Keyword: climate-change

Headline: Against threats of deletion by gov't, NASA Earth science data gets rescued by indie coders

Byline: INQUIRER.net

Published Date: 04:26 PM February 15, 2017

Section: technology

Word Count: 528

Content:

Image: INQUIRER.net stock photo

Image: INQUIRER.net stock photo

In an age where information can be more valuable than money, the loss of the former is nothing short of a catastrophe.

Thankfully, a group of dedicated individuals is backing up years'-worth of NASA Earth science data before they get purged by the current government.

What to do when the government deletes data

The Trump administration has made it quite public that climate change is not one of their concerns. This has lead members of the scientific community to fear the purging of federal climate data by the government. Wired reports that groups like DataRefuge and the Environmental Data and Governance Initiative have assembled teams of hackers, scientists, and students to start collecting the endangered data to save it outside government servers.

Teams of coders are currently hard at work building robust systems that would handle the monitoring of changes to government websites. These systems are keeping track of what has already been taken out, because apparently, the government has already started deleting data.

Data collection is proceeding at a methodical pace. Half of the group has set out web crawlers for easily-copied government pages to send the text to the Internet Archive – a digital library that contains hundreds of billions of webpage snapshots. More data-intensive projects are tagged for the other group's reference. This other group is referred to as "baggers" and they write custom script to scrape complicated data sets from patched-together federal websites.

"All these systems were written piecemeal over the course of 30 years. There's no coherent philosophy to providing data on these websites," says Daniel Roesler, chief technology officer at UtilityAPI and one of the volunteer guides for the Berkeley bagger group.

Despite the presence of hacker in the teams, DataRefuge prohibits outright hacking, so every time anyone hits a wall or "404 page not found," they will need to legally dig deeper, which takes time.

Data that has already been removed certainly poses a problem, but it is further compounded by the fact that they don't know when it happened or if someone else has already backed up the data.

Emptied databases on climate change

Some of the databases that have already been emptied include the Global Change Data Center's reports archive and a NASA atmospheric CO_2 dataset.

Plans for the long run are already being put together by two dozen or so of the most advanced software builders in the team. They are trying to create tools that would monitor and filter out changes as well as updates to the government websites.

"Climate change data is just the tip of the iceberg," says Eric Kansa, an anthropologist who manages archaeological data archiving for the non-profit group Open Context. "There are a huge number of other datasets being threatened, with cultural, historical, sociological information."

Subscribe to our daily newsletter

By providing an email address. I agree to the Terms of Use and acknowledge that I have read the Privacy Policy.

It is hoped that these dedicated individuals will be able to save most of the endangered data as these databases hold valuable insight into the current status of global climate change. Alfred Bayle/JB