MESSY DATA AND RELUCTANT USERS - THE TROUBLE WITH HEALTHCARE DATA

Sam Bail @spbail

Infiniteconf 2019

HI!

I'm one of the many Sams you've met today

PhD in semantic web, knowledge representation and automated reasoning

Spent 5 ½ years at Flatiron Health in NYC analyzing oncology data

Less big data, more artisinal handcrafted data

Germany > UK > US = Lots of different healthcare systems

sambail, one word.
@spbail

Healthcare data scientists in academia: "We built a cool model with this data!"

Healthcare data scientists in industry: "Well the data is a homogenous mess and HIPAA makes things hard and you really need domain experts and your users probably won't trust you..."

#MLConfNYC

10:20 PM - 29 Mar 2019

2 Retweets **12** Likes

2 12

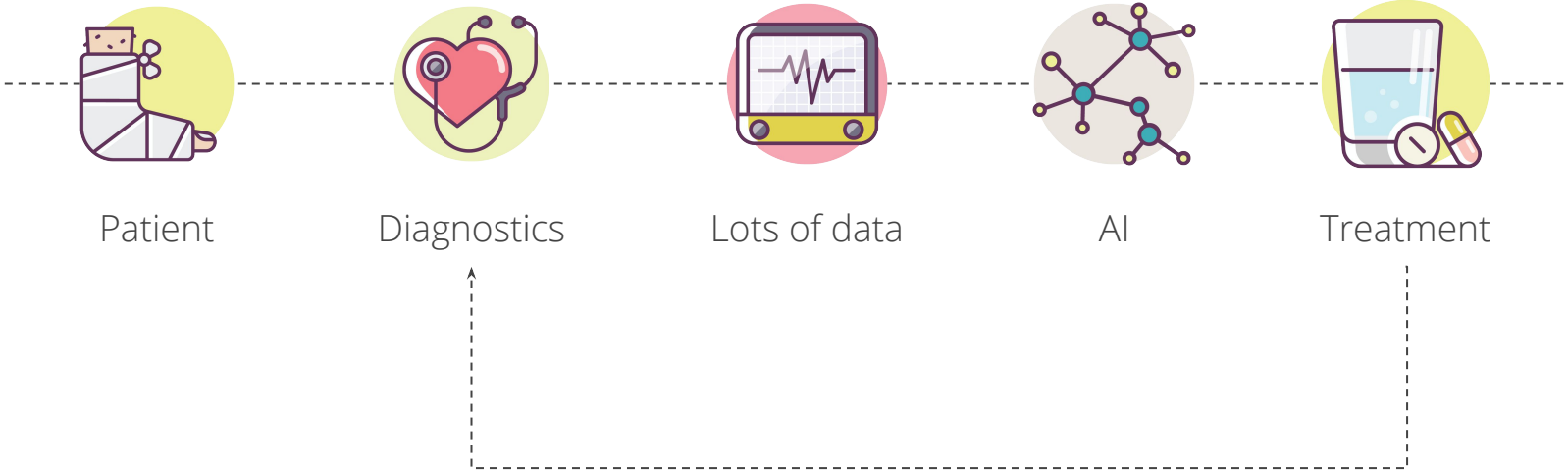
OUTLINE

1. The vision
2. The problem: Messy data
3. The other problem: Reluctant users
4. Small steps

1 - THE VISION

I, for one, welcome our robot overlords.

THE AI DOCTOR



HIGH HOPES

IBM Watson for Oncology is a prominent example of healthcare + AI in recent years

Starting in 2011, over fifty organizations announced Watson collaborations

By 2017, only five projects out of a sample of 24 had been launched

Reference [3,4]

HIGH HOPES

IBM Watson for Oncology is a prominent example of healthcare + AI in recent years

Starting in 2011, over fifty organizations announced Watson collaborations

By 2017, only five projects out of a sample of 24 had been launched

---

Babylon Health is a patient-facing app that provides an AI chatbot for triaging symptoms

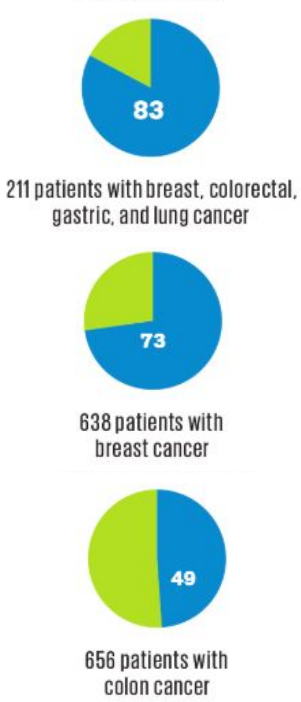
Babylon has two contracts with the NHS

In 2018, physicians voiced concerns about the accuracy of 10-15% of the bot's diagnoses

Reference [3,4,5]

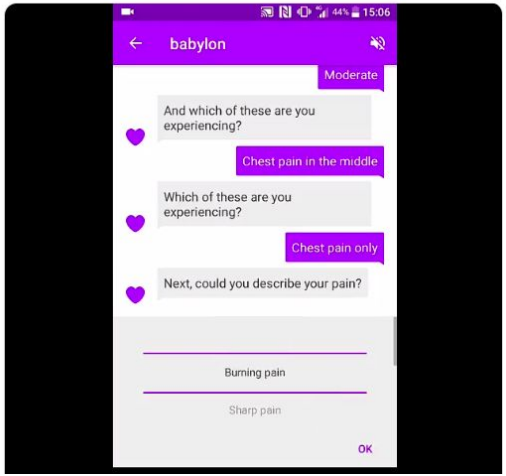
TECHNICAL CHALLENGES

> [The Watson-based Oncology Expert Advisor system] had accuracy scores ranging from **90 to 96 percent** when dealing with clear concepts like diagnosis, but scores of **only 63 to 65 percent** for time-dependent information like therapy timelines.



48yr old obese 30/day male smoker develops sudden onset central chest pain & sweating....

I say call 999, the Babylon App says see your GP...



Reference [3,4,10]

USER ACCEPTANCE CHALLENGES

Physician **agrees** with recommendation

It doesn't add value!

Physician **disagrees** with recommendation

It's wrong!

2 - THE PROBLEM: MESSY DATA

Healthcare data is hard! Let’s go shopping.

“HEALTHCARE DATA”

WORKING DEFINITION:

Any kind of “real-world” data that is generated as part of a patient’s and clinician’s interaction with data capturing software and medical devices, e.g. medical records, scans, lab and pathology reports, billing records, chat interactions, device data, etc.

DATA != DATA*

"CLASSIC" DATA SCIENCE		HEALTHCARE DATA
Structured logs, user data or unstructured data, e.g. social media, publications	--------	Lots of user-entered "structured" data and unstructured documents
Usually complete data	--------	Big gaps in data
"Self-generated" internal data	--------	Data from external sources
Ambiguity and acronyms	--------	Ambiguity and acronyms, lack of context, needs subject matter expertise
Fewer privacy concerns, data can be stored in "the cloud"	--------	Privacy is critical, accessing and storing data is *hard*

** somewhat based on my own view of the world*

JUST *HOW* MESSY?

"Structured" and unstructured data

Gaps in data

Data silos

Ambiguity in medical text

Privacy restrictions

"Structured": discrete database fields, might still allow free-text
Unstructured: Scanned letters, lab reports, faxes, physician notes

SAMPLE VISIT NOTE

Initial - CCC

Note Date: 11/08/16
Signed by (ORTHOPEDIC SURGEON), MD, PHD on 11/11/16 at 3:32 pm Affiliation: HOSPITAL

Active Medication list as of 11/08/16:
Medications - Prescription
FLUROSEMIDE – 20 mg daily
TYLENOL – OTC as needed

This is a first office visit to my clinic by Mr. XXXX, a very pleasant 57-year-old male patient, who sustained in 1993, as the result of a ski accident, a pelvic fracture with vertical shear that has healed in about an inch vertical shortening. Nevertheless, Mr. XXXX has had a remarkably active life. He exercises and has been managing very well over the last few years until recently when he has developed some groin type pain, very reminiscent of arthritic symptoms. Films obtained today confirmed that finding with some bone-on-bone contact and significant posttraumatic hip osteoarthritis.

He actually has a remarkably good gait. He has overall good strength. He has pain along the groin. He has a little bit of anterior medial pain that may be muscular in nature and even though he has a leg length discrepancy, he walks a very normal gait on exam. His extremity appears to be sensory intact and well perfused. He reports the typical symptoms of pain on initiation of motion, winter pain and pain at end of the day.

•••

JUST *HOW* MESSY?

"Structured" and unstructured data

Gaps in data

- Patients see multiple physicians
- EHR migrations
- Workflow changes

Data silos

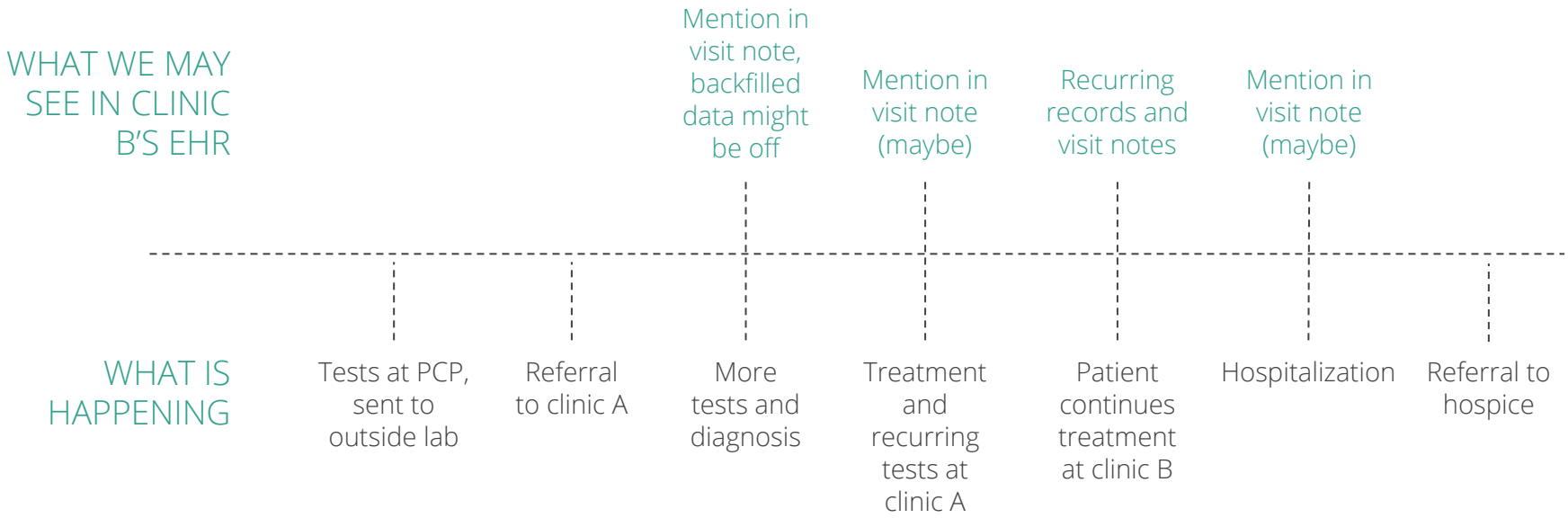
Ambiguity in medical text

Privacy restrictions

THE PATIENT JOURNEY*

** Heavily simplified and based on what I've seen in oncology - I'm not a doctor!*

THE PATIENT JOURNEY*



** Heavily simplified and based on what I've seen in oncology - I'm not a doctor!*

JUST *HOW* MESSY?

"Structured" and unstructured data

Gaps in data

Data silos

Ambiguity in medical text

Privacy restrictions

Data is hard to access
"No" data model or coding standards
Scaling beyond a single institution is hard

JUST *HOW* MESSY?

"Structured" and unstructured data

Gaps in data

Data silos

Ambiguity in medical text

Privacy restrictions

Heavy use of acronyms and abbreviations
Longitudinal data and sequencing is hard

Reference [2]

JUST *HOW* MESSY?

"Structured" and unstructured data

Gaps in data

Data silos

Ambiguity in medical text

Privacy restrictions

We can't just store data "in the cloud"
Linking data sets and mapping entities is limited
Sharing (and validating) data is hard

HOW DID WE GET THERE?

US HITECH ACT 2009: Encourage EHR adoption, but not interoperability

Data was an afterthought - meant for humans to look at ("Glorified paper")

UX was an afterthought - data entry is painful and encourages dictation

No incentive to document anything in structured form if it's not needed for billing

Reference [7,8]

THE TL;DR

Getting **clean** and **reliable** healthcare data as input for any kind of ML is hard.

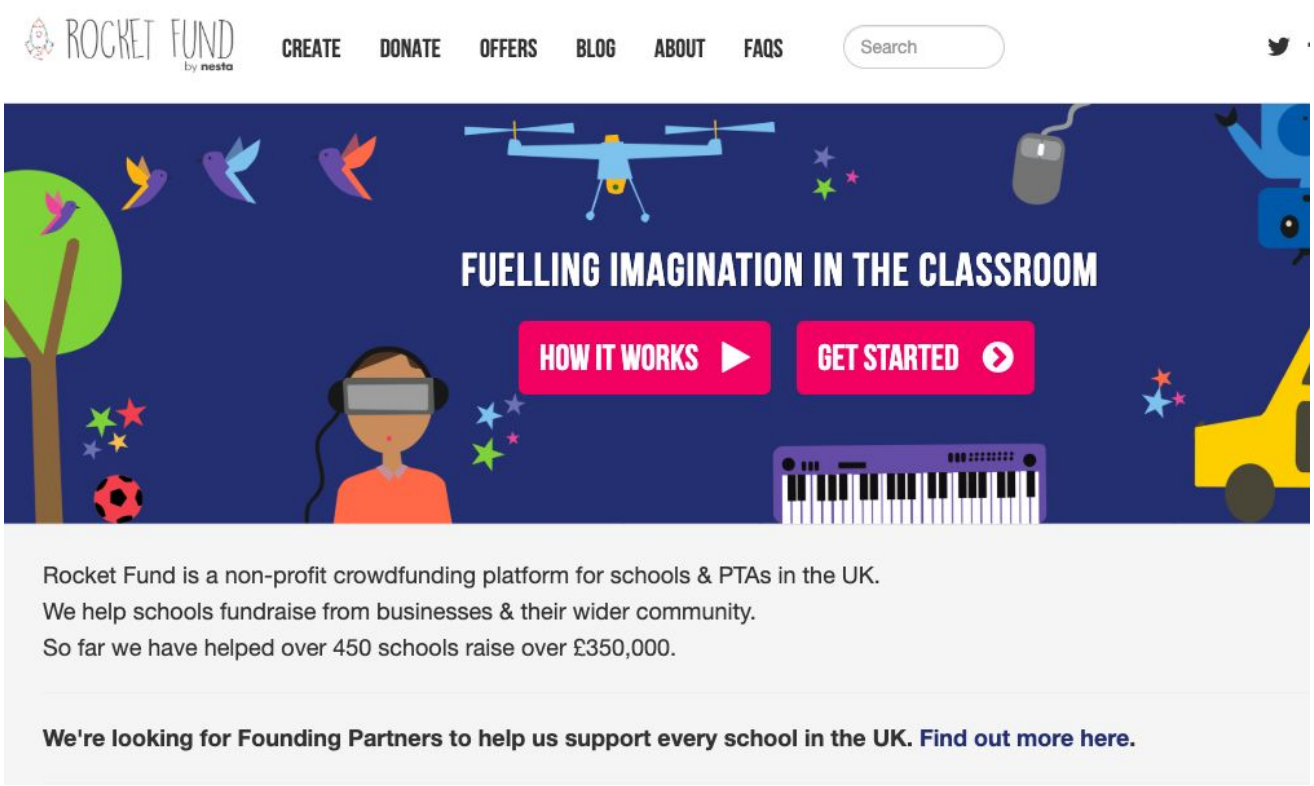
Scaling data access and standardization across the boundaries of a single institution is hard.

Reference [7,8]

3 - THE OTHER PROBLEM: RELUCTANT USERS

Or, “The Pencil Problem”

THE PENCIL PROBLEM



Reference [1]

THE PENCIL PROBLEM

Watercolour pencils for Erith School

By Ms Williams

Year 10 and 11 Students want to use watercolour pencils for their course work - can you help?

Art & Design

SUCCESSFUL
WE RAISED £320

Ysgol Llangelynnin rubber mulch

By Gareth Evans

For a playground equipment where the children can play safely

Other

£235 RAISED	9% FUNDED	PROJECT FINISHED

Hayes Park School needs flower & vegetable beds!

By Louise Prince

Help HPS make the environment more attractive by growing flowers & vegetables around the...

Other

£88 RAISED	56% FUNDED	5 DAYS TO GO

New Minibus

By Carl Fitter

We need to replace our ageing minibus.

Other

SUCCESSFUL
WE RAISED £2,131

Reference [1]

WHAT I'M TRYING TO SAY IS...

We haven't even solved the basics yet in healthcare.

GARBAGE IN, FRUSTRATION OUT

> “Three people will list the same diagnosis three different ways. [...]
>
> The problem lists have become a **hoarder’s stash**. [...]
>
> They’re long, they’re **deficient**, they’re **redundant**.”

Susan Sadoughi, “Why Doctors Hate Their Computers”

Reference [7]

“DOCTORS HATE THEIR COMPUTERS”

Slow data entry

Alert fatigue

Insights and then what?

Lack of transparency

“Most days, I will have done only around **thirty to sixty per cent** of my notes by the end of the day“

Susan Sadoughi, “Why Doctors Hate Their Computers”

Reference [7]

“DOCTORS HATE THEIR COMPUTERS”

Slow data entry

Alert fatigue

Insights and then what?

Lack of transparency

“Of roughly 350,000 medication orders per month, pharmacists were receiving pop-up alerts on **nearly half** of them“

Robert Wachter, “The Digital Doctor”

Reference [8]

“DOCTORS HATE THEIR COMPUTERS”

Slow data entry

Alert fatigue

Insights and then what?

Lack of transparency

“If we use AI to detect more spinal fractures, **we've now shifted the problem** to having to treat more patients“

Kerry Weinberg (Amgen), MLConf NYC 2019

“DOCTORS HATE THEIR COMPUTERS”

Slow data entry

Alert fatigue

Insights and then what?

Lack of transparency

“I would certainly want to see **some validation** to whether the synthetic data is representative of anything that would make sense”

Dr. Jonathan Chen, “Why Doctors Hate Their Computers”

Reference [8]

PRIORITIES

In 2018, we prototyped a simple **decision support** system for oncology clinics...

The main reaction from the **oncologists** who tested it was...

PRIORITIES

In 2018, we prototyped a simple **decision support** system for oncology clinics...

The main reaction from the **oncologists** who tested it was...

*"Can you just make the EHR **faster** please?" **

* *Paraphrased for dramatic effect*

THE TL;DR

It will take a lot of effort to convince clinicians that computers are helpful, not just painful.

4 - SMALL STEPS

Don’t give up just yet.

POTENTIAL WINS FOR AI + HEALTHCARE DATA*

CLINICIANS

Image processing
Annotating and diagnosing scans, e.g. Microsoft InnerEye

Practice workflows
Claim denial prediction, clinical trial matching

Value-based care
Predict and reduce hospitalizations

PATIENTS

Triaging (“digital nurse”)
Prevent hospital visits, e.g. Babylon, Sensely

Mental health
Easily accessible help, e.g. Woebot, Youper, (Talkspace)…

** Focused on applications that target clinicians and patients rather than researchers and biased by my own perspective*

BUT... THE PENCIL PROBLEM!

Lack of **interoperability** is not just a data problem - it's also a challenge for patients and clinicians

Administrative aspects of healthcare (scheduling, communication, billing, transport, wait times, etc) are still overwhelming for a lot of patients

Most patient portals have **terrible UX**, too - and the users may be older and less tech savvy

THERE ARE STILL PLENTY OF "SIMPLE" PROBLEMS TO SOLVE.

THANK YOU

Sam Bail @spbail
Recording will be available on the conference website

REFERENCES

- [1] Shiny moonshot technology will not save healthcare — yet
- [2] What Is the Role of Natural Language Processing in Healthcare?
- [3] How IBM Watson Overpromised and Underdelivered on AI Health Care
- [4] IBM's Watson supercomputer recommended 'unsafe and incorrect' cancer treatments, internal documents show
- [5] This Health Startup Won Big Government Deals—But Inside, Doctors Flagged Problems
- [6] Augmenting Mental Health Care in the Digital Age
- [7] Why Doctors Hate Their Computers
- [8] The Digital Doctor (excerpt here)
- [9] An Ingenious Approach To Designing AI That Doctors Trust
- [10] Dr Murphy on Twitter
- [11] Care.data and access to UK health records: patient privacy and public trust
- Thanks to Lucy Bridges (@linuxlucy) for a detailed overview of data flow in the NHS.