**Standards are needed for interoperability, data representation for data exchange**

- Standardization
  - Syntax
  - Semantics
  - Relevant context

# Example: Blood Glucose Instance

```
{
    "blood_glucose": {
        "value": 128
    }
}
```

```
{
    "blood_glucose": {
        "unit": "mg/dL",
        "value": 128
    },
    "effective_time_frame": {
        "date_time": "2024-02-05T08:25:00-07:00"
    },
    "temporal_relationship_to_meal": "fasting",
    "temporal_relationship_to_sleep": "on waking"
}
```









# Example: Measures of Sleep

| Name |
| --- |
| .. |
| {..} apnea-hypopnea-index-1.0.json |
| {..} arousal-index-1.0.json |
| {..} deep-sleep-1.0.json |
| {..} light-sleep-1.0.json |
| {..} sleep-episode-1.0.json |
| {..} sleep-onset-latency-1.0.json |
| {..} sleep-stage-summary-1.0.json |
| {..} snore-index-1.0.json |
| {..} time-in-bed-1.0.json |
| {..} total-sleep-time-1.0.json |
| {..} wake-after-sleep-onset-1.0.json |



This schema represents the amount of time between when a person starts to want to go to sleep and sleep onset



This schema represents the summary duration of all awakenings (i.e. minimum 15 secs) after sleep onset in a main sleep episode.

# Open mHealth, IEEE P1752 Family of Standards

Open mHealth (late 2011) developed schemas and tools

IEEE P1752 Open Mobile Data Working Group
Chair: Dr. Ida Sim, UCSF

- Std. 1752.1-2021 IEEE Standard for Open Mobile Health Data—Representation of Metadata, Sleep, and Physical Activity Measures Document + Open-source schemas in IEEE-sponsored repository
- P1752.2 Standard for Mobile Health Data: Representation of Cardiovascular, Respiratory, and Metabolic Measures

Schemas (templates) define how instance data should look like in terms of data types, properties, etc. (syntax) and what they represent (semantics)

# Overview of schemas

P1752.2

**Metabolic, e.g.,**

- blood glucose
- body weight
- food entry

1752.1

**Physical activity, e.g.,**

- physical activity
- (step count)

1752.1

**Metadata, e.g.,**

- header
- schema-id

Open mHealth

**Medication, e.g.,**

- med prescription, adherence percent
- patient med schedule

1752.1

**Environmental, e.g.,**

- ambient temperature

1752.1

**Sleep, e.g.,**

- total sleep time
- sleep episode
- WASO

1752.1

**Utilities, e.g.,**

- time, units of measure
- descriptive statistics

1752.1

**Survey**

- question, answer

P1752.2

**CardioRespiratory, e.g.,**

- pulse
- blood pressure
- respiratory rate
- O2 saturation

# Recap of key learnings

- The JupyterHealth Project
- Some characteristics of health data
- Introduction to the CGMSandbox:
  - Data streams
  - Visualizations
  - Overlays
- Wearables, patient generated data, and digital biomarkers
- Opportunities and challenges
- Standards for data exchange

# References

- JupyterHealth website - https://jupyterhealth.org/
- DBDP website - https://www.dbdp.org/
- IEEE 1752.1 repository - https://opensource.ieee.org/omh/1752
- Open mHealth website - https://www.openmhealth.org/

To receive announcements from the project, go to https://groups.google.com/g/jupyter-health/ and ask to join