

Open Data on Amazon Web Services

UW Cloud Day

Jed Sundwall, AWS Open Data Global Lead

12 November 2015

Why does AWS care about open data?

Open data is data that can be used by anyone for any purpose for free.

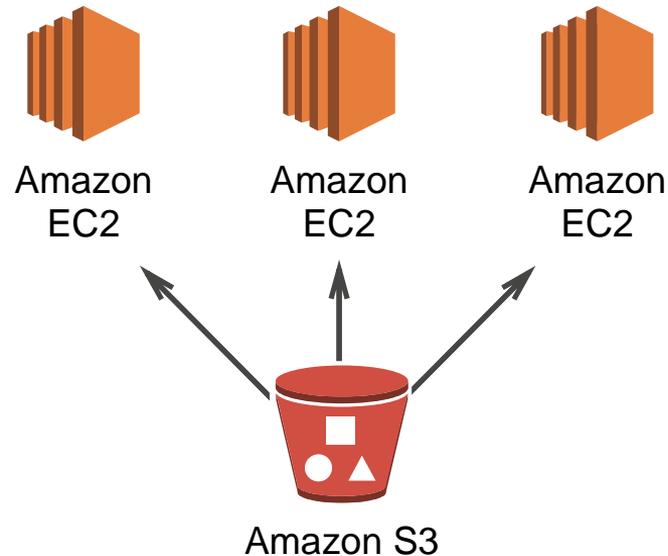


ZOO NIVERSE

Many of our customers in the scientific community and in industry, rely on quality open data as much as they rely on our computing, storage, and other web services.

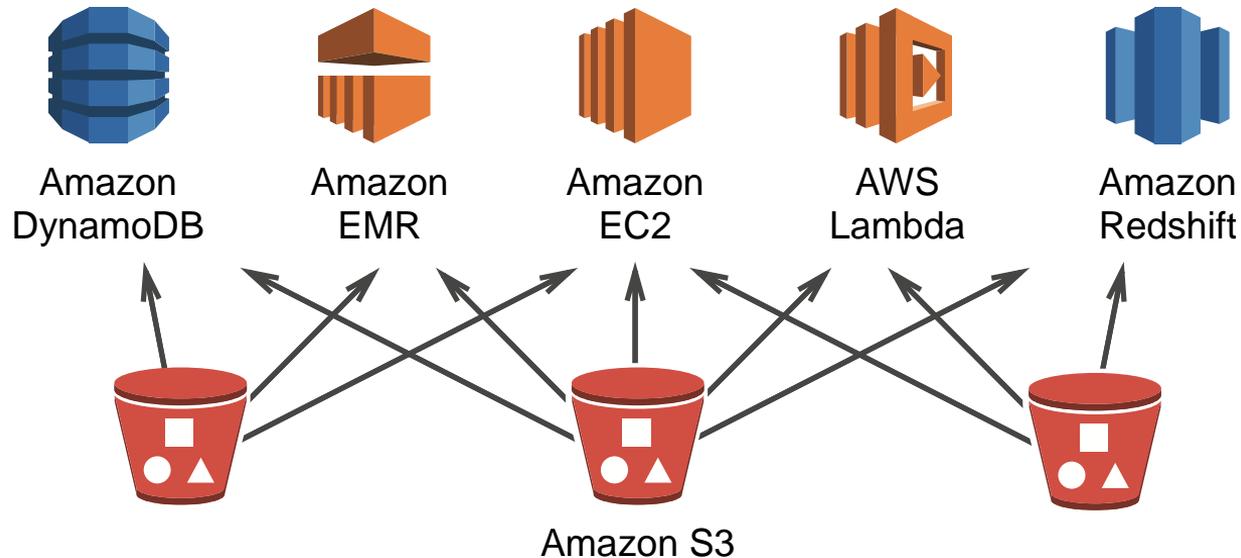
The power of open data in the cloud

Making data open on AWS enables more innovation by making data available for rapid access to our flexible and low-cost computing resources.



The power of open data in the cloud

Making data open on AWS enables more innovation by making data available for rapid access to our flexible and low-cost computing resources.





1-click deployment to launch, on multiple regions around the world



Pay-as-you-go pricing

Data Integration



Advanced Analytics

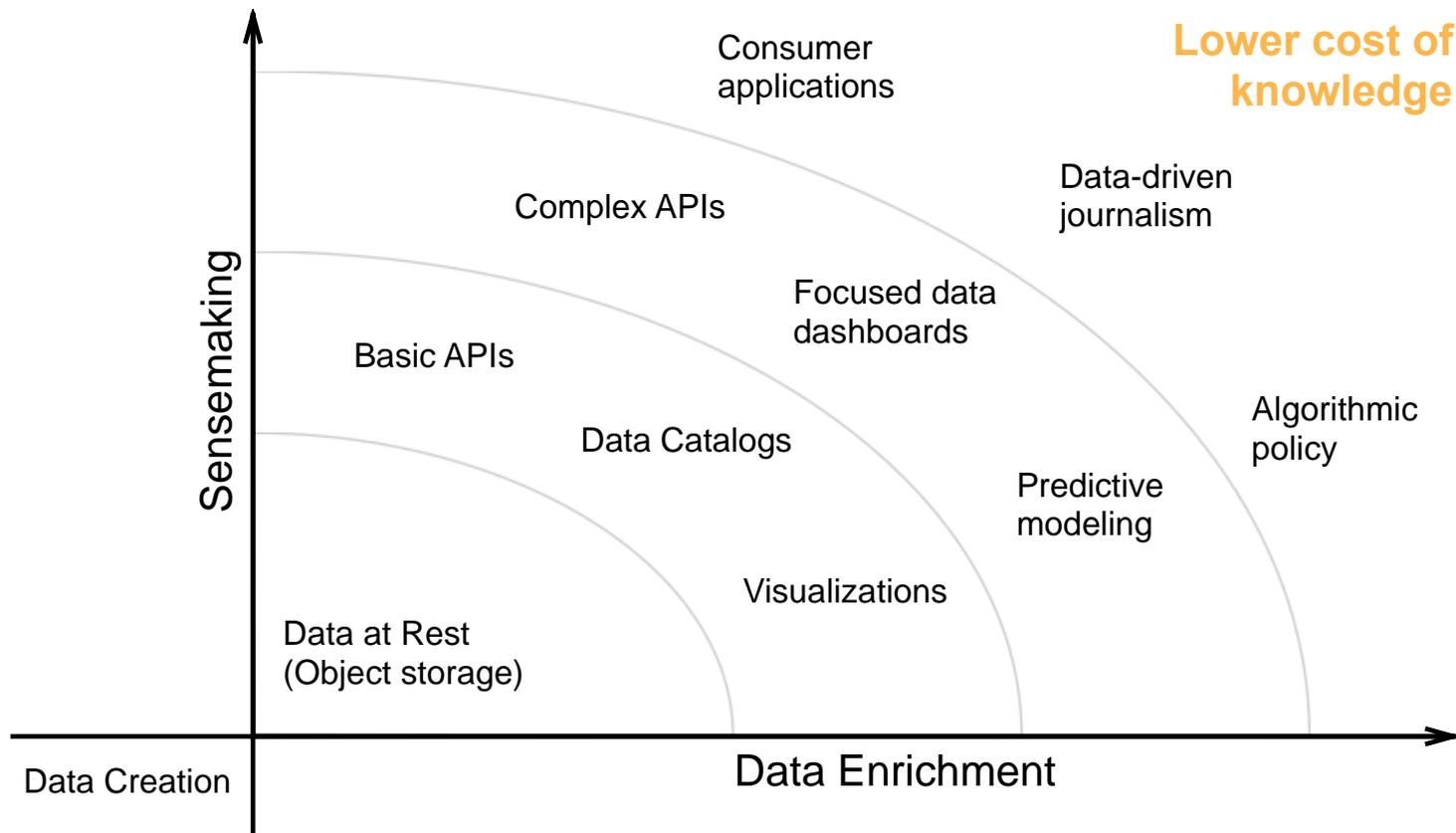


Analysis & Visualization

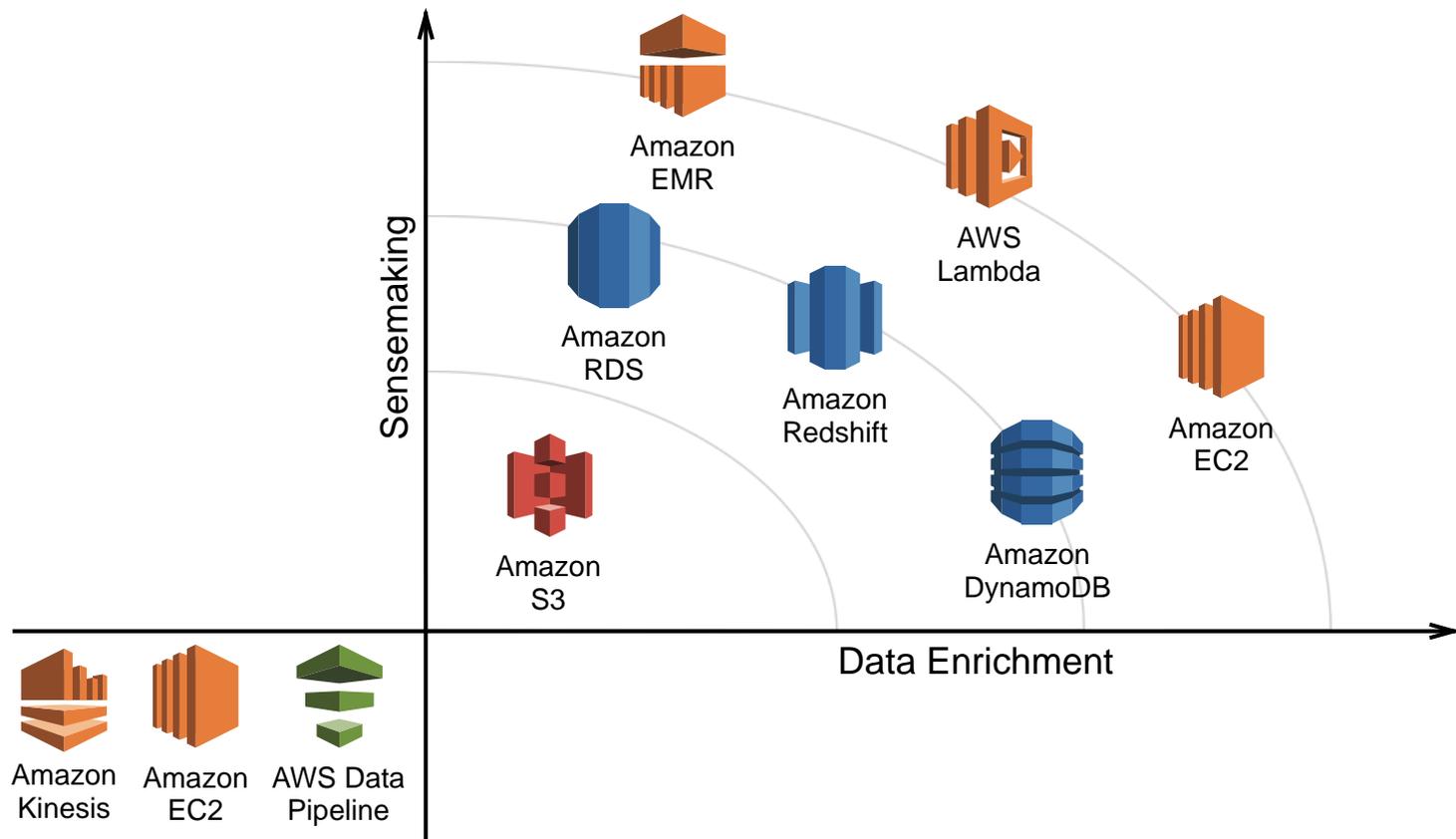


<http://bit.ly/awsAnalytics>

Open data as a platform



Open data as a platform



An Amazonian approach to open data

Two ideas that inform how we approach public data sets:

- Work backwards from the customer
- Eliminate undifferentiated heavy lifting

Working Backwards

- Think about data sets as products
- Seek out valuable data by listening to customer needs
- Consider real-world use cases for the data
- Consider the size of the user community or market opportunity

Undifferentiated heavy lifting

“...data must be organized, well-documented, consistently formatted, and error free. Cleaning the data is often the most taxing part of data science, and is frequently 80% of the work.”

— *Data Driven* by DJ Patil and Hilary Mason

Undifferentiated heavy lifting

“...data must be organized, well-documented, consistently formatted, and error free. Cleaning the data is often the most taxing part of data science, and is frequently 80% of the work.”

— *Data Driven* by DJ Patil and Hilary Mason

We ask: **How can we get rid of that 80%?**

Public datasets on AWS

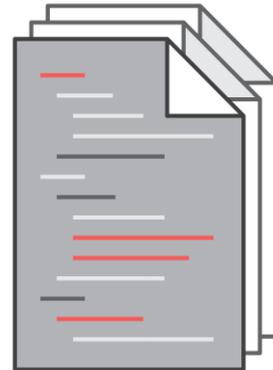
To enable more innovation, AWS hosts a selection of datasets that anyone can access for free. Data in our public datasets is available for rapid access to our flexible and low-cost computing resources.



Life Sciences
1000 Genomes Project



Earth Science
Landsat on AWS



Internet Science
Common Crawl Corpus