Healthcare Data Exchange Framework (HDEF)
scalable economy of secure information and services

A pilot project executed within Research Collaboration Agreement between US FDA and IBM Watson Health Group

HDEF pronounces as "hā-def"
What should a good story have:

- **Challenge:** what is the problem we are trying to solve?
- **Setup:** what do we have now?
- **Development:** what actions did we take?
- **Resolution:** how did we solve the problem?
- **Poetic culmination:** what did we learn?
- **Inspiring afterthoughts:** what will be next?
- **Great title:** “...”
Challenges: age of data explosion.

• **We all know FDA mission:** in brief, we are here to ensure safety and efficacy of medical products (before, during and after the product is in the market).

• **We need data:** preclinical, clinical, EMR, nutrition, wellness, pre- and post-market, statistical, genomics, metabolomics, adverse events, patient level, summary ... you name it.

• **Lack of standards:** not harmonized data and processes, ad hoc solutions are found on case by case, insufficient provenance.

• **Infrastructure needed:** lacking network throughput, always not enough compute and storage power, lack of expertise in new computational methodologies.

• **Inefficiencies:** administrative, legal barriers, and some “19th-century-cool” legacy technologies driving document and data flows.

• **Security keywords:** FedEx Truck, printed paper documents OCR, emails, unencrypted hard-drives, anonymization is considered security technology.

• **Ownership and access:** limited and cumbersome access to data through complicated DUA, IRB approvals, legal agreements, etc.

After all of this: our FDA scientists deserve a medal for doing this amazing work.
Challenges: current processes

- **slow**: takes weeks to receive the data due to inefficient document flow process, thin network pipes between EDRs and FDA campus, and arcane delivery methods

- **non-secure**: delivery is performed using insecure network packets, unencrypted hard drives and paper documents

- **week provenance**: tracking of data provenance is by accountability and trust not by design

- **not scalable**: moving large -omics data is challenging, frequently patched with *ad hoc* solutions. It takes more time to get -omics data than to analyze.
Challenges:

- Lack of incentives
- Lack of funding
- Fear of change
- Fear of burden
Setup: age of technology explosion.

• **Modern networking:** the world is going 100GB (100 times faster) with its Internet-2s, fiber optic channels, parallel transfer procedures, using modern https, sftp and other protocols.

• **Modern platforms:** high performance storage, archival and computing on massive, distributed, parallel, cloud computing environments such as **HIVE**, Galaxy, Seven Bridges, Nexus, many others …

• **E-doc:** tens if not hundreds of solutions for secure document workflows, electronic signature systems, processual and **smart contract** systems, etc …

• **Security:** novel paradigms of encryption, **blockchain**, distributed data lakes, private/public key systems, **hyper-ledger**, etc …

• **Bioinformatics:** innovative evolutionary algorithmics, artificial intelligence, stochastic modeling, natural language processing, HIVE algebraic attractors, modeling and simulations, data standardization protocols, **honeycomb typing**, data janitorial services …

A perfect setup: we have the Lego pieces, all we need is the builder.
Setup technology explosion.

- **Modern networking:** the world is going 100GB (100 times faster) with its Internet-2s, fiber optic channels, parallel transfer procedures, using modern https, sftp and other protocols.

- **Modern platforms:** high performance storage, archival and computing on massive, distributed, parallel, cloud computing environments such as HIVE, Galaxy, Seven Bridges, Nexus, many others …

- **E-doc:** tens if not hundreds of solutions for secure document workflows, electronic signature systems, processual and **smart contract** systems, etc …

- **Security:** novel paradigms of encryption, **blockchain**, distributed data lakes, private/public key systems, **hyper-ledger**, etc …

- **Bioinformatics:** innovative evolutionary algorithmics, artificial intelligence, stochastic modeling, natural language processing, HIVE algebraic attractors, modeling and simulations, data standardization protocols, honeycomb typing, data janitorial services …

A perfect setup: we have the Lego pieces, all we need is the builder.
HDEF: what is required

- **security**: to ensure protection of private, proprietary data
- **speed of transfer**: to ensure fast and efficient on demand transfer
- **ease of electronic document flow**: to facilitate administrative and legal processes
- **transaction auditability**: to enable tracing and logging every single transaction
- **horizontal scalability**: to enable big, continuous, heterogeneous data
- **vertical scalability**: to enable expansion of features with new data science services
Big picture

- **individuals** as owners of their wellness data
- **insurance provider**
- **hospital owner of legacy data**
- **clinical trial in hospital setting**

**data aggregator**

**hospital**

**patient data owner**

**pharma data owner**

**databank**

**EMR bank**

**omics bank**

**HDEF blockchain**

**FDA reviewer**

**consumers**

1. **#1**
2. **#2**
3. **#3**
4. **#4**
5. **#5**

**SEM**

**deposition**

**search**

**request**

**contract negotiation**

**permission validation**

**download**

**research labs**

**insurance provider**

**pharma**
Mold arthroplasty Hip Joint
Resurfacing Hip Joint
Osteotomy Hip

patient 1
CT01-134711 21 White female
Mold arthroplasty Hip

patient 2
CT01-112358 32 Asian male
Resurfacing Hip Joint

...
...
...
...
...
...

patient N
CT01-31415 White female
Osteotomy Hip

USUBJID age race gender location
procedure

-------

Demographics domain

-------

Procedure domain

-------

DM PR EX AE CE PE

-------
Precision medicine lets us treat patients based on their genetic makeup: one chromosome and a single gene at a time. For example, a gene for a good response to a cancer drug like Atezolizumab might be present in a patient’s genetic makeup. However, a blue eye gene might indicate that a car-T based cell cancer immunotherapy is not going to work for this patient. This patient might also have a prevalence for Alzheimer potential patient for Mediterranean disease: autoimmune reaction potential after 40. It is advised to avoid metal on metal hip arthroplasty devices. Tens of thousands of features to be extracted about every patient: predictive, prognostic, diagnostic. For instance, Huntington disease patient might be a carrier of rare genetic trait.
Resolution: it works

- **fast**: takes minutes to receive the data through parallel concurrent network pipes using modern protocols of download, DUA contracts, electronic signatures

- **secure**: uses novel data & process delocalization paradigms from HIVE and blockchain technology in order to ensure unbreakable, uninterpretable, heuristic distribution of information

- **provenance**: tracking of data provenance is by design enforced by policy, auditable and traceable

- **scalable**: type and size agnostic peta-scale platform optimized for large data storage, transformation, and analysis.

- **cloud compatible**: may extend to FedRAMP approved cloud and compliment the hardware with end-to-end solution
Resolution:

- technology works – welcome to 21st century
- platform is deployed and is being expanded, current participants: FDA, IBM, GWU, Harvard, Cornell, small businesses
- we can move any type of data and practically any size of data, we can move and host not just data, but processes also
- and we did it based on open source codebase: anyone can participate
- we can do more with this - in-silico, virtual cohort, pragmatic and continuous trials, harmonization, standardization, data janitorial and compounding services, data brokership and derivation services
- roundtable on HDEF is to be held this Oct 26 with a large conference and release to follow up early next year
Poetic culmination: what else we can do?

- Patient and advocacy services
- Content enriched data and patient search
- Low quality data
- Data janitorial services
- Clean data
- Data transformation services
- EMR databank services
- Big data bank services
- Data registries
- Data brokerage services
- Data analytics services
- Consumer
Inspiring afterthoughts?

- liberation of patient data through patient ownership of data
- enabling legacy data reuse and value recovery
- optimization and then transformation of the regulatory review process entirely
- direct patient to registry, patient to doctor, patient to scientist relations
- ...
- the next bullets are limited only by your imagination
- ...
- ...