

BIG DATA ANALYTICS FOR EARTH OBSERVATION – WHAT ARE THE OPPORTUNITIES?

Jon Blower

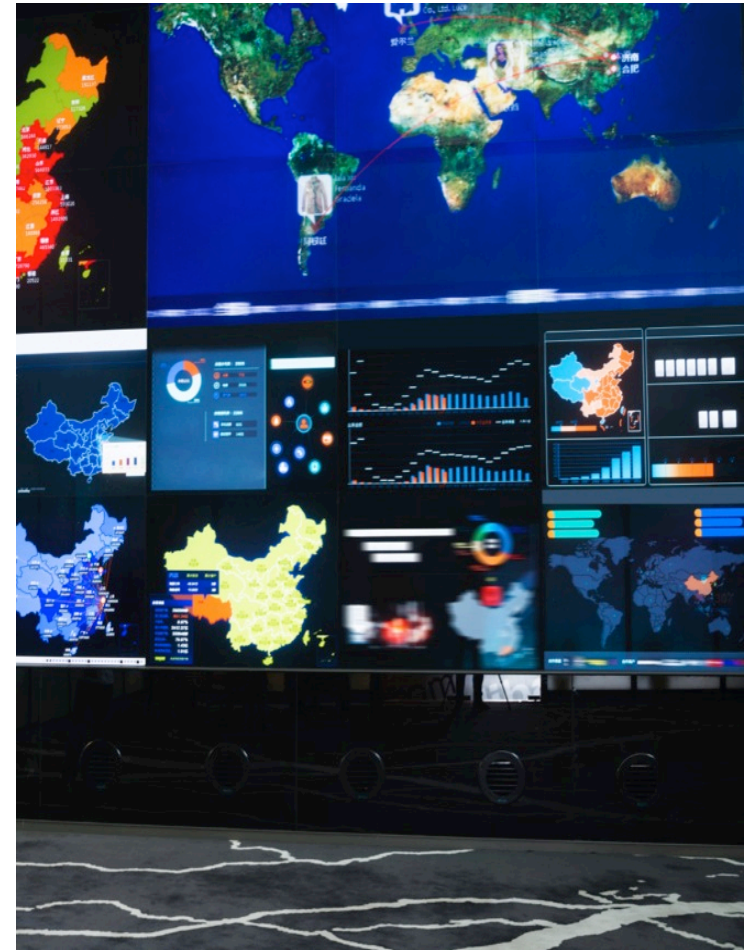
CTO

Institute for Environmental Analytics

THE INSTITUTE FOR ENVIRONMENTAL ANALYTICS



- Centre for big data analytics focused on environmental challenges of many kinds
 - Hosted at University of Reading, UK
- Skills in remote sensing, high-performance computing, mathematical and scientific expertise in range of areas
- Driven by user challenges, most of whom are not “traditional” EO users
- Portfolio of early-stage “demonstrator” projects to test new ideas
- Supported by world-leading scientific research and a broad partnership





DELIVERING VALUE FROM BIG DATA

I THINK "BIG DATA" MEANS...

A collection of technological and cultural shifts that:

- overcome some key limits to innovation
- make us part of a much bigger community

OPPORTUNITY 1: HOSTED PROCESSING

HOSTED PROCESSING

- EO data becoming too big to move around
- Increasingly need to do processing “next to the data”
- Many “side benefits” of doing this:
 - Reduces need for large institutional resources
 - Creates environments suitable for collaboration
 - Potential to offer user-customised products

ESA THEMATIC EXPLOITATION PLATFORMS



→ A SHARED VIRTUAL ENVIRONMENT FOR FINDING AND USING EARTH OBSERVATION DATA – LEARN MORE



TEP Communities

The logo for the geohazards TEP community, featuring an orange square icon with a white topographic map symbol, the text "geohazards" in white, and "tep" in a smaller font below it.

geohazards
tep

The logo for the coastal TEP community, featuring a blue square icon with a white wave symbol, the text "coastal" in white, and "tep" in a smaller font below it.

coastal
tep

The logo for the forestry TEP community, featuring a green square icon with a white tree symbol, the text "forestry" in white, and "tep" in a smaller font below it.

forestry
tep

The logo for the hydrology TEP community, featuring a blue square icon with a white water drop symbol, the text "hydrology" in white, and "tep" in a smaller font below it.

hydrology
tep

The logo for the polar TEP community, featuring a blue square icon with a white snowflake symbol, the text "polar" in white, and "tep" in a smaller font below it.

polar
tep

The logo for the urban TEP community, featuring a red square icon with a white bar chart symbol, the text "urban" in white, and "tep" in a smaller font below it.

urban
tep

PROBA-V Mission Exploitation Platform



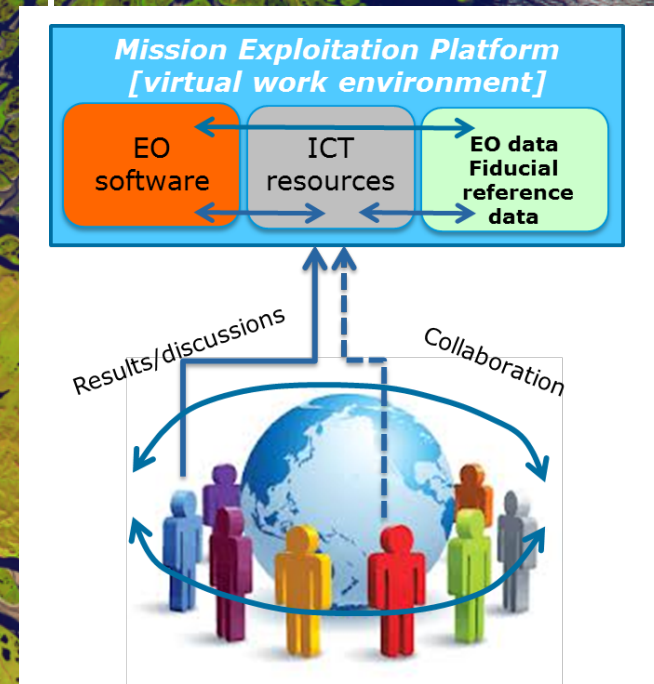
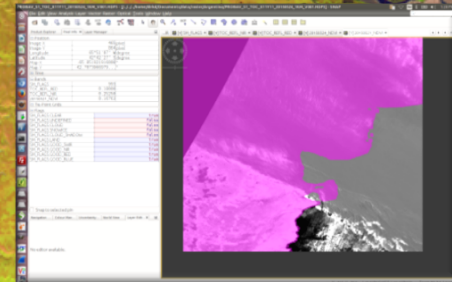
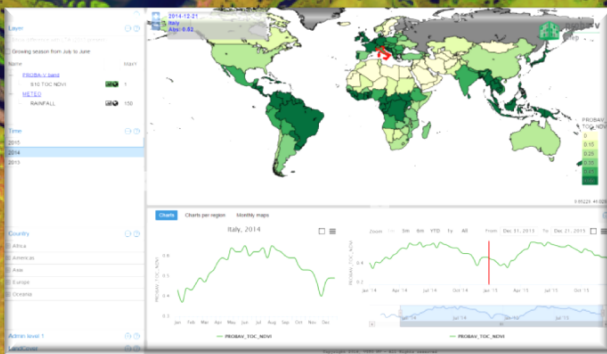
proba-v
mep

<http://proba-v-mep.esa.int/>

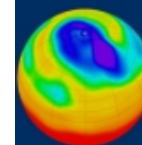
Contact: erwin.goor@vito.be



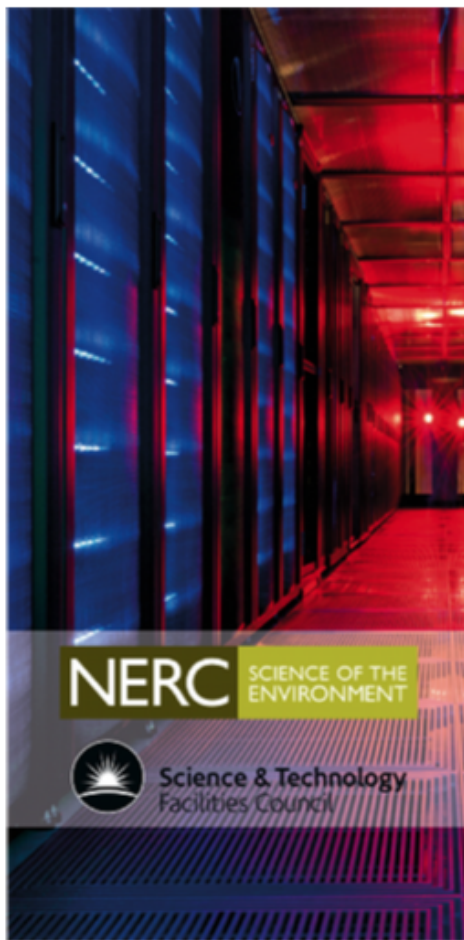
- ❑ Full-resolution viewing and Time Series exploration
- ❑ On-demand processing services for derived products
- ❑ Data analytics on time series
- ❑ R&D platform for application developers



THE JASMIN DATA CLOUD



Centre for Environmental
Data Archival
SCIENCE AND TECHNOLOGY FACILITIES COUNCIL
NATURAL ENVIRONMENT RESEARCH COUNCIL



- Part of UK's “collaborative ground segment”
- In-place access to Sentinel and other environmental data (weather, climate, ...)



Sentinel-2 on AWS

Sentinel-2 data is available for anyone to use via Amazon S3.

About the data

Data structure

Browse through data

Accessing the Data

Featured uses

Contact us

Sentinel-2 data is available for anyone via Amazon S3, either over Internet or within AWS. All Sentinel-2 scenes are made available, often within hours of production.

Earth observation data provided by the [Sentinel-2](#) satellites are revolutionizing the market of space applications. Free, full and open access to data with very short revisit times, high spatial resolution, and good spectral resolution can benefit several sectors - agriculture, environmental and land-change monitoring, natural disaster response, insurance and others.

The Sentinel-2 mission is a land monitoring constellation of two satellites (Sentinel-2A was launched on 23 June 2015 and Sentinel-2B will follow in the second half of 2016) that provide high resolution optical imagery and provide continuity for the current SPOT and Landsat missions. The mission will provide a global coverage of the Earth's land surface every 10 days with one satellite (and 5 days with 2 satellites), making the data of great use in on-going studies.

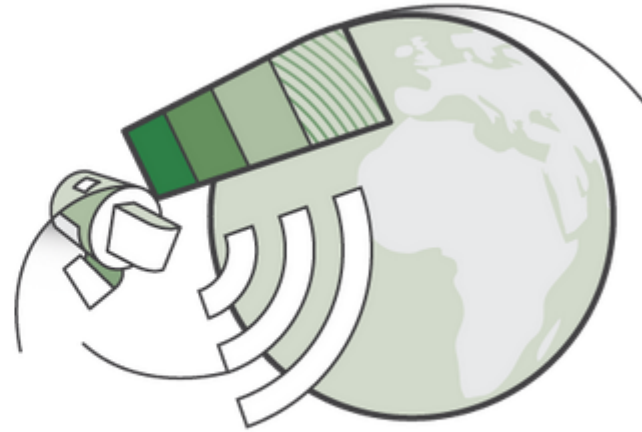
Landsat on AWS

Landsat 8 data is available for anyone to use via Amazon S3. All Landsat 8 scenes from 2015 are available along with a selection of cloud-free scenes from 2013 and 2014. All new Landsat 8 scenes are made available each day, often within hours of production.

The [Landsat](#) program is a joint effort of the [U.S. Geological Survey](#) and [NASA](#). First launched in 1972, the Landsat series of satellites has produced the longest, continuous record of Earth's land surface as seen from space. NASA is in charge of developing remote-sensing instruments and spacecraft, launching the satellites, and validating their performance. USGS develops the associated ground systems, then takes ownership and operates the satellites, as well as managing data reception, archiving, and distribution. Since late 2008, Landsat data have been made available to all users free of charge.



<https://aws.amazon.com/public-data-sets/landsat/>



Google Earth Engine Playground interface showing a script, its execution results, and a map visualization.

Scripts | Docs | Playground.js | Get Link | Save | Run | Reset

```
1 Imports (5 entries)
2 // load the most recent MODIS composite
3 var modis = ee.Image(imageCollection
4   .sort('system:time_start', false)
5   .first());
6
7 // print metadata to the console
8 print(modis);
9
10 var sld = "\
11   <RasterSymbolizer>\
12   <ContrastEnhancement><Normalize/></ContrastEnhancement>\
13   <ChannelSelection>\
14     <RedChannel>\
15       <SourceChannelName>sur_refl_b01</SourceChannelName>\
16     </RedChannel>\
17     <GreenChannel>\
18       <SourceChannelName>sur_refl_b04</SourceChannelName>
```

Inspector | Console | Tasks

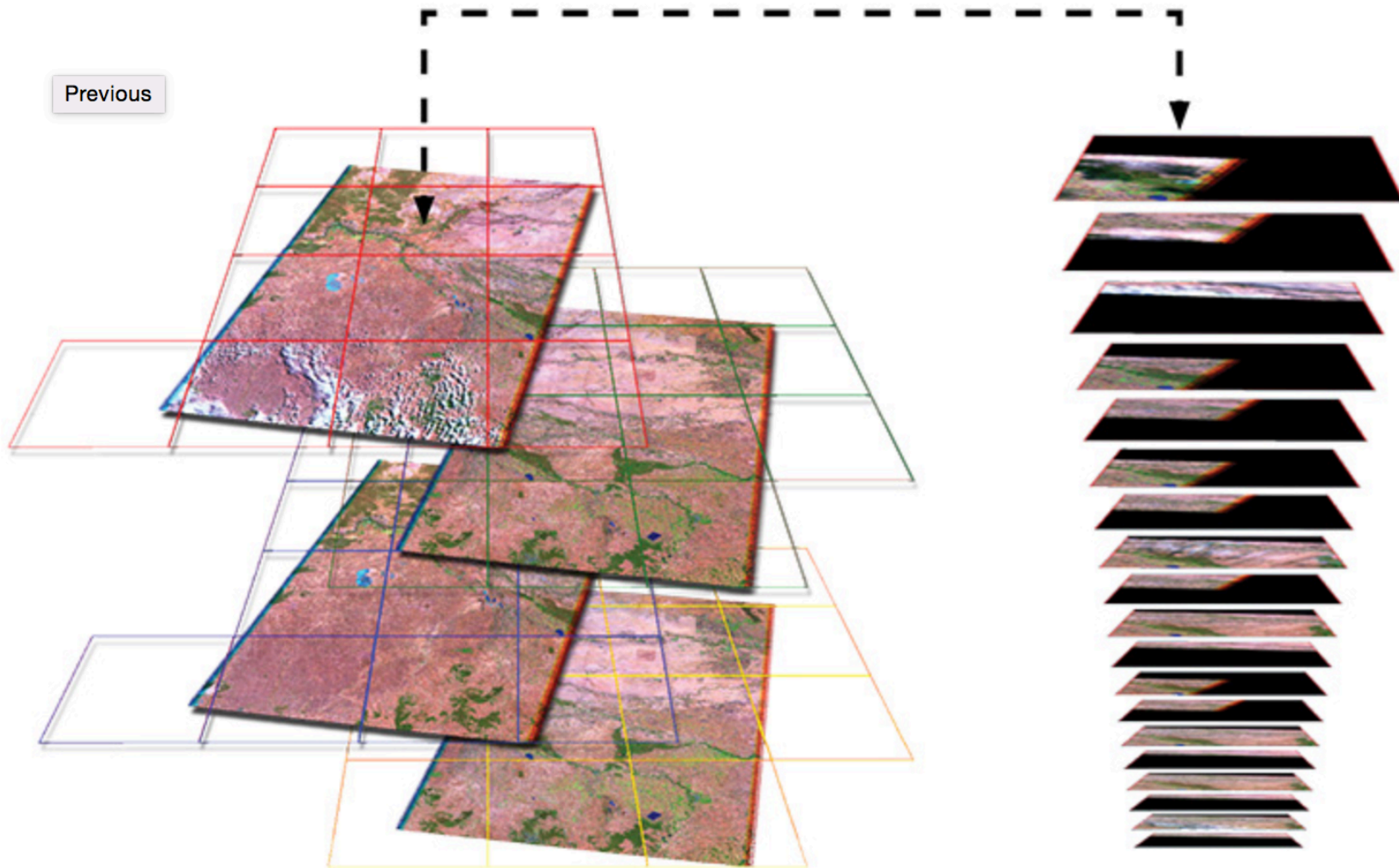
- Point (13.54, 23.56) at 20Km/px
 - Pixels
 - MODIS composite: Image (3 bands)
 - DEM: Image (2 bands)
 - Objects
 - MODIS composite: Image (3 bands)
 - type: Image
 - bands: List (3 elements)
 - properties: Object (5 properties)
 - DEM: Image NOAA/NGDC/ETOP01 (2 bands)
 - type: Image
 - id: NOAA/NGDC/ETOP01
 - version: 1406914481423000
 - bands: List (2 elements)
 - properties: Object (17 properties)

Map visualization showing a global view with a dark blue overlay. The map includes a scale bar (0-1000 km) and a "Layers" panel.

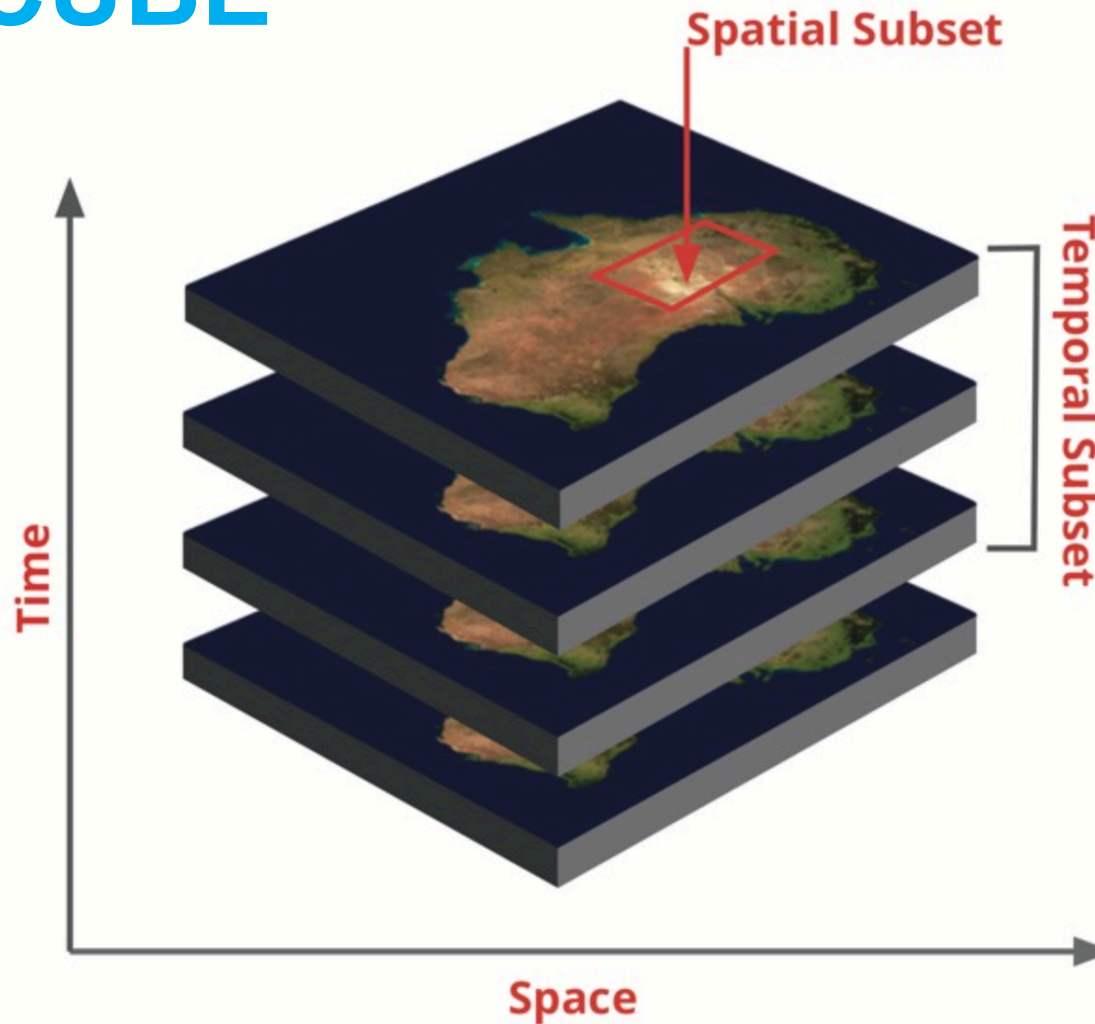
DELIVERING VALUE FROM BIG DATA

AUSTRALIAN GEOSCIENCE DATA CUBE

<http://www.datacube.org.au>



AUSTRALIAN GEOSCIENCE DATA CUBE



<http://www.envsys.co.uk/index.php/projects/data-cubes-australian-agriculture/>

AUSTRALIAN GEOSCIENCE DATA CUBE



Home > Hazards > Flood > Portal > Floods Public Search >

Australian Government Geoscience Australia **Floods Public Search** Quick Links

Water Observations from Space

OR

By selecting a point:

Legend

- Water not detected.
- Water detected in 1% of observations (includes flooding and misclassified shadows).
- Water detected in 5% of observations (includes intermittent water bodies).
- Water detected in 20% of observations (includes water bodies that often dry out).
- Water detected in 50% of observations.
- Water detected in 80% of observations (permanent water bodies).
- Water detected always.

A satellite map of Australia showing water detection data. The map is overlaid with a color-coded grid representing the percentage of observations where water was detected. The colors range from red (1% detection) to blue (80% detection). Major water bodies and rivers are clearly visible, with higher detection rates (yellow, green, blue) concentrated in these areas. National parks are labeled on the map, including Stirling Range National Park, Grampians National Park, Ashburton Downs National Park, and Wetland National Park. A scale bar indicates 20 km.

Copyright Disclaimer Privacy Accessibility Information Publication Scheme Freedom of Information Floods Public Search Feedback

DELIVERING VALUE FROM BIG DATA

OPPORTUNITY 2: IMPROVED ALGORITHMS

IMPROVED RETRIEVAL ALGORITHMS

- Computational resources have often limited us to simple algorithms
- Images often processed independently of each other
- Access to more powerful compute/data resources and processing frameworks can change this:
 - Consider whole timeseries of images simultaneously
 - Avoid need for data reduction “pre-processing” steps
 - Preserve more information in processing chain (e.g. uncertainty)
 - Combine datasets more rigorously and integrate models and observations (data assimilation)

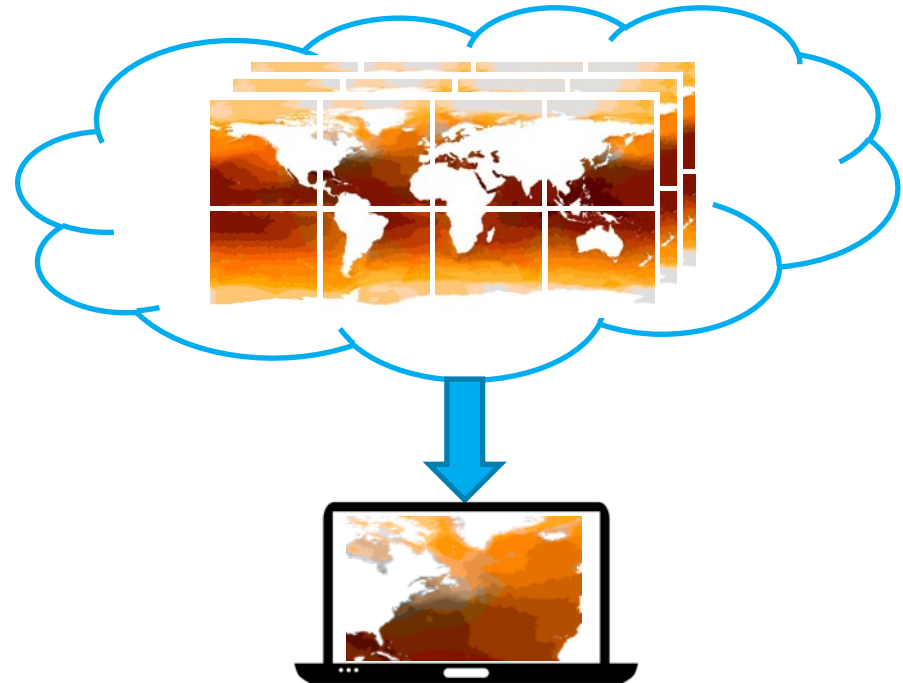
IMPROVED DATA ANALYSIS ALGORITHMS

- Currently we program at quite a low level
 - We tell the computer step by step how to solve a problem
 - We understand how to program “small data” problems
 - e.g. those that fit into typical RAM
- New software libraries and frameworks are being developed to produce “scalable” algorithms that work on very large data sets
 - Hadoop, Spark, etc.
- But ideally we want algorithms that work on both “small data” and “big data” transparently

EXAMPLE: BIG DATA ANALYTICS USING DASK

- Data infrastructure provides “data cube” as a set of files
- Write simple analysis script in Python
 - Use Dask to treat files as single virtual dataset
 - Dask automatically accesses and subsets *only the required files*
- Scales up and down transparently

Calculate mean sea surface temperature over certain region:



```
dataset = getDataset("whatever")  
sst = dataset["sst"]  
result = da.mean(sst[0, :450, :]).compute()
```


OPPORTUNITY 3: A MORE "DATA-HUNGRY" CULTURE

A “DATA-HUNGRY” CULTURE

- Increased demand for data from government and industry
- Increased publication of data, e.g. government Open Data
- Growth of “data science” as a discipline and career

- Rise of “hackathons”, apps and more “unofficial” uses of data
- Appearance of data from “unusual” sources

- Great opportunities for creativity and innovation!

England Makes 3D Data of the Entire Country Free After Minecrafters Ask For It

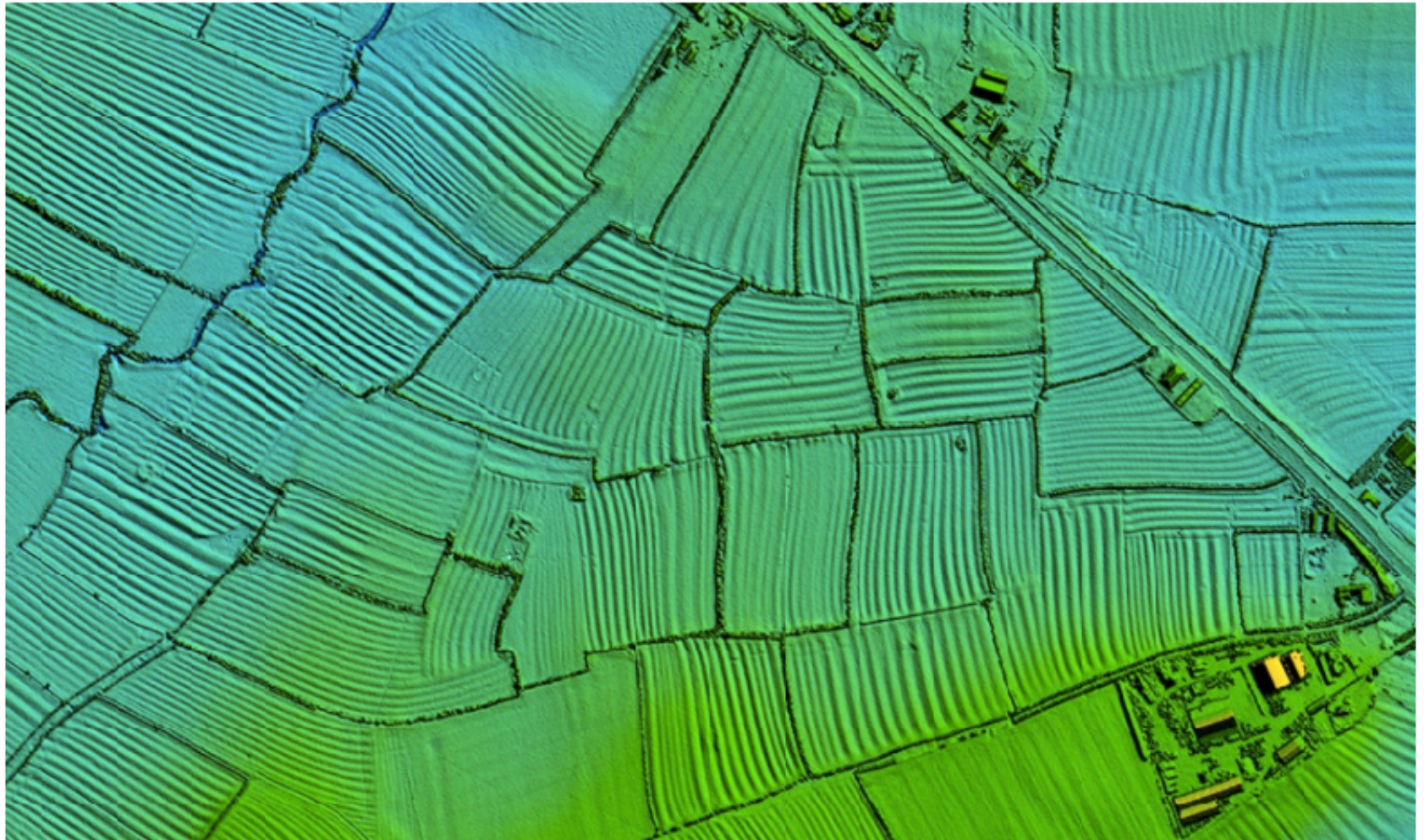
gizmodo.com

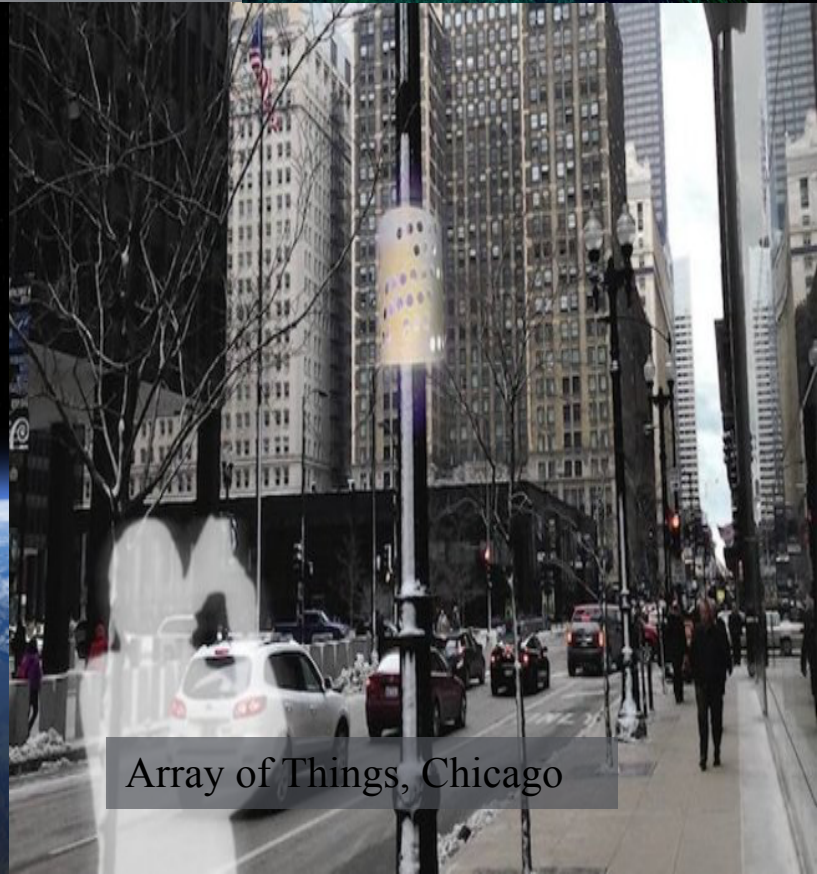
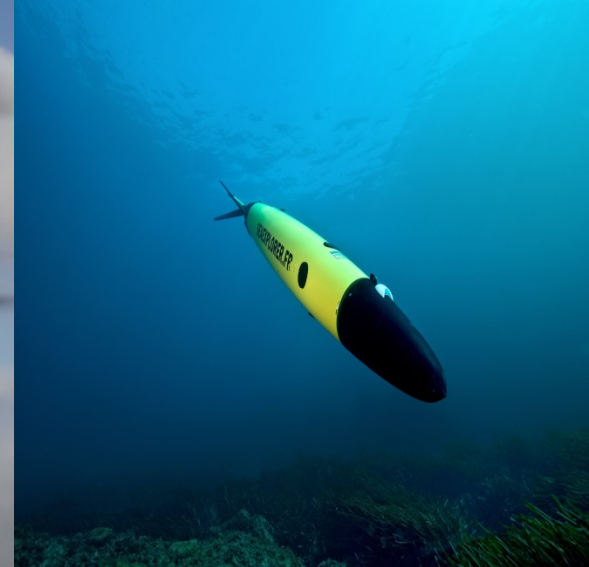


[Kelsey Campbell-Dollaghan](#)

Filed to: FREE LASERS 10/14/15 2:40pm

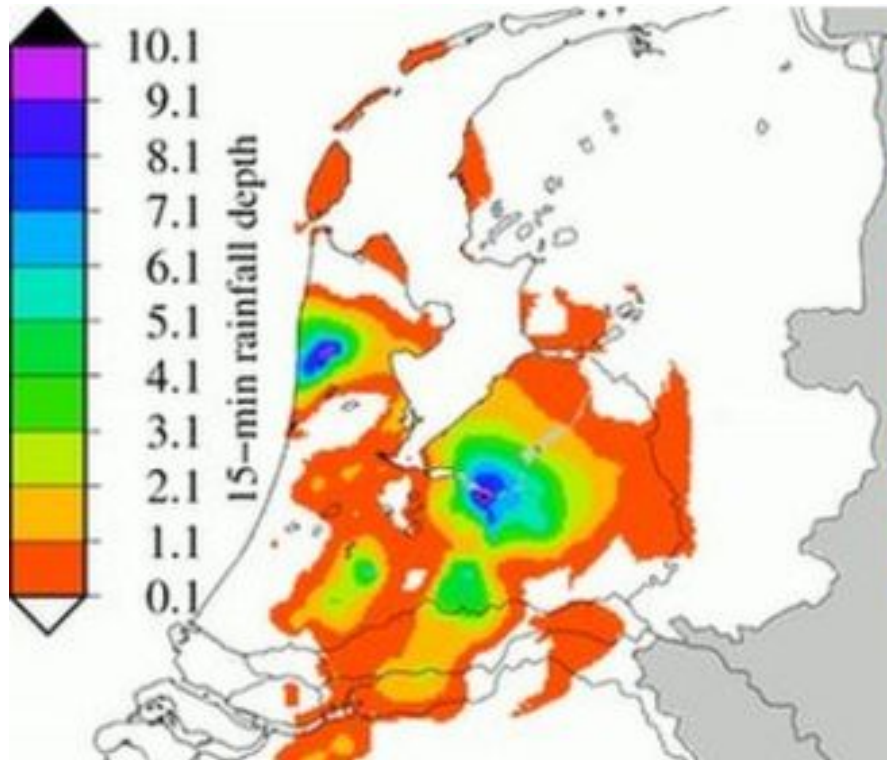
26,169 🔥 7 ☆ ▾



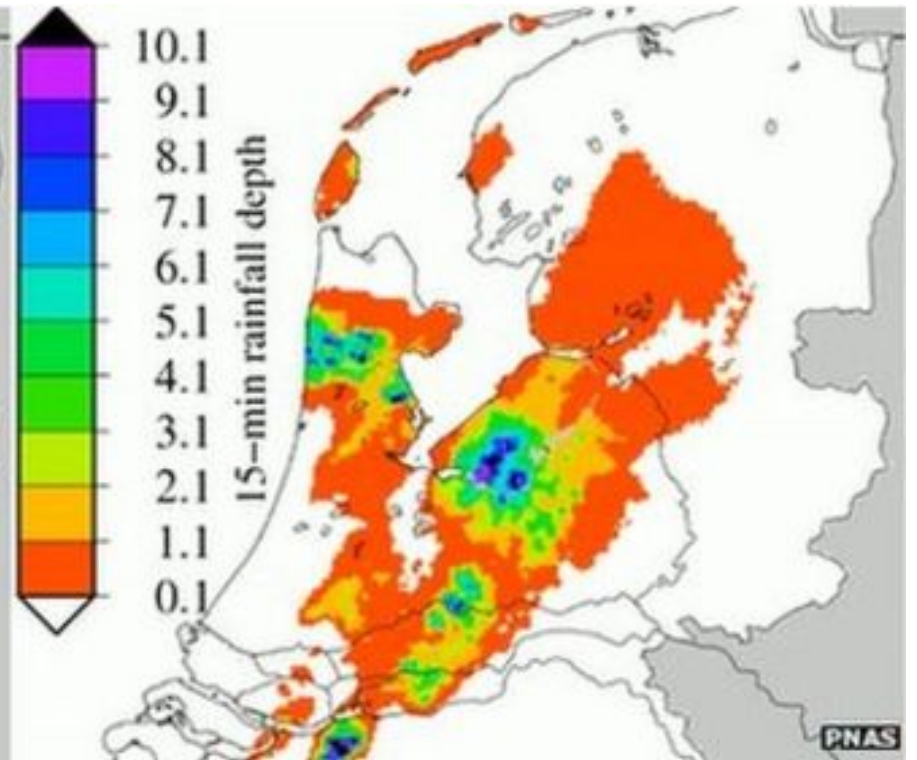


Array of Things, Chicago

RAIN DETECTION USING MOBILE PHONE SIGNAL STRENGTH DATA



mobile phone data

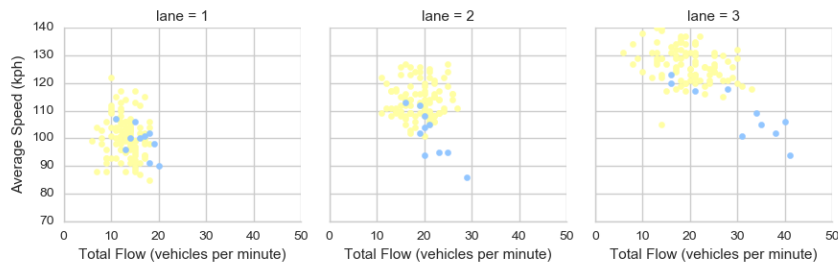
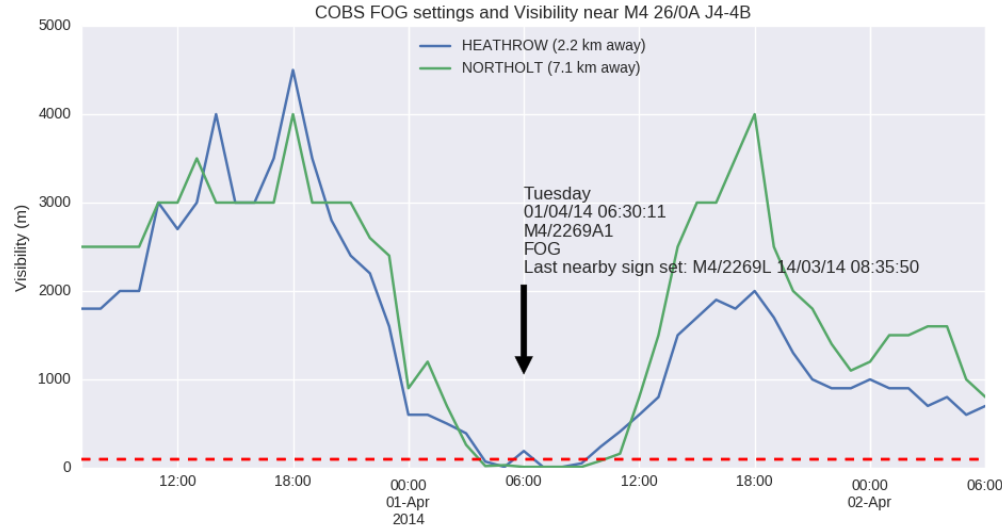
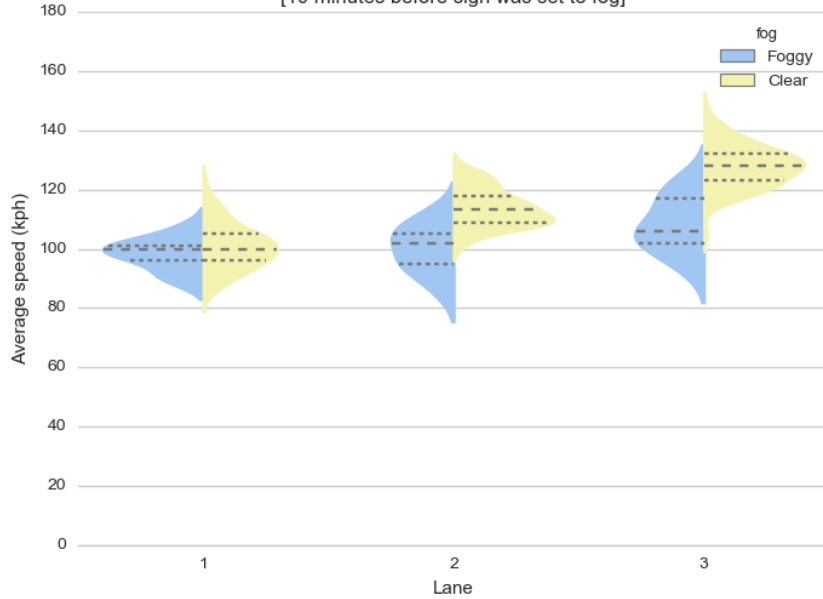


radar + gauges

Overeem et al
doi:10.1073/pnas.1217961110

FOG DETECTION USING TRAFFIC DATA

M42_6277B (2013-09-25 07:02:28 to 2013-09-25 07:12:28)
[10 minutes before sign was set to fog]



OPPORTUNITY 4: THE WEB AS A GLOBAL ENABLER

THE WEB AS A GLOBAL ENABLER

- “Data hungry” culture is driving innovation in many fields
- The Web is the obvious common platform for sharing and collaboration
- EO data doesn’t always “play nicely” with the wider Web
 - Data behind specialist portals
 - Data formats are binary, complex and unfamiliar to most
- EO data is “**on** the web” but not really “**part of** the web”
- We can (and must!) break EO data out of its silo



About 65,100,000 results (0.36 seconds)

Showtimes for The Martian

Oct 14 Thu, Oct 15

All times	Morning	Afternoon	Evening	Night			
Regal Gallery Place Stadium 14 - Map							
Standard	1:00pm	4:30	6:00	8:00			
3D	11:30am	12:30pm	2:45	3:45	7:00	9:25	10:20
AMC Loews Georgetown 14 - Map							
Standard	11:00am	2:15pm					
3D	12:00pm	1:00	3:15	4:15	6:30	7:30	9:45
	10:40						
AMC Courthouse Plaza 8 - Map							
Standard	4:45pm	9:00					
3D	12:30pm	1:30	3:45	7:00	8:00	10:15	


All times are in ET

More showtimes

The Martian

2015 film

8.4/10 · IMDb
93% · Rotten Tomatoes
81% · Metacritic



When astronauts blast off from the planet Mars, they leave behind Mark Watney (Matt Damon), presumed dead after a fierce storm. With only a meager amount of supplies, the stranded visitor must utilize his wits and spirit to find a way to survive on the hostile planet. Meanwhile, back on Earth, membe... [More](#)

Release date: October 2, 2015 (USA)
Director: Ridley Scott
Running time: 2h 22m
MPAA rating: PG-13
Adapted from: The Martian

Available on

YouTube From \$14.99

The Martian (2015) - IMDb
www.imdb.com/title/tt3659388/ Internet Movie Database
★ ★ ★ ★ Rating: 8.3/10 - 79,977 votes
Directed by Ridley Scott. With Matt Damon, Jessica Chastain, Kristen Wiig, Kate Mara. During a manned mission to Mars, Astronaut Mark Watney is presumed ...
[Full Cast & Crew](#) - [Kate Mara](#) - [Ridley Scott](#) - [Poster](#)

The Martian (film) - Wikipedia, the free encyclopedia
[https://en.wikipedia.org/wiki/The_Martian_\(film\)](https://en.wikipedia.org/wiki/The_Martian_(film)) Wikipedia
Plot. When the Ares III manned mission to Mars is hit by an intense Martian storm, astronaut Mark Watney is lost and presumed dead. With the lives of her crew at ...
[The Martian \(Weir novel\)](#) - [Michael Peña](#) - [Naomi Scott](#) - [Drew Goddard](#)

The Martian (Weir novel) - Wikipedia, the free encyclopedia
[https://en.wikipedia.org/wiki/The_Martian_\(Weir_novel\)](https://en.wikipedia.org/wiki/The_Martian_(Weir_novel)) Wikipedia
The Martian is a 2011 science fiction novel and the first published novel by American author Andy Weir. It was originally self-published in 2011 after which ...
Author: Andy Weir **Publisher:** Crown Publishing Group
Pages: 369 **Cover artist:** Eric White

In the news

Box office: 'Pan' bombs, 'Martian' stays No. 1
CNN - 5 hours ago
... loss for Warner Bros. after opening to \$15.5 million. "Martian" once again topped box office.

US box office: The Martian soars into blockbuster territory as Pan plummets
The Guardian - 3 hours ago

'Pan' lacks magic at the box office; 'The Martian' soars
Fox News - 18 hours ago

[More news for the martian](#)

Critic reviews View 6+ more






An enthralling and rigorously realistic outer-space survival story in which Matt Damon plays a NASA botanist stranded on the Red Planet after a sandstorm forces his crewmates to abort mission. Full review
Peter Debruge · [Variety](#)

You won't find a space epic that's more fun to geek out at than The Martian. Full review
Peter Travers · [Rolling Stone](#)


Anchored by another great turn from Matt Damon, The Martian mixes smarts, laughs, weird character bits and tension on a huge canvas. The result is Scott's most purely enjoyable film for ages. Full review
Ian Freer · [Empire](#)

At its heart, The Martian is an unapologetically stirring celebration of our ability, as a species, to solve even the most daunting problems via rational thought, step by step by step. Full review
Mike D'Angelo · [A.V. Club](#)

Cast View 10+ more

 Matt Damon Mark Watney	 Jessica Chastain Melissa Lewis	 Kristen Wiig Annie Montrose	 Kate Mara Beth Johansson	 Michael Peña Rick Martinez
---	---	---	---	---

People also search for View 15+ more



Search for "The Martian", Google shows:

- Facts about the film
- Cast
- Showtimes
- Reviews
- Related films

This is Linked Data in action!

(powered by schema.org vocabulary)

Why can't we do the same for EO data?

[Introduction](#)[Site Structure](#)[Breadcrumbs](#)[Sitelinks Searchbox](#)[Your Site Name in Results](#)[Authorized Presence](#)[Corporate Contacts](#)[Logos](#)[Social Profile Links](#)[Creative Works](#)[Articles](#)[Courses](#)[Music](#)[Recipes](#)[Reviews](#)[TV & Movies](#)[Videos](#)[Commerce](#)[Local Businesses](#)[Events](#)[Products](#)[Science](#)[Science Datasets](#)

Science Datasets

The web contains specialized repositories for datasets in many scientific domains: life sciences, earth sciences, material sciences, and more. Similarly, many governments maintain repositories of civic and government data. However, much of that structured data is not readily available to search engines, which must extract the data from HTML pages in order to provide search services to users. When webmasters provide [structured markup](#), they enable search engines to “understand” this metadata, which in turn improves data discovery, leading scientists to the information they need for their work.

For example, consider this dataset that describes [historical snow levels in the Northern Hemisphere](#). This page contains basic information about the data, like spatial coverage and units. Other pages on the site contain additional metadata: who produces the dataset, how to download it, and the license for using the data. With structured data markup, these pages can be more easily discovered by other scientists searching for climate data in that subject area.



Dataset markup is available for you to experiment with before it's released to general availability.

When you implement the markup, you'll see previews in the Structured Data Testing Tool. You won't, however, see your datasets appear in Search.

Let Google know that you added structured markup to your dataset by providing your details in our partner interest form.

[EXPRESS INTEREST](#)

What qualifies as a dataset?

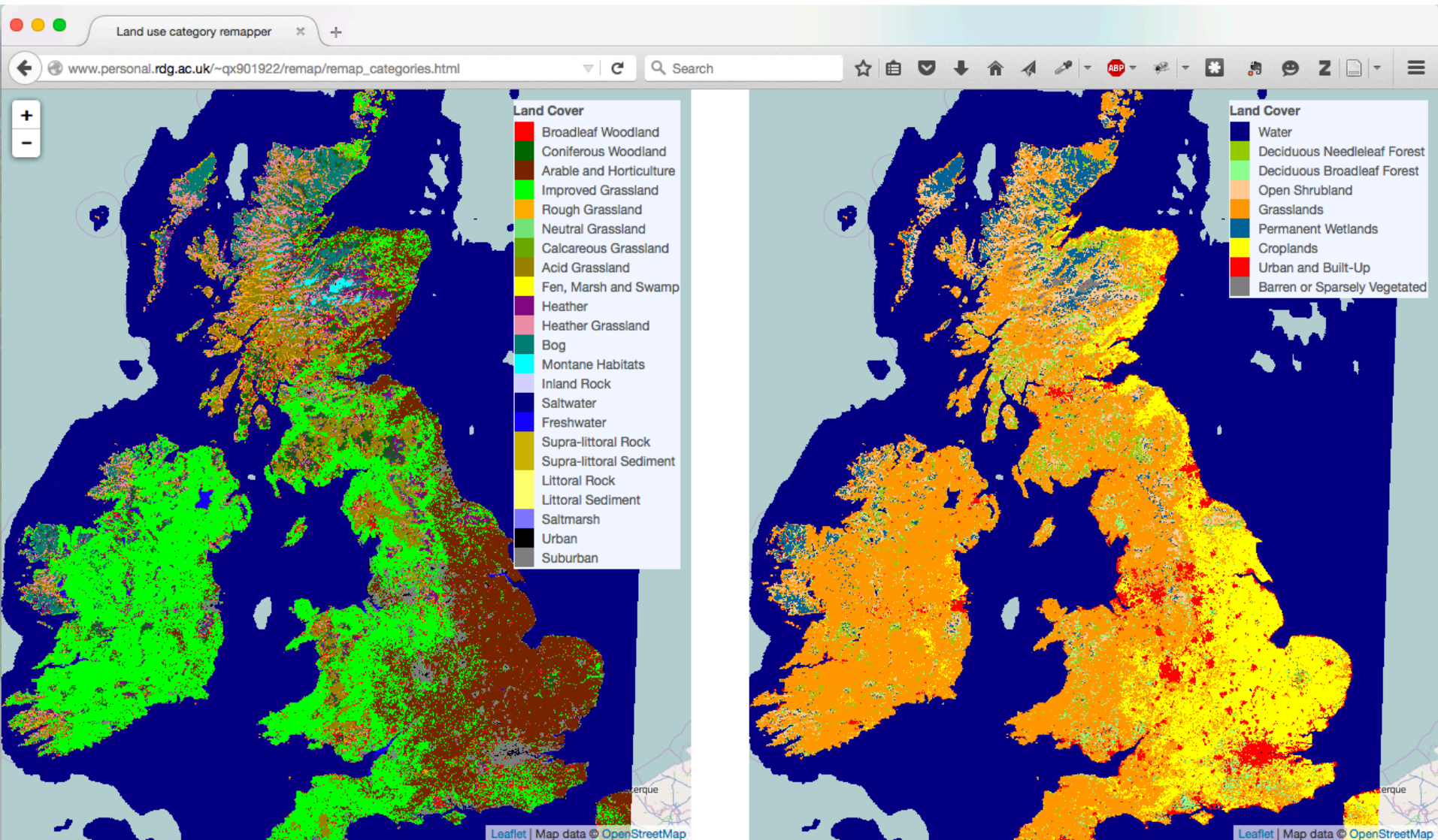
For purposes of inclusion, we take a broad view of what qualifies as a dataset:

Contents

[What qualifies as a dataset?](#)[Mark up your dataset descriptions](#)[Basic dataset properties](#)[Data catalog properties](#)[Download information properties](#)[Temporal coverage](#)[Spatial coverage](#)[Citations and publications](#)[Provenance and license information](#)

Interactive, in-browser reclassification of land cover maps

youtu.be/dxfmTkBdn90



CONCLUSIONS

- What does “Big Data” really mean for the EO community?
 - Access to new data and processing resources
 - Opportunities to create *collaborative* platforms
 - Opportunities to move past limits of “legacy” algorithms
 - Opportunities to find *and empower* new users for EO data...
 - ... and to use the Web more effectively as a means to reach them
- Two “grand challenges” for us all:
 - How do we “demystify” EO data for new users?
 - How can we improve education and training?

Thank you!

j.blower@the-iea.org
[@Jon_Blower](https://twitter.com/Jon_Blower)

www.the-iea.org
[@env_analytics](https://twitter.com/env_analytics)