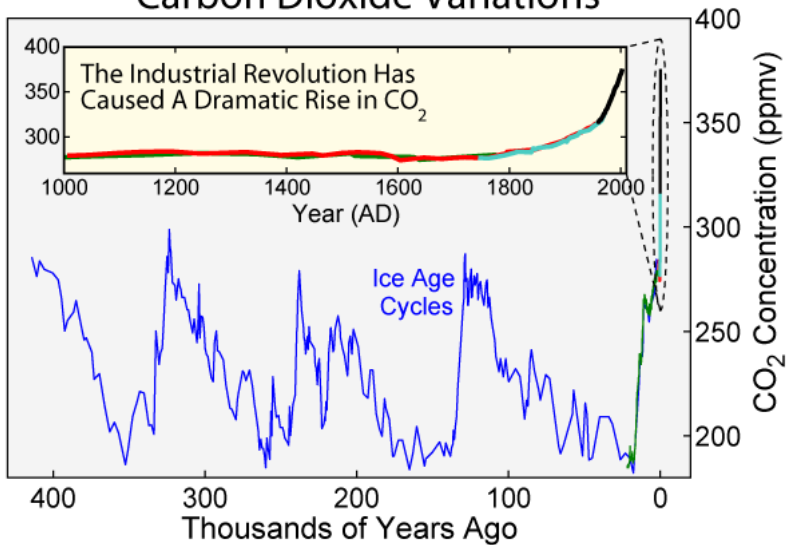# THE AZIMUTH CLIMATE DATA BACKUP PROJECT

John Baez, Jan Galkowski, Sakari Maaranen
Scott Maxwell and over 627 others

October 25, 2017

Carbon Dioxide Variations

The Industrial Revolution Has Caused A Dramatic Rise in CO_2

Ice Age Cycles

By December 2016 the situation was clear.

Trump's choice to head the Environmental Protection Agency, Scott Pruitt, called himself a

"leading advocate against the EPA's activist agenda"

He had sued this agency 14 times.

Trump's choice to head the Environmental Protection Agency, Scott Pruitt, called himself a

"leading advocate against the EPA's activist agenda"

He had sued this agency 14 times.

Trump's choice for the Department of Energy, Rick Perry, claimed

"we have been experiencing a cooling trend"

and

"there are a substantial number of scientists who have manipulated data so that they will have dollars rolling into their projects".

On December 10th, 2016, meteorologist and reporter Eric Holthaus tweeted:

Scientists: Do you have a US .gov climate database that you don't want to see disappear?

Add it here: https://docs.google.com/spreadsheets/d...

Please share.

December 11th, my friend the statistician Jan Galkowski emailed me:

> *John, so I have just started trying to mirror all of CDIAC [the Carbon Dioxide Information Analysis Center]. We'll see. I'll put it in a tarball, and then throw it up on Google. It should keep everything intact. Using WinHTTrack. I have coordinated with Eric Holthaus via Twitter, creating, per your suggestion, a new personal account which I am using exclusively to follow the principals.*
>
> *Once CDIAC is done, and checked over, I'll move on to other sites.*

> *Alarmed that decades of crucial climate measurements could vanish under a hostile Trump administration, scientists have begun a feverish attempt to copy reams of government data onto independent servers in hopes of safeguarding it from any political interference.*

> *The efforts include a "guerrilla archiving" event in Toronto, where experts will copy irreplaceable public data, meetings at the University of Pennsylvania focused on how to download as much federal data as possible in the coming weeks, and a collaboration of scientists and database experts who are compiling an online site to harbor scientific information.*

By December 16th, Sakari Maaranen offered me his help. He's a computer systems architect in Finland:

> *I have several terabytes worth in Helsinki (can get more) and a gigabit connection.*

Other computer experts joined our team, especially Scott Maxwell of Google. They copied data from US government databases to servers in Germany.

Other computer experts joined our team, especially Scott Maxwell of Google. They copied data from US government databases to servers in Germany.

I started a Kickstarter campaign to raise funds on December 28th. It ended on January 31st. We got $20,427 of donations from 627 people. Our finances are completely transparent, available online.

Other computer experts joined our team, especially Scott Maxwell of Google. They copied data from US government databases to servers in Germany.

I started a Kickstarter campaign to raise funds on December 28th. It ended on January 31st. We got $20,427 of donations from 627 people. Our finances are completely transparent, available online.

We decided to back up 30 terabytes of data. We finished in April. But this data costs about $300/month to hold on commercial servers — with backups in case something breaks!

(One server overheated and broke in June, but everything was okay.)

So, I contacted Steven Mandeville-Gamble at UCR. On January 11th, Danna Gianforte wrote:

> *I am committing to you that one way or another, I will ensure you have access to the storage you need to save our precious climate data from the new administration. If you can save it temporarily through your efforts, I will ensure we get you a set of UCR server space to move it to soon. Does that work?*

So, I contacted Steven Mandeville-Gamble at UCR. On January 11th, Danna Gianforte wrote:

> *I am committing to you that one way or another, I will ensure you have access to the storage you need to save our precious climate data from the new administration. If you can save it temporarily through your efforts, I will ensure we get you a set of UCR server space to move it to soon. Does that work?*

We decided to use Syncplicity, the new UCR cloud system. Downside: unlike the servers in Germany, this is not open worldwide, and the data is not there yet. Upside: it has institutional support.

Our data is currently available at three commercial servers:

ftp://www01.azimuthproject-kickstarter.org/pub/

ftp://www03.azimuthproject-kickstarter.org/pub/

ftp://www04.azimuthproject-kickstarter.org/pub/

Some of what we've got:

- NASA GISTEMP website: 103 gigabytes
- NOAA Carbon Dioxide Information Analysis Center data: 294 gigabytes
- NOAA Carbon Tracker website: 513 gigabytes
- Global ARGO data, NOAA: 3015 gigabytes

and dozens more, listed on our project's website.

What's up now?

The Environmental Data and Governance Initiative is monitoring the EPA's website, watching the words 'climate change' disappear from their webpages, archiving backups, and publishing reports.

So far no US government *databases* seem to have disappeared.

EPA employees are being forced to attend forced to attend anti-leaking classes, and, this week, prevented from speaking about the new climate change report.

Many groups have backed up climate data. Nick Santos at U.C. Davis has created Climate Mirror, a portal for these backups.

Many groups have backed up climate data. Nick Santos at U.C. Davis has created Climate Mirror, a portal for these backups.

However, it's still hard to know everything that's been backed up, or *not yet* backed up. Professional archivists like the California Digital Library are getting involved — see the CDL's End of Term Web Archive — but apparently more for *websites* than *databases*.

Many groups have backed up climate data. Nick Santos at U.C. Davis has created Climate Mirror, a portal for these backups.

However, it's still hard to know everything that's been backed up, or *not yet* backed up. Professional archivists like the California Digital Library are getting involved — see the CDL's End of Term Web Archive — but apparently more for *websites* than *databases*.

In short: we've made some progress backing up US government climate databases, but we do not have a well-indexed, professional archive.