



Data Revolutions, AI and Health

Franz Och

och@humanlongevity.com

PRISME Forum, May 2015



Extend the Healthy Human Life Span





(Slightly over-simplistic)

Two kinds of fields:

- Data revolution happened in the past
- Data revolution will happen soon



The Data Revolution in Machine Translation (and related fields)

and how it relates to disrupting human health





Machine Translation 10 years ago

- Low usage
- Few languages
- Very poor quality
- Rule-based machine translation
 - Linguist experts writing rules
- Challenges:
 - Complexity/nuances/variation of natural language
 - High cost of building systems



- 500+ million users
- 90+ languages

HUMAN LONGEVITY,

NC.



- Much better quality
- Speech/Image Translation





Statistical Machine Translation How does it work?





What sparked the data revolution in MT?





In Health?





Data revolutions: What happens?

What happens:

- Data is central
- Everything becomes statistics data science
- Simplicity
- Feedback loops
- Compute, compute, compute
- Capturing the long tail
- Science fiction capabilities



More Data – Better Quality





Sequencing Cost and Time

Celera

Human Longevity, Inc.



	Per genome:		
\$100M	67,000x cheaper	\$1 500	
Q I I	90x faster	~ ,500	
9 months		\rightarrow 3 days	



Indenter I street Mentaleter

Quantitative Phenotyping

NeuroQuant	®		CorTechs Labs, In .4690 Executive Dr, #25 San Diego, CA, 92
Age-Related Atrop	hy Report		Tel:(858)459-970
PATIENT INFORMATION			
Patient ID: 005_S_0221	Patient Name: Smith, John Jr.		ex: M
Accession Number: 093951-1	Referring Physician: Jones, Steven MD,PhD		xam Date: 2014/02/22 10:18:17 AM
MORPHOMETRY RESULTS			
Brain Structure	Volume (cm²)	% of ICV (5%- 95% Normative Percentile*)	Normative Percentile*
Hippocampi	5.83	0.35 (0.45-0.60)	< 5
Lateral Ventricles	56.14	3.32 (0.84-3.46)	92

. ...

0 00 10 10 0 071

OF



From Data to Knowledge



- 1. More Data
- 2. Quality Data
- 3. Integrated Data



Integrated Data – Break up Silos

Clinical Information/Samples

Microbiome



Evolutionary Rapid Prototyping

HUMAN

NC.

LONGEVITY.





REVIEW



The answer is 17 years, what is the question: understanding time lags in translational research

Zoë Slote Morris¹ • Steven Wooding² • Jonathan Grant²

¹Institute of Public Health, University of Cambridge, Cambridge CB2 0SR, UK ²RAND Europe, Cambridge CB4 1YG, UK Correspondence to: Jonathan Grant. Email: jgrant@rand.org



Everything becomes data science



"Software is eating the world"

"Machine Learning is eating the world"



System requirements:

- Terabytes of models
- Latencies in ms range
- Massive number of requests

System architecture:

- Distributed models and representations
- Cloud-computing



Old MT:

Only cost-effective for a few big languages

New MT:

- 90+ languages (only bottleneck is data)
- Generic training to support language
- Supporting a language is (almost) free





Time for one-person trials

Precision medicine requires a different type of clinical trial that focuses on individual, not average, responses to therapy, says **Nicholas J. Schork**.

Nature, April 2014



Feedback loops



Example: speech recognition

- Using data improves system
- More usage more data
- Realistic data



What's next?

Dramatic progress, eg:

- Speech recognition
- Picture understanding
 - ImageNet
 - Zero-Shot Learning

Multi-task learning

Describes without errors



A person riding a motorcycle on a dirt road.

Describes with minor errors



Two dogs play in the grass.



A group of young people playing a game of frisbee.



Two hockey players are fighting over the puck.

Combining many different knowledge sources

Data + Feedback = AI





There will be a data revolution

Everything will change

Huge opportunities for human health







- **1.** More Data
- 2. Quality Data
- **3. Integrated Data**

Thank You!



