


Bioinformatics Institute (BII) Scientific Conference 2023

**Creating value from
biomedical data**

Sebastian Maurer-Stroh
ED BII
sebastianms@bii.a-star.edu.sg



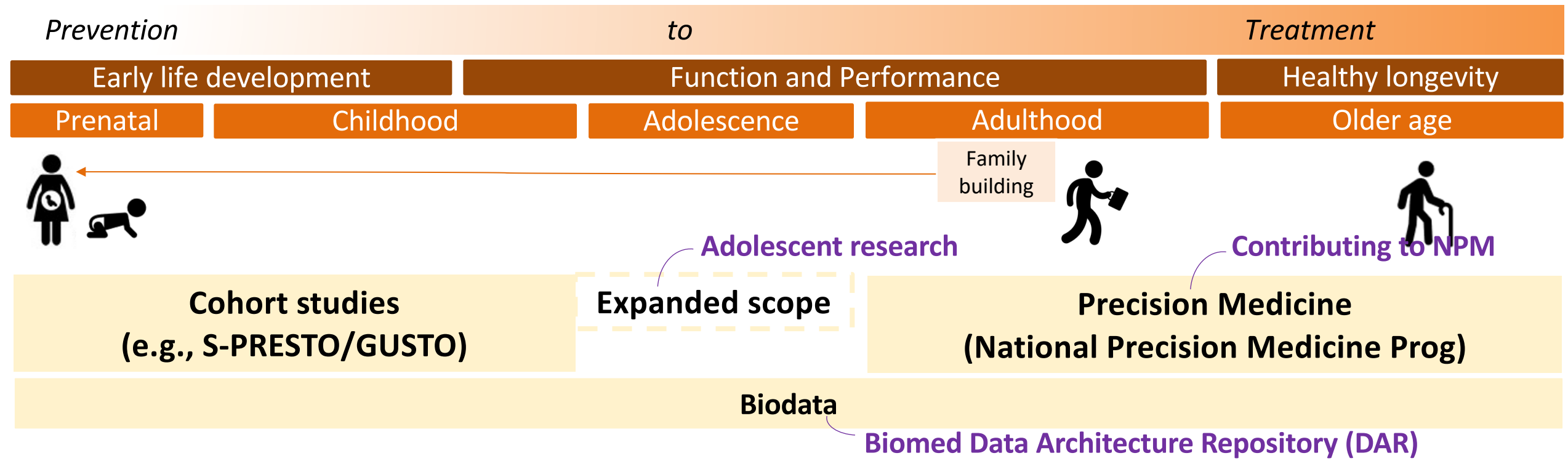
2021	\$35M total
2022	\$110M total
2023	\$235M goal

Baseline investment sum
for cohorts hosted by BII

One simple ask for all of you:

Let us **create value** from
biomedical data together

To address national health challenges & needs, A*STAR is increasingly directing efforts towards population health and well-being

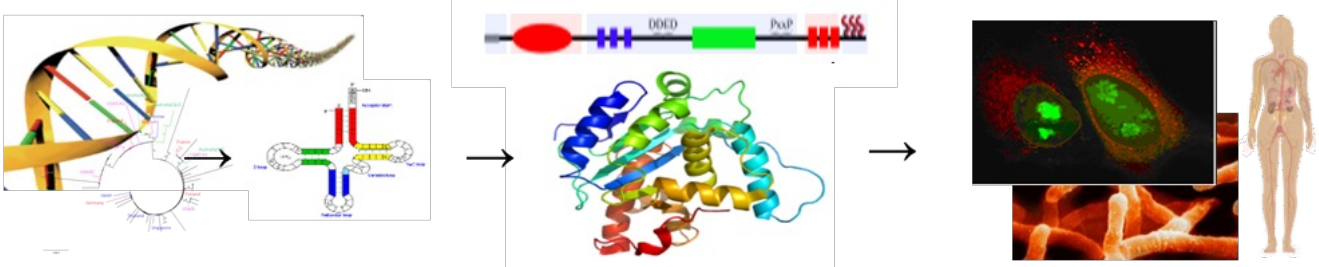


Building on existing strengths in Early Life Research (e.g., S-PRESTO/GUSTO), Genomics (Precision Medicine) and Biodata capabilities (Biomed DAR), BMRC will **expand our scope to address broader Population Health challenges for Singapore**

A*STAR BII's Winning Strategy

Mission Creating Value from **Bi**omedical Data

Genomics at GIS,
other biodata at BII



requires unique domain expertise
(mix of biologists and computer scientists)

A*STAR's BII – A Simple Formula



What is our strategy?
What is the “cathedral”
we are building in BII?

BII

Creating Value from Biomedical Data

RI's Strategy

Overall Goals in 5 years

A*STAR and BII as the Biomedical Data Analysis partner of choice for the whole ecosystem

Thrust 1: Data Hub

Biomed DAR++

Thrust 2: Data Analysis

Increase collaboration capacity

Thrust 3: People Development

HBMS SchoOL → IBIDA, Talents

**Key Programme 1
Epidemic Preparedness**

COVID → Disease X with ID labs and GISAID

**Key Programme 2
Digital Health**

Digital Pathology, Cardio → SingHealth HTP/DHA

**Key Programme 3
Human Potential**

ATTRaCT + PRISM, PRECISE, 3D AI, GUSTO with SICS

BIODATA-highway

Why?

Biomedical data is very valuable (e.g. >\$110M value hosted in 2022)

For who?

Singapore and the world

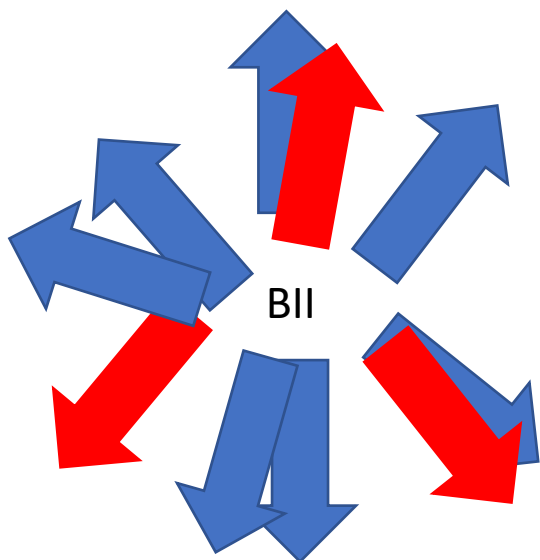
How?

Extend usable life span of high value biodata as secure datahub

Upcycle biodata with computational analysis and tool development for actionable insights

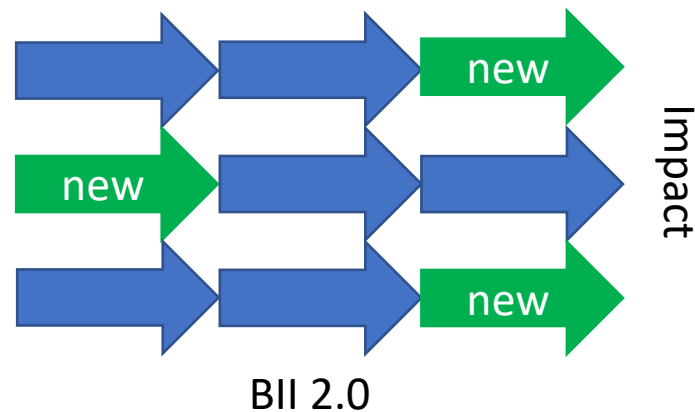
Changes from FY22 and old BII:

- Grow groups in new focus areas (Datahub, Syn Data, Bio-ChatGPT)
- Bring in SICS data management and software development team
- Pivot groups (including phasing out all wetlab activities)
- Restructure admin and create specific leader/owner/pm roles

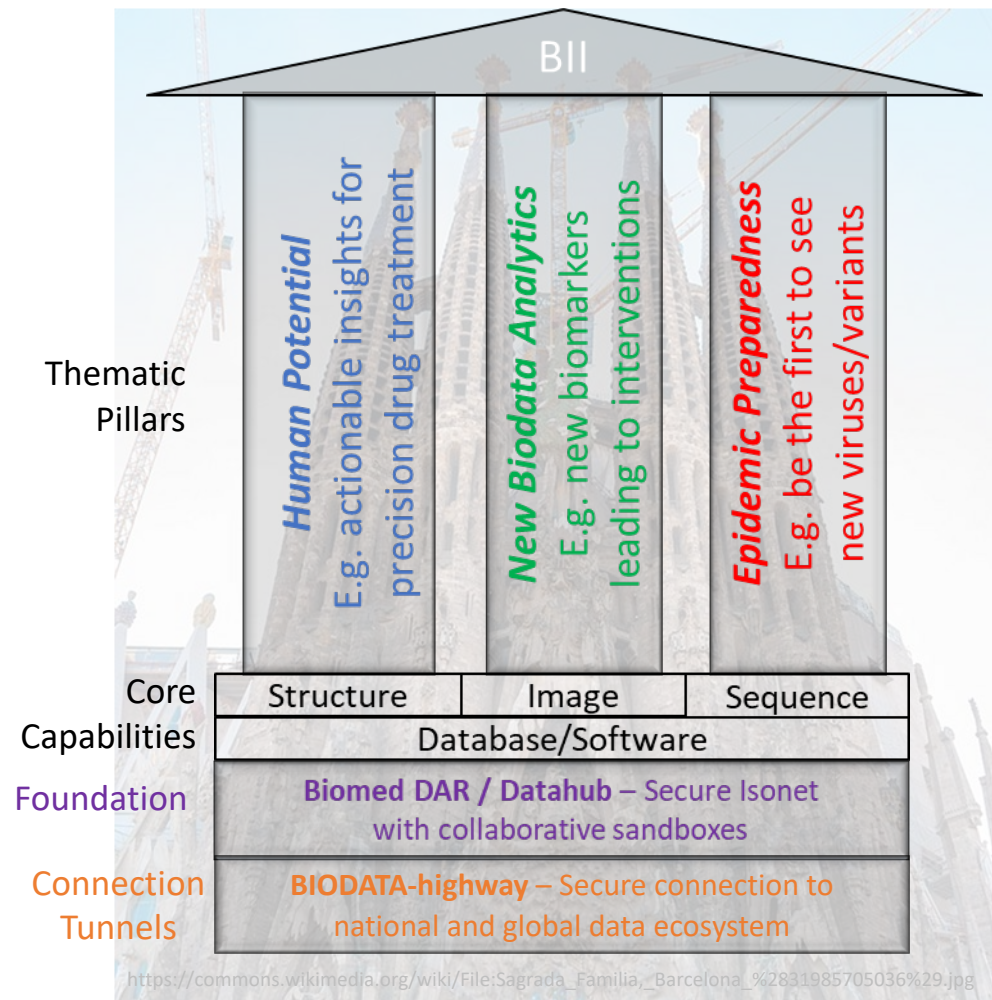


2023 aim:

- Everyone knows their role
- Everyone is connected to input and intended output



IMPACT



Getting everyone on board

Road to Impact

MISSION. We advance science and develop innovative technology to further economic growth and improve lives.

- How to measure our success? Papers? Industry money? Lives saved?
=> **IMPACT**
- What isn't IMPACT and why are we doing this? Mental models
- What is IMPACT?
- **STRATEGY** How can we organize ourselves towards IMPACT?
- **VISION** for the future

*a** CREATING GROWTH. ENHANCING LIVES



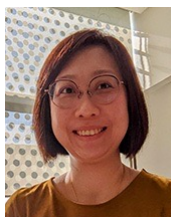
Why BII?

The only place with >100 Bioinformaticians within 4000km radius

BII

Critical mass needed to enable structured career path for **Bio**informaticians

- IBIDA course (300 people trained)
- Bilateral workshops (EDDC, SICS, SERI)
- Young talents: 4 new first time grant holders (OF-IRG, CDA, ID-HTCO, AI3-HTCO)
- **Connecting Bioinformaticians across RIs (25 joint appointees across 9 RIs)**
- **New group leads 2022/2023:**



WOO Xing Yi
(senior PI, Research
data integration)

100% BII



KOH Winston
(junior PI, Bio-
chatGPT)

BII-ISCE2 75-25% (new 2023)



MUKKESH Kumar
(Data Science
Manager)

BII-SICS 75-25% (new 2023)



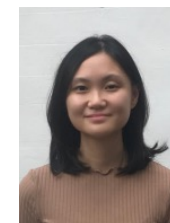
LAU Mai Chan
(Asst PI, Spatial
Immunomics)

BII-SIgN 60-40%



WANG Dennis
(senior PI, Bioinformatics
platform)

SICS-BII 50-50%



YEO Hui Ting Grace
(Asst PI, GIS-BII fellow
for Spatial Omics)

GIS-BII 60-40% (new 2023)

We should always ask,
what can we do better?

4 months course: 12 weeks +6h per day and 4 weeks full-day
= feels like aging 4 years in experience

88 case study discussions

9 plenary lectures

5 simulations

General Management Programme



Andy Wasynczuk

negotiation



Linda Hill



Scott Mayfield

finance

performance

leadership

accounting

marketing



V.G. Narayanan

purpose

innovation



Kash Rangan

strategy



Rafael Di Tella

economics



Stefan Thomke



Jan Rivkin

The case study method

Example: “Steve Jobs at Apple”

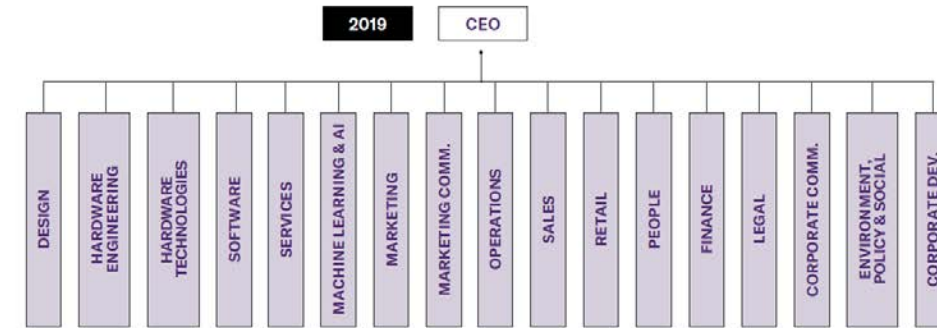
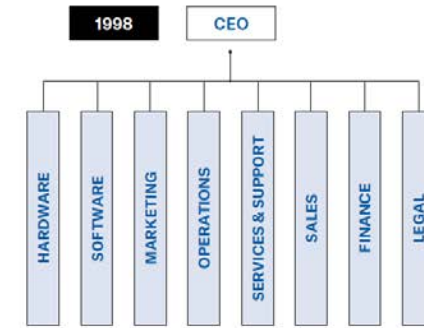
- Why has Apple been so successful?
- Is there as systematic "approach" to how Apple is managed (the "Apple Way")? If yes, how would you characterize this “system” and its “elements” (strategy, innovation, leadership, and execution)?



Example: “Steve Jobs at Apple”

- Imagine that you are an Apple competitor. How would you attack the company? Are there any weaknesses?

- What if Apple would buy your company, what would they do?



I am executive director of a research institute with 120 staff developing computational solutions for life science research and healthcare applications. I have identified 3 emerging areas that would be highly promising to venture into but only have seed funds to kickstart one of them. Intermediary and traditional alternative grants are slow to appear.

The 3 areas are using the following emerging technologies:

- digital twin via synthetic data
- ChatGPT
- quantum computing

They can create impact in the biomedical space (new findings and monetization) but the development time to impact differs:

- synthetic data (work already started, several competitors, first impact in 6-months)
- ChatGPT (fastest time to team, some competitors, first impact in 1 year)
- quantum computing (longer for future disruption, few competitors, impact in 3-5 years)

My questions for your feedback please are:

Should spread resources and pursue them all at once with prolonged development time?	0
Should focus resources on only one for targeted push, and which one? Finish first one?	2.2
Drop earlier work and jump on the next big thing?	1.8
Do bet for long-term disruption?	2

Emerging Areas in Computational Biotechnology (size = impact)

← Less

More →

In the field of computational solutions for life science research and healthcare applications, I have identified 3 emerging areas that would be highly promising to venture into.

My questions for your feedback please are:

Should spread resources and pursue them all at once with prolonged development time?	0
Should focus resources on only one for targeted push, and which one? Finish first one?	2.2
Drop earlier work and jump on the next big thing?	1.8
Do bet for long-term disruption?	2

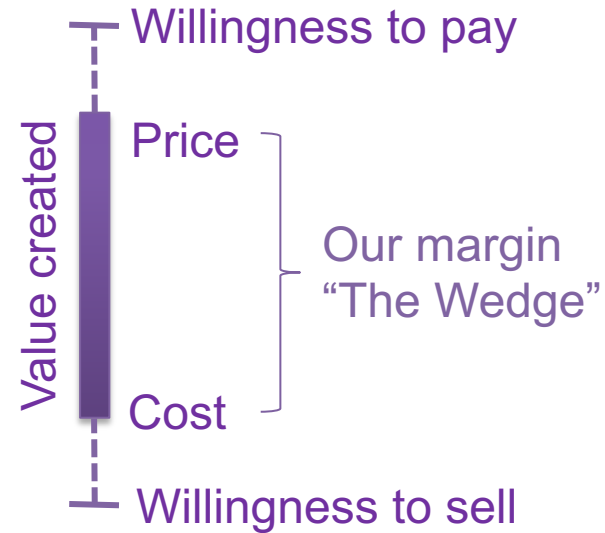
Emerging Areas in Computational Biotechnology (size = impact)

Simplicity is the ultimate sophistication

Classic: Microsoft Re-Designs the iPod Packaging
<https://www.youtube.com/watch?v=EUxnJraKM3k>



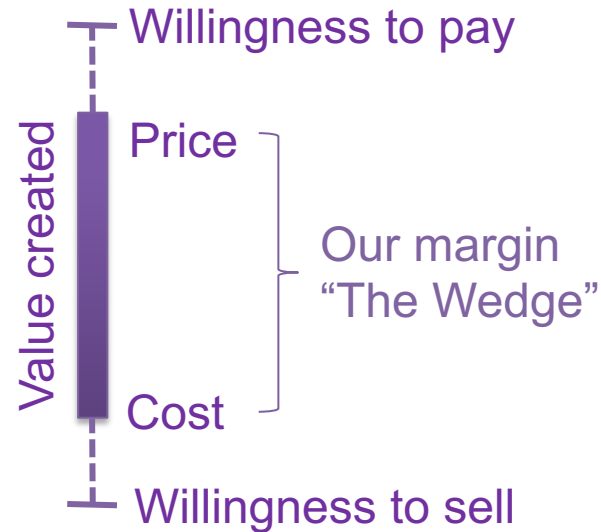
How can we as BII increase the value created from data?



Value Stick



How can we as BII increase the value created from data?



Value Stick

Different models to increase the wedge:

E.g. **Apple** (superior products)

E.g. **Best Buy** – both angles!

E.g. **Walmart** (reduce cost)



Strategy: Value Loop/Flywheel to grow Value Wedge

Weak loop:

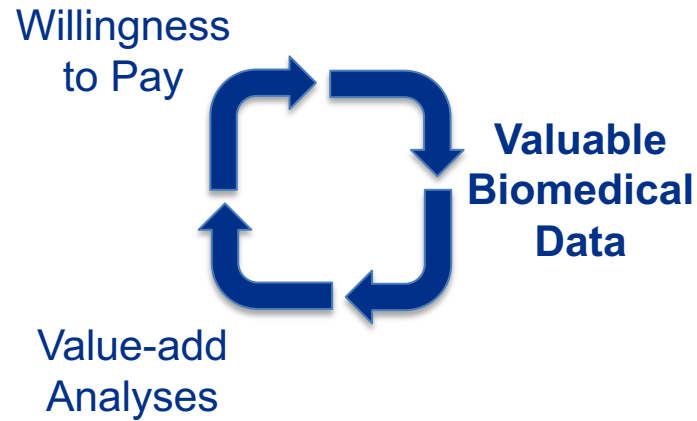
Lower price ->
more data ->
lower cost



Value Stick

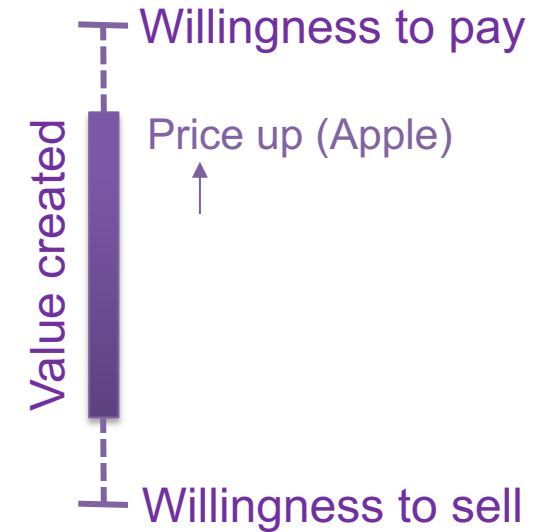


Value Loop/Flywheel to grow Value Wedge FASTER



Strong loop:

Add value ->
higher WTP ->
more data



Value Stick

Adding Actionable Insights

Example COVID-19:
data + analysis = impact

BII

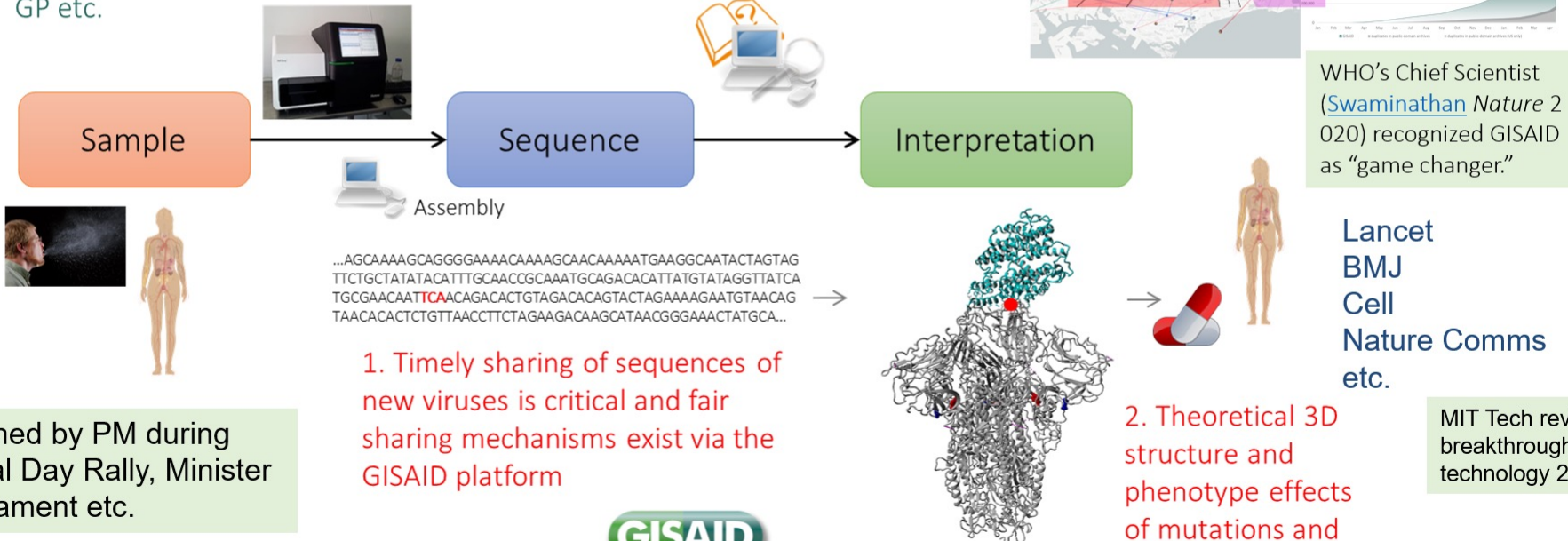
Creating Value from Biomedical Data

One A*STAR: BII works with ID Labs, SlgN, GIS, EDDC, DxDhub, IHPC, I2R

Hospital, GP etc.

Lab

Bioinformatics



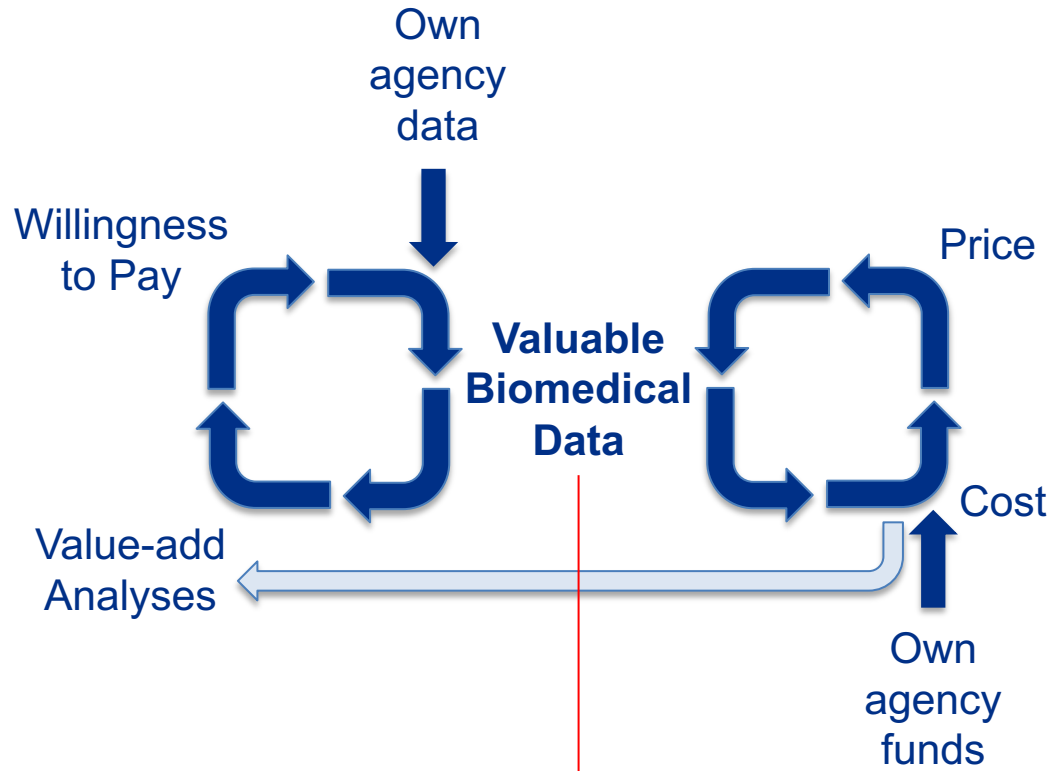
3. Comparison with other strains' sequences to trace global and local transmission



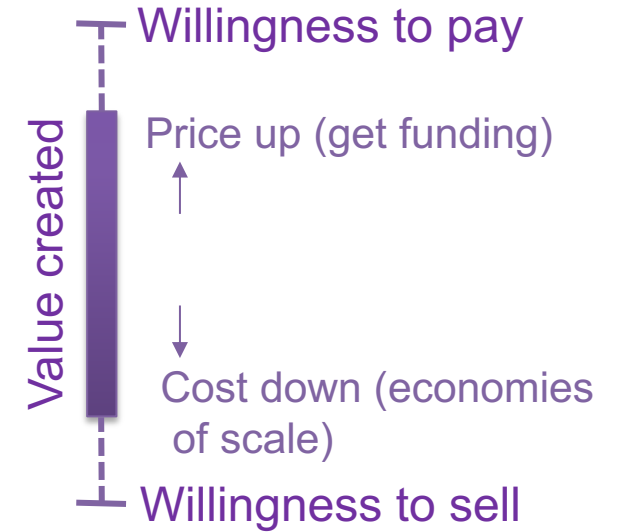
BII/GIS has global (GISAID, WHO, CEPI) and national role (NCID, MOH) in points 1-3



Value Loop/Flywheel to grow Value Wedge



2021 \$35M total
 2022 \$110M total
 2023 \$235M predicted



Value Stick

Market-share is not equal to profitability and growth is a choice (raising prices can slow growth and increase profitability)



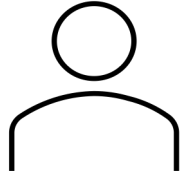
How to disrupt and survive disruption? Competitive Strategy

Blockbuster (physical DVD rental shops) vs Netflix (DVD-by-mail)

- **Blockbuster:** incumbent inertia, original business doing well, *“DVD-by-mail is ridiculous, you will never get to us”*
- **Netflix (DVD-by-mail):**
 - indirect attack in area new to incumbent,
 - growing the market by reaching new customers (turning problem into virtue: latest releases less available, recommendation engine buffers load and suggests unexpected new diversity)



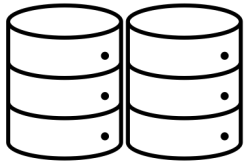
Barriers for Free Use of Health-Related Data Research



Patient re-identification is possible with access to few data elements



Regulatory, ethical and legal data protection hurdles



Extensive process is required for transferring and sharing of datasets for novel insights.

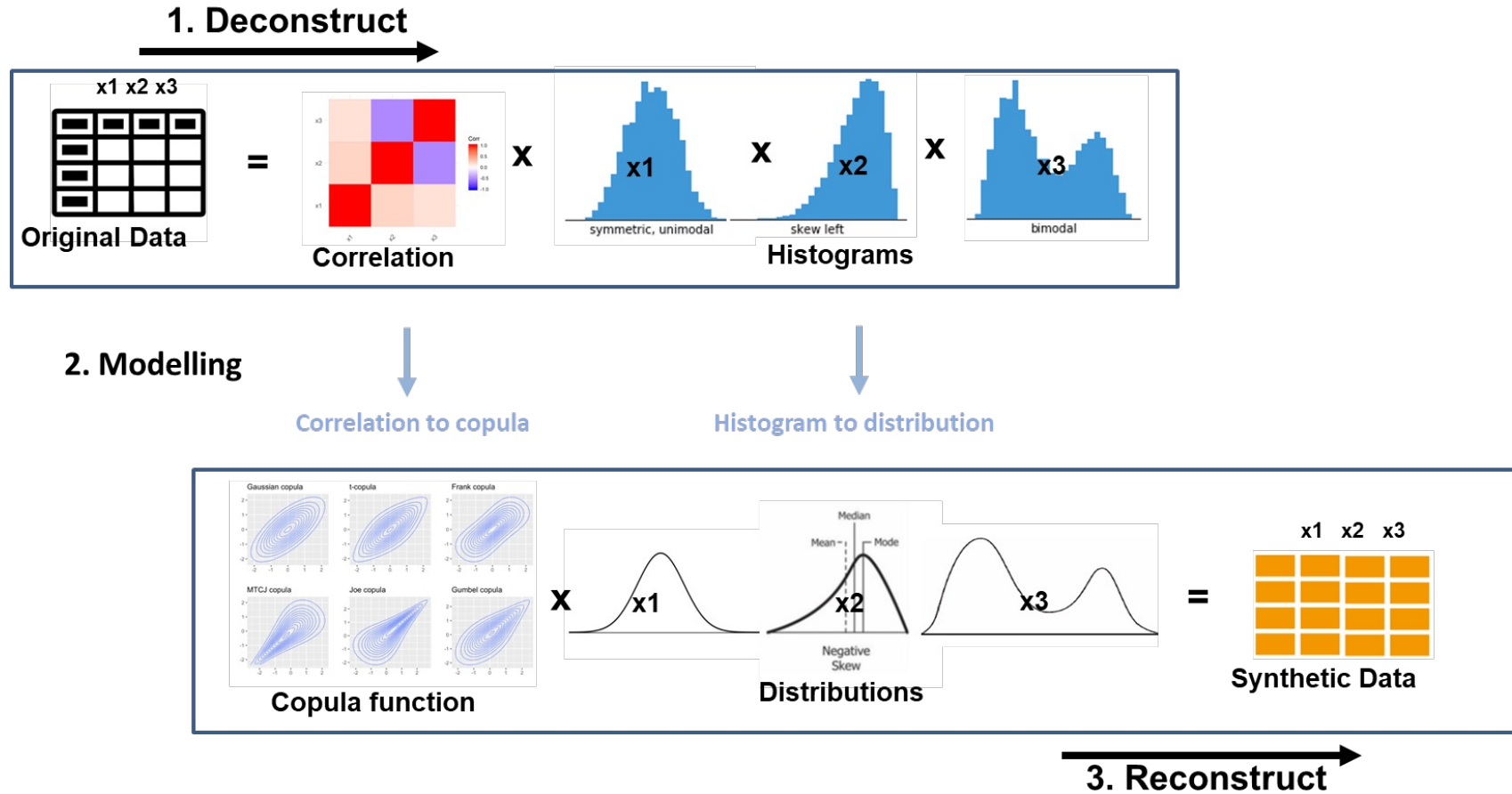


Heterogeneity of data increases difficulty of implementing high-end computing tools



What if ... we could remove sensitive aspects of the data but keep its biomedical value? **Synthetic data**

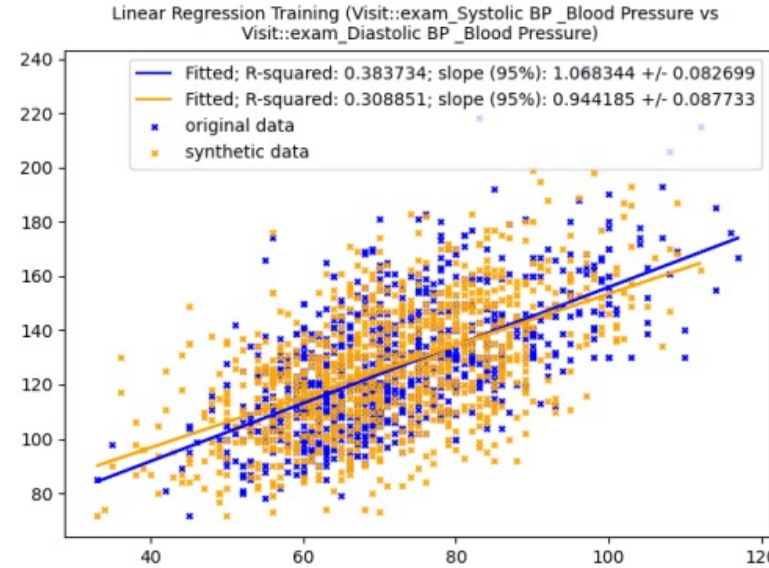
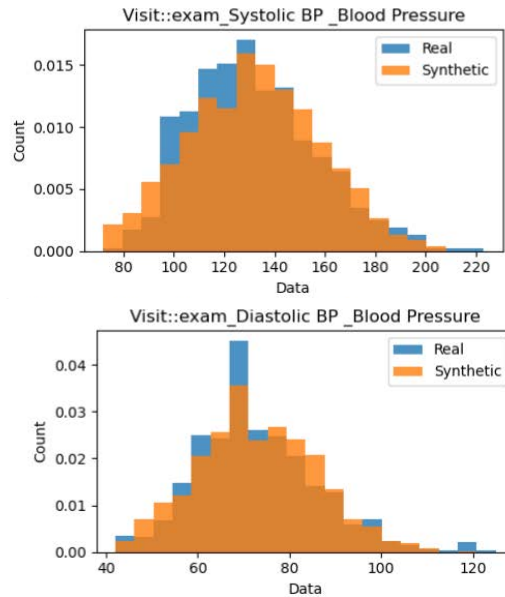
Data abstraction with preservation of correlation and data distributions:



We are not the first but early and have new technology expanding the realism of synthetic data



Synthetic data is “real” (value for some purposes)



Linear correlation of systolic with diastolic blood pressure

For example, we show on a real data set that the privacy-safe synthetic version still finds back the relevant clinical correlations.

Therefore, synthetic data can be used to find and develop models about diseases and treatments for patients but without the strings of not being able to share and use widely. This has the potential not only to complement but even to leap-frog classical research.



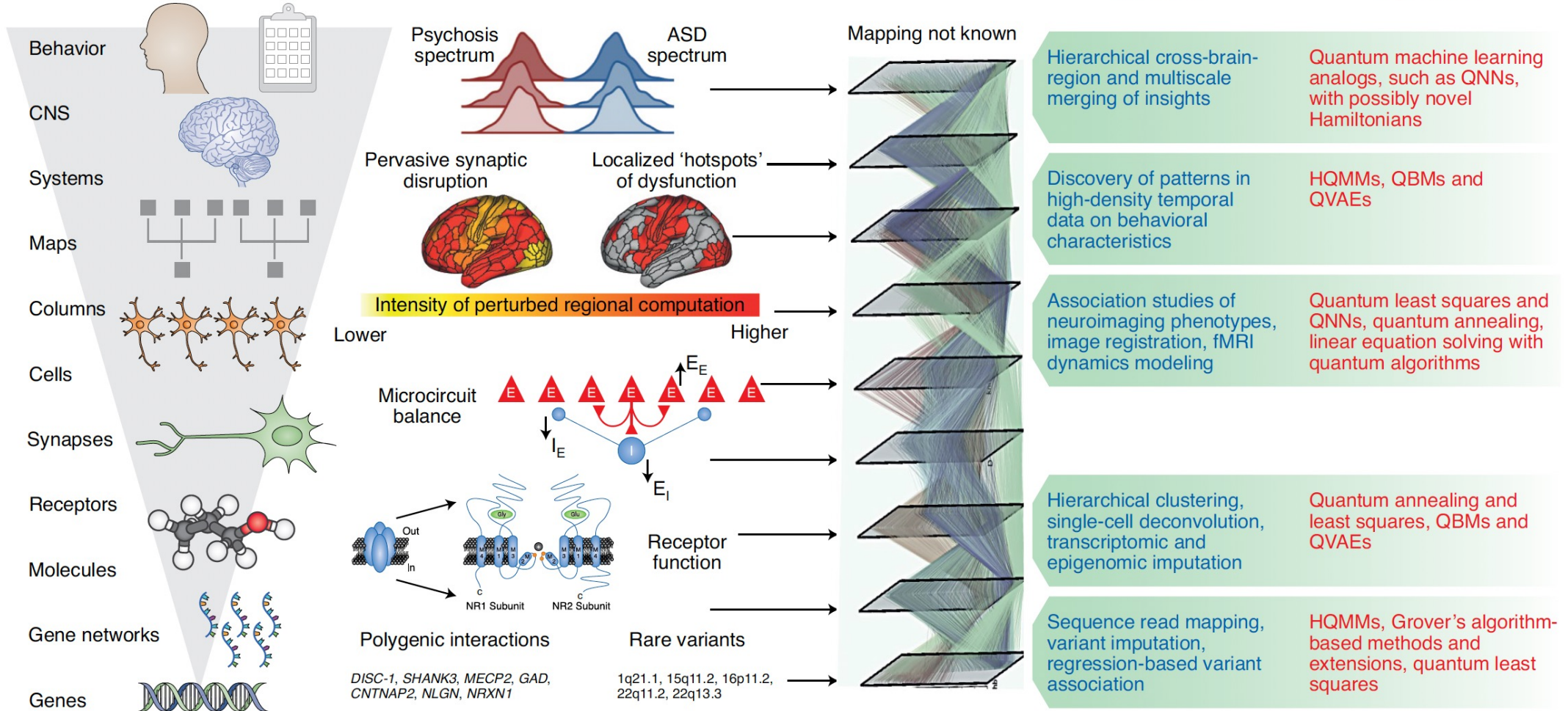
How to disrupt and survive disruption? Competitive Strategy

Blockbuster (physical shops) vs Netflix (DVD-by-mail) vs Netflix (streaming)

- **Blockbuster:** incumbent inertia, original business doing well, *“DVD-by-mail is ridiculous, you will never get to us”*
- **Netflix (DVD-by-mail):**
 - indirect attack in area new to incumbent,
 - growing the market by reaching new customers (turning problem into virtue: latest releases less available, recommendation engine buffers load and suggests unexpected new diversity)
- **Netflix (streaming):**
 - Once technical solution on par in quality, many other advantages would overwhelm old model
 - CEO saw risk of disruption early and chose to **self-disrupt** although original business was doing well then

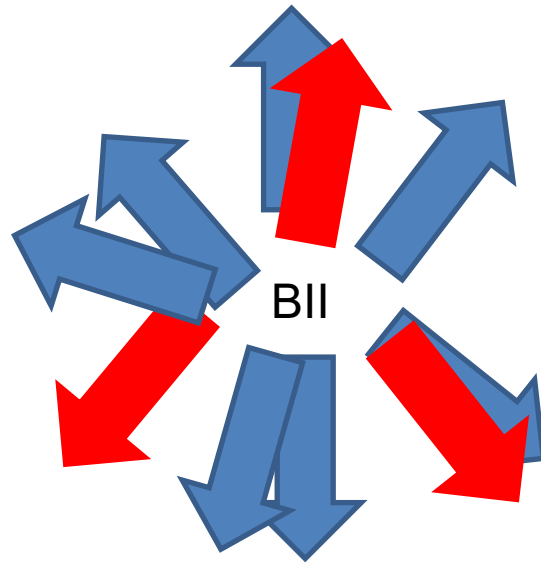
AI/Quantum enabling algorithms for drug and mechanism discovery

Computational complexity across level of analysis: many-to-many mapping problem



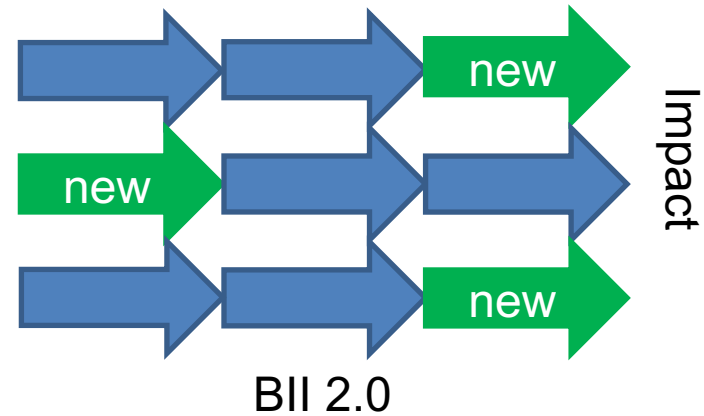
One simple ask for all of you:

Let us **create value** from
biomedical data together



2023 aim:

- Everyone knows their role
- Everyone is connected to input and intended output



Thank You!