



Datos Abiertos en AWS

<https://opendata.aws>

Jed Sundwall, Global Open Data Lead
jed@amazon.com

¿Por qué AWS se preocupa por los Bases de datos públicas?

El uso compartido de datos en AWS permite que una gran comunidad en crecimiento de investigadores, emprendedores y empresas que utilizan la nube de AWS accedan a ellos.



Muchos clientes de AWS suministran datos al público para acelerar la investigación y el desarrollo de productos.



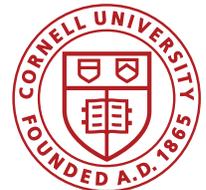
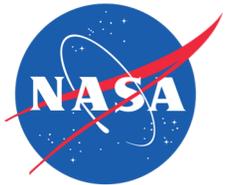
Muchos clientes de AWS utilizan datos compartidos en AWS para crear nuevos productos y servicios.

El programa de datos abiertos de AWS amplía el acceso a datos para su análisis en la nube.

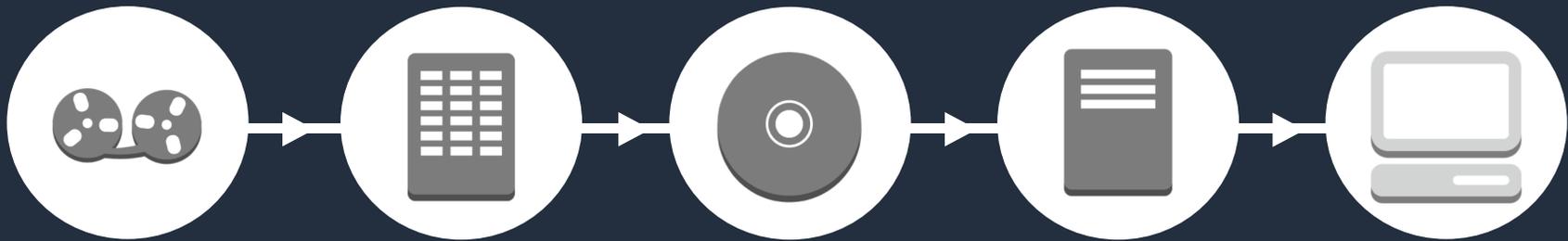
<https://opendata.aws>

AWS Public Datasets

<https://registry.opendata.aws>



Adquisición tradicional de datos



"...los datos deben estar organizados, documentados correctamente, con formato coherente y no deben contener errores. La limpieza de datos suele ser la parte más difícil de la ciencia de datos y, con frecuencia, representa el **80 % del trabajo**".

- "Data Driven" de DJ Patil y Hilary Mason

Quehaceres indiferenciados

Adquisición de datos en la nube



Ventajas del uso compartido de datos en la nube



La comunidad
global de usuarios



Mayor velocidad
de investigación



Nuevos
servicios y
herramientas



Menor costo
de investigación

AWS Public Datasets

<https://registry.opendata.aws/>

The screenshot shows a web browser window with the URL `registry.opendata.aws/tag/satellite-imagery/`. The page title is "Registry of Open Data on AWS" and the AWS logo is in the top right. The main content is divided into two columns. The left column has an "About" section, a search bar with the text "Search datasets (currently 10 matching datasets)", and an "Add to this registry" section. The right column features a "Sentinel-2" section with tags for "earth observation", "satellite imagery", "gis", "natural resource", "sustainability", and "disaster response". Below the tags is a paragraph about the Sentinel-2 mission, a "Details" link, and a "Usage examples" section with four bullet points. Below that is a "Landsat 8" section with similar tags, a paragraph about the satellite, a "Details" link, and a "Usage examples" section with two bullet points.

Registry of Open Data on AWS

About

This registry exists to help people discover and share datasets that are available via AWS resources. [Learn more about sharing data on AWS.](#)

See [all usage examples for datasets listed in this registry](#) tagged with **satellite imagery**.

Search datasets (currently 10 matching datasets)

You are currently viewing a subset of data tagged with **satellite imagery**.

Add to this registry

If you want to add a dataset or example of how to use a dataset to this registry, please follow the instructions on the [Registry of Open Data on AWS GitHub repository](#).

Unless specifically stated in the applicable dataset documentation, datasets available through the Registry of Open Data on AWS are not provided and maintained by AWS. Datasets are provided and maintained by a variety of third parties under a variety of licenses. Please check dataset licenses and related documentation to determine if a dataset may be used for your application.

Sentinel-2

[earth observation](#) [satellite imagery](#) [gis](#) [natural resource](#) [sustainability](#) [disaster response](#)

The **Sentinel-2 mission** is a land monitoring constellation of two satellites that provide high resolution optical imagery and provide continuity for the current SPOT and Landsat missions. The mission provides a global coverage of the Earth's land surface every 5 days, making the data of great use in on-going studies. L1C data are available from June 2015 globally. L2A data are available from April 2017 over wider Europe region and globally since December 2018.

[Details](#) →

Usage examples

- [Sentinel Hub WMS/WMTS/WCS Service by Sinergise](#)
- [Python package for working with Sentinel-2 AWS data by Sinergise](#)
- [Exploring the Chile wildfires with Landsat and Sentinel-2 imagery by Timothy Whitehead](#)
- [Sentinel Playground by Sinergise](#)
- [EOS Land Viewer by Earth Observing System](#)

[See 16 usage examples](#) →

Landsat 8

[earth observation](#) [satellite imagery](#) [gis](#) [natural resource](#) [sustainability](#) [disaster response](#)

An ongoing collection of satellite imagery of all land on Earth produced by the Landsat 8 satellite.

[Details](#) →

Usage examples

- [Sentinel Playground for Landsat by Sinergise](#)
- [Integrate imagery from the full Landsat archive into your own apps, maps, and](#)



AWS Cloud Credits for Research

Date





AWS has been supporting global research projects through the Cloud Credits for Research program since 2009.

- AWS believes that researchers are key drivers of technological innovation
- AWS Cloud Credits for Research is aimed at helping researchers move their research endeavors to the cloud so they can innovate more quickly and at lower cost



We support researchers who seek to:

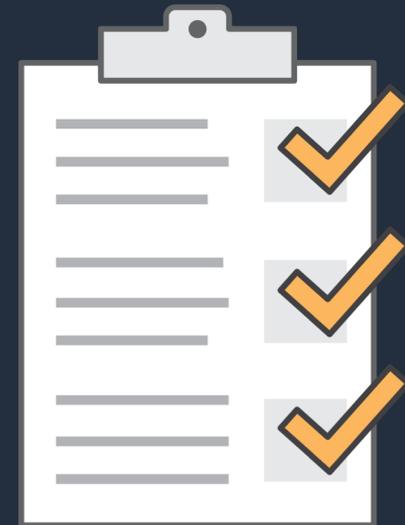
- **Build reusable tools** to facilitate their future research and the research of their community
- **Perform pilot, proof of concept, or benchmark tests** evaluating the efficacy of moving research workloads or open data sets to the cloud
- **Train a broader community** on the usage of cloud for research workloads via workshops or tutorials

How do I apply?

Fill out an application online at: <http://aws.amazon.com/research-credits/>

Include the following information in your proposal:

1. Brief description of problem to be solved.
2. Proposed AWS solution
3. Plan for sharing outcomes created during project.
4. Any potential future use of AWS beyond grant duration by individual research group or broader community.
5. Names of any AWS employees you have been in contact with.
6. Any AWS Public Data Sets to be used in your research.
7. Keywords.



Include a budget estimate by using the simple monthly calculator:
<https://calculator.s3.amazonaws.com/index.html> (be sure to multiply by 12!)

Frequently Asked Questions

What services are **NOT** covered by credits?

Reserved Instance purchases, Mechanical Turk, Marketplace purchases, Business or Enterprise Support, Subscription Fees. **Most of new services are available, but please check with our team if you are unsure.**

For how long are the credits valid?

Research credits expire after **1 year** from the date of redemption. **Please review onboarding emails.**

Can I apply for an extension or for additional credits?

There is no way to extend the expiration date of credits, so you will need to apply for additional credits. Because these credits come from a new budget year, they are evaluated against all new requests.

Frequently Asked Questions

When will I hear if my application was successful?

We accept and review applications on a quarterly basis. Typical notification dates are below:

Submission Date	Notification Date
January 1 - March 31	June
April 1 - June 30	September
July 1 - September 30	December
October 1 - December 31	March

World-class Research

Powered by the AWS Cloud

US National Institutes for Health develops an AWS-hosted web service for microbiome analysis

- NIH collaborators used AWS credits to develop Nephele, a platform to facilitate identification of which species of bacteria are found in a biological sample
- Nephele enables scientists across the world to upload DNA microbiome sequence data and obtain a list of corresponding microorganisms
- Nephele uses AWS Lambda to spin up EC2 instances and custom genome analysis tools along with user-submitted sequence data



AWS was the most mature of the cloud platforms evaluated and it had the most number of research users. **Amazon also offered funding to some of the Nephele collaborators through its research grant program**, which helped get some development efforts off the ground...

- GenomeWeb article, February 2016

International Centre for Radio Astronomy Research uses AWS to analyze crowd-sourced data

- The Australian International Centre for Radio Astronomy Research (ICRAR) has developed a community computing initiative called theSkyNet
- theSkyNet allows ICRAR to use spare CPU cycles volunteered by the public to simulate a supercomputer
- ICRAR uses Amazon Route53 to route users to theSkyNet website and uses Amazon EBS to store upwards of 400 GB worth of imaging data monthly as it is processed by the community
- By using AWS to manage crowd-sourced CPU cycles, ICRAR has the compute capacity to analyze between 400 and 500 galaxies simultaneously



The scalability of AWS has been enormously helpful. I can add more capacity as I need it with minimal fuss. Using AWS allows us to process upwards of 150 GB of sky images and store more than 400 GB of imaging data every month.

-Kevin Vinsen, ICRAR

Penn State uses AWS to make biotech research available to the scientific community

- The Biological Engineering Department wanted to give biotech researchers an easy way to share data and run computationally intensive simulations on DNA
- Moved from an office server to AWS to make Penn State's design methods and optimization algorithms available to researchers all over the world
- Scaled to meet the demands of more than 6,000 users, who have designed more than 50,000 synthetic DNA sequences
- Reduced time to market so that research findings are more immediately available to the scientific community
- Eliminated the need to buy, power, and maintain expensive hardware



We can create an algorithm that can take 32, 64, 128 different cores, and make it accessible to anyone in the world.

Howard Salis

Ph.D., Assistant Professor, Penn State



Code.org scales to support 20 million students for Hour of Code campaign on AWS

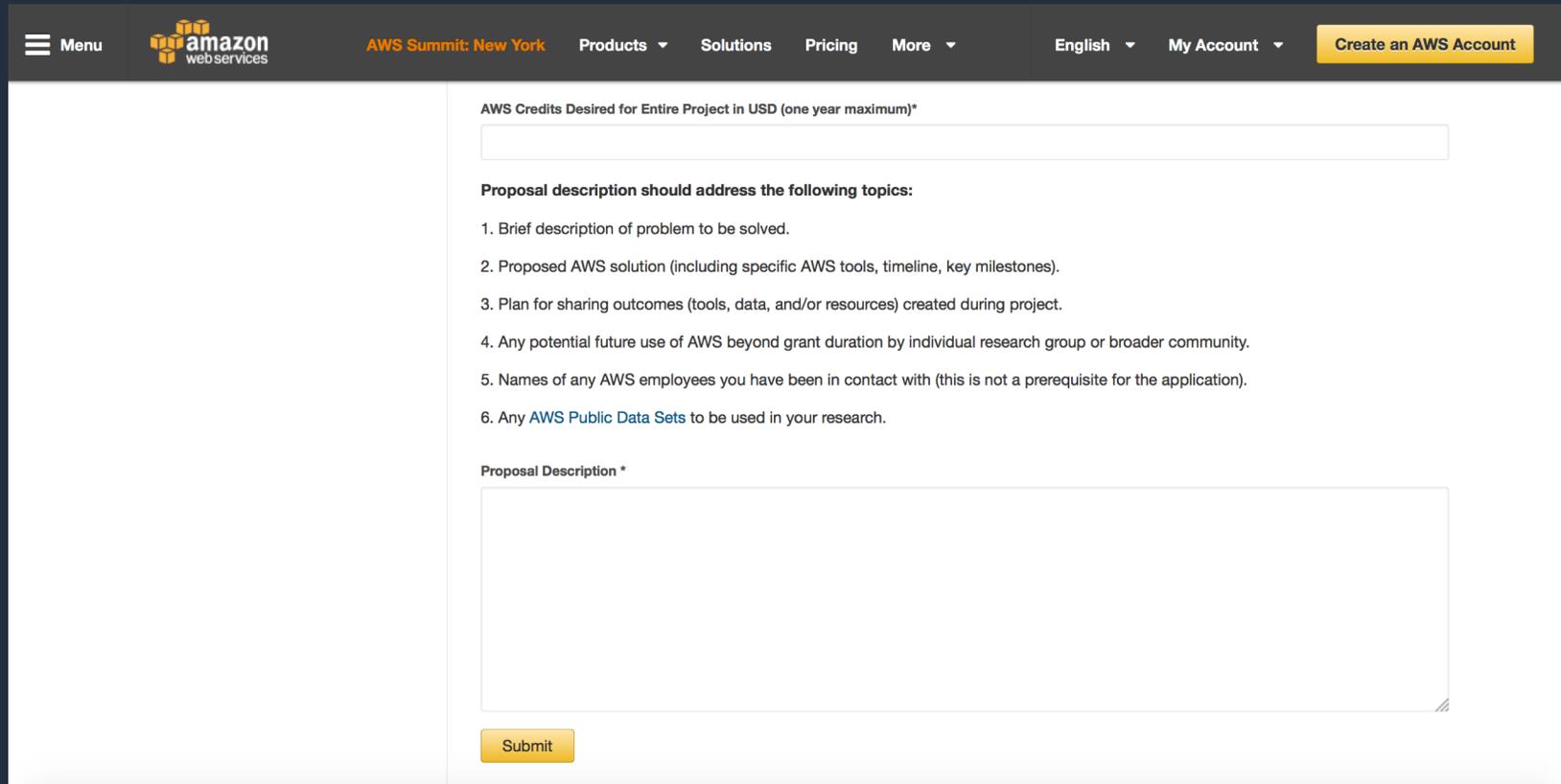
- Code.org wanted to launch its website to coincide CSEdWeek and promote its Hour of Code campaign
- The company turned to AWS to build an environment with high availability and redundancy
- AWS provided the elasticity to keep the website running when traffic spiked to 20 million student programmers in one hour
- AWS also provided the scalability to spin down instances even as traffic grew and save on costs



Running on AWS kept our application running when traffic spiked from zero to 20 million student programmers in one

^{hour}
-Geoffrey Elliott, Code.org

Visit our website for more information and to submit an application:
<http://aws.amazon.com/research-credits/>



The screenshot shows the AWS Research Credits application form. At the top, there is a navigation bar with the AWS logo, a menu icon, and links for 'AWS Summit: New York', 'Products', 'Solutions', 'Pricing', 'More', 'English', 'My Account', and a 'Create an AWS Account' button. The main content area contains a form with the following elements:

- A text input field labeled 'AWS Credits Desired for Entire Project in USD (one year maximum)*'.
- A section titled 'Proposal description should address the following topics:' followed by a numbered list of six items:
 1. Brief description of problem to be solved.
 2. Proposed AWS solution (including specific AWS tools, timeline, key milestones).
 3. Plan for sharing outcomes (tools, data, and/or resources) created during project.
 4. Any potential future use of AWS beyond grant duration by individual research group or broader community.
 5. Names of any AWS employees you have been in contact with (this is not a prerequisite for the application).
 6. Any [AWS Public Data Sets](#) to be used in your research.
- A large text area labeled 'Proposal Description *'.
- A 'Submit' button at the bottom of the form.