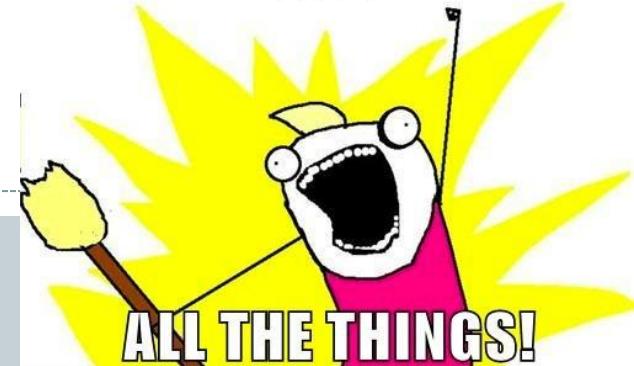


Social Media as Data Source

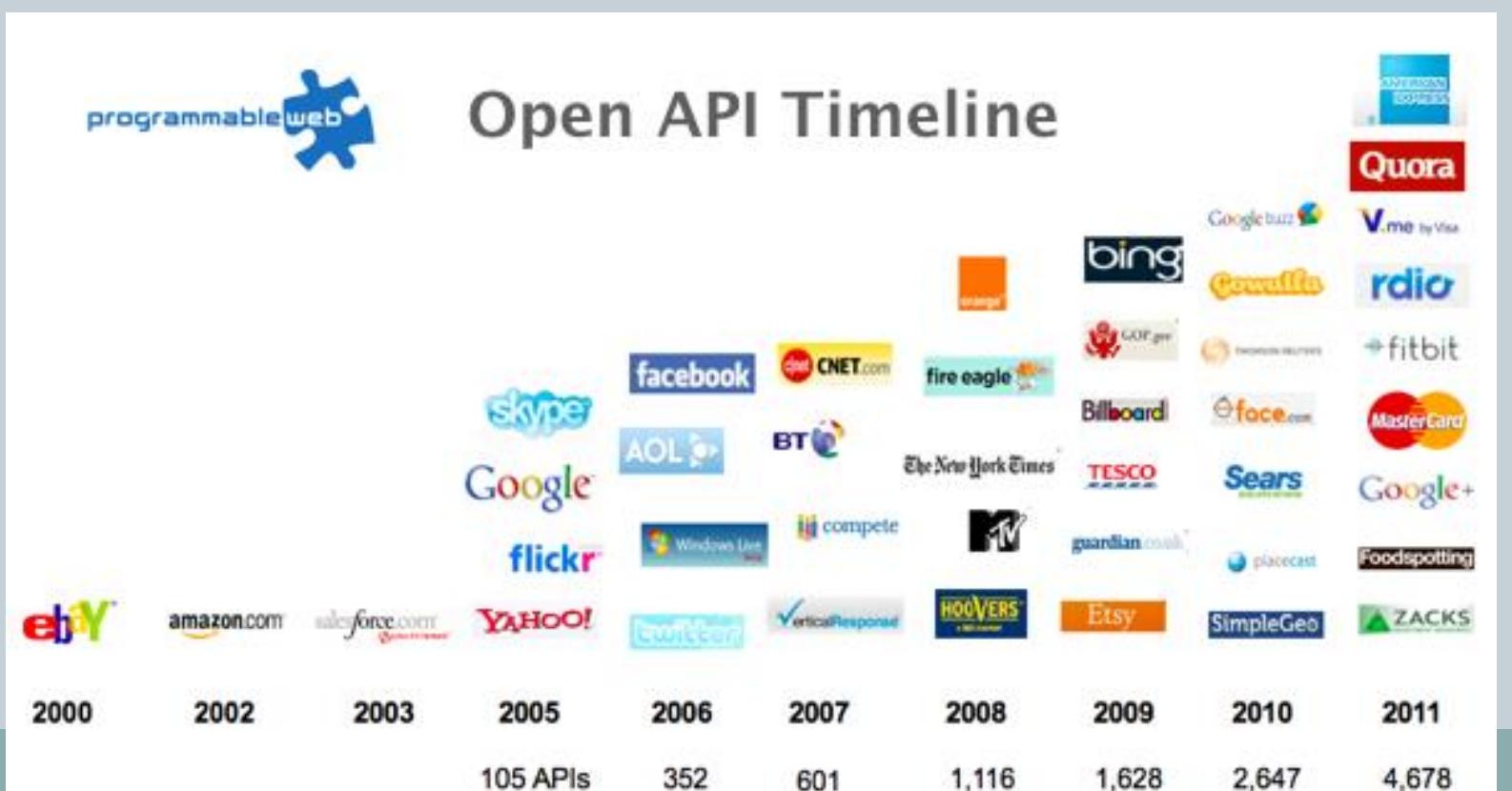


SU 5050
LECTURE 7
JESSICA L. MCCARTY, PH.D.

Definitions



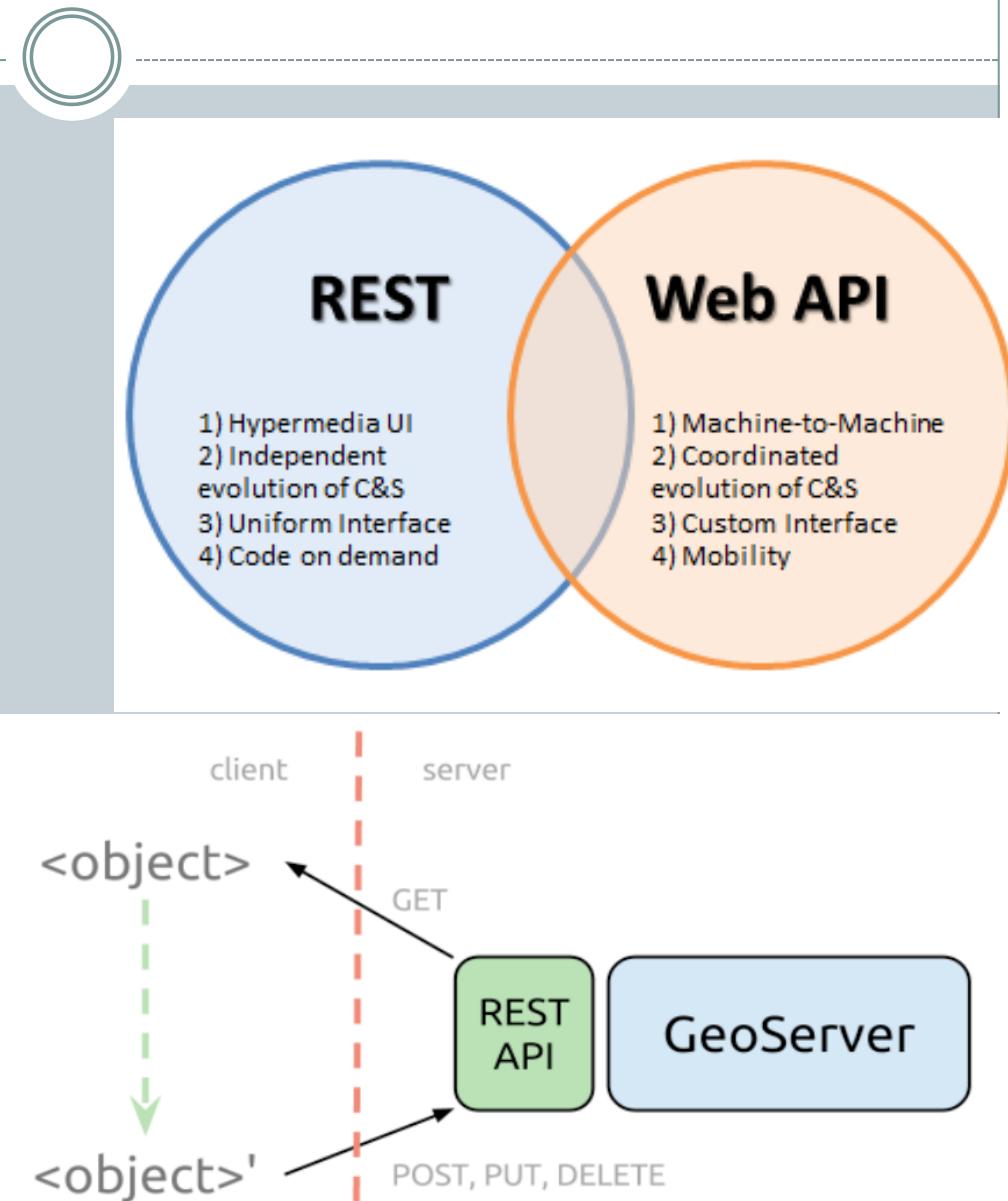
- API
 - Application programming interface



Definitions

- RESTful API

- Representational state transfer (REST)
- Abstraction of the WWW architecture
- Does not require client to know anything about structure of the API



Facebook



- <https://developers.facebook.com/>

Facebook Analytics for Apps

Understand how your customers use your app across all of their devices.

[Learn More](#)



App Monetization
Monetize with ads and publisher tools
[Learn More](#)

App Invites
Let people recommend your app to friends
[Learn More](#)

Social Plugins
Easily make your app or website social
[Learn More](#)

Messenger
Grow your app by powering conversations
[Learn More](#)

Discover More Products

[All](#) [iOS](#) [Android](#) [Websites](#) [Media Publishers](#) [Games](#)

Apps Ads
Grow your app by reaching a larger audience and increasing your installs.

Analytics for Apps
Build better apps by understanding your users through funnels, cohorts, segmentation and attribution.

App Monetization
Monetize your app with ads from Facebook and LiveRail.

Audience Network
Monetize by embedding ads from Facebook's 2 million advertisers in your app.

Messenger
Grow and engage your app by powering the conversation

Parse
The complete mobile app platform - Core, Push and Analytics.

Sharing

Twitter



- <https://dev.twitter.com/>
- <http://gnip.com/>

The screenshot shows the Twitter Flight website, which is a promotional page for "Twitter's Developer Conference". The page features a large Twitter logo and the word "flight" in a large, white, sans-serif font. Below the title, it says "Twitter's Developer Conference" and "Wednesday, October 21, 2015". A prominent blue button labeled "REGISTER NOW >" is visible. At the bottom, there are sections for "Build the Best Apps" (Fabric logo), "Tap into Twitter" (Twitter logo), and "Gain Unique Insight" (GNIP logo). On the right side, there is a "Tweets" sidebar showing a tweet from TwitterDev (@TwitterDev) dated September 25, 2015.

Twitter / Developers

Search English

flight

Twitter's Developer Conference

Wednesday, October 21, 2015

REGISTER NOW >

Build the Best Apps

Tap into Twitter

Gain Unique Insight

Tweets

TwitterDev @TwitterDev 25 Sep

At @overheir? questions about the Twitter API or @fabric for your hacks? Visit @andypiper in the main hall! [fota15](#) [pic.twitter.com/kqqlnkAUlg](#)

Show Photo

Instagram



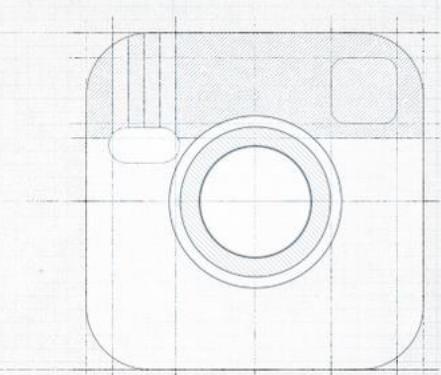
- <http://instagram.com/developer/>
- <http://worldc.am/>

Hello Developers.

The first version of the Instagram API is an exciting step forward towards making it easier for users to have open access to their data. We created it so that you can surface the amazing content Instagram users share every second, in fun and innovative ways.

Build something great.

[Register Your Application](#) then [dive into the documentation](#)



Big Data Services



- Besides GNIP....
- <http://datasift.com/platform/datasources/>
- <http://topsy.com/>
- <http://www.programmableweb.com/news/two-great-social-data-platforms-how-datasift-and-gnip-stack/brief/2014/02/10>

Python and Social Media APIs



- <http://www.pythonforbeginners.com/api/list-of-python-apis>



Human Gold Standard



Human User



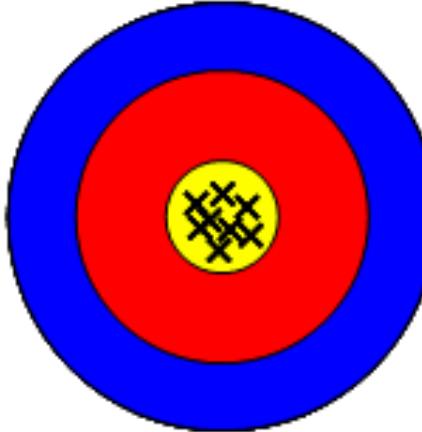
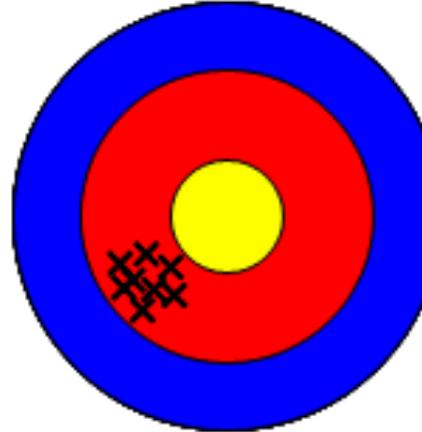
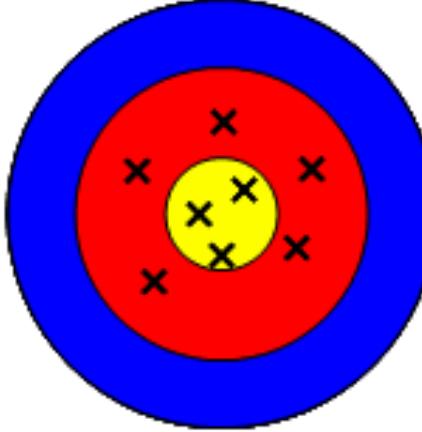
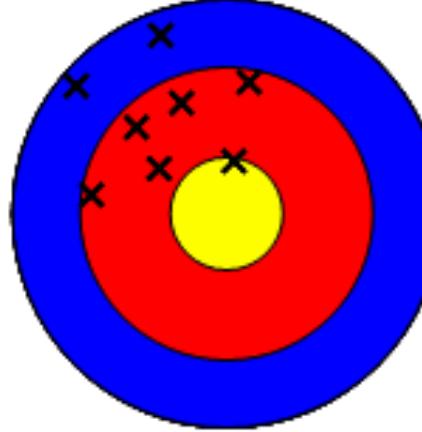
- Data mining algorithms often trying to extract patterns, locations that would be obvious to a human evaluator given a small scale
 - NLP
 - Geolocation
- Gold Standard = predefined evaluation
 - Rubric



Human Gold Standard



- Becomes a predefined evaluation of the results of your data mining algorithm given its performance vs a human evaluator
- Geolocation vs. Human-assigned location from context
- NLP POS-tagging vs. Human-assigned POS tagging

	Accurate	Inaccurate (systematic error)
Precise		
Imprecise (reproducibility error)		

Different types of evaluation



- Intrinsic vs. Extrinsic
- *Intrinsic* considers an isolated algorithm and characterizes performance with respect to gold standard.
- *Extrinsic* characterizes the performance of many versions of the same type of algorithm within a systems, i.e., comparing accuracies.

For the Midterm



- <http://ijcsi.org/papers/7-5-181-186.pdf>
- Under Links

IJCSI International Journal of Computer Science Issues, Vol. 7, Issue 5, September 2010
ISSN (Online): 1694-0814
www.IJCSI.org

181

A New Approach for Evaluation of Data Mining Techniques

Moawia Elfaki Yahia¹, Murtada El-mukashfi El-taher²

¹College of Computer Science and IT
King Faisal University
Saudi Arabia, Alhasa 31982

²Faculty of Mathematical Sciences
University of Khartoum
Sudan, Khartoum 11115

Abstract
This paper tries to put a new direction for the evaluation of some techniques for solving data mining tasks such as: Statistics, Visualization, Clustering, Decision Trees, Association Rules and Neural Networks. The new approach has succeed in defining some new criteria for the evaluation process, and it has obtained valuable results based on what the technique is, the environment of using each techniques, the advantages and disadvantages of each technique, the consequences of choosing any of these techniques to extract hidden predictive information from large databases, and the methods of implementation of each technique. Finally, the paper has presented some valuable recommendations in this field.

Keywords: Data Mining Evaluation, Statistics, Visualization, Clustering, Decision Trees, Association Rules, Neural Networks.

2. Data Mining Overview

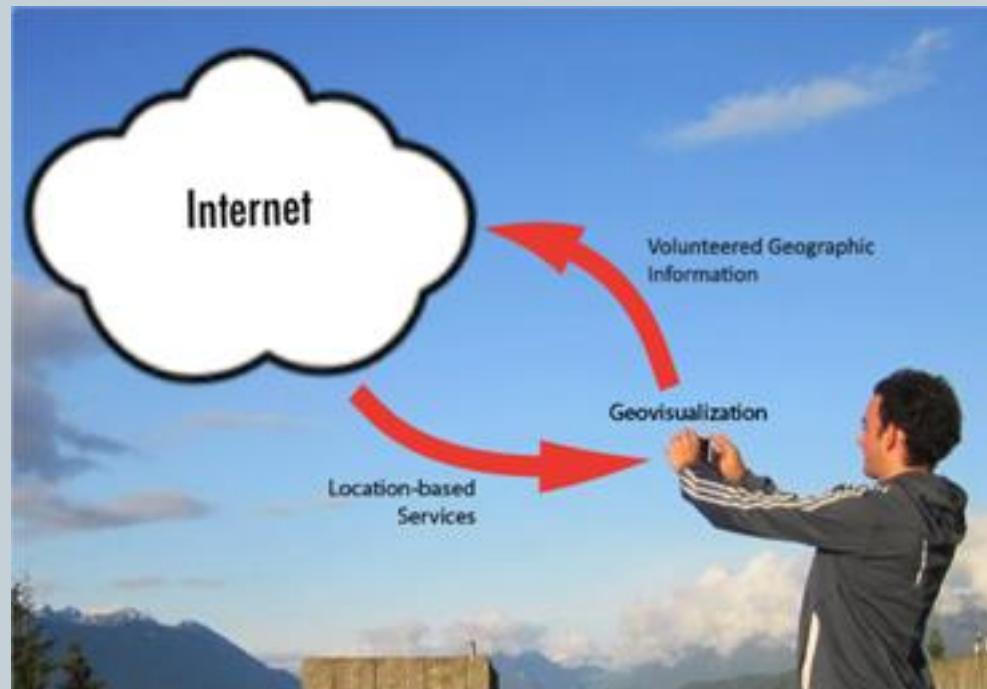
Data mining, the extraction of hidden predictive information from large databases, is a powerful new technology with great potential to help companies focus on the most important information in their data warehouses [6]. Data mining tools predict future trends and behaviors allowing businesses to make proactive knowledge driven decisions. Data mining tools can answer business question that traditionally were too time consuming to resolve. They scour database for hidden patterns, finding predictive information that experts may miss because it lies outside their expectations.

For Your Data Mining Projects



- You will be required to have an evaluation technique (accuracy assessment/validation/evaluation)
- Data mining techniques often lend themselves to human gold standard
 - Whereby a human evaluator is used to assess the accuracy, precision, and/or validity of the results or some component of the results

Volunteered Geographic Information



Geography 2.0: Human Sensor Network



- Streams of Environmental Information
 - Distributed sensors through wireless networks with geolocation capabilities.
 - Differentiated by intent of providers of data.
- Volunteered geographic information (VGI)
 - Interactive web apps, mobile apps, and citizen science projects with geocoding (e.g., addresses / GPS) and geotagging capabilities
- Ambient geographic information (AGI)
 - Social networking and *user-generated content* provides the opportunity to harvest geographic information
 - ✖ Themes and Context

Example VGI Projects



- Weather and climate
 - GLOBE (student/teacher/citizen science)
 - www.globe.gov
- Biota
 - Christmas Bird Count, Breeding Bird Survey
 - www.pwrc.usgs.gov/BBS/
 - Phenology network
 - www.usanpn.org/

Geotagging



- Wikimapia
 - www.wikimapia.org
 - Provides “Gazeteer” function
 - >15 million places tagged
 - Can include colloquial names and be edited by anyone

OpenStreetMap



- www.openstreetmap.org
- Direct competitor to Google Maps
- Highly biased
 - Editors mainly male, highly educated, techie jobs
 - More strip clubs than daycare centers
 - <http://www.fastcoexist.com/3032543/how-street-maps-can-be-sexist>

Geospatial Data: GIS, RS, GPS, Field Data



GISystem vs. GIScience

GISystem – software

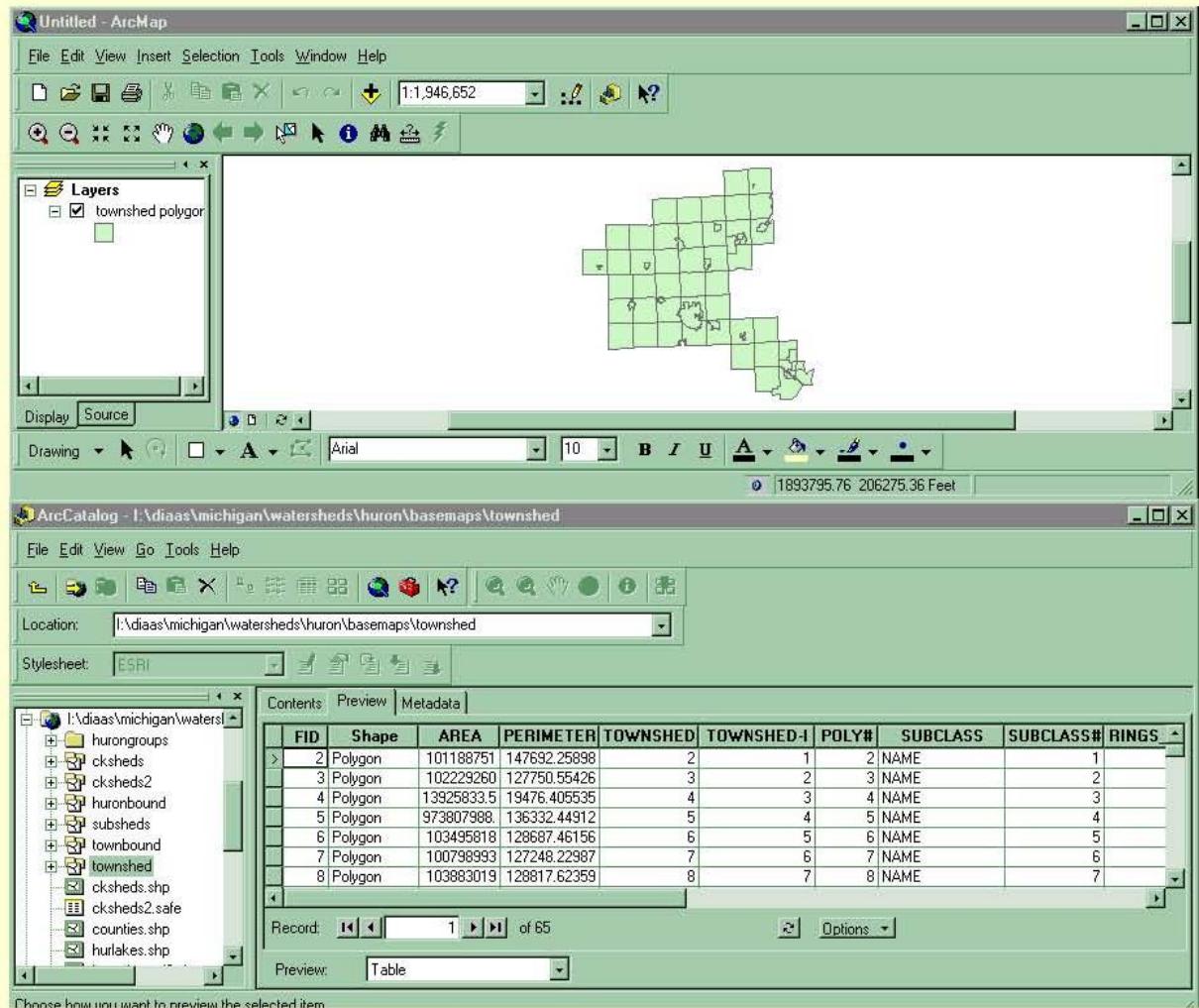
GIScience – science behind the software

e.g. database structures, qualitative
spatial reasoning, spatial web services,
spatial information management and
distribution.....

Multiple Faces of GIS

GIS is used as

- a database and data repository
- a graphic visualization tool
- a tool for exploratory and confirmatory analysis
- a tool for predictive modeling
- an environmental management tool



Definition

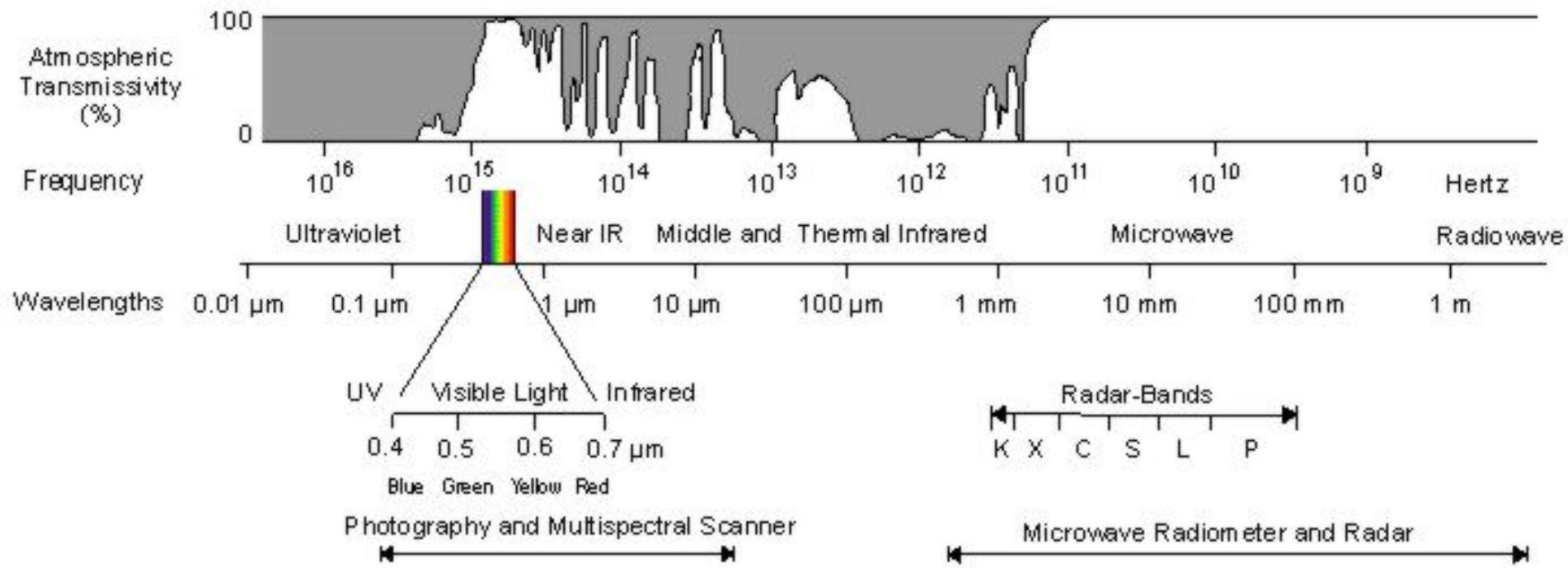


- Remote Sensing
 - “...the science and art of obtaining information about an object, area, or phenomenon through the analysis of data acquired by a device that is not in contact with the object, area, or phenomenon under investigation.”
 - Lillesand and Kiefer, p. 1

Electromagnetic Energy at Surface



- Energy is emitted from sun in varying wavelengths
- Some wavelengths are absorbed by the atmosphere



30-meter resolution

1-meter resolution



Image © 2005 DigitalGlobe

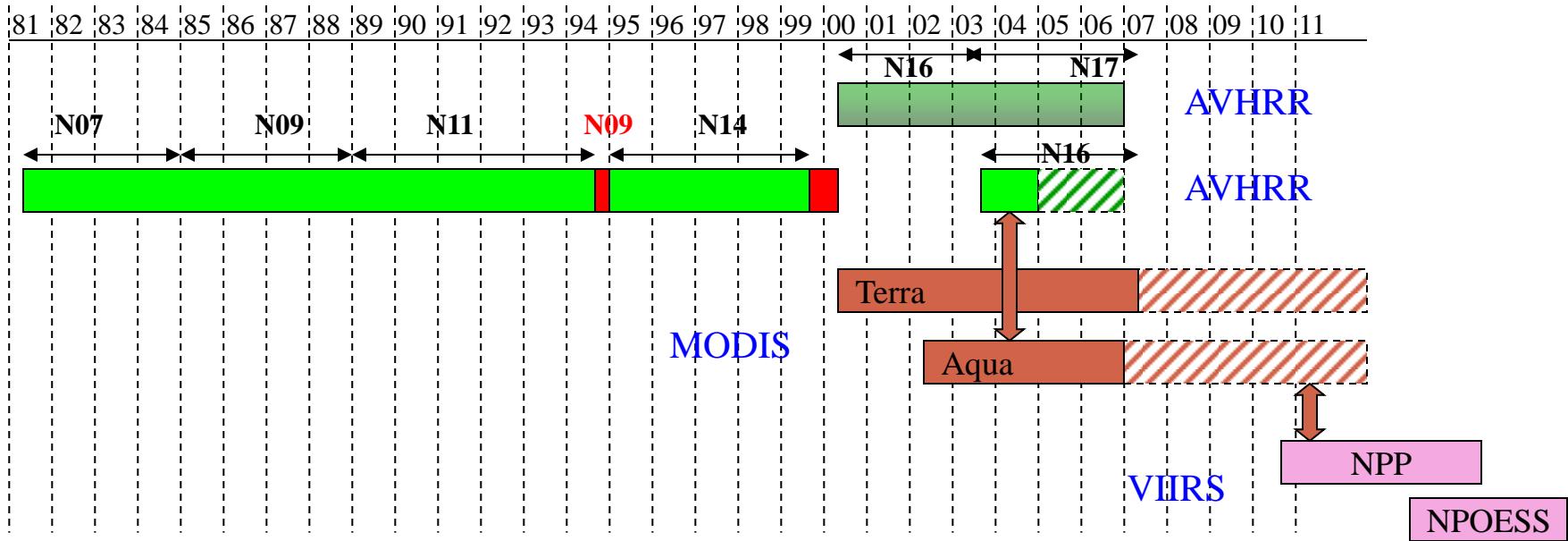
© 2005 Google™

Pointer 41°56'32.61" N 83°37'03.62" W

Streaming ||||||| 100%

Eye alt 4901 ft

Land LTDR (long term data records)



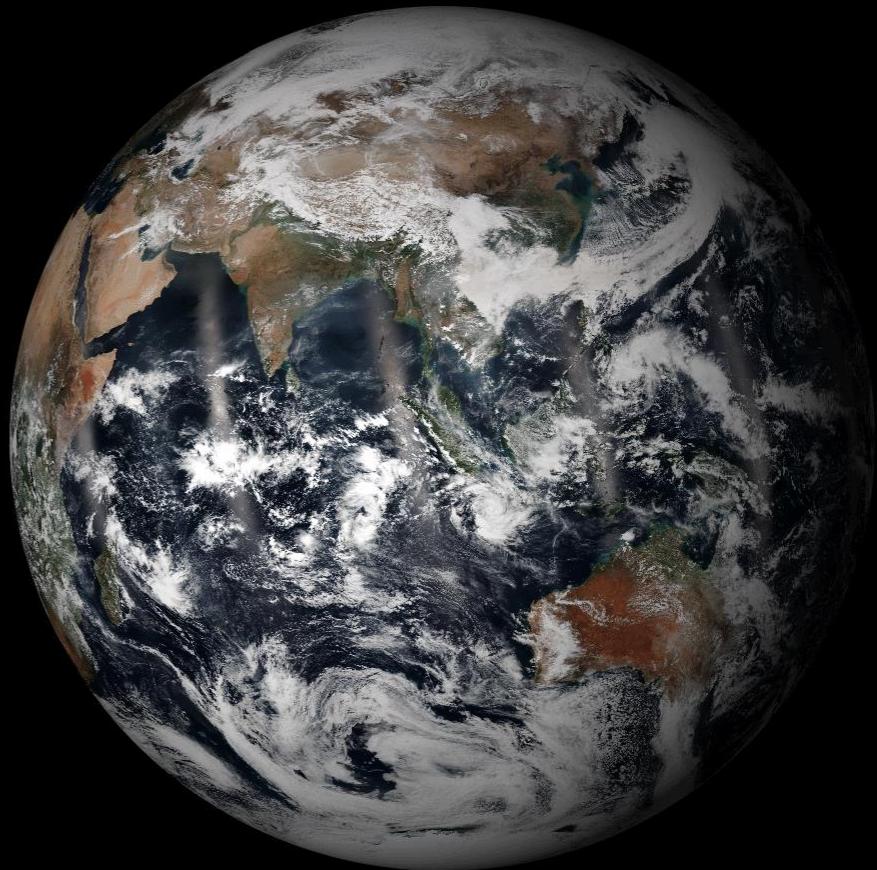
AVHRR (GAC) 1982-1999 + 2003-2006

MODIS (MO(Y)D09 CMG) 2000-present

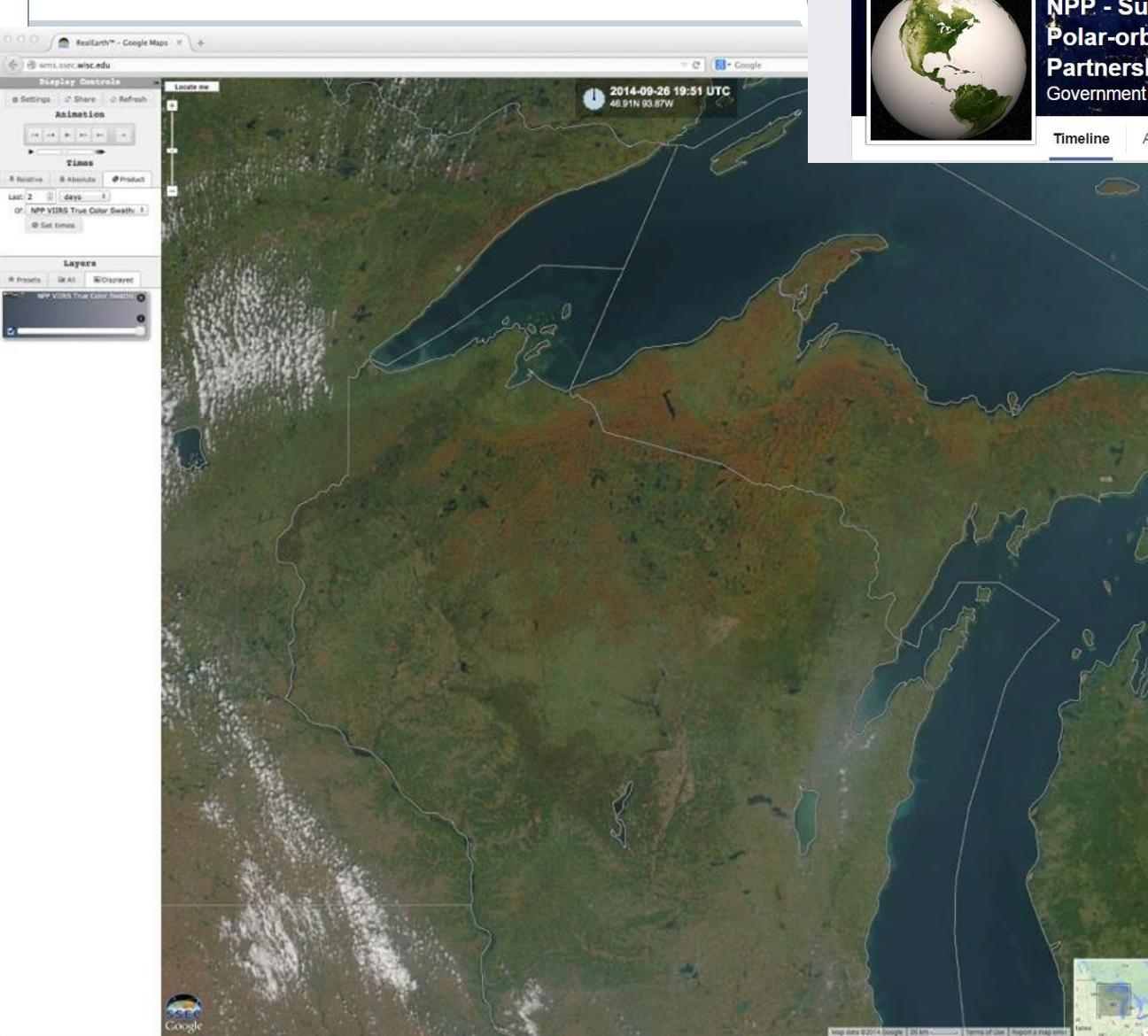
VIIRS 2010 – 2020 (global gridded comparison products)

GOAL Seamless Data Record

VIIRS (2011 – present)



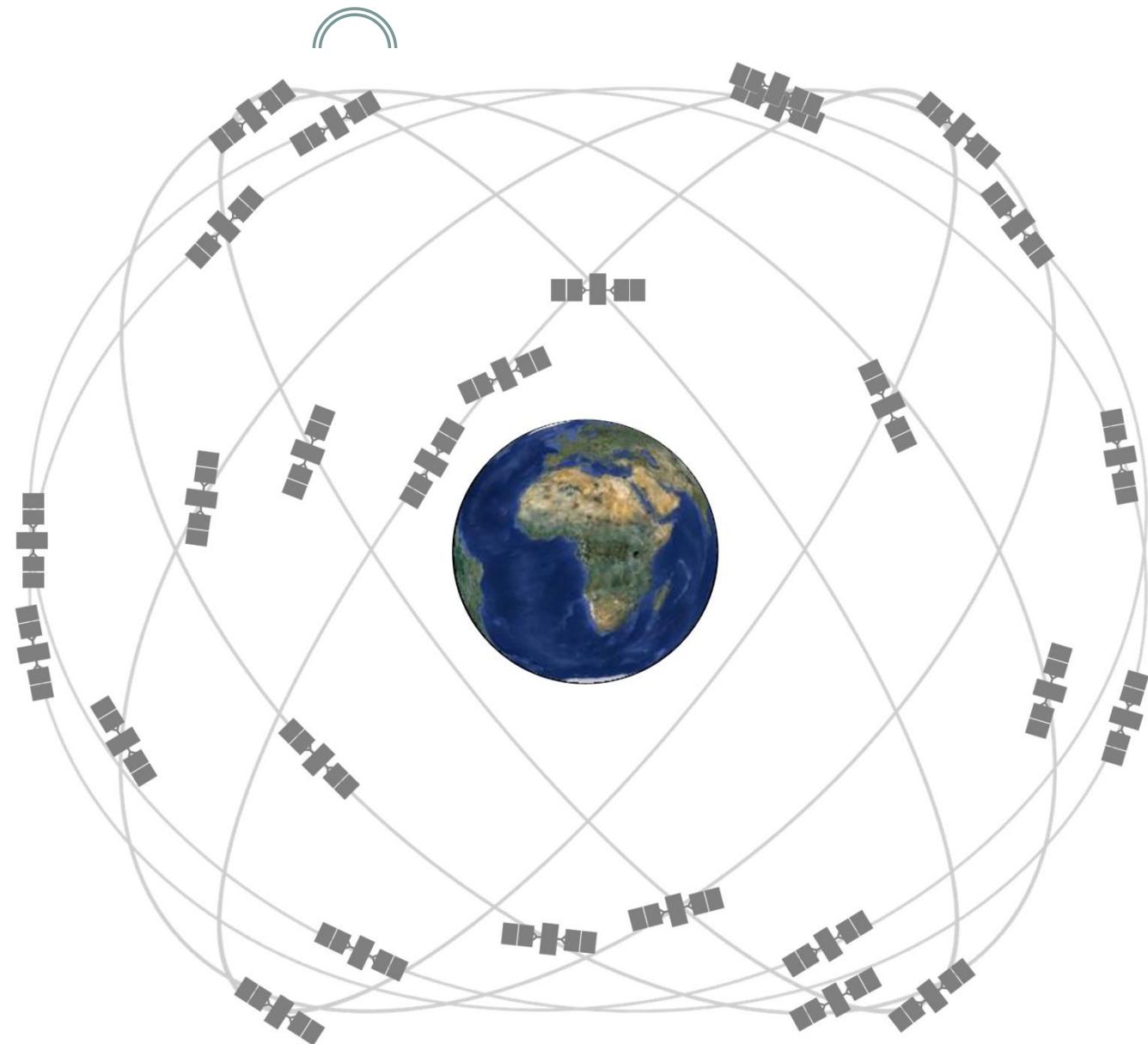
VIIRS Suomi NPP



http://www.nasa.gov/mission_pages/NPP/news/suomi.html

Global Positioning System (GPS)

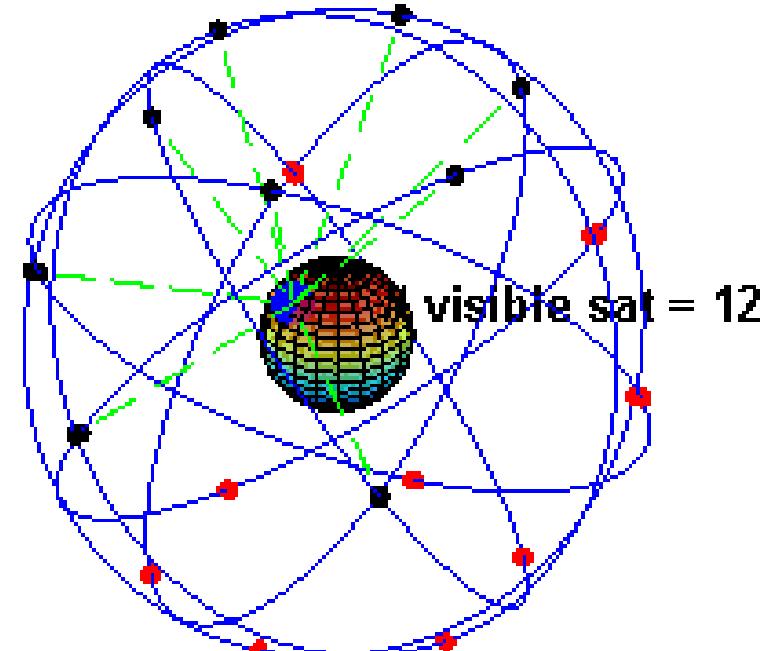
[http://www.gps.gov
/multimedia/images
/constellation.jpg](http://www.gps.gov/multimedia/images/constellation.jpg)



Global Positioning System (GPS)



- 24 GPS satellites (21 plus 3 spares) orbit at 20,350 km
- Each satellite
 - Knows its location
 - Has an atomic clock
- Radio signals broadcast by each satellite
 - Carrier Phase: L1 (19cm), L2 (24cm) wavelengths
 - Code Signals: C/A, P
 - Military Signal: P(Y) received on both L1 and L2 [encrypted]
- Carrier Phase (w/in millimeters) more precise than Code Phase (C/A = meters, P = decimeters)





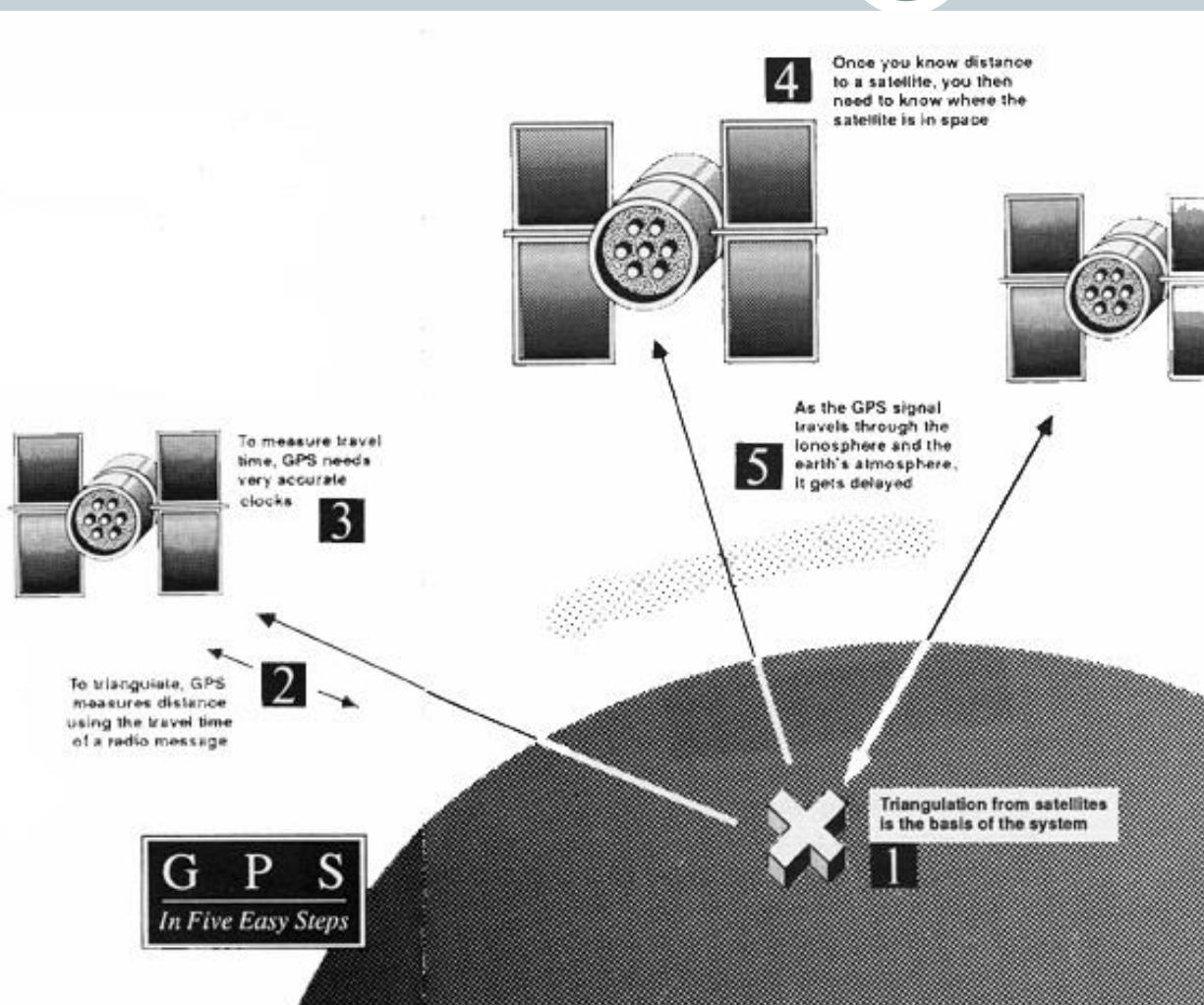
GPS Receivers



- Recreational (\$100 - \$500)
 - Designed for receiver use only, can manually record positions or some have data cables
 - > 7 meters accuracy
- GIS-grade (\$1500 - \$15,000)
 - Designed for data to be downloaded and used in GIS
 - 1-3 meter accuracy with differential correction
- Survey-grade (> \$30,000)
 - With real-time differential correction using interferometry
 - Designed specifically for engineering schematic creation
 - Uses carrier phase for sub-cm accuracy (e.g., +/- 2 cm)



GPS



- The idea is to use the satellites to locate ourselves anywhere on the earth

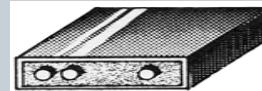
How does it work?

- Pseudo-random code (i.e., radio signal) generated by satellite and decoded by receiver
- Receiver compares signals and determines time it took to travel from satellite to receiver
 - Speed of light is a physical constant, so we can determine distance

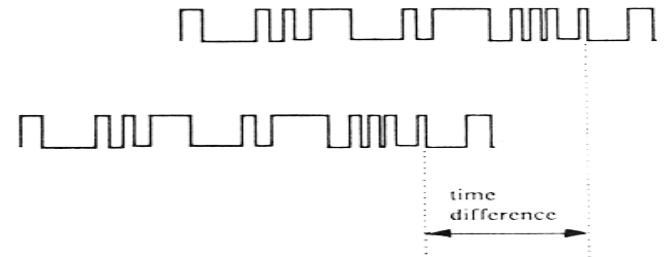
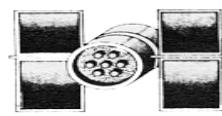
Distance = speed of light * travel time

Code Phase

Pseudo-code from
receiver



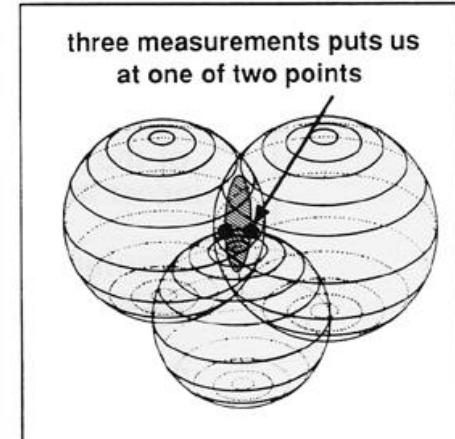
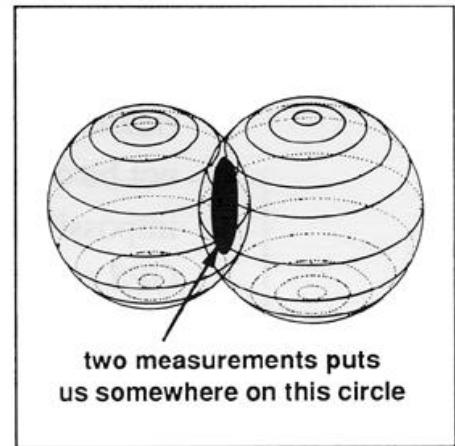
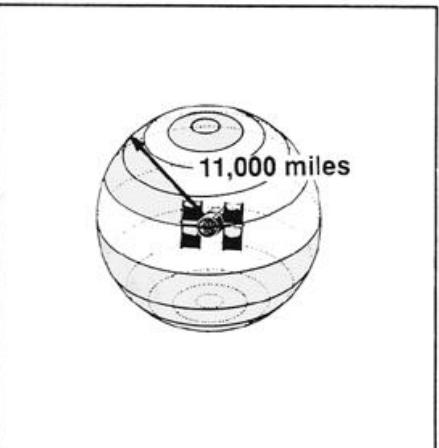
Pseudo-code from
satellite



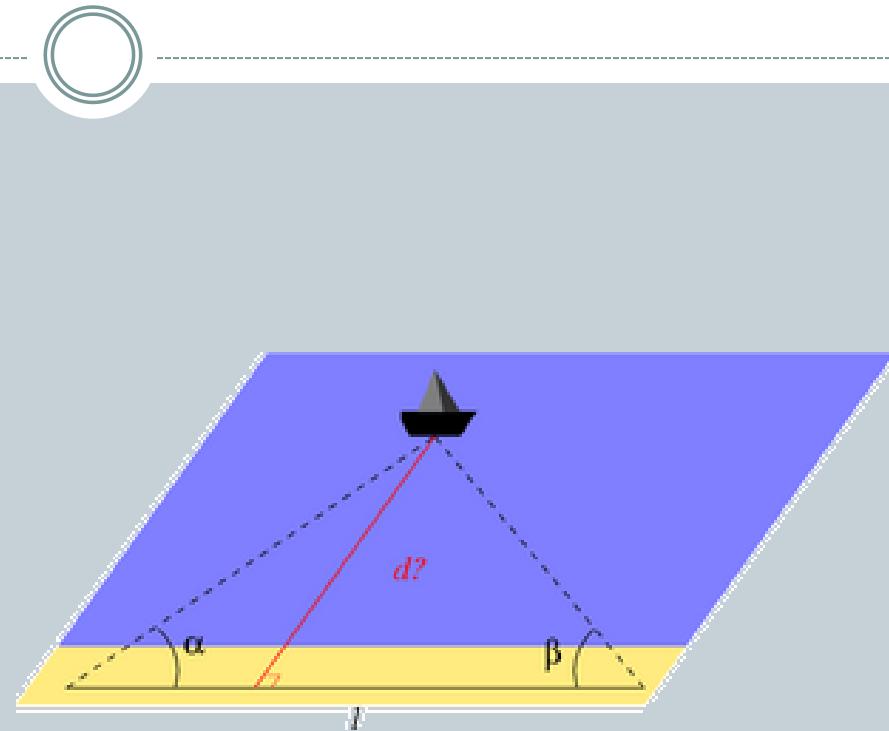
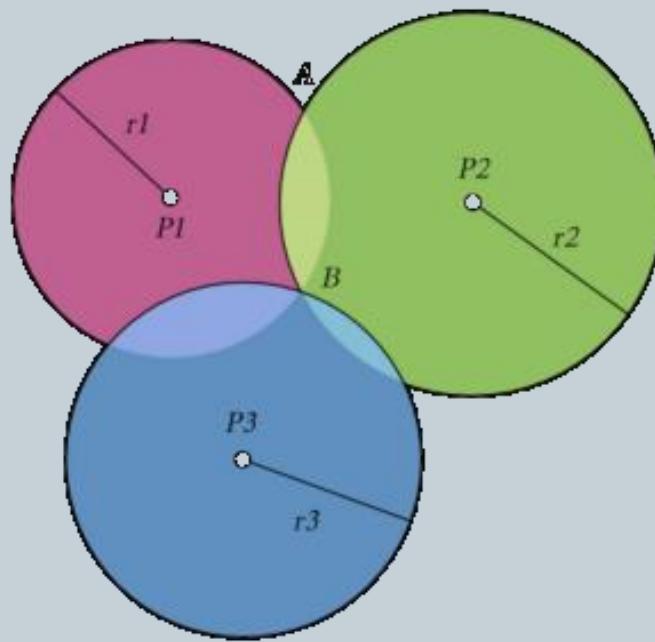
Locating Ourselves



- Knowing distance to one satellite locates us on a sphere
- Two satellites locates us on a circle
 - Three satellites locates us to one of two points, one of which can be easily determined as correct



Trilateration vs. Triangulation



[Wikipedia.com](https://en.wikipedia.org)

GPS uses trilateration, process is commonly described as triangulation because is easier to conceptualize

Field Data



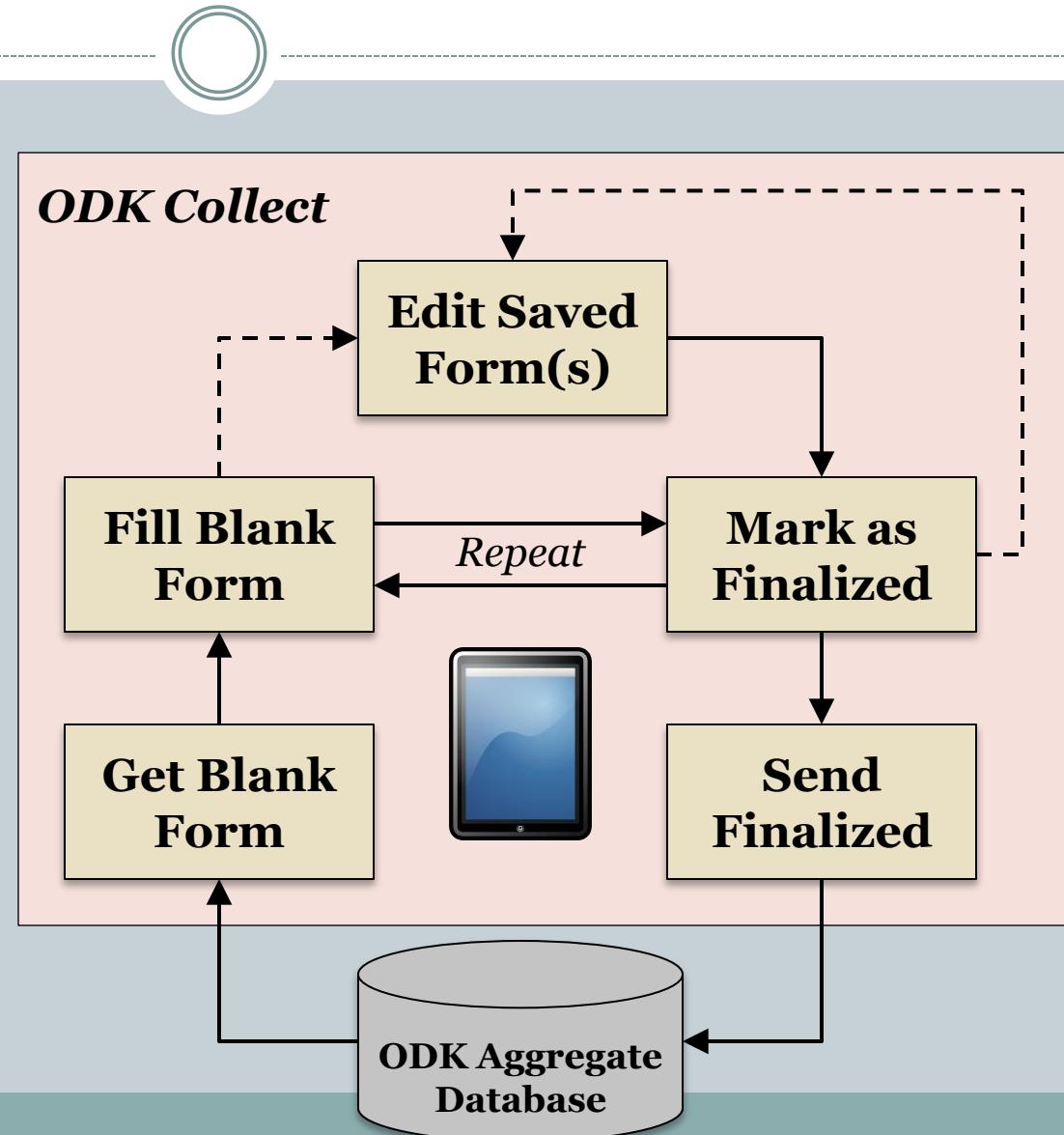
RUN AWAY!!!



"Marvelously
zany humor."
— NEWSWEEK

Data Collection Platforms

- Collect field data **offline** from an Android mobile device (**ODK Collect**)
- Upload field data, whenever convenient, to any enterprise database (**ODK Aggregate**)
- Used by Jane Goodall, The Carter Center



Example Data Collection Tools

ODK Collect > Main Menu

ODK Collect 1.2(1012)
Data collection made easier...

Fill Blank Form

Edit Saved Form

Send Finalized Form

Get Blank Form

Delete Saved Form

ODK Collect > Birds

observation (1)

What bird did you see?

Some birds have included audio or video

Golden eagle

Kingfisher

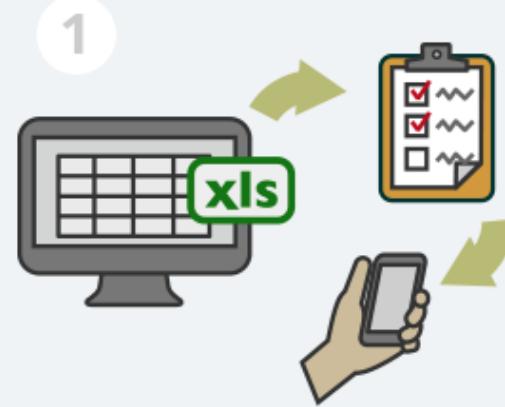
ODK Collect > Forest Plot Survey

Plot type

Other Data Collection Solutions

formhub

provides a hosted platform for users to upload and manage their XForms as **XLSForms**. Users can create data collection forms for Open Data Kit or Enketo easily in Microsoft Excel.



Design

Author surveys quickly and easily in Excel and have instant access on your Android phone. No advanced technical degree required.



Collect

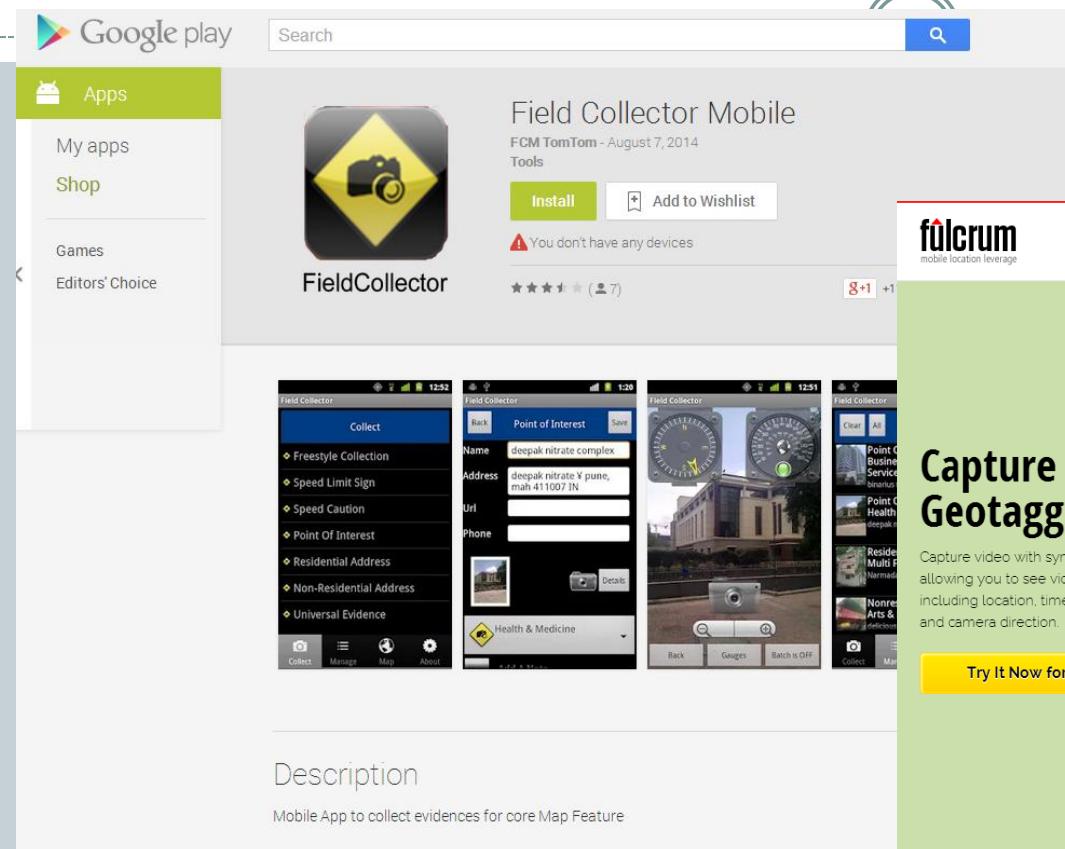
Easily distribute your survey on an Android device or on the web.
Data connection not needed.



Analyze

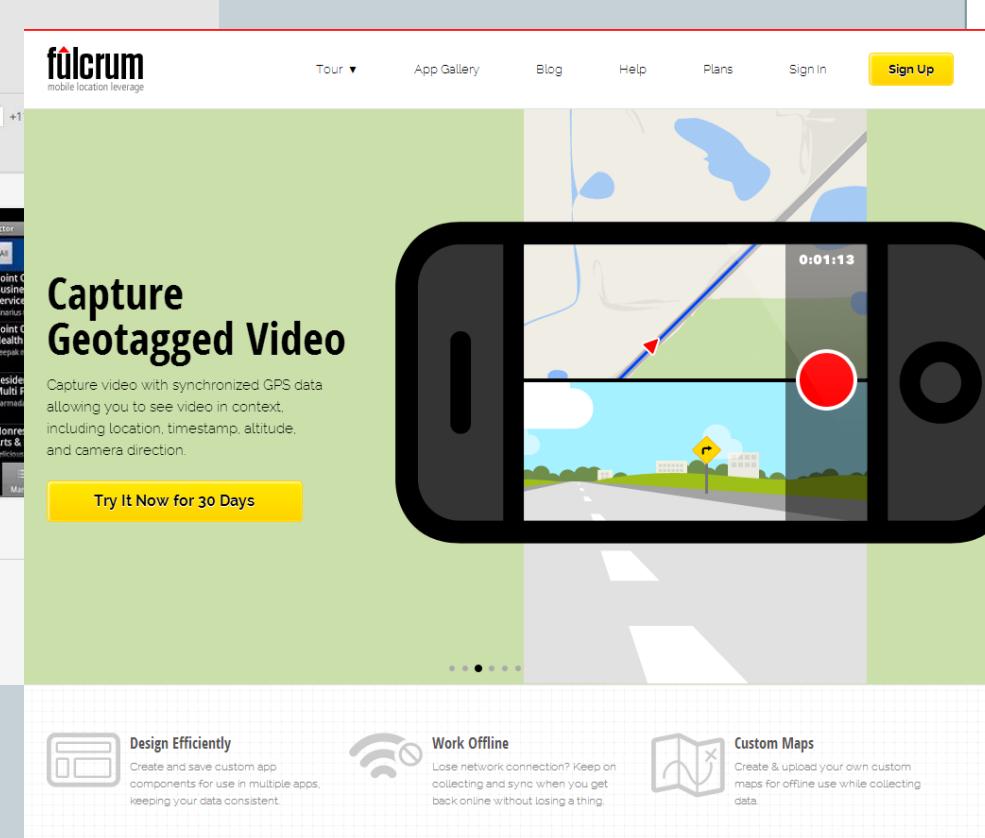
Visualize your data as it is collected, and gain understanding with our powerful insights tools.

Other Apps and Tools



The screenshot shows the Google Play Store interface. On the left, there's a sidebar with 'Google play' at the top, followed by 'Apps' (selected), 'My apps', 'Shop', 'Games', and 'Editors' Choice'. The main area displays the 'Field Collector Mobile' app page. At the top, there's a search bar with a magnifying glass icon. Below it, the app's icon (a yellow diamond with a camera) is shown, followed by the app name 'Field Collector Mobile' and its developer 'FCM TomTom'. A date 'August 7, 2014' and category 'Tools' are also visible. There are two buttons: 'Install' (green) and 'Add to Wishlist' (white). A message 'You don't have any devices' is displayed. Below these are the app's rating (4 stars from 7 reviews) and a 'Get' button.

Field Collector Mobile
FCM TomTom - August 7, 2014
Tools
Install Add to Wishlist
You don't have any devices
4 ★★★★ (7)
Get



The screenshot shows the fulcrum website. At the top, there's a navigation bar with 'Tour ▾', 'App Gallery', 'Blog', 'Help', 'Plans', 'Sign In', and a 'Sign Up' button. Below the navigation, there's a large green section with the heading 'Capture Geotagged Video'. It includes a subtext: 'Capture video with synchronized GPS data allowing you to see video in context, including location, timestamp, altitude, and camera direction.' A yellow 'Try It Now for 30 Days' button is located below this text. To the right, there's a large image of a smartphone displaying a video feed with a map overlay, showing a road and buildings. Below this image, there are three circular dots indicating more content. At the bottom, there are three sections: 'Design Efficiently' (with an icon of a document and a gear), 'Work Offline' (with an icon of a signal tower), and 'Custom Maps' (with an icon of a map).

fulcrum
mobile location leverage
Tour ▾ App Gallery Blog Help Plans Sign In Sign Up
Capture Geotagged Video
Capture video with synchronized GPS data allowing you to see video in context, including location, timestamp, altitude, and camera direction.
Try It Now for 30 Days

Design Efficiently
Create and save custom app components for use in multiple apps, keeping your data consistent.

Work Offline
Lose network connection? Keep on collecting and sync when you get back online without losing a thing.

Custom Maps
Create & upload your own custom maps for offline use while collecting data.

- [Environmental Data Collection Clearinghouse](#)

Motor City Mapping

https://www.motorcitymapping.org/#t=overview&s=detroit&f=all

MOTOR CITY MAPPING About Contribute Training Sign in

DETROIT

Overview Explore Blexts Search

All Publicly Owned Privately Owned

375,838 properties surveyed

Structures vs Lots Structure Condition Structure Occupancy

Type	Count
structures	262,073
lots	113,758

Condition	Count
good	218,718
fair	30,129
poor	8,603
sugg. demo	4,423

Occupancy	Count
occupied	202,548
maybe	10,837
unoccupied	48,468

Lots with dumping: 6,121 Structures with fire damage: 6,808 Structures needing boarding: 27,702

https://play.google.com/store/apps/details?id=com.makeloveland.blexting

Google play

Search

Apps My apps Shop Games Editors' Choice

Blexting Loveland Technologies - August 10, 2014 Communication

Install Add to Wishlist

You don't have any devices

★★★ (1)

Recommend this on Google

Examples from my work

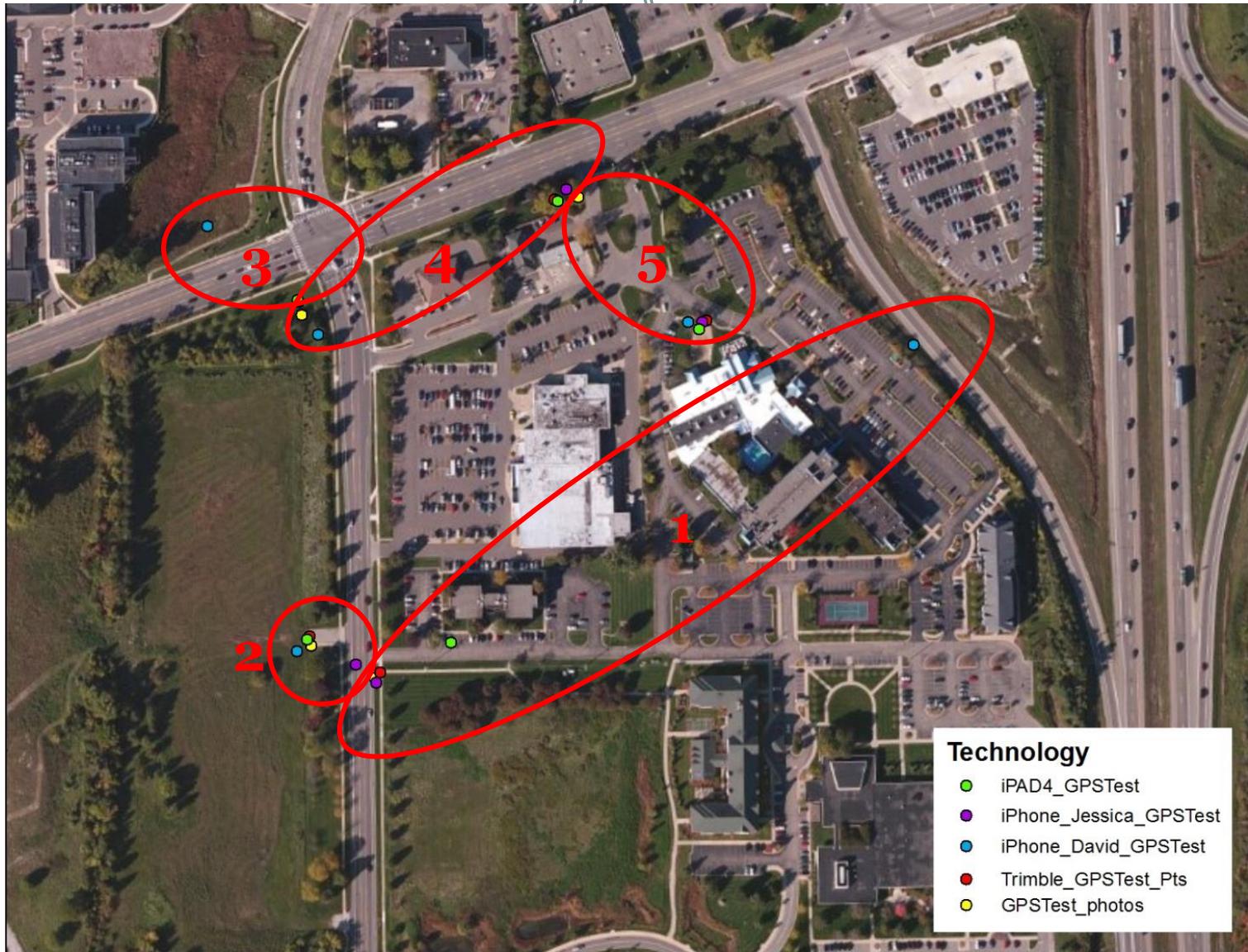


- http://mtri.org/food_security.html
- Mobile and web apps for crop condition monitoring in sub-Saharan Africa
- iPhone and Android



F

GPS Test 1



GPS Test 2



FormEntry



Iphone 4s

FormEntry App

No Wifi, No Data, No
Bluetooth

Voice and Text Only*

Trimble GPS

GeoExplorer
2008 Series

Sub-meter

<i>Latitude</i>	<i>Longitude</i>	<i>Latitude</i>	<i>Longitude</i>	<i>Site #</i>
42.30274652	-83.69259364	42.302807	-83.692511	1
42.30294525	-83.69303134	42.302995	-83.693003	2
42.30472754	-83.69306571	42.304705	-83.693075	3
42.30523108	-83.69109839	42.305243	-83.691265	4
42.30458479	-83.69027998	42.304608	-83.690262	5

Distance between collects (m ; Straight Line)
9.67
5.91
2.71
13.79
3.06

* All field collects with FormEntry app should be
done by standing 10 m inside of each field

Mean 7.03
Median 5.91

FormEntry



- Lat/Long precision

- Multi-scale models require 5 levels of significance (decimal degrees)

Site #	Latitude Precision Agreement (Decimal Place)	Longitude Precision Agreement (Decimal Place)
1	3	4
2	4	4
3	4	4
4	4	3
5	3	4

Mean	3.6	3.8
Median	4	4
Mode	4	4