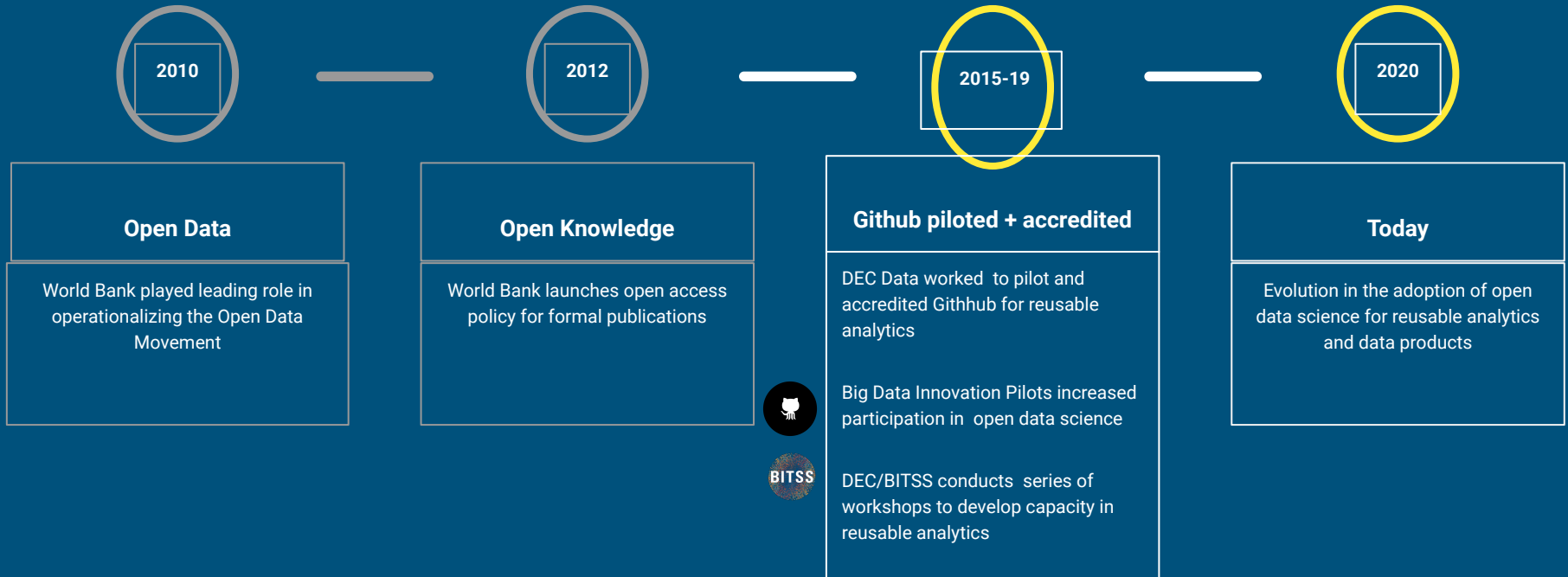


An aerial photograph of the World Bank building, a large modern structure with a distinctive perforated facade. The image is overlaid with a semi-transparent blue filter. A yellow L-shaped graphic element is positioned in the top-left corner, and a blue horizontal line is centered below the title. The title text is in a bold, yellow, sans-serif font.

# Evolution of Open Analytics at the World Bank

Trevor Monroe,  
Analytics & Tools Unit  
Development Data Group, World Bank  
[@trevmon28](#)

# Evolution of Open Data/Analytics at WB



# COVID has accelerated use of Open Tools, Practices

pp...*engineering practices are transforming research*  
- Susan Athey, Stanford



Machine Learning  
Susan Athey, Professor  
Economics, Stanford University

Machine Learning

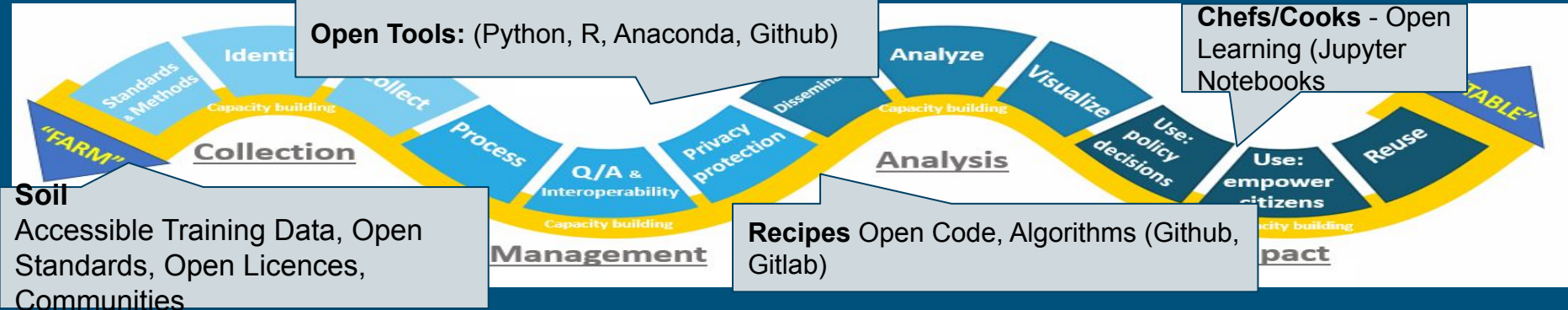
pp...*free data is cheaper than ever*  
- Jed Sundwall, AWS



## Open, Reusable, Collaborative, Iterative Data Products

- *Adoption of open languages, libraries, frameworks* in production of data products, (e.g. R, Python, Pandas, Scikit, Spark, Dask)
- *Use of collaborative code repositories* to manage data products/projects remotely asynchronously (Git, Github, Gitlab)
- *Building on work of open source community, licenses, standards* (MIT, Apache, MZL, COG, STAC, DDI)
- *Delivering data products with code and tutorials* use the products (Github, Jupyter, Binder, Medium)

# “Farm to Table” metaphor for Development Data



Open public commons of assets, tools, resources can amplify Capabilities, better soil, methods and skills throughout the value chain

# WB Mobility Task Force - COVID Analytics



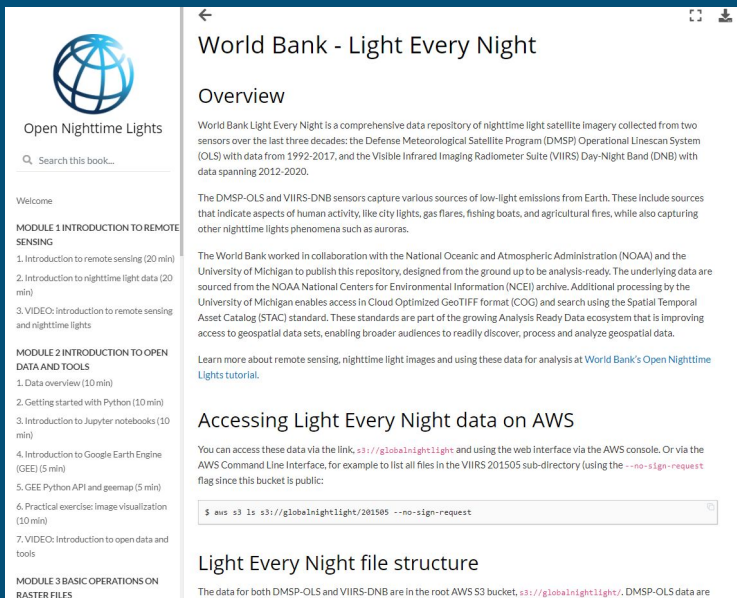
World Bank + Others



## Doing Data Products Differently:

- **Open Algorithms:** WB developed ready made algorithms to process mobile data into measures from open source community (Flowminder)
- **Dynamic Collaboration:** Github repo used to support implementation in collaboration with vendors/academics/MNOs
- **Containerization:** open technologies (hadoop, Spark, Kubernetes, Docker) provisioned on MNO networks for data collection/processing
- **Open Knowledge** - measurement algorithms and training collateral from Country NSO workshops published as public goods
- <https://github.com/worldbank/covid-mobile-data>

# Open Nighttime Lights



The screenshot shows the 'World Bank - Light Every Night' book interface. It features a globe icon and a search bar. The table of contents includes:

- MODULE 1 INTRODUCTION TO REMOTE SENSING
  - 1. Introduction to remote sensing (20 min)
  - 2. Introduction to nighttime light data (20 min)
  - 3. VIDEO: Introduction to remote sensing and nighttime lights
- MODULE 2 INTRODUCTION TO OPEN DATA AND TOOLS
  - 1. Data overview (10 min)
  - 2. Getting started with Python (10 min)
  - 3. Introduction to Jupyter notebooks (10 min)
  - 4. Introduction to Google Earth Engine (GEE) (5 min)
  - 5. GEE Python API and geemap (5 min)
  - 6. Practical exercise: Image visualization (10 min)
  - 7. VIDEO: Introduction to open data and tools
- MODULE 3 BASIC OPERATIONS ON RASTER FILES

The 'Accessing Light Every Night data on AWS' section contains the following text:

You can access these data via the link, `s3://globalnightlights` and using the web interface via the AWS console. Or via the AWS Command Line Interface, for example to list all files in the VIIRS 201505 sub-directory (using the `--no-sign-request` flag since this bucket is public):

```
$ aws s3 ls s3://globalnightlights/201505 --no-sign-request
```

The 'Light Every Night file structure' section contains the following text:

The data for both DMSP-OLS and VIIRS-DNB are in the root AWS S3 bucket, `s3://globalnightlights/`. DMSP-OLS data are

## OPEN DATA + ANALYTICS IN ACTION:

- **Open Collaboration:** Project team uses cloud and Github repository to support implementation (WB, NOAA, UMich, New Light, AWS)
- **Open Standards:** data published in analysis ready format (COG/STAC) on cloud
- **Open Knowledge** - algorithms, and training collateral are public goods (Github, Jupyter books)
- <https://worldbank.github.io/OpenNightLights/welcome.html>



# Other WB commons for Open Analytics



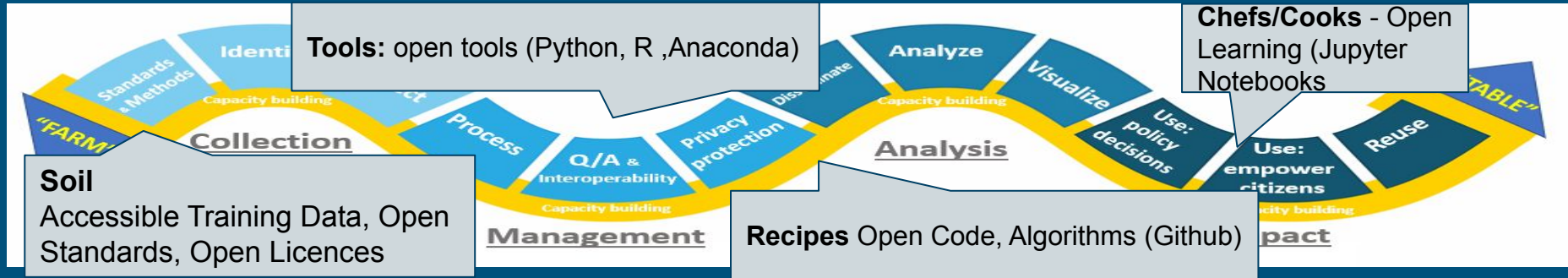
- ***Standards and tools to enrich Data and Analytics Documentation and Discovery*** - much of the information about data and models are buried in text.



- ***World Bank Open Source Licence*** for software tools (coming soon for survey solutions <https://mysurvey.solutions/en/>)



# Toward a robust enabling environment for Open Development Data Products



What can Funders of public sector data projects do?	What can Implementing Agencies do?
<ul style="list-style-type: none"> <li>● Prioritize projects that strengthen reusable, scalable public goods</li> <li>● Understand the open data science commons for sector/domain</li> <li>● Raise awareness of program managers in the practices, resources in open source</li> </ul>	<ul style="list-style-type: none"> <li>● Use and contribute the open data science commons through project implementation</li> <li>● Deliver the data product along with open analytical tools and tutorials</li> </ul>



# LEARN MORE

WB <a href="#">Open Data Toolkit</a>	Resources to understand open data programs	<a href="http://opendatatoolkit.org">http://opendatatoolkit.org</a>
BITSS - Guide to Reusable Open Analytics	Resources to support computational reproducibility in research	<a href="https://bitss.github.io/ACRE/">https://bitss.github.io/ACRE/</a>
Data in Action Toolkit	Resources for designing and scoping data science products and projects	<a href="https://worldbank.github.io/Data-in-Action/">https://worldbank.github.io/Data-in-Action/</a>
World Bank Github	Applied data products, tools and learning resources for development	<a href="https://github.com/worldbank">https://github.com/worldbank</a>
World Development Report	Focus on Data for Better Lives	<a href="https://www.worldbank.org/en/publication/wdr2021">https://www.worldbank.org/en/publication/wdr2021</a>
Open Data Product Resources	Curated list of open data, tools, resources	<a href="https://bit.ly/3uc46rb">https://bit.ly/3uc46rb</a>
Tools for Data Science	Coursera Course on open tools	<a href="https://www.coursera.org/learn/open-source-tools-for-data-science">https://www.coursera.org/learn/open-source-tools-for-data-science</a>

# OPEN DATA PRODUCT *STARTER KIT*

\* FOR MORE TOOL SEE ODSC | FOSS4G

## GENERAL PURPOSE

## RESOURCES | TOOLS | PRACTICES

Open Languages, Analysis

[Python](#) | [R](#) | [Gitub](#) | [GitLab](#) | | [Jupyter](#) | [Anaconda](#) | [Kaggle](#) | [Cookie Cutter Data Science](#) | [Docker](#) | [BDAS](#) | [BITTS](#) | [PySal](#) |

Interoperability

[Tidy Data](#) | [Slippy](#) | [GeoStat](#) | [World Pop – GRID3](#) | [NIST](#) | [Analysis Ready Data](#) | [IPUMS](#) | [SDMX](#) | [DDI](#) | [DublinCore](#) | [Schemas.org](#) | [ONS-ML](#)

Ethics & Privacy Preserving Methods

[UN Handbook of Privacy Preserving Methods](#) | [Differential Privacy](#) | [UNGP Risk Assessment Tool](#) | [Deon DrivenData Checklist](#)

Rapid Data Collection, Tasking, Labeling

[MTURK](#) | [AWS Sagemaker Ground Truth](#) | [Figure8](#) | [Hive Data](#) | [Samasource](#) | [Qualtrics Prolific](#) | [Premise](#) | [Native](#) | [Snorkle](#) | [Kobo](#) | [Qfield](#) | [RapidPro](#) | [SurveyCTO](#) | [Survey Solutions](#) | [Geo-referencing data for ML](#)

Data Catalogs

[WB Data Catalog](#) | [Google Data Search](#) | [Open Street Map](#) | [Enigma](#) | [AWS Open Data](#) | [WorldPop](#) | [Kaggle](#) | [Awesome Satellite Data](#)

## STEP

## SATELLITE

## MOBILE

## TEXT

**COLLECTION** - High Frequency Data Collection; Data Labelling;

[Label Maker](#) | [Cumulus](#) | [MTURK](#) | [Hive Data](#)

[Flowkit](#) | [UTSDC](#) | [OPAL](#) | [Positium](#) | [FB MTK API](#)

[Google API](#) | [Twitter decahose stream](#) | [FB Marketing API](#) | [LinkedInDevAPI](#) | [GDELT](#) | [Factiva Content API](#) |

**DISCOVERY** – Training Data Sets; Rich Context Data Search; Knowledge Products | Primitives

[Earth on AWS](#) | [GeoNet](#) | [Nasa Earth Science](#) | [Maxar Open Data](#) | [Planet Explorer](#) | [GEOSTAT](#) | [Global Change Master Directory](#) | [Radiant ML Hub](#) | [Carto Observatory](#) | [WSF](#) | [ELA](#) | [Sentinel-Hub](#) | [PopGrid](#) | [STAC](#) | [COG](#) | [ASL SpaceNet](#) | [UCI](#)

[Cubiq](#) | [Mapbox Telemetry](#) | [SmartGraph \(M-SDK\)](#) | [Orbital Insights Go](#) | [OpenMobile](#) | [Mirage](#) | [Uber-D](#) | [OMF](#) | [OpenCellID](#) |

[Coleridge Rich Context Scholarly articles API](#) | [Kaggle-Text](#) | [Awesome-NLP](#)

**INTEGRATION & ANALYTICS**

Analytic tools; privacy preserving methods; integration frameworks

[Grid3](#) | [GEE](#) | [eo-learn](#) | [Solaris](#) | [Raster Vision](#) | [Pangeo](#) | [RoboSat](#) | [Sat StatePlay](#) | [G-DIF](#) | [GeoPandas](#) | [Orbital Insights Go](#) | [PDAL](#) | [Sentinal Toolbox](#) | [Orfeo](#) | [mosaicJSON](#) | [GeoNode](#) | [SNAP](#) | [RasterFoundry](#) | [PPovRepo](#) | [Spfeas](#)

[Bandicoot](#) | [Flowkit](#) | [UT-CDR Analysis Kit](#) | [Mobile Privacy Model](#) | [Movesense](#) | [Grid3](#) | [Scikit-Mobility](#)

[BERT](#) | [GLUE](#) | [GPT-2](#) | [Microsoft Turing](#) | [Pytorch](#) | [StanfordNLP](#) | [Stanza](#) | [NLTK](#) | | [OpenNLP](#)



# Thank You!

Trevor Monroe,  
Analytics & Tools Unit  
Development Data Group, World Bank Group  
[@trevmon28](#)