

# バイオサイエンスデータベース、 10年の軌跡とこれから ～話題提供～

瀬々 潤

株式会社ヒューマノーム研究所 代表取締役社長

<https://humanome.jp>

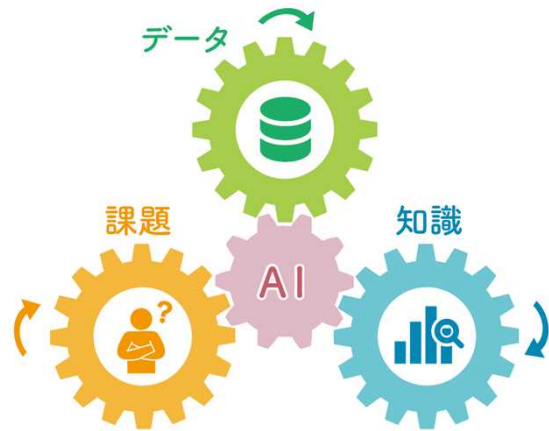
2020年10月5日

株式会社ヒューマノーム研究所

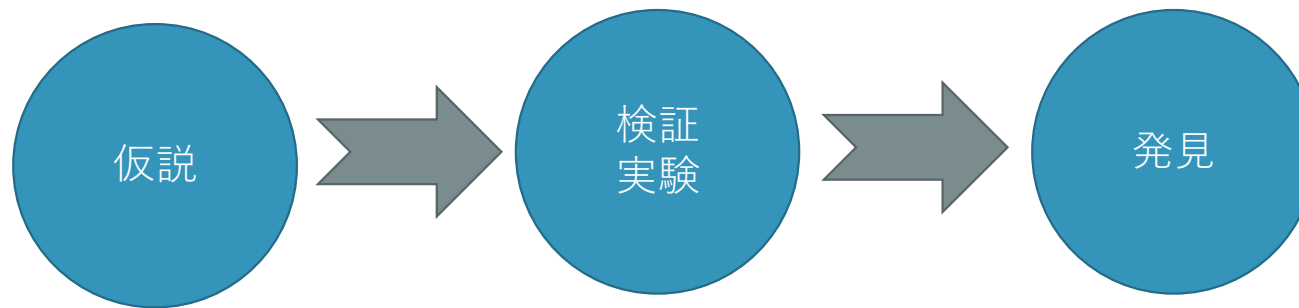
Licensed under a Creative Commons 表示4.0国際ライセンス  
©2020瀬々 潤 (株式会社ヒューマノーム研究所)



AIと最新テクノロジーで健康社会の実現を目指します



- 1 笑顔あふれる生活
- 2 予防医療の実現
- 3 生活スタイル改善

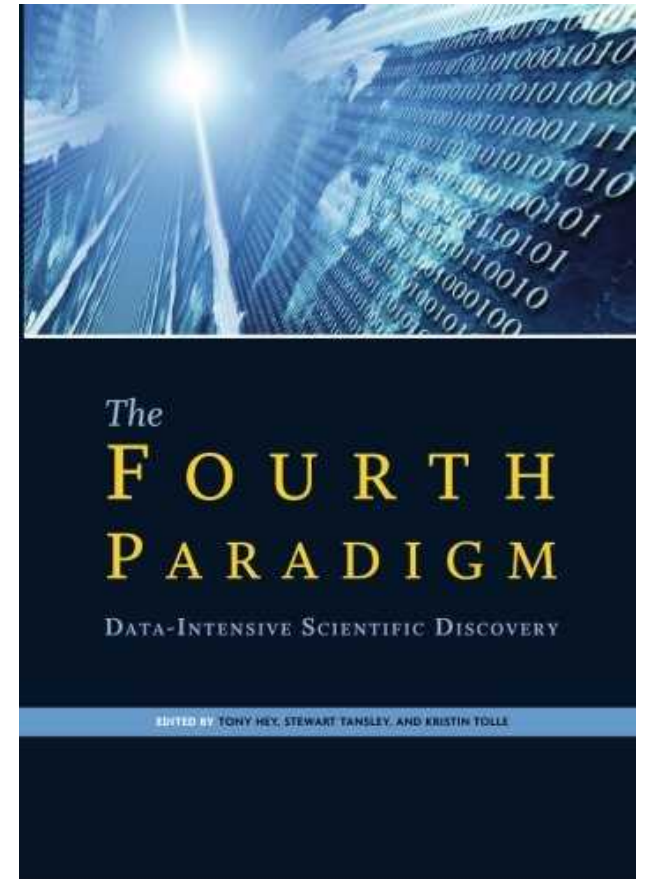


## データベース整備の重要性

# 生命科学では、より複雑な現象解明が求められる

偶然発生した異常例を発見し、原因究明が多くの生命科学の手法。このスキーマは、原因遺伝子の少ない単純例には有効。しかし、より複雑な現象には無効。生命科学の「行き詰まり」を解消しなければならない。

- このパラダイムシフトは、生命科学だけではない
- 業界を超えて、苦勞している。DXは一例。
- Data-intensive Science の推進が言われてきた。
- 生命科学の、汎用性へのチャレンジ。



# データベースは科学の水

The Economist Topics Current edition More

Regulating the internet giants

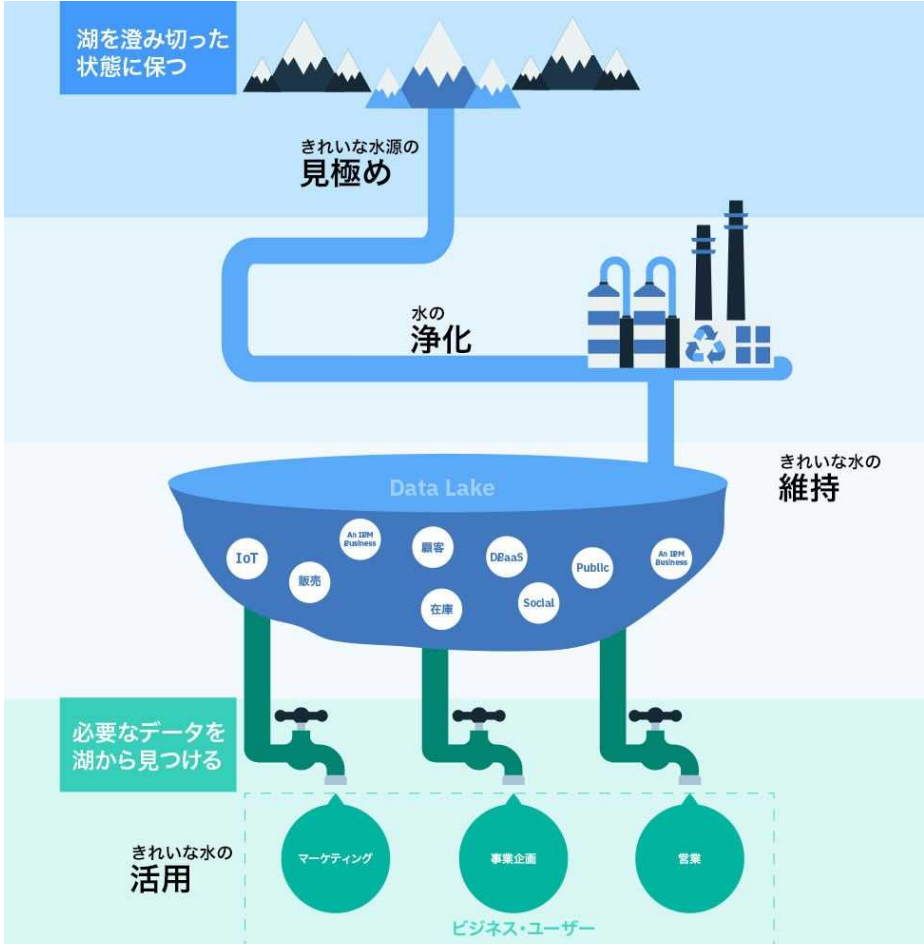
## The world's most valuable resource is no longer oil, but data

The data economy demands a new approach to antitrust rules

Print edition | Leaders >

May 6th 2017

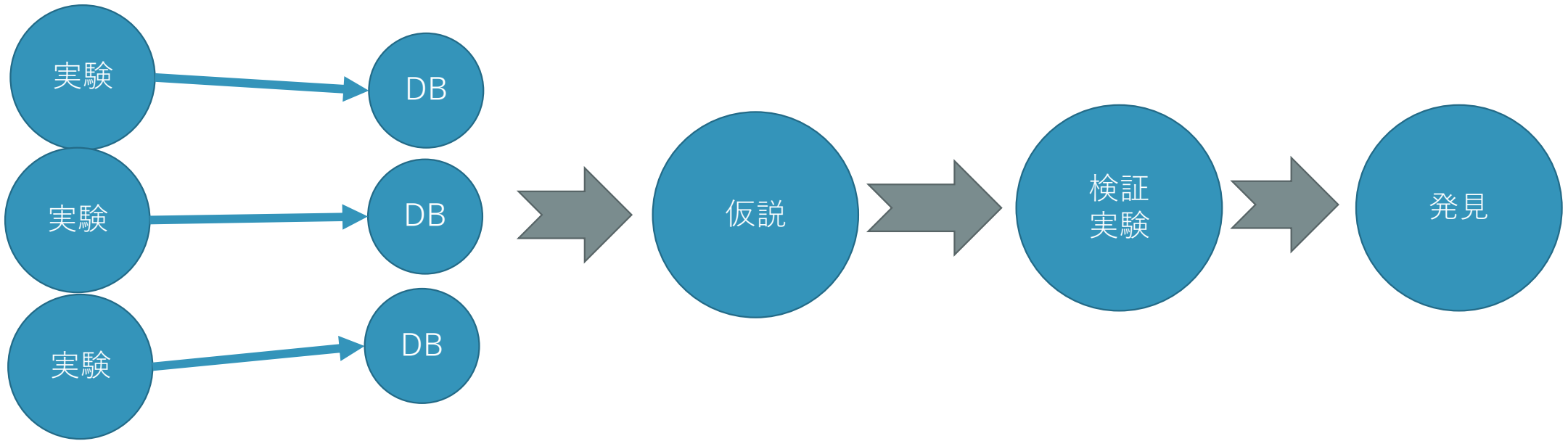
Twitter Facebook LinkedIn Email Print



<https://www.ibm.com/jp-ja/analytics/use-cases/governing-data-lake>

Licensed under a Creative Commons 表示4.0国際ライセンス ©2020瀬々潤 (株式会社ヒューマノーム研究所)





論文？実験？画像？

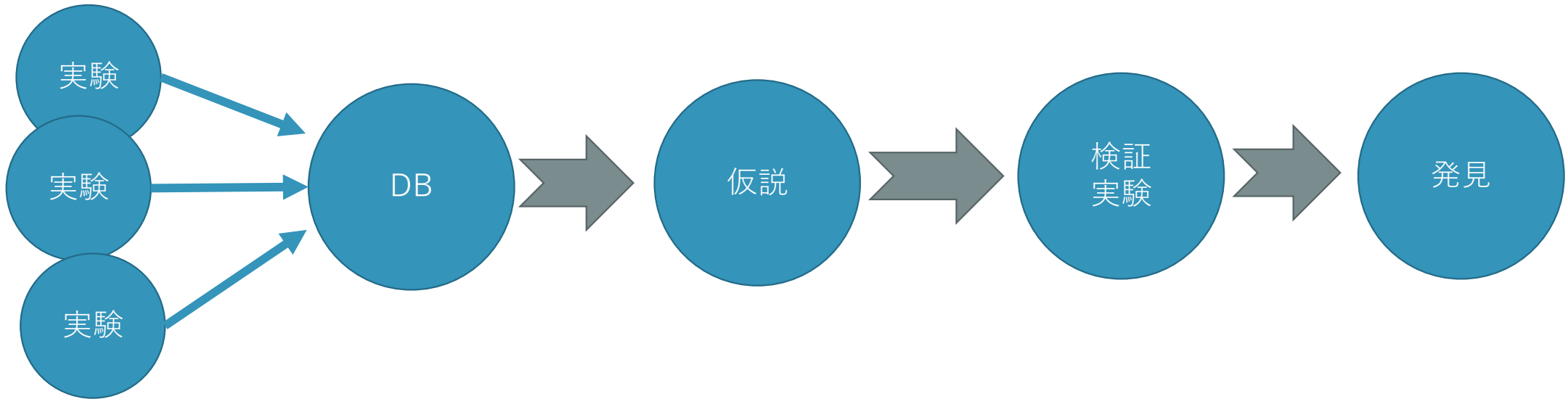
## 生命科学のデータベース整備は必ずしもコスパが良くない

### 多様な実験を対象とするので、汎用DBが困難

Microsoft Office の様に、一つのツールが何千万人・何億人も使うようなものであれば、投資効率が低い。一方、今まで生命科学で作られてきたデータベースは、特定の研究室の大規模実験用だったり、特定の現象専用だったり、特定の種に依存していたり・・・

- DBの汎用化が、投資効果をあげるポイント
- そもそも、今の生命科学実験は「キット化」が進んでいる
- 構造的なデータを貯めるだけであれば、方策があるのでは？







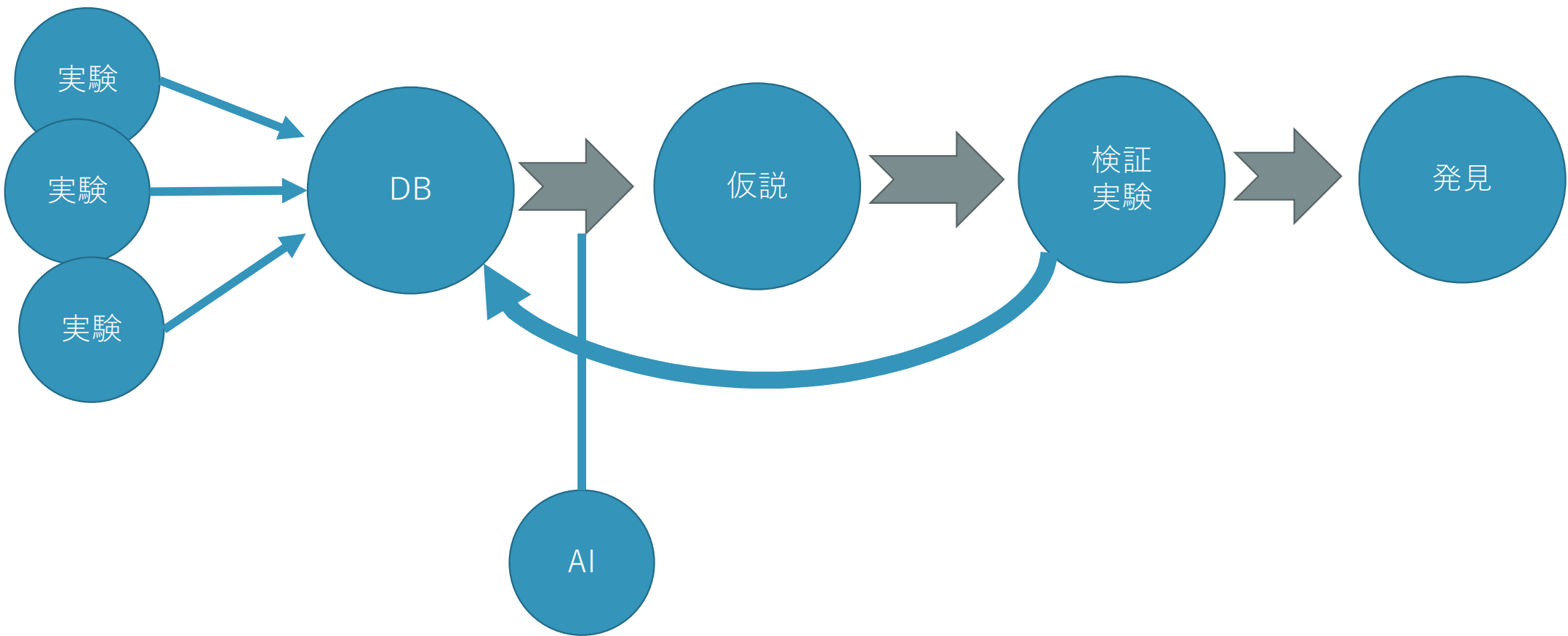
## 目標設定による異分野融合の必要性

# DBは、作者の意図が入った構造化情報である

DBとは、データの蓄積と同時に、データを見つけやすくするもの。DB化には、何らかのデータの構造化が必要。構造化は、作成者の意図が入る。一定のお作法はあるが、未来を洞察できる作成者のDBは、使いやすい。

とりあえず、データを貯めようというデータレイクの考え方も広まったが・・・

- データベースの設計には、知識と経験が要求される
- ドメイン知識が無い人のデータベースは、必要な情報が必要な知識で蓄積されない
- 本当に良いデータベース（+UI）は、自然と知識の相乗効果を誘導できるのでは？



## 発見型のデータベースとは何か？

### データが単に集まっても活用は難しい

- 81万件のビッグデータ
- このデータが活かされるには何が必要？

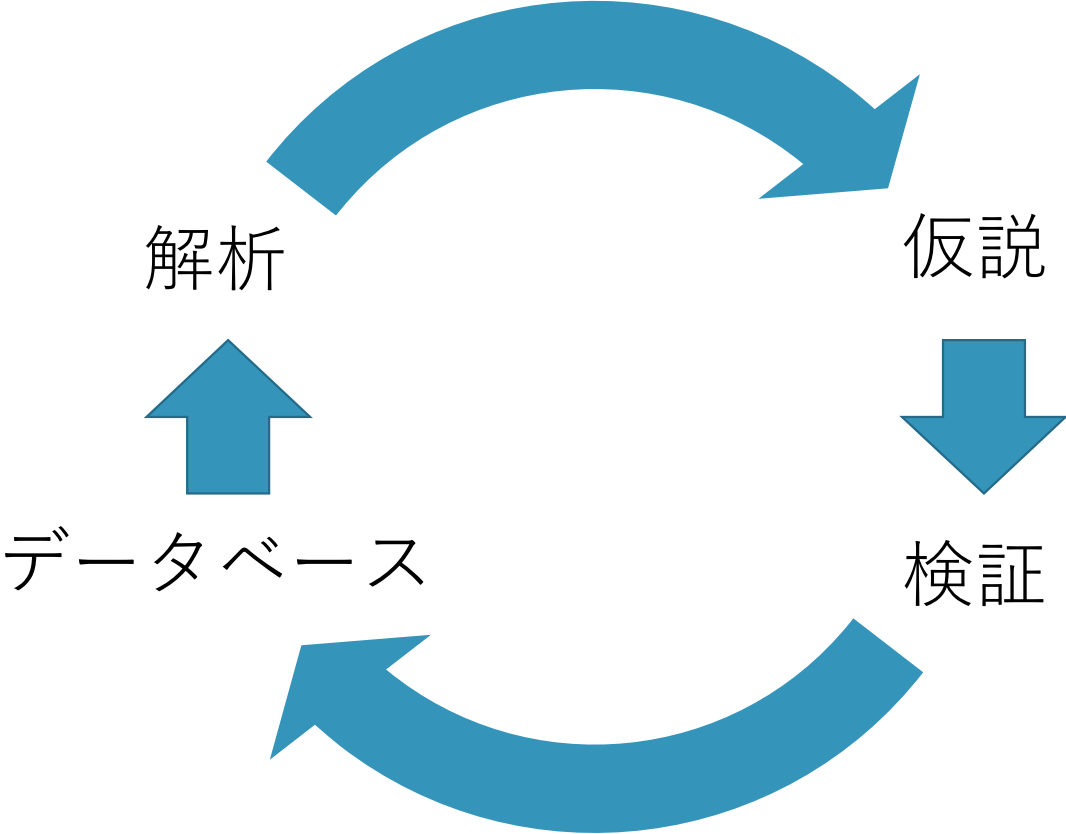
The screenshot shows the Data.gov interface for the 'U.S. Chronic Disease Indicators (CDI)' dataset. The page features the CDC logo, a 'Federal' tag, and navigation links for 'Organizations' and 'Datasets'. The main content area displays the dataset title, metadata (Updated: August 10, 2020), and 'Access & Use Information' which states the dataset is public. Below this, there is a 'Downloads & Resources' section with a 'Download' button and a 'CSV' file icon with '821 views'.

<https://catalog.data.gov/dataset/u-s-chronic-disease-indicators-cdi>

Licensed under a Creative Commons 表示4.0国際ライセンス  
©2020瀬々潤 (株式会社ヒューマノーム研究所)



# DB人材と解析人材は異なる

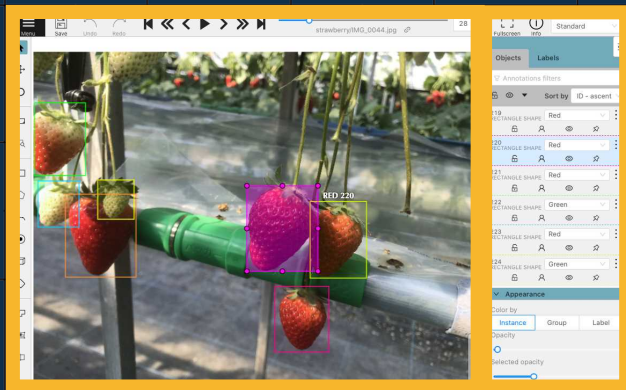
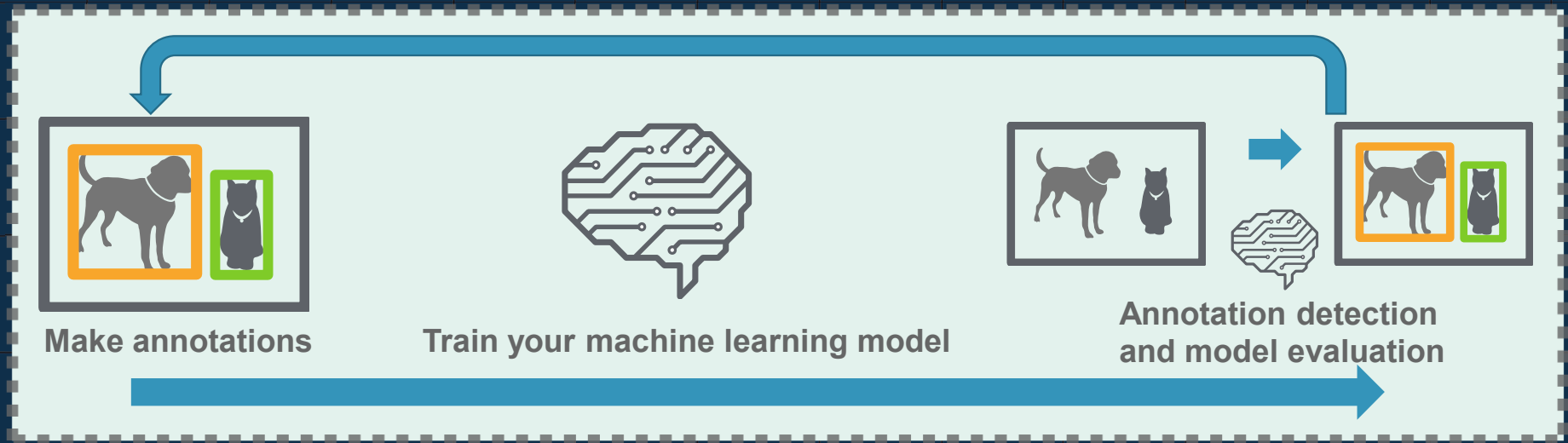




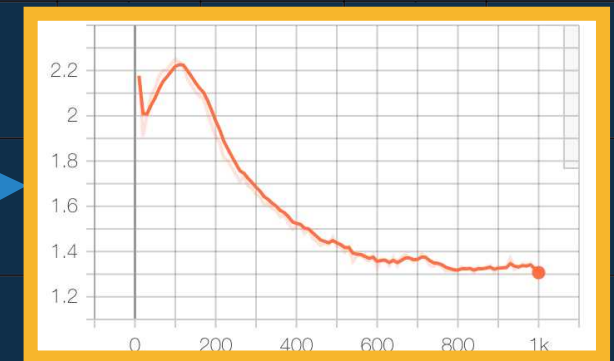
Humanome Lab  
ヒューマノームラボ

# Humanome Eyes: No-CODE AI

eyes@humanome.jp



Models			
PreTrained Models			
Framework	Name		Labels
Pre-Trained	Mask RCNN Object Detector		Supported labels
Your Models			
Status	Name	Trained	Labels
finished	model_19_2020.07.07.01.01.22	July 7th 2020	Supported labels
finished	<b>Your Model</b>	July 6th 2020	Red
finished	model_16_2020.07.06.14.49.50	July 6th 2020	Red Green



Licensed under a Creative Commons 表示4.0国際ライセンス  
©2020瀬々潤 (株式会社ヒューマノーム研究所)



## データベースから仮説を生成する科学の時代

長年言われていることであるが、生命の多様性に取得データの内容が未だ追いついていない。仮説が生成できるデータベースのあり方の検討が必要。

## データベースを作るのは「おまけ」研究ではない

データはナマモノ。貯めて閲覧できるだけでは、料理されずに腐っていく。自然とデータが溜まり、活用される循環システムを作る研究が必要。

## 実験人材、DB人材、解析人材、特徴が異なる

育成パスは色々だと思うが、情報系にも基礎に興味がある人と応用に興味がある人など色々。人材育成には、多様な評価システムも重要。