## Scalable Information Systems for Ag in Africa and Beyond

November 17, 2016 William Wu QED | <u>http://qed.ai</u>



## OUTLINE

- Introduction
- Problems: Data Scarcity
- Technology Solutions



## INTRODUCTION



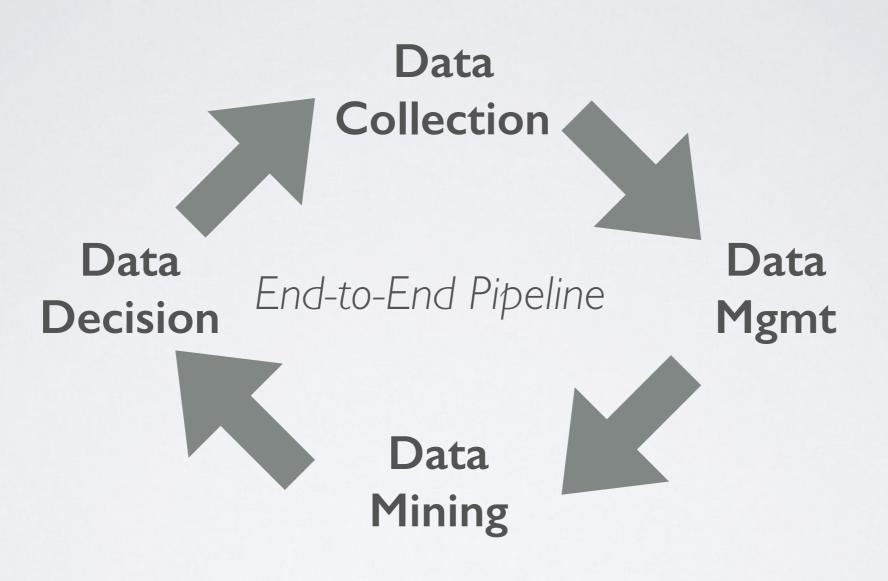
### William Wu

Ph.D. EE, M.S. Math (Stanford) B.S. EECS (Berkeley) Ex-NASA JPL



### Jiehua Chen ("JC")

Ph.D. Statistics, M.S. Econ (Stanford) B.Sc Math and Physics (Tsinghua) Ex-Columbia Univ., Earth Institute





"Deploy or die!" — MIT "quod erat demonstrandum" — Euclid

### **Global Development**

"Help the world's poorest people lift themselves out of poverty."

## Ag Dev

- 3/4 of world's poorest subsist on ~0.5 ha of farmland.
- unproductive soils, plant disease and pests, drought, ill livestock, credit, market access, pricing, land reform
- malnutrition, child mortality

### BILL& MELINDA GATES foundation

GLOBAL DEVELOPMENT

Agricultural Development

**Emergency Response** 

Family Planning

Financial Services for the Poor

**Global Libraries** 

Integrated Delivery

Maternal, Newborn & Child Health

Nutrition

Polio

Vaccine Delivery

Water, Sanitation & Hygiene

BILL& MELINDA GATES foundation







soil + landscape info is incomplete and outdated

agronomy projects for maize, cassava, cereals, etc. inputs and advice to support ~400K SHFs in SSA



tech consultant



## PROBLEMS: DATA SCARCITY



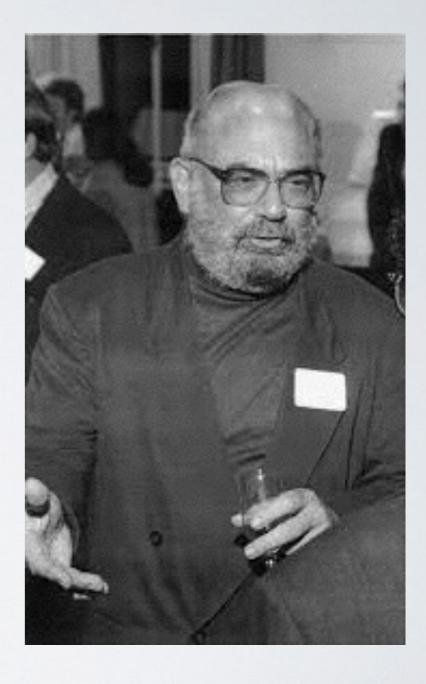
## "WELCOMETOTHE MACHINE"





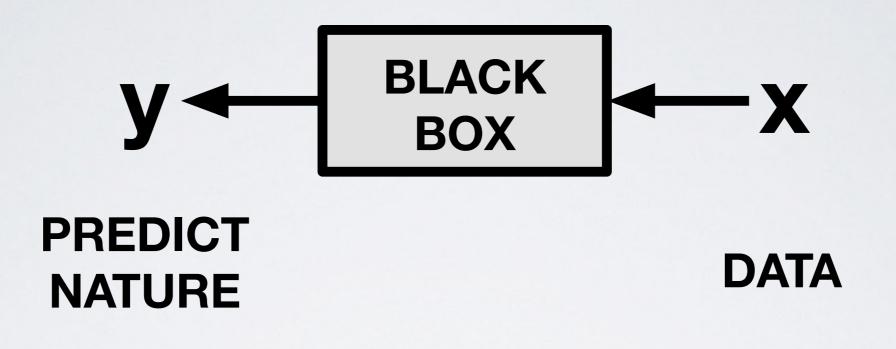
# LEO BREIMAN (1928-2005)

- Left academia to be full-time consultant in industry (13 yrs).
- "Statistical Modeling: The Two Cultures" (2001)





"Goal: Use data to predict and to get information about the underlying data mechanism."





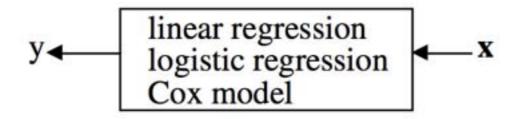


### The Data Modeling Culture

The analysis in this culture starts with assuming a stochastic data model for the inside of the black box. For example, a common data model is that data are generated by independent draws from

response variables = f(predictor variables, random noise, parameters)

The values of the parameters are estimated from the data and the model then used for information and/or prediction. Thus the black box is filled in like this:



 $f(X) = \beta_0 + \sum_{j=1}^{p} X_j \beta_j$ 

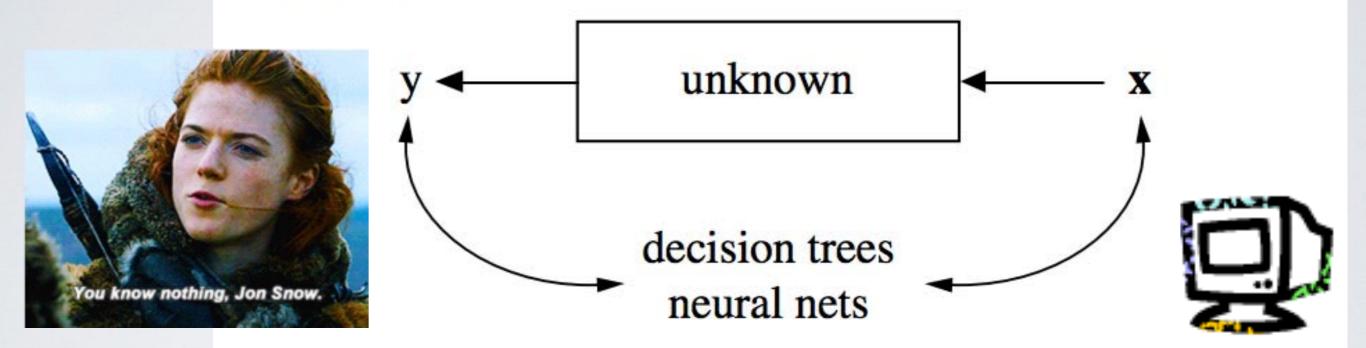
$$\log \frac{p(x)}{1 - p(x)} = \beta_0 + x \cdot \beta$$

Model validation. Yes-no using goodness-of-fit tests and residual examination. *Estimated culture population.* 98% of all statisticians.

qec IIII

### The Algorithmic Modeling Culture

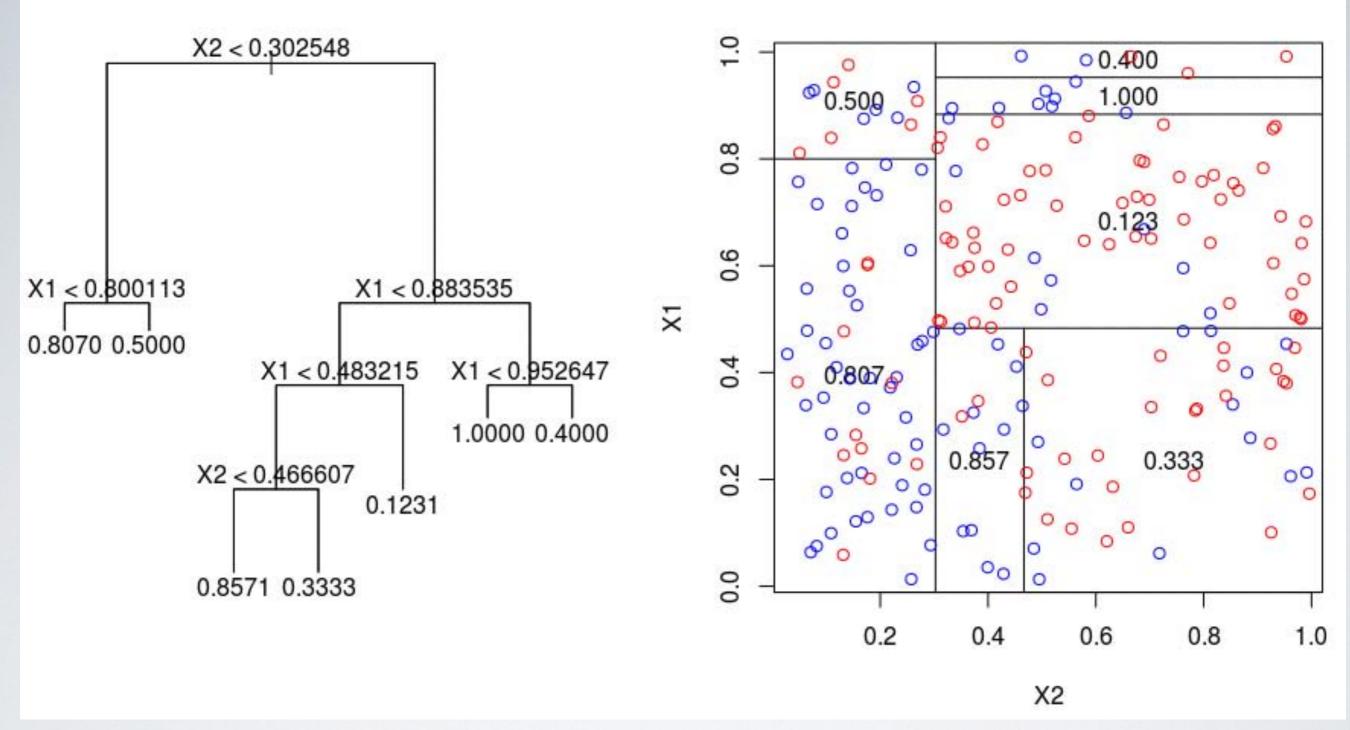
The analysis in this culture considers the inside of the box complex and unknown. Their approach is to find a function  $f(\mathbf{x})$ —an algorithm that operates on  $\mathbf{x}$  to predict the responses  $\mathbf{y}$ . Their black box looks like this:



*Model validation*. Measured by predictive accuracy. *Estimated culture population*. 2% of statisticians, many in other fields.



\_ec

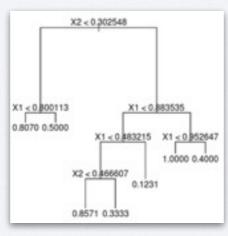


- Decision Tree: repeatedly partition data into nested classes
- Random Forest: make many decision trees from random subsets of the data, and average them

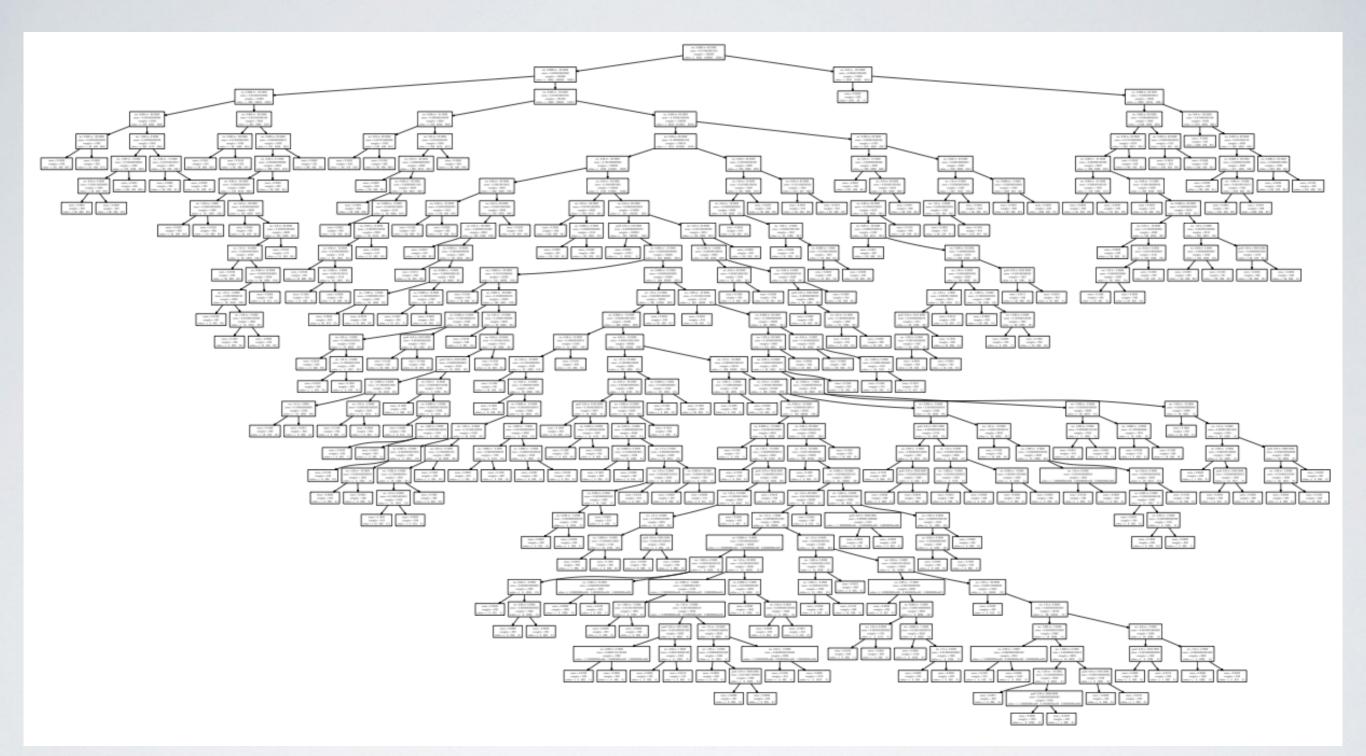
qed

I:III:



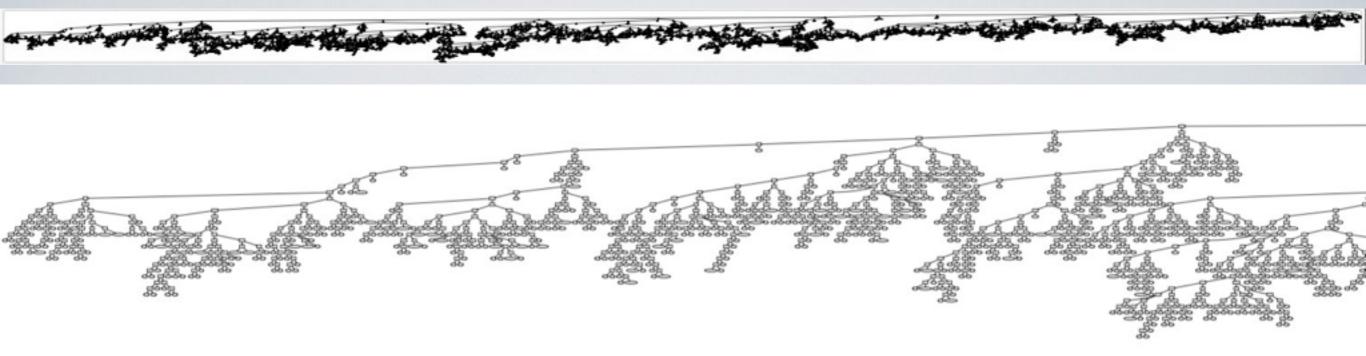


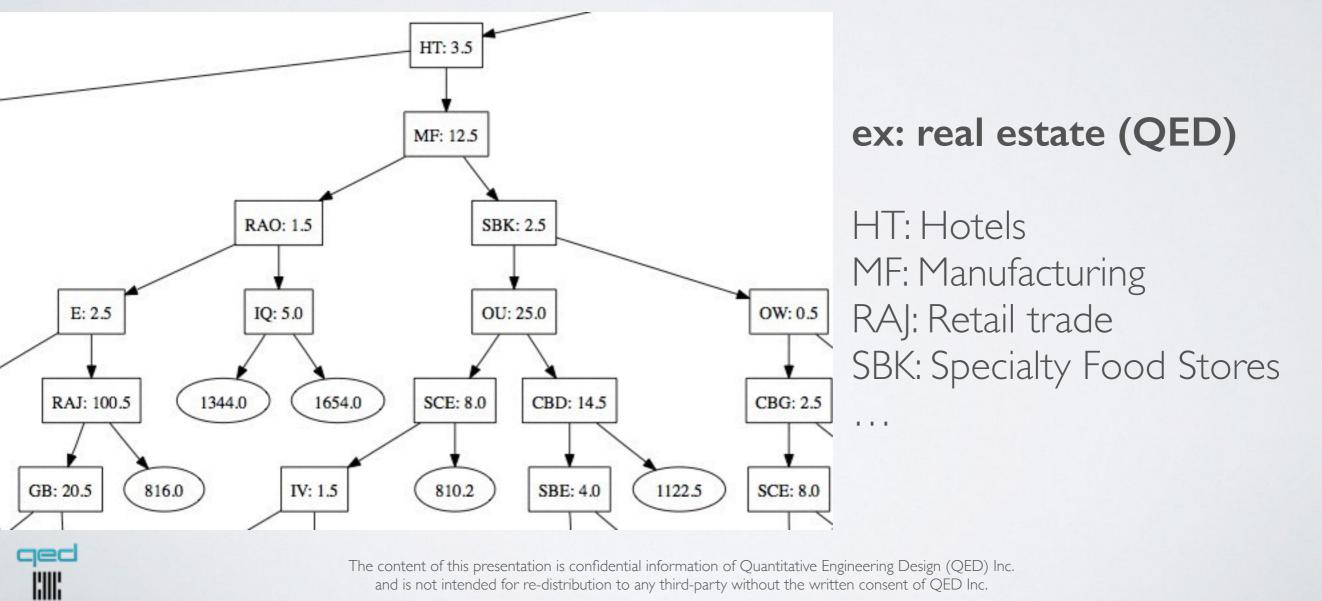




ex: predictive image compression (W.Wu)



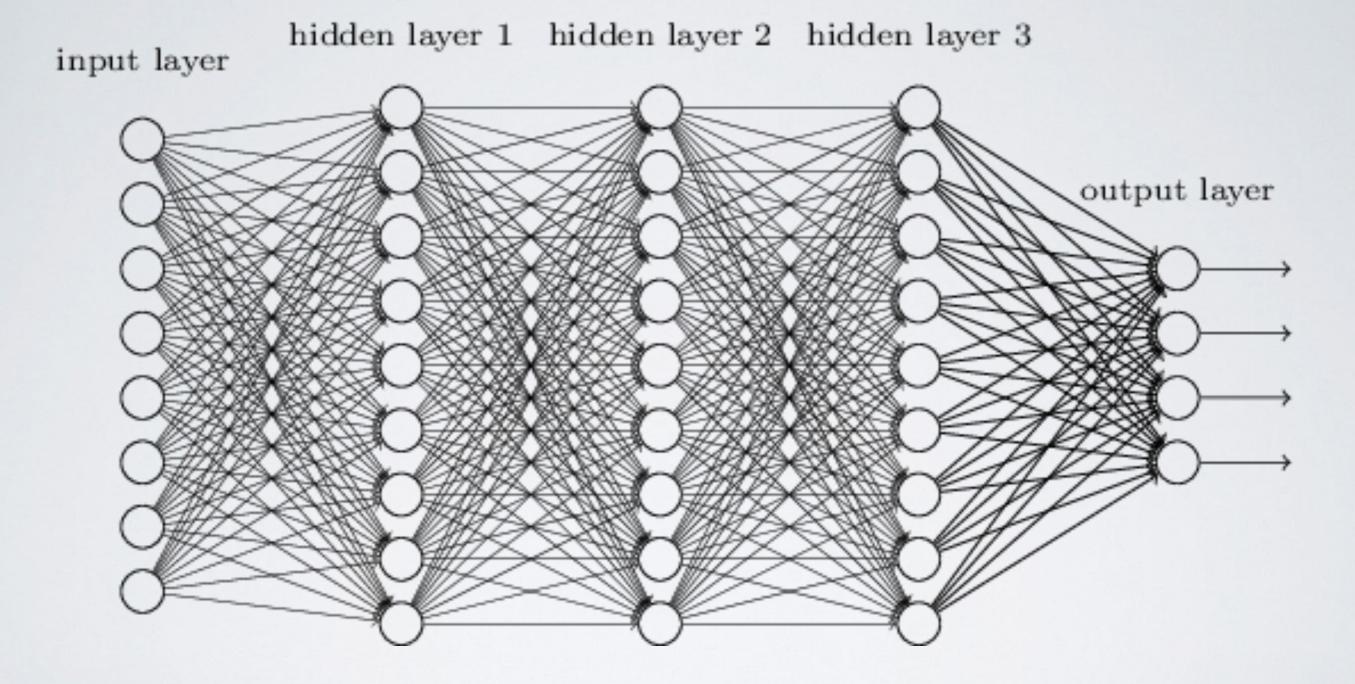




and is not intended for re-distribution to any third-party without the written consent of QED Inc.

### neural networks (1953)

qed



- 2016: Machines beat traditional data models and humans:
  - handwriting transcription
  - movie recommendation
  - spam classification
  - online advertising

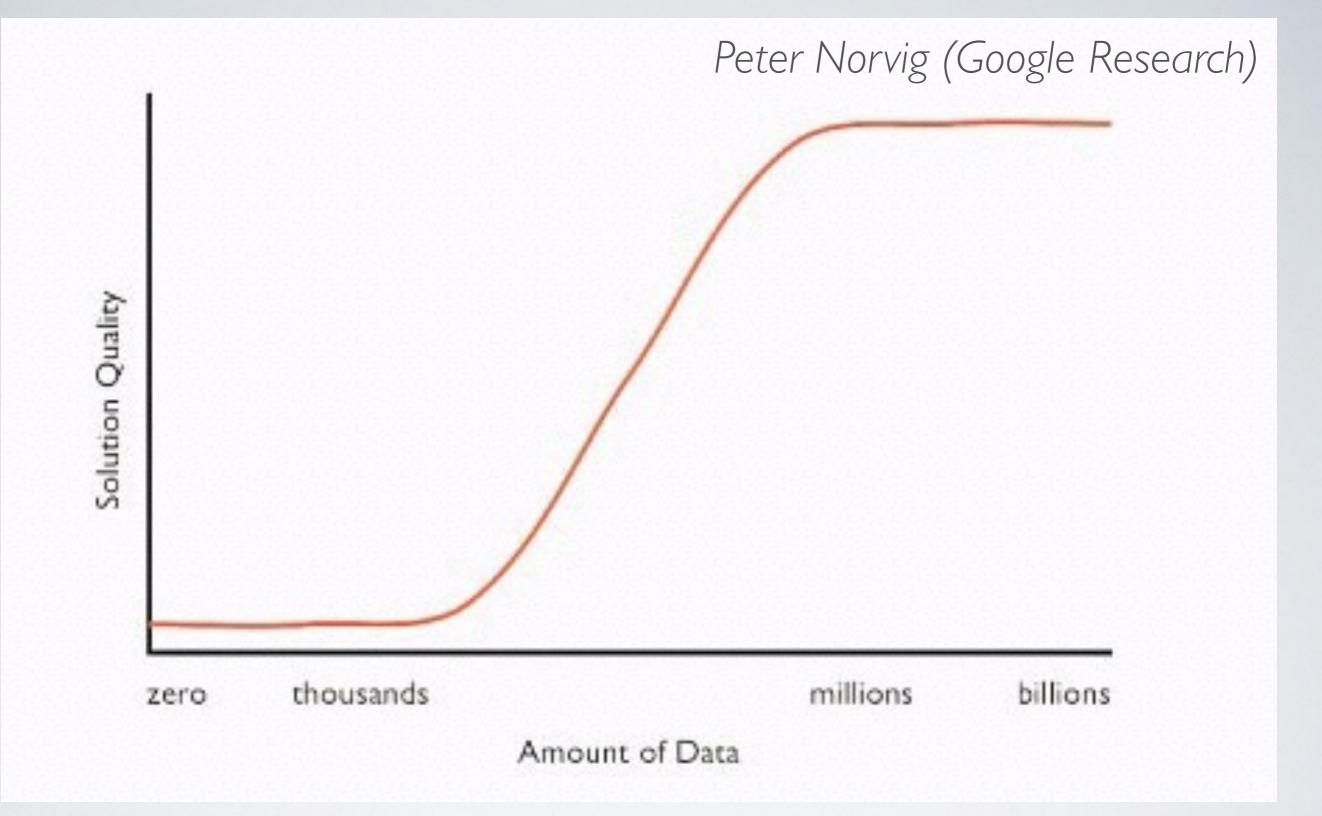
- insurance policies
- breast cancer analysis
- face detection
- atari, chess, GO



• What's the catch?

qed

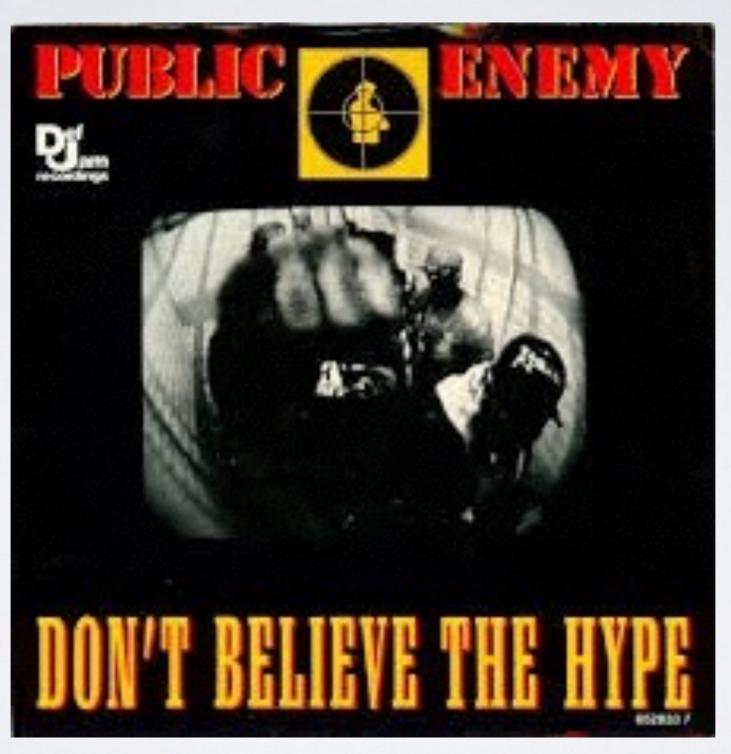
!:III:



• Requires ENORMOUS data (TB/PB).



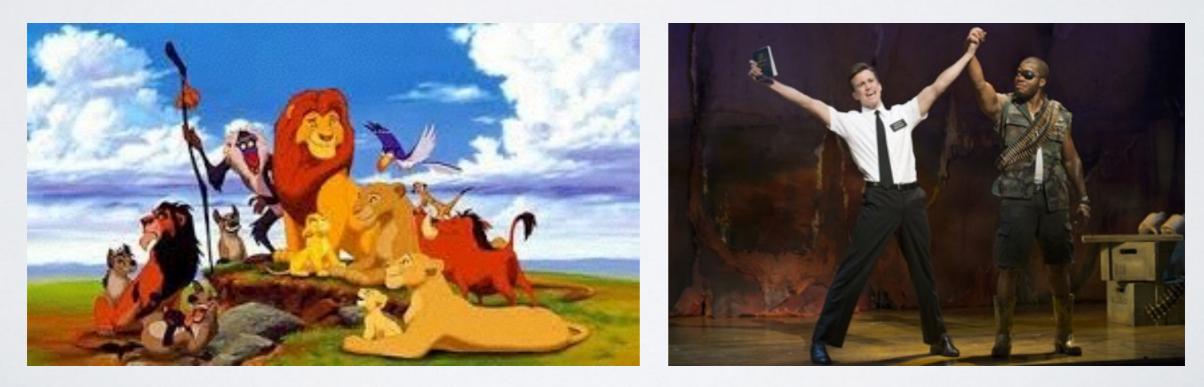
## "DON'T BELIEVETHE HYPE"



qed IIII

## AG FANTASY

- food security ("9B by 2050")
- small-holder farmers
- "big data"
- machine learning and maps





## AG REALITY

• Hans Jenny:

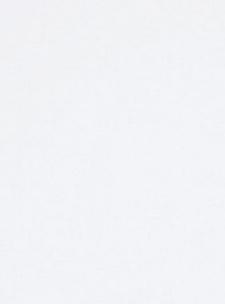
s = f(cl, o, r, p, t, ...)

- yield =  $G \times E \times M \times F$
- SHF input = NA
- "databases" = .doc, .xls
- black box = Excel





# SOLUTIONS:

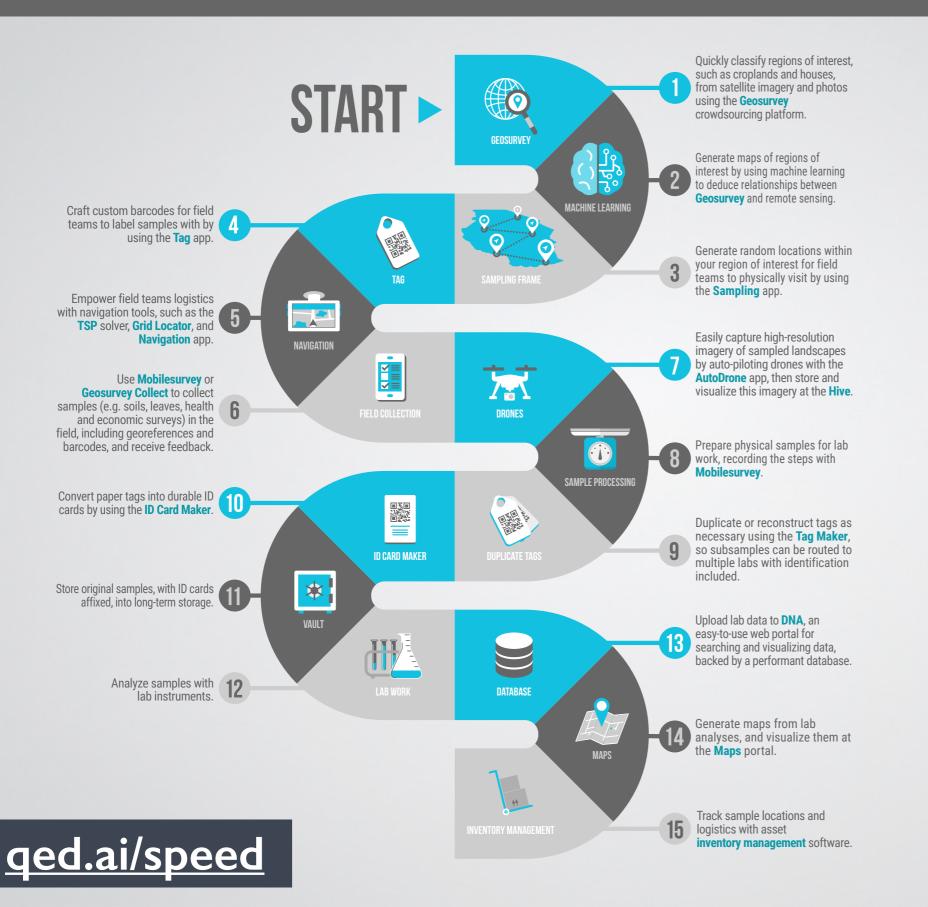




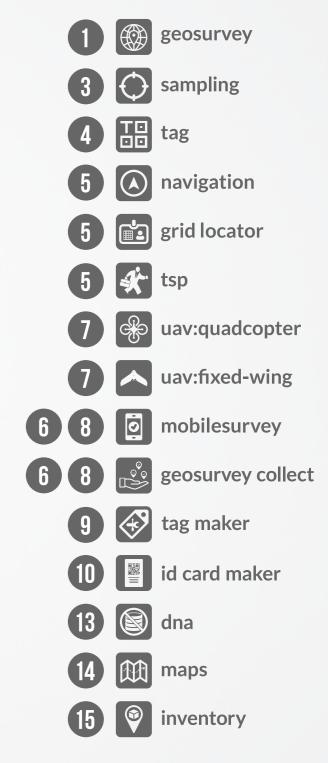


### SCALABLE PIPELINE FOR ENGINEERING OF ENORMOUS DATA

End-to-end technology solutions to support systematic geospatial data collection and analysis.



**QED APP LEGEND** 



# SUB-PROBLEMS

- determine ROI
- sampling frame
- navigate to sampling sites
- georeferenced extraction
- lab analyses
- database
- mapping (spatial-temporal modeling)
- decisions

qed

!:III:

## GEOSURVEY



- problem:
  - map croplands and HS for soil sampling
  - Govt. Statistics Division story
- solution: crowdsourced geospatial analysis



### https://geosurvey.qed.ai/admin/survey/resultchange/839878/





### [geosurvey.qed.ai]



### https://geosurvey.qed.ai/admin/survey/resultchange/59472/





[geosurvey.qed.ai]

The content of this presentation is confidential information of Quantitative Engineering Design (QED) Inc. and is not intended for re-distribution to any third-party without the written consent of QED Inc.

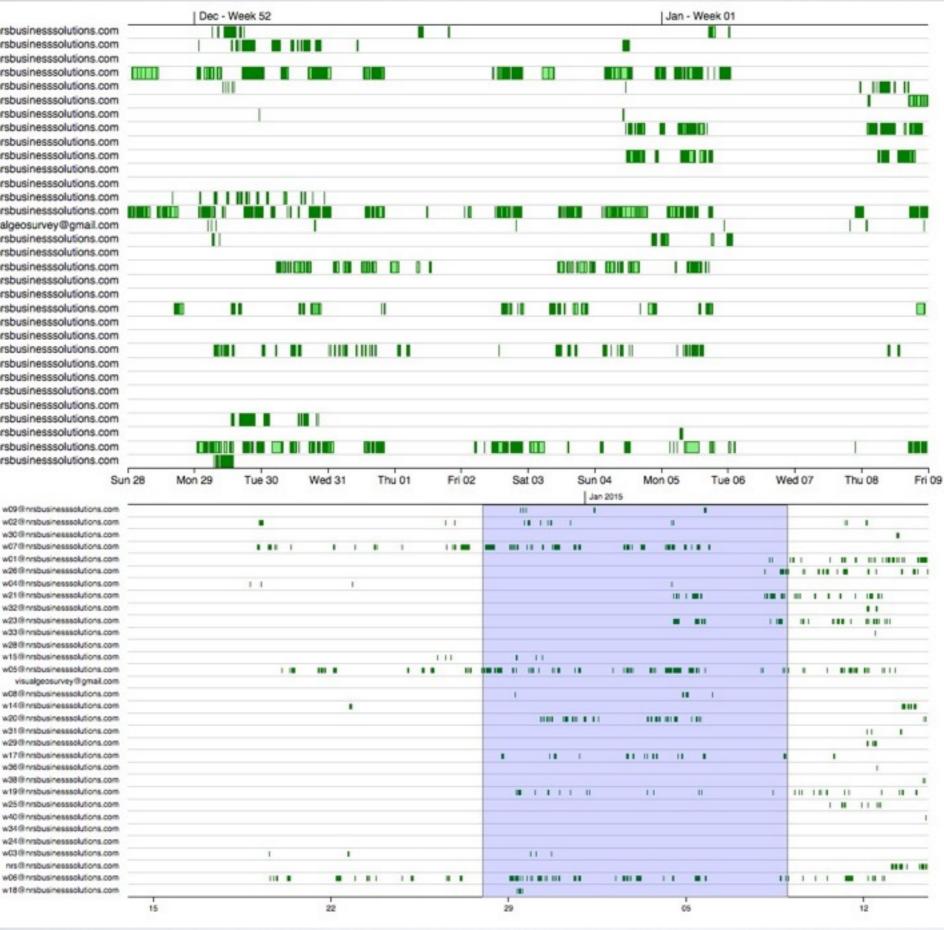
qec IIII

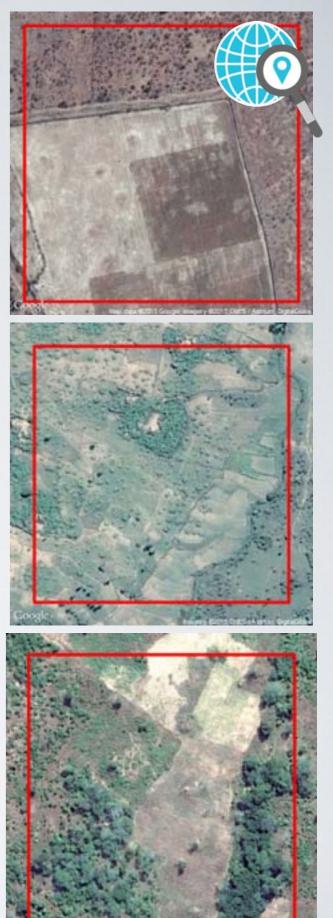
w09@nrsbusinesssolutions.com w02@nrsbusinesssolutions.com w30@nrsbusinesssolutions.com w07@nrsbusinesssolutions.com w01@nrsbusinesssolutions.com w26@nrsbusinesssolutions.com w04@nrsbusinesssolutions.com w21@nrsbusinesssolutions.com w32@nrsbusinesssolutions.com w23@nrsbusinesssolutions.com w33@nrsbusinesssolutions.com w28@nrsbusinesssolutions.com w15@nrsbusinesssolutions.com w05@nrsbusinesssolutions.com

visualgeosurvey@gmail.com w08@nrsbusinesssolutions.com w14@nrsbusinesssolutions.com w20@nrsbusinesssolutions.com w31@nrsbusinesssolutions.com w29@nrsbusinesssolutions.com w17@nrsbusinesssolutions.com w36@nrsbusinesssolutions.com w38@nrsbusinesssolutions.com w19@nrsbusinesssolutions.com w25@nrsbusinesssolutions.com w40@nrsbusinesssolutions.com w34@nrsbusinesssolutions.com w24@nrsbusinesssolutions.com w03@nrsbusinesssolutions.com nrs@nrsbusinesssolutions.com w06@nrsbusinesssolutions.com w18@nrsbusinesssolutions.com

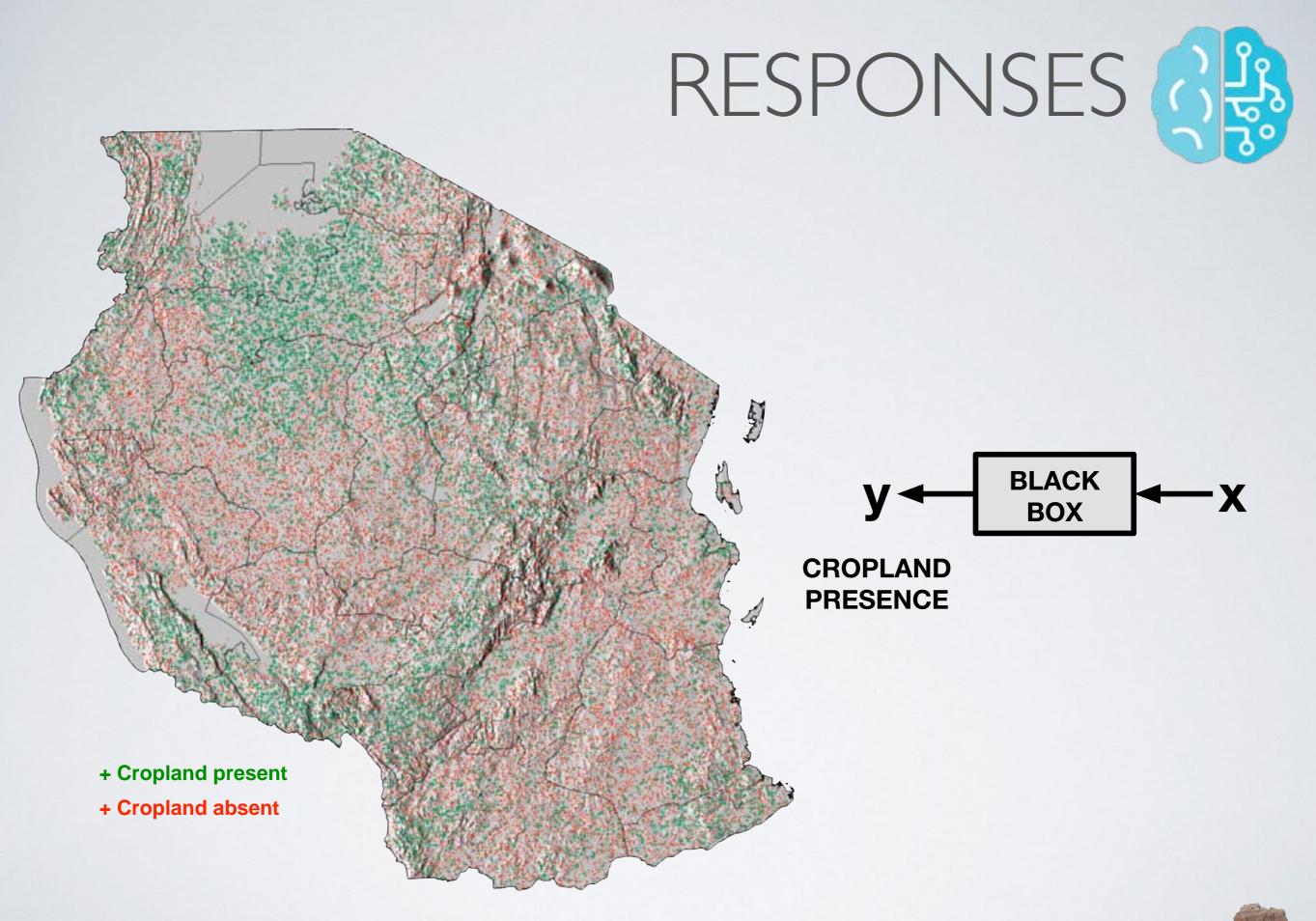
> w02@nrsbusinessolutions.com w30@noshusinessank.tions.com w07@nrsbusinesssolutions.com w01@nrsbusinessolutions.com w26@nnsbusinesssolutions.com w04@nrsbusinessolutions.com w21@nrsbusinesssolutions.com w32@nrsbusinesssolutions.com w23@nrsbusinesssolutions.com w33@nrsbusinesssolutions.com w28/II nrsbusinessolutions.com w15@nrsbusinessaolutions.com w05@nnsbusinessolutions.com visualgeosurvey(il gmail.com w08@nrsbusinesssolutions.com w14@nrsbusinesssolutions.com w2018 nrsbusinessolutions.com w31@nnsbusinesssolutions.com w29 @ nrsbusinesssolutions.com w17@nrsbusinesssolutions.com w36/Enrabusinesssolutions.com w38/8 nrsbusinessolutions.com w19@nrsbusinesssolutions.com w25/II nrsbusinessack/tions.com w40@nrsbusinesssolutions.com w34@nnsbusinesssolutions.com w24/8 nrsbusinessolutions.com w03@nrsbusinesssolutions.com nrs@nrsbusinesssolutions.com w06/II nrsbusinesssolutions.com w18/9 nrsbusinessolutions.com

[geosurvey.qed.ai]









[geosurvey.qed.ai]

nc. AfSIS

## PREDICTORS



CROPLAND PRESENCE

Remote sensing is everywhere.

REMOTE

SENSING

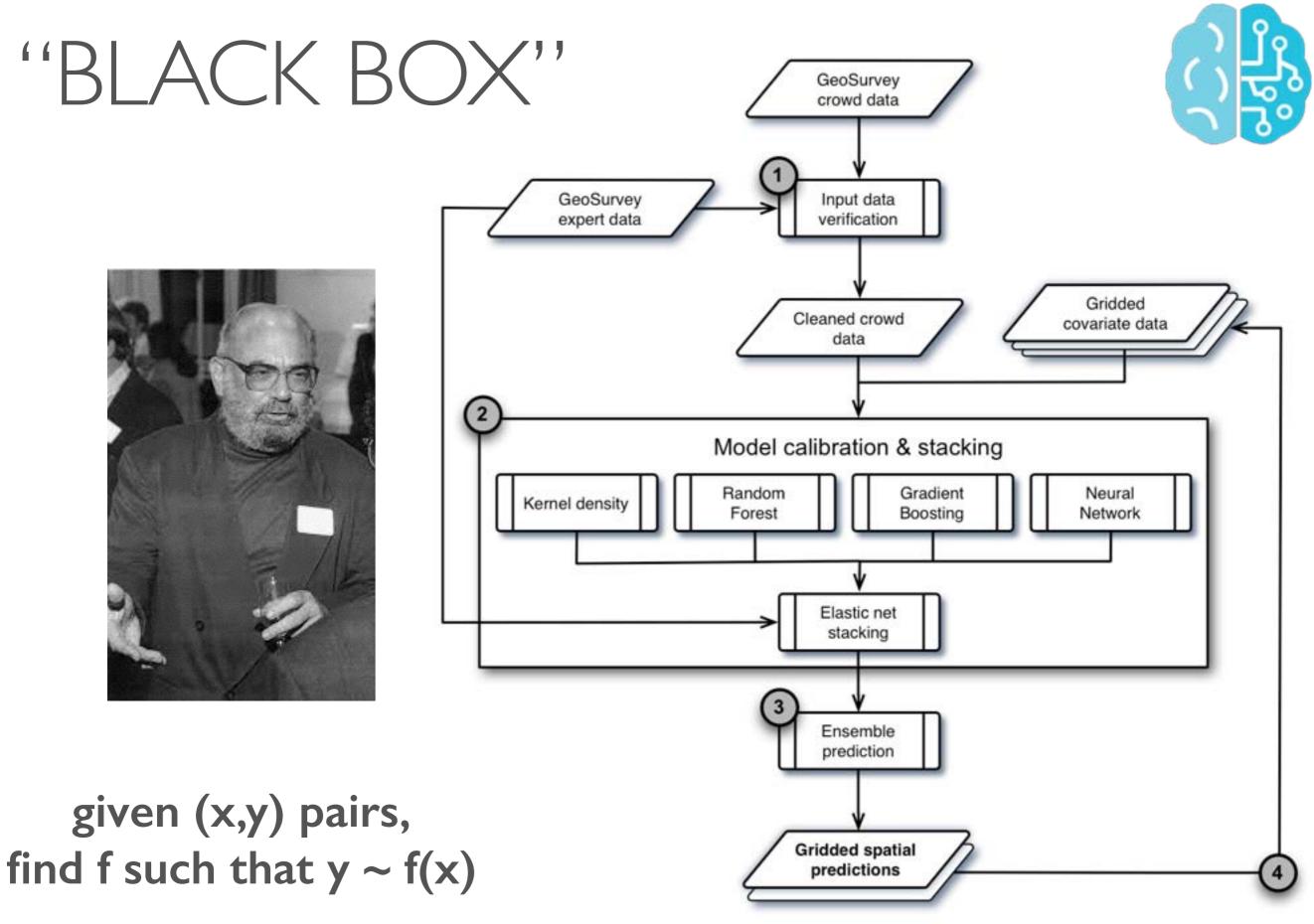
maps <sub>ged.ai</sub>	Search	Q Home Maps Sign Up Log In Terms of Use P	
Browse L	ayers	()	
Type here	٩		
YEAR		Long-term Average Reflectance Band 2, 2000-2014	
1978	0		
1979	0	Char.	
1980	0	Long-term Average Reflectance Band 1, 2000-2014	
1981	0	W And	
1982	0		
Show more		Enhanced Vegetation Index Average, 2010	
IONTH		The MODIS MOD1301 Enhanced Vegetation Index is an index measurement of greenness. Index values closer to 1 indicate an abundance of vegetation	
January	0	whereas values closer to 0 indicate scarce vegetation. This image contains the annual average of all EVI observations in 2010.	
February	O	Q Africa ≣ 2010 Q yearly average	
March	0		
April	0	Land surface temperature (LST) night time series monthly	
May	0	average, February	
Show more_		The MODIS MYD11Az Land Surface Temperature (LST) is a measure of ground temperature of the Earth's surface in degrees Cetsius. This image contains the	
PATIAL RESOLU	TION	time series monthly average for February, using all available February LST Night observations from February 2003 - February 2015.	
30000m		Africa  February Q monthly average	
250m			
1000m	63	Normalized Difference Vegetation Index Time Feries	
5000m	Ø	Normalized Difference Vegetation Index Time Series Average, July	
500m	60	The MODIS MOD1301 Normalized Difference Vegetation Index (NDVI) is an index measurement of greenness. Index values closer to 1 indicate an abundance of	
Show more_		vegetation whereas values closer to 0 indicate scarce vegetation. This image	
		contains the time series monthly average for July, using all available July NDVI observations from July 2000 - July 2015.	
Africa	633	Q Africa ■ July Q monthly average	



Reflectance Blue Band 3 Time Series Average, October The MODIS MOD1301 Reflectance Blue Band 3 data product is used in the development of the vegetation indices. Despite being the noisiest band, the blue band is particularly usefule in reducing atmospheric disturbances in the vegetation indices. This image contains the time series monthly average for October, using all available October reflectance blue observations from October 2000 - October 2014

### [maps.qed.ai]





The content of this presentation is confidential information of Quantitative Engineering Design (QED) Inc. and is not intended for re-distribution to any third-party without the written consent of QED Inc.

qed IIII



### Predicted cropland presence; IM points

- croplands
- houses
- roads
- phones n' drones

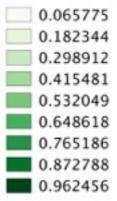


The content of this presentation is confidential information of Quantitative Engineering Design (QED) Inc. and is not intended for re-distribution to any third-party without the written consent of QED Inc.



#### Legend

geosurvey\_h2o\_crp\_predictions



## GLOBAL ACCESS







Home / Admin / Surveys / Edit Survey / Edit questionnaire

## QUESTIONNAIRE

Delete

Human settlements present?	Yes/No/Don't Know	2
Croplands present?	0%/1-50%/51-100%/Don't Know	A 🖹
Woody cover > 60% ?	Yes/No/Don't Know	1
Place a pindrop on all the farms.		Ø
Drag questi	ions to change their order.	
	Place a pindrop on all the farms.	

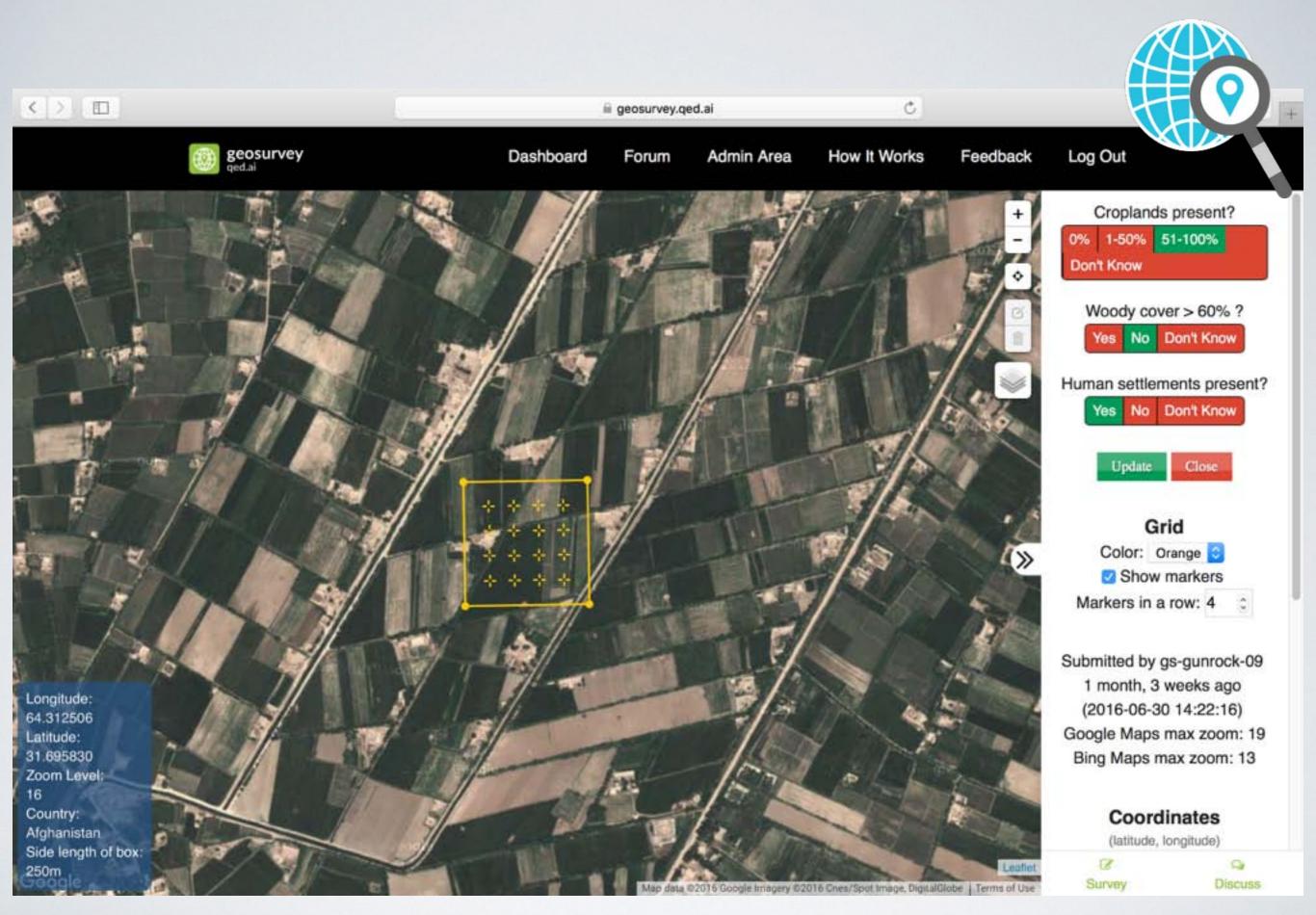
### easy to set up arbitrary surveys in arbitrary regions

[geosurvey.qed.ai]

The content of this presentation is confidential information of Quantitative Engineering Design (QED) Inc. and is not intended for re-distribution to any third-party without the written consent of QED Inc.

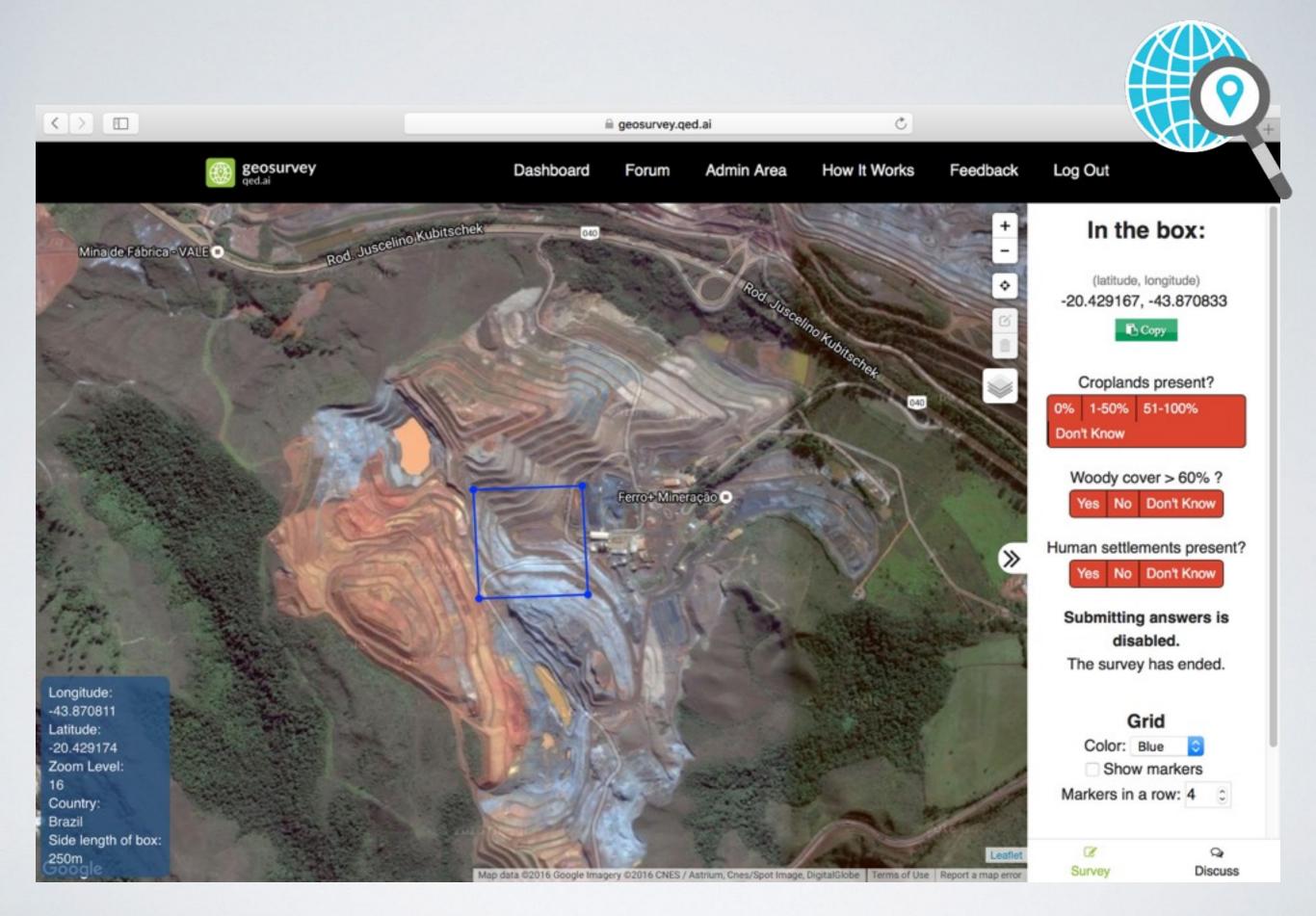
Hi, v





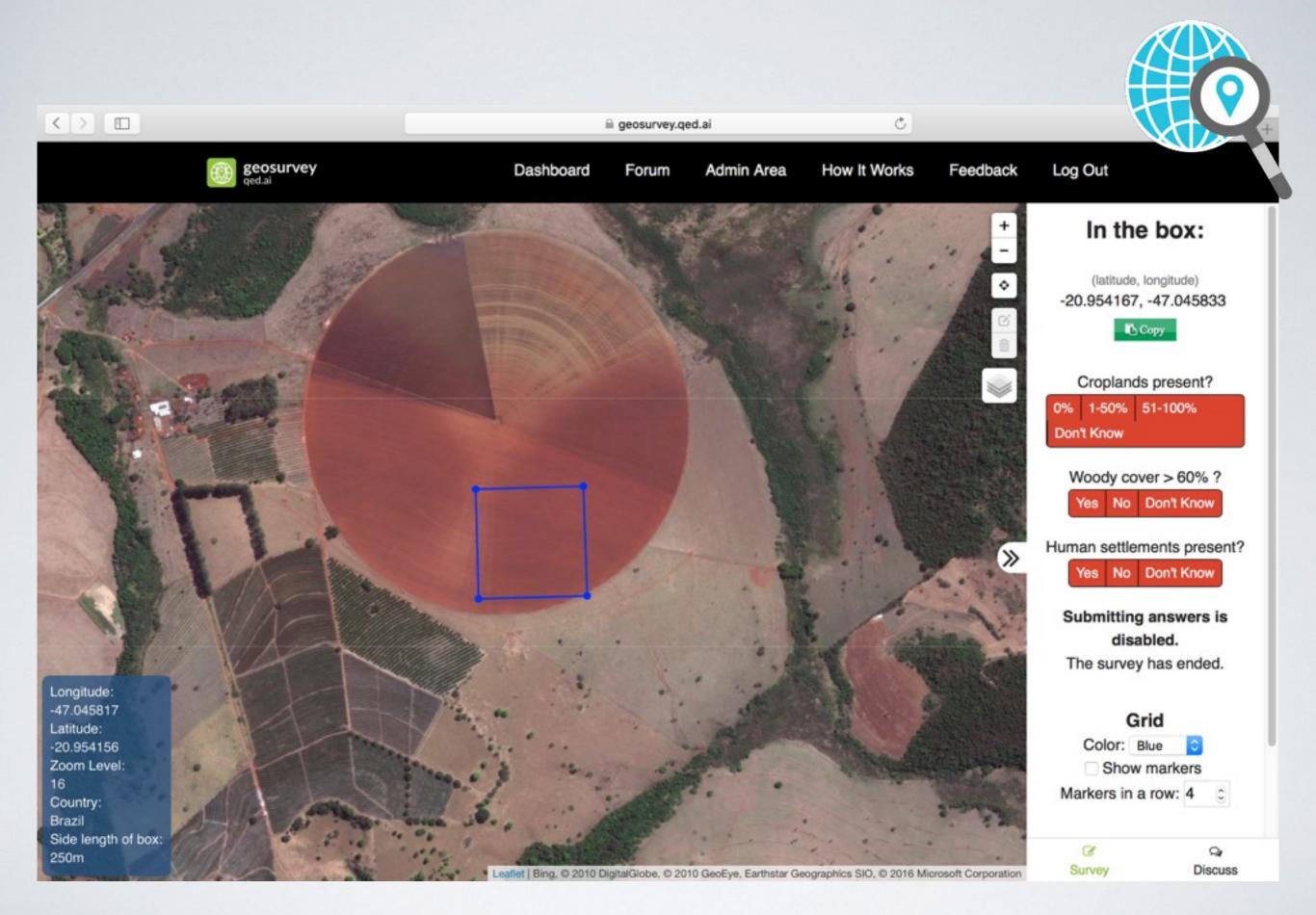
### [geosurvey.qed.ai]





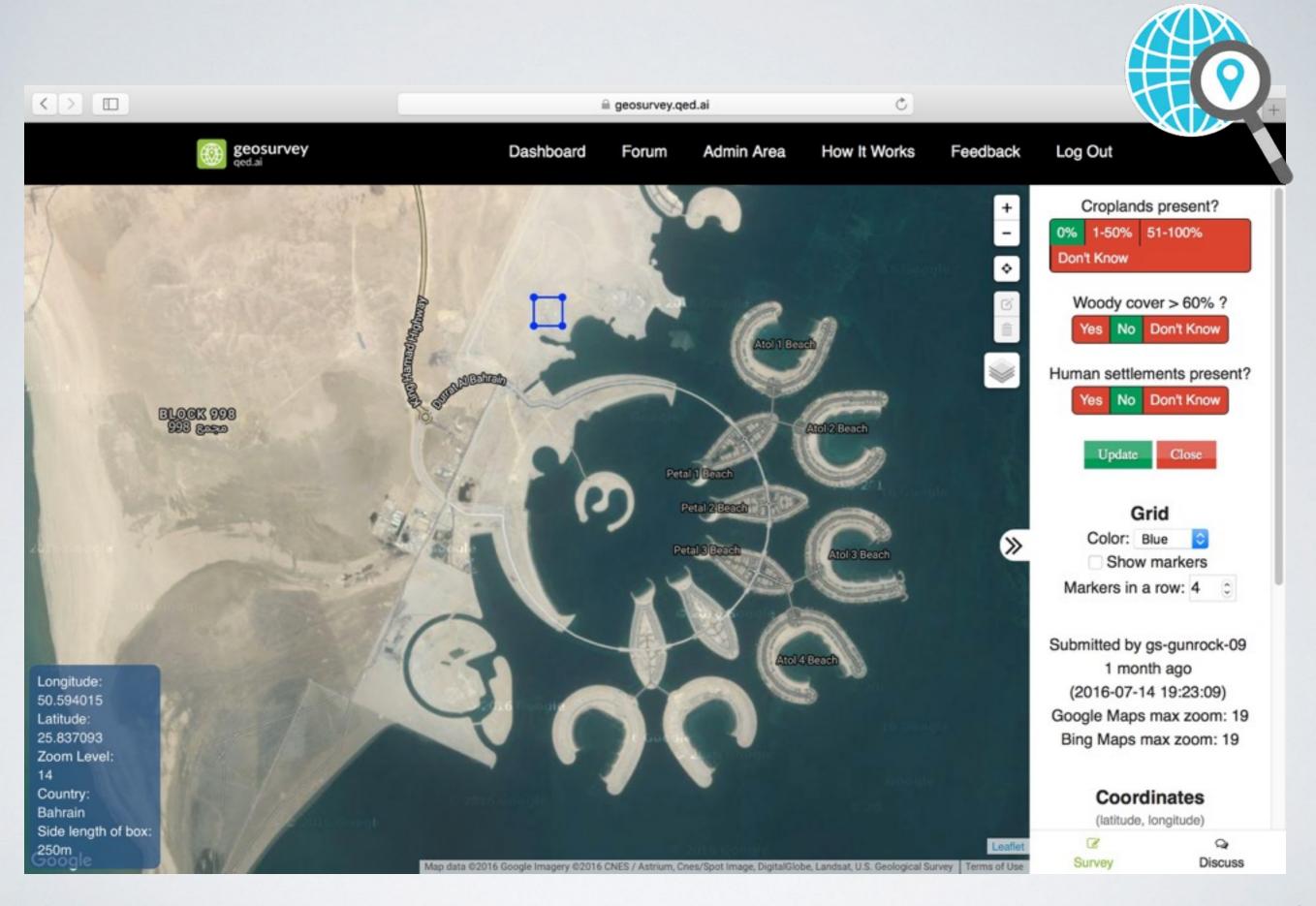
### [geosurvey.qed.ai]





### [geosurvey.qed.ai]

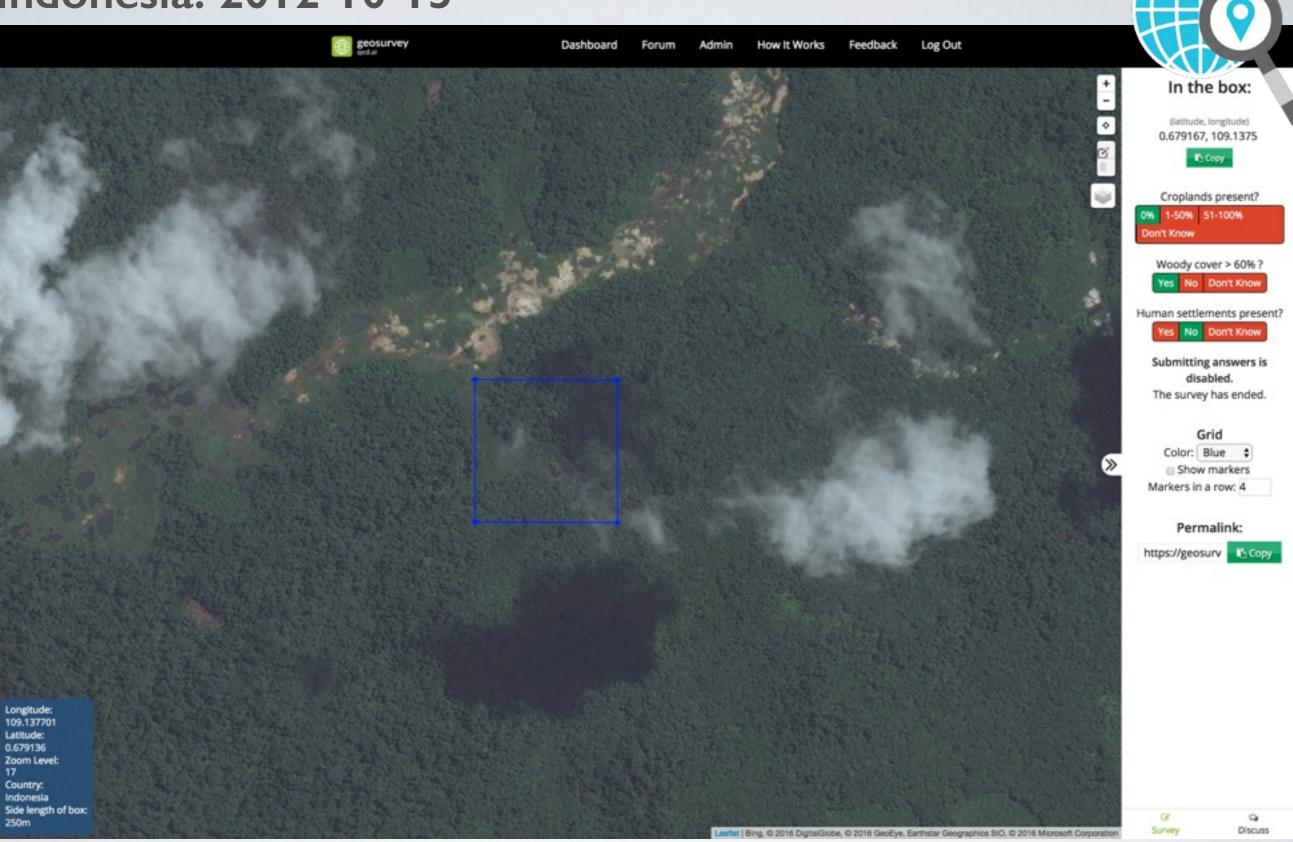




#### [geosurvey.qed.ai]



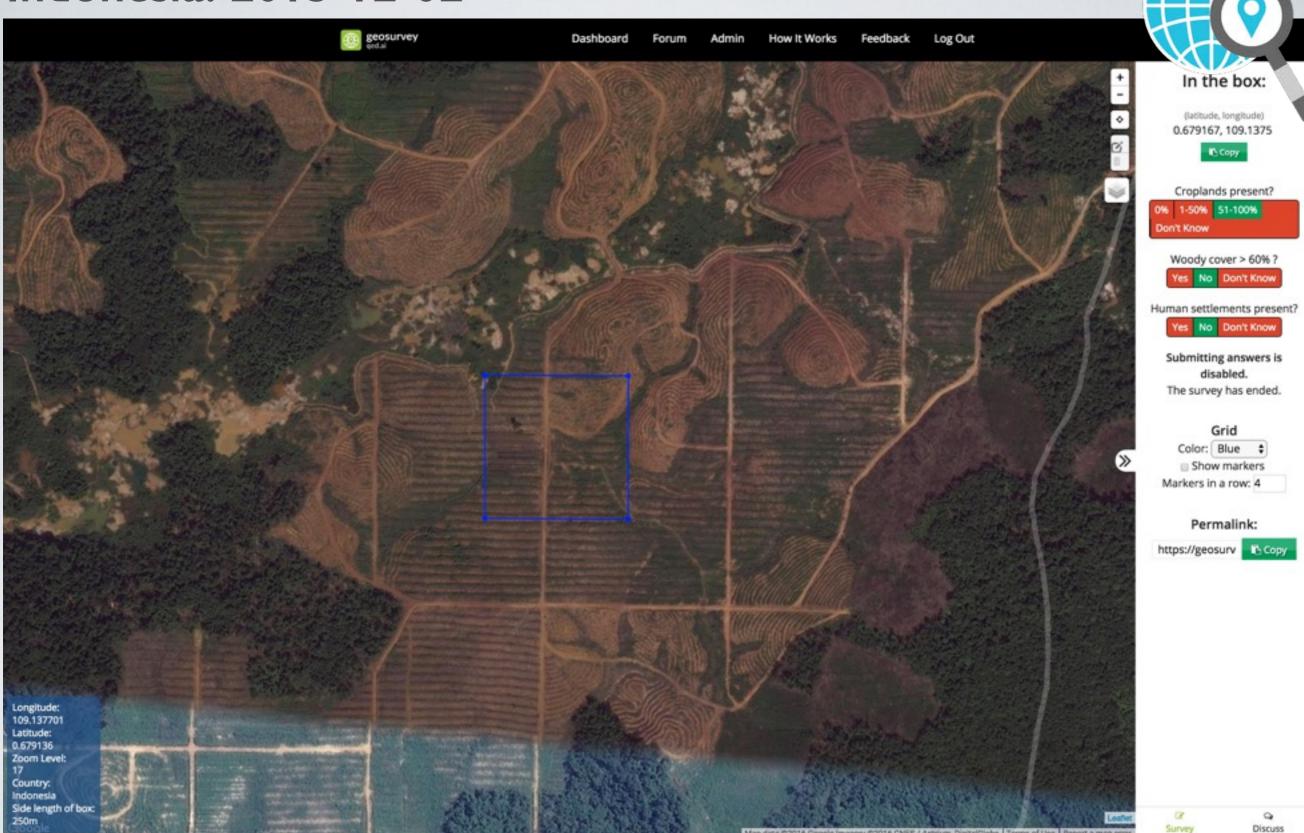
### Indonesia: 2012-10-15



## [geosurvey.qed.ai]



## Indonesia: 2015-12-02



## [geosurvey.qed.ai]



# SUB-PROBLEMS

- determine ROI
- sampling frame
- navigate to sampling sites
- georeferenced extraction
- lab analyses
- database
- mapping (spatial-temporal modeling)
- decisions

qed

!:III:

		0
• • • AFSIS sampling ×		Stimmer M
← → C fí 🗋 sampling.qed.ai	di 🗞 🕁 🤞	
	Sampling Tool (manual)	
		coints within ROI
	Center	
	-3,22	Browse
	# of samples	
	5	
	# of 1k by 1k grids sampled	
	5	
GID=E223-S882-61 ×	Generate Sampling Points	
	Download CSV Download	Kml
P AN ANT AND A PARAMENTAL AND A AN ANALY AND	Download flight Kml files	Download GPX
BAR SANGAR AND	•	
A REAL AND A		and and
	State areas and	
and the second sec		
The second secon		

Generate randomized sampling locations that are hierarchically clustered to minimize transportation costs.

# [sampling.qed.ai]

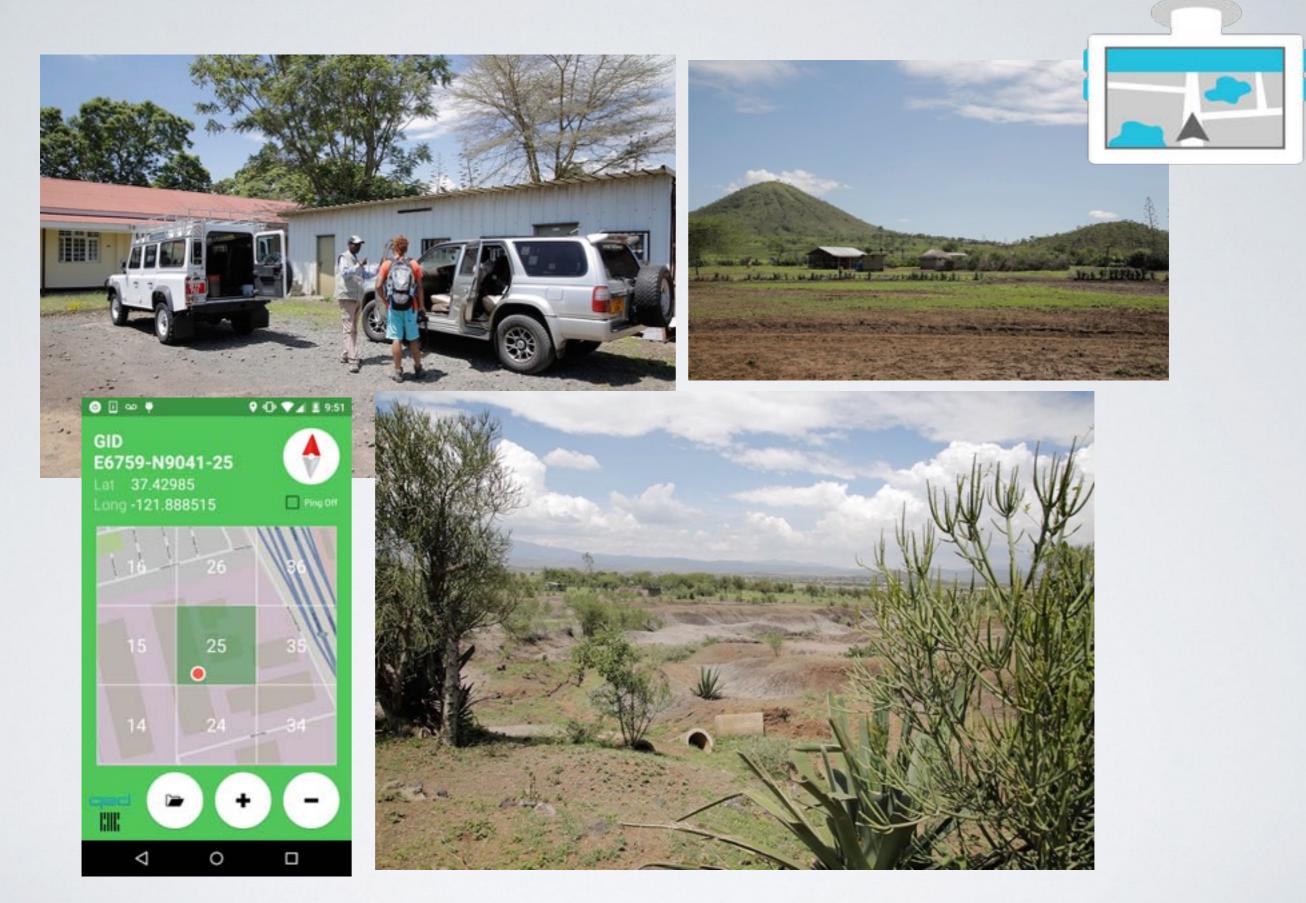


# SUB-PROBLEMS

- determine ROI
- sampling frame
- navigate to sampling sites
- georeferenced extraction
- lab analyses
- database
- mapping (spatial-temporal modeling)
- decisions

qed

!:III:



## [fieldwork]



# SUB-PROBLEMS

- determine ROI
- sampling frame
- navigate to sampling sites
- georeferenced extraction
- lab analyses
- database
- mapping (spatial-temporal modeling)
- decisions

qed

!:III:

Coc Units:	de sh	ieet g	gene	rator					
mm			\$	Codes		Preview			
1				Image type:					
Pag	е			QRCode	\$	V//////			
Dime	nsions			Code type:	Length:		🔍 🔍 tag, kutabiri, com/ahwe		
	1010113			Random String	8		← → C fi  tag.kutabi	Lcom/sheet/	QQ (2) 2
Width: 210	mm	Height: 297	mm	Prefix (e.g. country o	ode):		CERSON CERT		
		- Carlo		max 10 chars					부
Marg	ins			Hidden prefix:			100 C C C C C C C C C C C C C C C C C C	60 - S	85
	Тор:						回防ち	回供知	
	0	mm		Laurent			41m8Hz66	42mRHz66	-
Left:		Right		Layout			2:2		3:2
0	mm	0	mm	No. of codes:	No. of copies:			目流目	
	0			1	2		 25563	123 <b>7</b> 23	163
	Bottom: 0	mm		Columns:	Rows:			IN STATE	
				3	4		Elwic.		
				Code padding:	Label: Number:	5	4:1	10g85eW2 4:2	5:1
				0	Show Show				
				-			一元別に	元別是	125
				Rulers			10.00	10.00	
				Horizontal: Vertical:	and the second se				
				Show Show	14		DiroYDOS	DiroYDOS	

Pre-generate customizable QR codes for labeling samples.

# [tag.qed.ai]





# [fieldwork]

#### MobileSurvey

- · Easily deploy surveys onto Android phones, using Excel spreadsheets.
- · Record georeferences, timestamps, pictures, and barcodes.
- Work in zero-wifi areas, caching data until network returns.
- Regular maintenance coupled with scalable geospatial visualizations and fraud detection.

#### Sign Up!

[mobilesurvey.qed.ai]

The content of this presentation is confidential information of Quantitative Engineering Design (QED) Inc. and is not intended for re-distribution to any third-party without the written consent of QED Inc.



ODK Collect > Soil sa..

Scan sample label

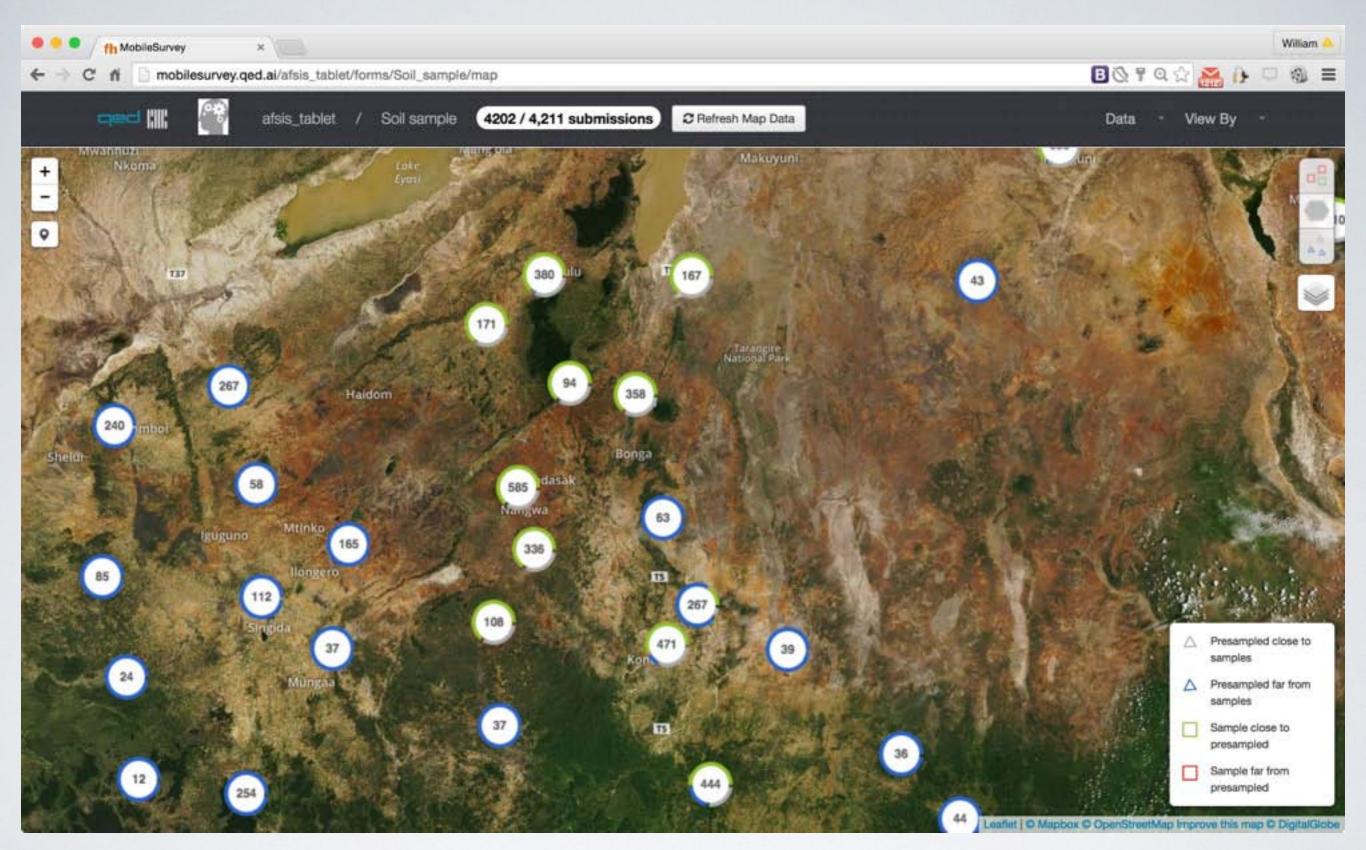
Bag and tag

MI T 4 17:03

(6)

TZ-hSkLbc





Build surveys with Excel and deploy to Android. Cache data until internet returns. Fraud detection and geospatial visualization.

[mobilesurvey.qed.ai]

# CROP ANALYSIS



• Huge yield losses annually from disease and nutrient deficiency. Needs fast intervention.





# <u>GEOSURVEY COLLECT</u>



- Field operators take pics.
- Diagnoses remotely provided on Geosurvey w/ interactive questionnaires.





#### 🚾 🔟 👸 🎔 🗳 🕺 🚳 🕈 🏶 💎 4G 🖌 🛢 10:02



#### geosurvey collect

#### https://geosurvey.qed.ai

#### New sample

Samples

Upload log



		(	? 〒 52%∎22:2
8		Samples	
All	Pending	Uploaded	Failed
	Maize Doctor Demo -	Cob	
	04 Oct 2016	lat: 46.922	
VB.	04:14:10 PM	Ing: 7.390	
YE	Uploaded	mg. 7.590	
Contraction of the local distance of the loc	Maize Doctor Demo -	Cob	
1000	04 Oct 2016	lat: 46.922	
A 10	04:13:39 PM	Ing: 7.390	
	Uploaded		
	Maize Doctor Demo -	Tassel	
	04 Oct 2016	lat: 46.921	
ALC: NOT	04:07:35 PM	Ing: 7.390	
	Uploaded		
1000	Maize Doctor Demo -	All Leaves	
13 mm (m)	04:07:00 PM	lat: 46.921	
110		Ing: 7.390	
	Uploaded		
2 1 B. B. B.	Maize Doctor Demo -	All Leaves	
100101	04 Oct 2016	lat: 46.921	
10.00	04:06:26 PM	Ing: 7.390	
BRAUS I	Uploaded	2017) (2020) Contra	
ALC: NO	Maize Doctor Demo -	Cob	
COLUMN TO AND	04 Oct 2016	lat: 46.920	
CONTRACTOR OF	04:03:58 PM	Ing: 7.389	
10	Uploaded		
	Maize Doctor Demo -	Lower Leaves	
1. 1. 1. 1. 1.	04 Oct 2016	lat: 46.920	
-	03:59:28 PM	Ing: 7.389	
198	Uploaded	mg. 7.569	
	Maize Doctor Demo -	Lower Leaves	
100	04 Oct 2016	1.1.17.000	

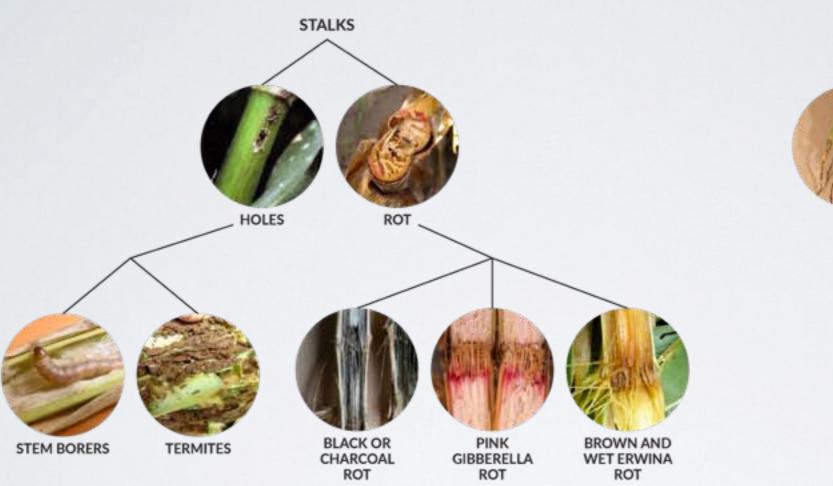




# **Maize Health Decision Tree v.1.0**

#### Scientific Direction: Dr. David Guerena (1AF)

QED | http://qed.ai



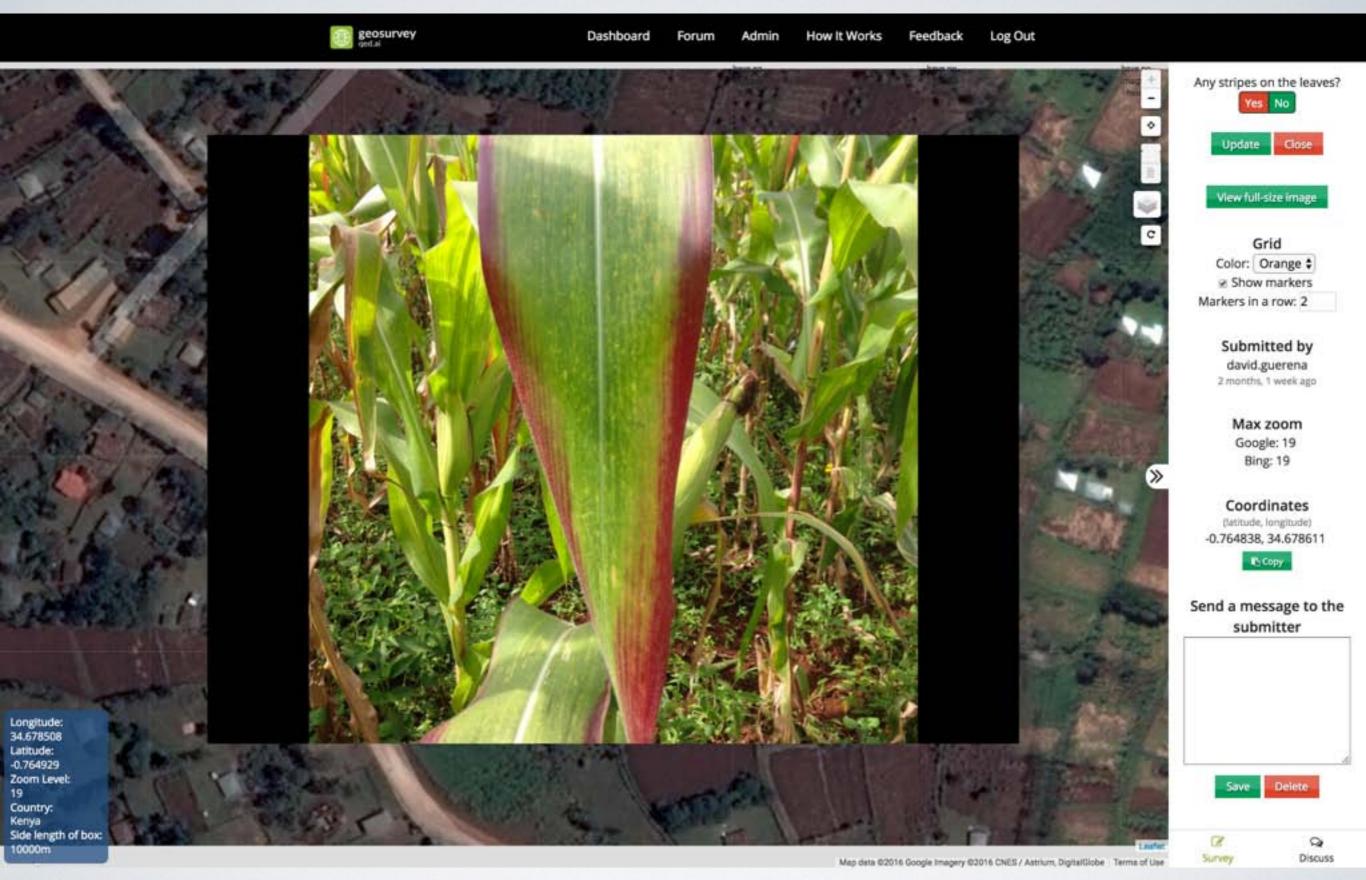


CUTWORMS

TERMITES



#### [geosurvey:collect] weblink: maize



### [geosurvey:collect]

The content of this presentation is confidential information of Quantitative Engineering Design (QED) Inc. and is not intended for re-distribution to any third-party without the written consent of QED Inc.

qed IIII



# https://youtu.be/SP0HTqlmv-Q

The content of this presentation is confidential information of Quantitative Engineering Design (QED) Inc. and is not intended for re-distribution to any third-party without the written consent of QED Inc.

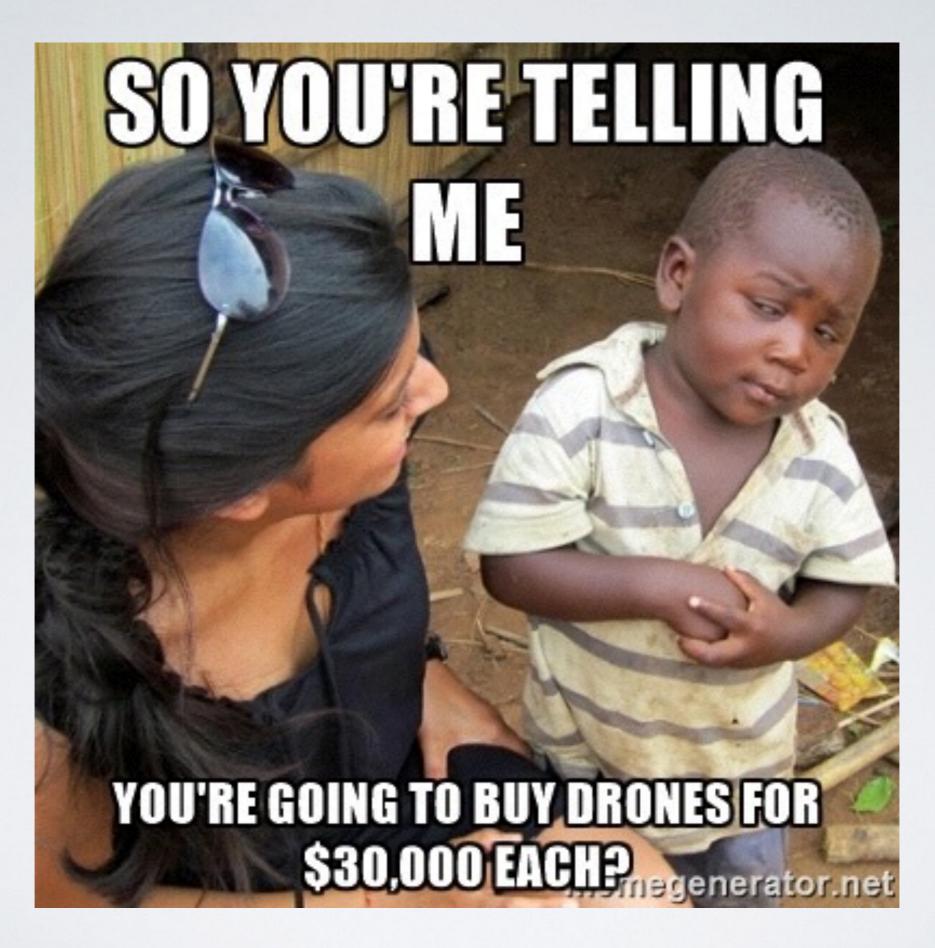
Afsis

### weblink: uav-orthomosaic





### [uav.qed.ai]





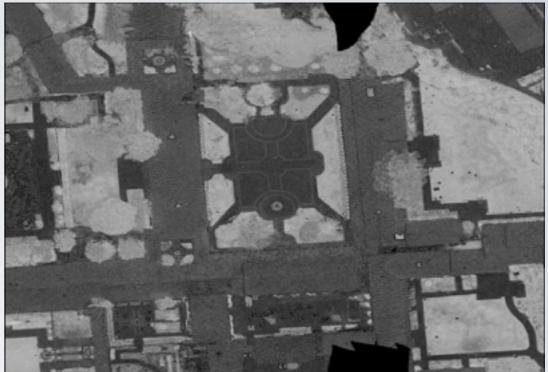


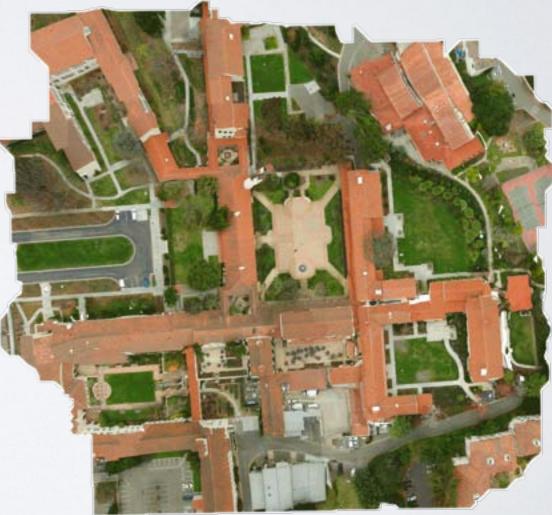
र्चिश्चन	<ul> <li>Image: Image: Im</li></ul>	\$ •⊡• ❤ ∡ ₤ 12:0
AutoDrone Design flight plans for your drone to fly autonomously offline.	Connection Disconnected Flight Status None	Battery Duknown Mission None
AutoDrone QED Tools Everyone This app is incompatible with your device.	USE DRON Center Lat	Center Lon
Add to Wishlist \$0.99 Buy	Flight Height	Constraint
nated orthomosaic flight ontrol. Works offline.	(m) Ascent Speed (max 4m/s)	Lateral Speed (max 14m/s)
99 cents.	$\triangleleft$	0

[uav:auto-drone]



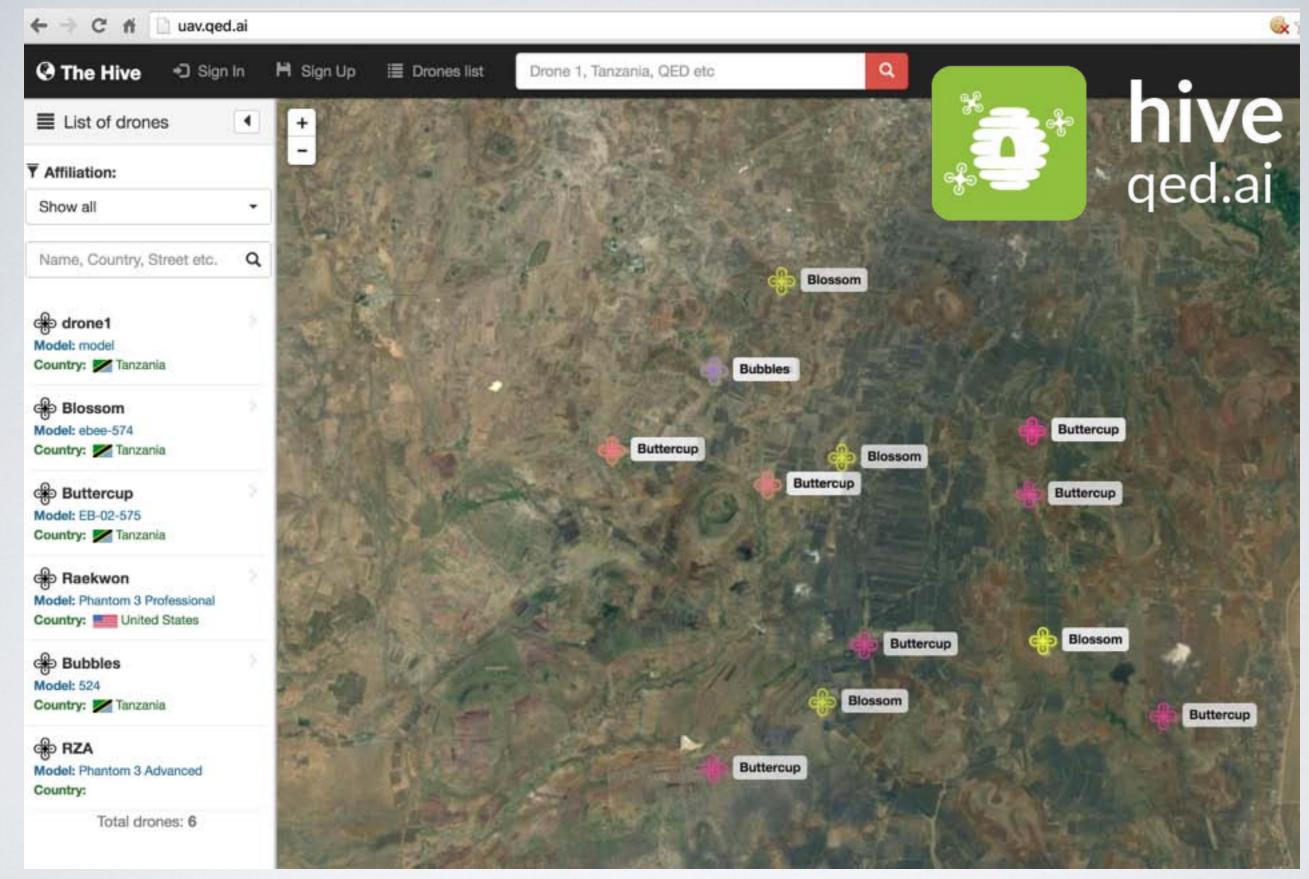






# [uav:auto-drone]





Application: Store and visualize orthomosaics of UAVs in the cloud.











# https://youtu.be/3wj6pvXsgOA

[uav:methodman]

# SUB-PROBLEMS

- determine ROI
- sampling frame
- navigate to sampling sites
- georeferenced extraction
- lab analyses
- database
- mapping (spatial-temporal modeling)
- decisions

qed

!:III:

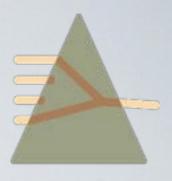




# [soil:preparation]







y = wet chemistry (\$\$\$) x = spectroscopy (\$) SPECTROSCOP





given lots of (x,y) pairs, find f such that y = f(x)



https://www.kaggle.com/c/afsis-soil-properties

# SUB-PROBLEMS

- determine ROI
- sampling frame
- navigate to sampling sites
- georeferenced extraction
- lab analyses
- database
- mapping (spatial-temporal modeling)
- decisions

qed

!:III:

← → C △ ③ afsisdb.qed.a	i/cabinet/wetchemistry/1933/cha	ange/	Q 🕁 🍰	🎽 🤉 👘 🧄 🚺 🔶 👯 🗄	
Ged.ai	Friday, 7th October 2016 14:44		Welcome, AfSIS	DB. Change password   Log out	
A Home	Home - Cabinet - Wet (	Chemistry icr076582			
Q Search					AFSIS DATA
Authentication and Authorization	Group: *	AfSIS	\$	Save	AFOIO DATAI
> Cabinet	SSN: *	icr076582	?	Save and continue	
Carbon and Nitrogen	Public:	Yes	•	editing	
Georeferences		A public sample is accessible to	all users.	Save and add another	
Laser Diffraction Particle Size Analyses	EC:	0.193	7 dS m^-1	Delete	
PXRF		Soil electrical conductivity (soil: v volume basis)	vater ratio of 1:2 weight to	Tools	
Soil Moisture Spectroscopy OPUS records	ExAc:	0.43	? cmolc kg^-1		
Wet Chemistry		Exchangeable Acidity		⊘ History	
	ExBas:	9.973306631	? cmolc kg^-1	Add Wet Chemistry	
<ul> <li>CSV Imports and Exports</li> </ul>		Sum of Mehlich-3 bases (Ca, Mg	, K, Na)		
> Info	M3 AI:	1670.0	? mg kg^-1		
> Invites		Extractable Aluminum concentral	tionby Mehlich 3 extraction		
> Log of Permission Changes	M3 B:	0.041	? mg kg^-1		
Spectroscopy OPUS		Extractable Boron concentration	by Mehlich 3 extraction		
records downloads	M3 Ca:	1590.0	? mg kg^-1		
al Metrics		Extractable Calcium concentration	on by Mehlich 3 extraction		
	M3 Cu:	0.77	? mg kg^-1		
		Extractable Copper concentration	n by Mehlich 3 extraction		
	M3 Fe:	94.9	? mg kg^-1		
	M3 K:	169.0	? mg kg^-1		
		[			
	M3 Mg:	180.0	? mg kg^-1		

Application: Cloud DB of soil information. RDBMS performance, off-site backup, batch upload, input validation, geospatial search, API support.

## [afsisdb.qed.ai]

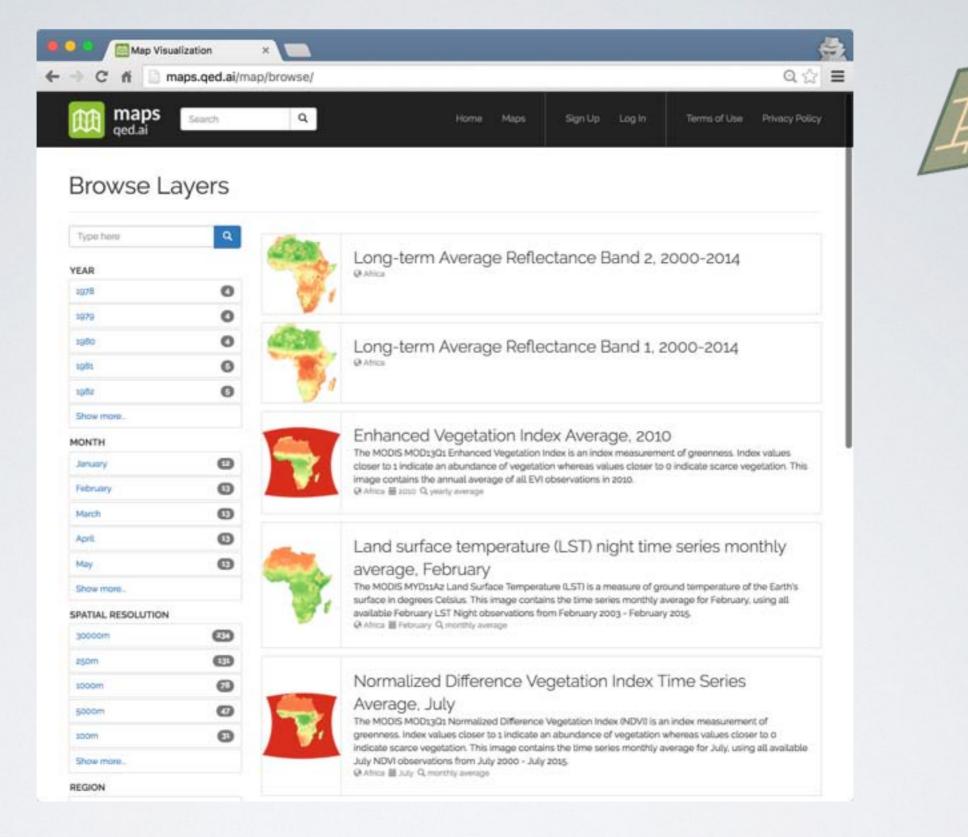


# SUB-PROBLEMS

- determine ROI
- sampling frame
- navigate to sampling sites
- georeferenced extraction
- lab analyses
- database
- mapping (spatial-temporal modeling)
- decisions

qed

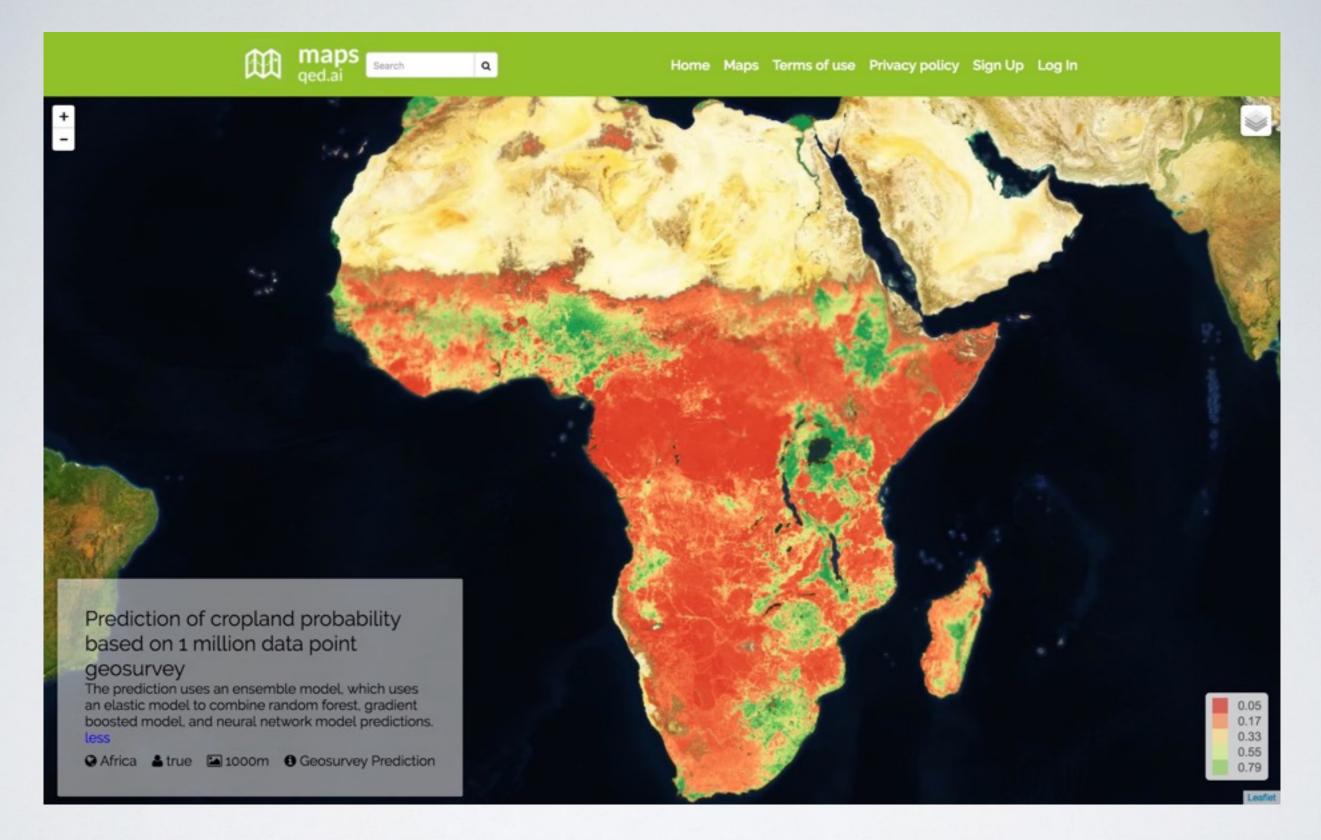
!:III:



Scalable viz. and mgmt. of remote sensing layers and geospatial predictions.

## [maps.qed.ai]





#### Scalable viz. and mgmt. of remote sensing layers and geospatial predictions.

## [maps.qed.ai]



# SUB-PROBLEMS

- determine ROI
- sampling frame
- navigate to sampling sites
- georeferenced extraction
- lab analyses
- database
- mapping (spatial-temporal modeling)
- decisions ...

qed

!:III:

# CONCLUSION







### **Soil Erosion**

Nutrient Imbalance

**Plant Disease** 

**Financial Services** 

Market Access

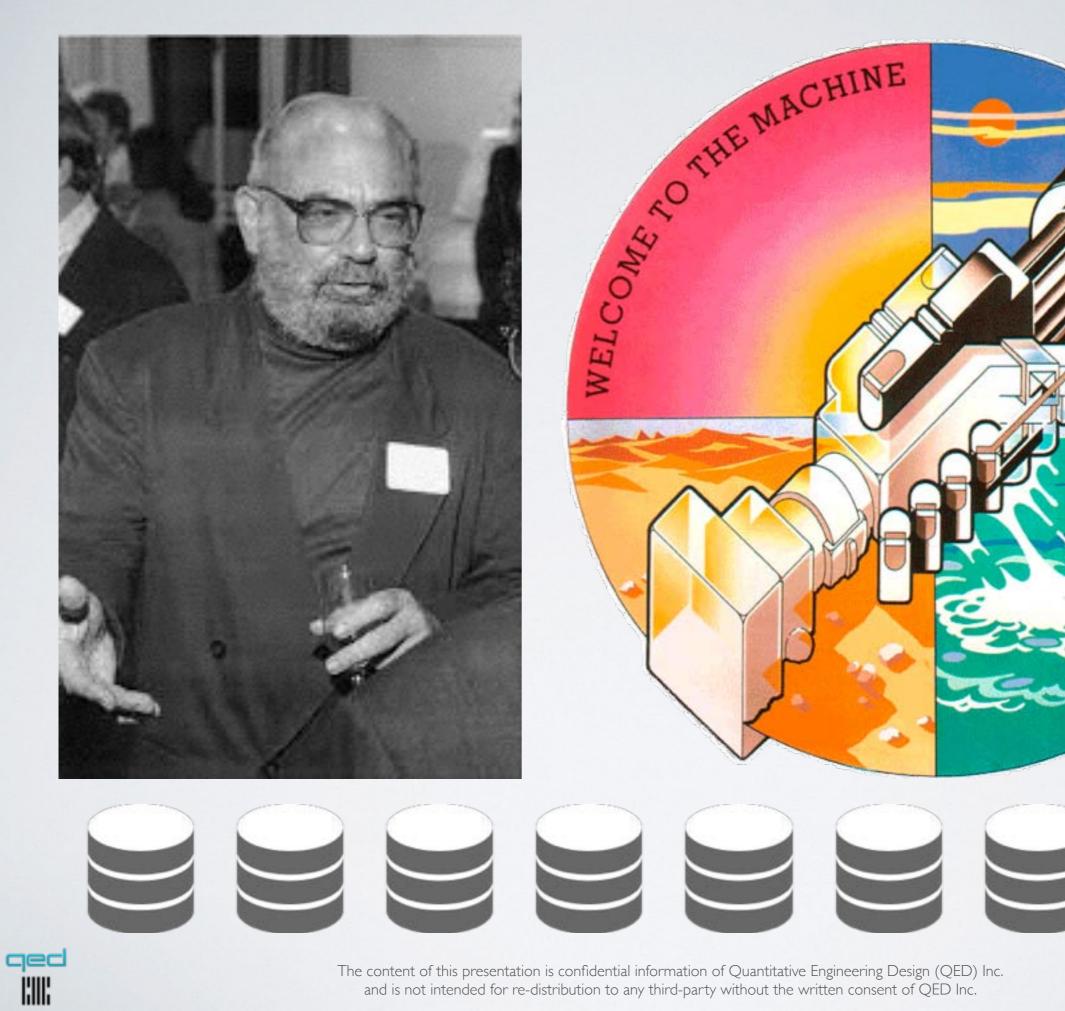
Yield Gaps

**Child Mortality** 

Malnutrition

# **Climate Change**

qed IIII



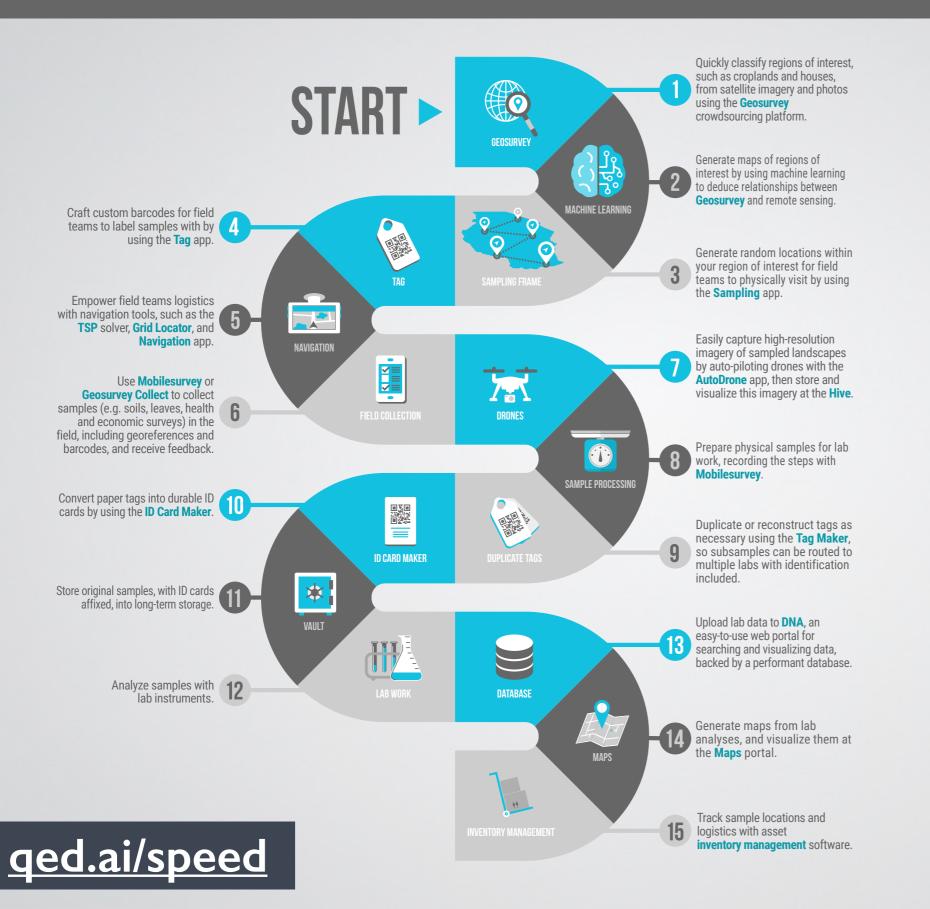
The content of this presentation is confidential information of Quantitative Engineering Design (QED) Inc. and is not intended for re-distribution to any third-party without the written consent of QED Inc.

 $\leq$ 

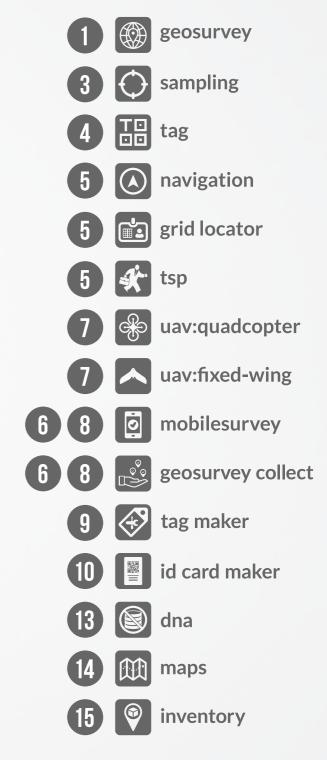


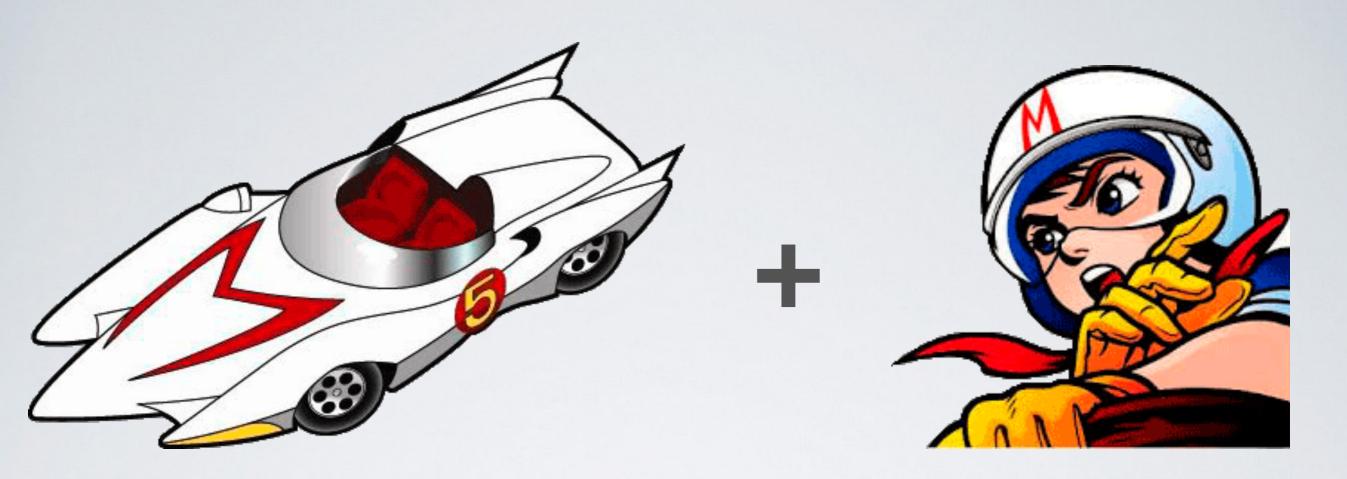
#### SCALABLE PIPELINE FOR ENGINEERING OF ENORMOUS DATA

End-to-end technology solutions to support systematic geospatial data collection and analysis.









#### Technology

Driver = ?

## "War is ninety percent information." - Napoleon Bonaparte

## "In God we trust; all others must bring data!" - William Edwards Deming



# THANKS FOR LISTENING!



To get in touch:

info@qed.ai

http://qed.ai

**#TweetQED** 

