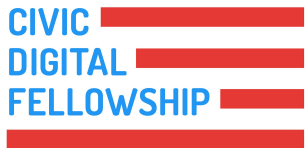


Data Analysis and Automation

CIT STRIDES

Nick Weber — Program Manager

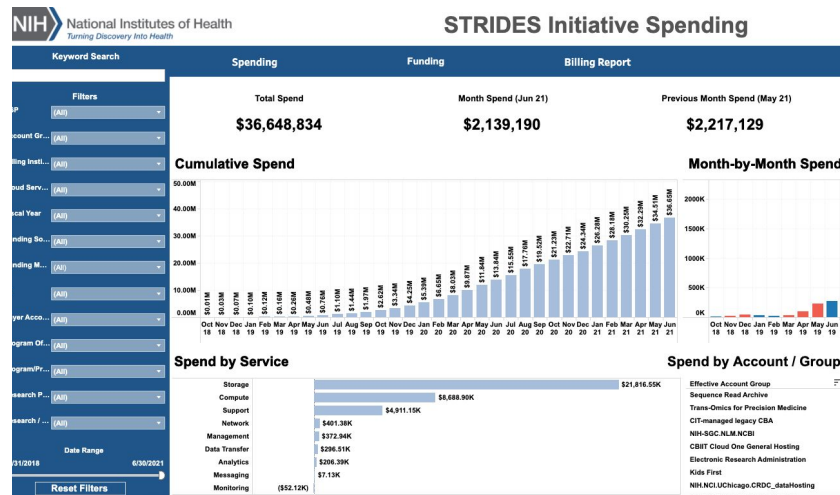
Joel Peterson — Cloud Architect



Annika Lin
Georgetown University
Economics and Computer Science

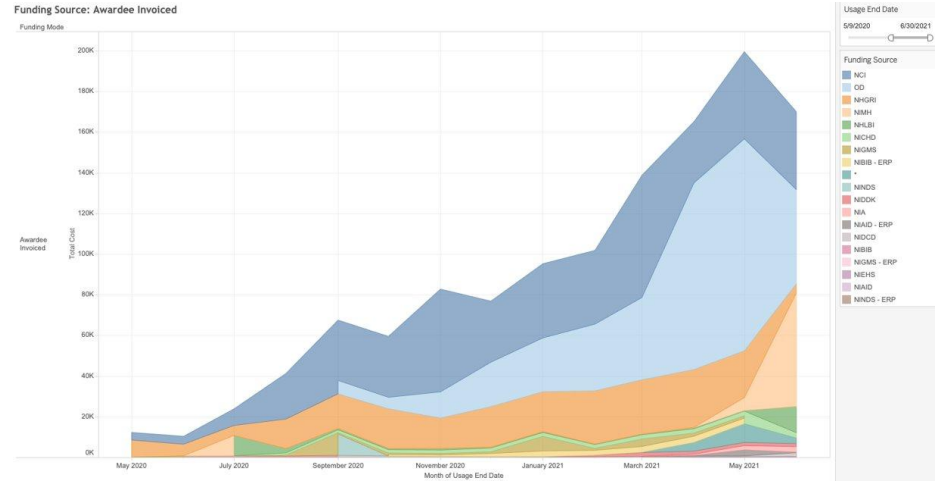
Critical: Cost and Usage Reporting

- Research programs participating in STRIDES spend +\$2M/month
- **Ensures stewardship** of federal **funds** and **sustainability** of the program
- Need to **track costs and balances** at a granular level and report to funding entity



Analyzing Cost and Usage Data

- Created **data visuals with Tableau** and presented trends and insights to STRIDES team and ODSS
- Investigated **cloud billing and usage** data on a **granular level** to **identify trends** among service categories and organization types



Initial Billing Report Process

1) Create PDFs

- **Manually** sets filter for every recipient in Tableau
- **Not scalable** for over 500 reports
- Level of Effort: **50 hours/month**

2) Send PDFs

- PowerAutomate emails PDFs to recipients
- Takes **16 hours/month**

Total Time: ~**66 hours/month** (or ~800 hours/year)

Account Name	Account ID	Month of Invoice	Analysis	Compute	Data Transfer	Management	Notes
14-ANINGUJHO	14-ANINGUJHO	Nov 15	\$0.00	\$0.01	\$0.00	\$0.00	
		Dec 15	\$0.00		\$0.00	\$0.00	
		Jan 20	\$0.00	\$32.14	\$0.00	\$0.00	
		Feb 20	\$0.00	\$16.44	\$0.01	\$0.00	
		Mar 20	\$0.00	\$22.08	\$0.01	\$0.00	
		Apr 20	\$0.00	\$21.99	\$0.10	\$0.00	
		May 20	\$0.00	\$10.43	\$0.01	\$0.00	
		Jun 20	\$0.00	\$141.30	\$0.07	\$1,410.34	
		Jul 20	\$0.00	\$0.23	\$0.01	\$0.00	
		Aug 20	\$0.00	\$16.47	\$0.01	\$0.00	
		Sep 20	\$0.00	\$246.09	\$0.00	\$0.00	
Total			\$0.00	\$5,02.14	\$0.01	\$1,410.34	

Automated Process with RStudio

- 600 lines of code with 4 files to create and email reports



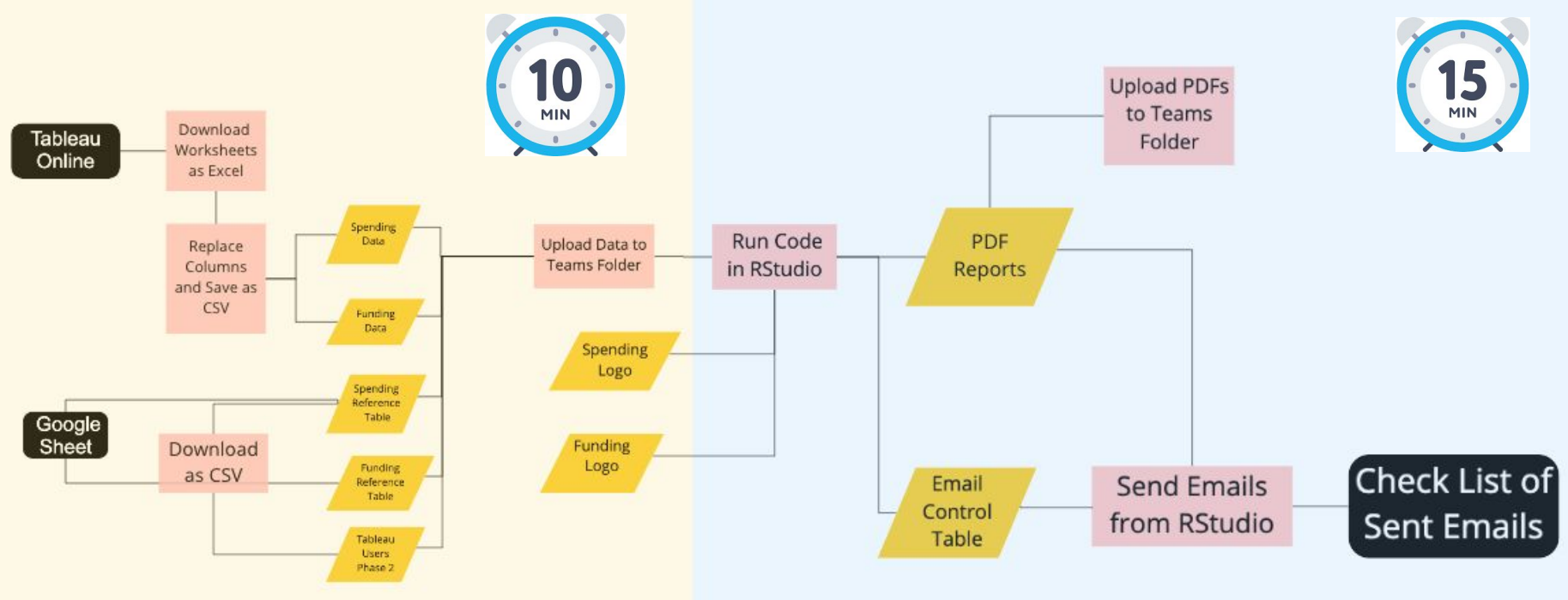
Microsoft365R Package

- Enables access from R to data stored in Teams including uploading and downloading files
- Ability to send emails via Outlook



Automated Billing Report Process

Saves 787 hours and ~\$150k Annually



Email with RStudio

NIH

Outlook

Search

Global.Access.Funding.2021.06.09.518.pdf

Download

Print

Save to OneDrive

STRIDES Initiative
National Institutes of Health

Cloud Funding, Expenses, and Balances by CAN Since Inception
as of June 2021

AWS Expenses

CAN Group	Funding Source	Funding Amount	Total Spending	Remaining Funds
8013100	NIDDK	\$6,000	\$360.89	\$5,639.11
8018985	Common Fund	\$8,000,000	\$1,329,337.53	\$6,670,662.47
8019044	NHBLI	\$0	\$99,738.42	(\$99,738.42)
8019987 + 8339253	OD/OER/eRA	\$3,500,000	\$1,746,692.98	\$1,753,307.02
8020497	NINDS/DIR	\$881	\$0.00	\$881.00
8336556	NINDS/DIR	\$3,000	\$4,032.59	(\$1,032.59)
8028453	OD/OIR	\$26,000	\$9,893.94	\$16,106.06
8029601	NIHGS	\$25,000	\$7,477.79	\$17,522.21
8037893	NIA	\$30,000	\$1,554.58	\$28,445.42
8469737	NIA	\$424,286	\$26,451.31	\$397,834.69
8039295	NINDS	\$925,929	\$6,236.49	\$919,692.51
8040481	ODSS	\$6,355,500	\$10,460,114.59	(\$4,104,614.59)
8044255	ODSS	\$1,832,535	\$904.62	\$1,831,630.38
8044257	ODSS	\$3,527,500	\$42,178.33	\$3,485,321.67
8044659	ODSS	\$25,000	\$0.00	\$25,000.00
8046539	ODSS	\$90,000	\$35.33	\$89,964.67
8047917	ODSS	\$100,000	\$0.00	\$100,000.00
8041139	NHGRI	\$205,000	\$77,580.71	\$127,419.29
8041157 + 8482505	NCI/CBIIT	\$1,488,349	\$2,020,513.66	(\$532,164.66)
8042009	NIMH	\$250,000	\$178,003.62	\$71,996.38
8044725	OD	\$1,000,000	\$0.20	\$999,999.80
8045143	NLM/LHC	\$70,000	\$0.00	\$70,000.00
8330304	NLM/LHC	\$45,000	\$12,077.38	\$32,922.62
8320301	NLM/OCCS	\$468,000	\$300,560.94	\$167,439.06
8326638 + 8368573	NEI/OAM/ITMB	\$24,400	\$25,092.58	(\$692.58)
8326651	NEI/DIR/OGVFB	\$26,000	\$0.31	\$25,999.69
8326758	NEI/LI	\$600	\$0.28	\$599.72
8361036	NIAID - OCICB	\$340,000	\$415,298.73	(\$75,298.73)
8363434	NIBIB	\$20,000	\$1,912.45	\$18,087.55
8368698	NIAID	\$0	\$35,713.15	(\$35,713.15)
8369138	NIAID - DCR	\$40,000	\$9,843.65	\$30,156.35
8374927	NICHD	\$300	\$308.87	(\$8.87)
8475185	NHLBI	\$1,000,000	\$3,969,885.12	(\$2,969,885.12)
Total		\$29,849,280	\$20,781,801.07	\$9,067,478.93

STRIDES Monthly Spending Report for June 2021

LA

Lin, Annika (NIH/CIT) [F]

Tue 8/3/2021 12:25 PM

To: Lin, Annika (NIH/CIT) [F]; al1494@georgetown.edu

Global.Access.Funding.20...
122 KB

Hi,

Thank you for your participation in the NIH STRIDES Initiative. Please see the attached report of funding drawdown for cloud account expenses paid for using a Direct CAN Citation through the month of June 2021. The report lists totals by cloud service provider and CAN. In addition to the attached PDF, you can also view and analyze cloud spending data using the NIH STRIDES Tableau dashboard tool. This dashboard will help you manage your cloud funding balances by CAN group and Cloud Service Provider (e.g. AWS or GCP). You can select either a graphical or table view of the data. You can use filters to examine the data by CAN group, Cloud Service Provider, or Funding Source. A few important points:

To sign in to the dashboard, use this link: https://prod-us-east-1-online-tableau.com/tables/view/NIHSTRIDESSpendingAnalysis_16190224897250/CANFundingDashboard-ExecTableView?origin=card_share_link&embedded-in

Please use this email address to sign in: lna10@nih.gov; al1494@georgetown.edu

You do not need to enter a password on the Tableau sign in screen. Access is federated with your home institution's login, so you will be able to use your normal sign in credentials.

Please let us know if you have any questions or issues.

Thank you.

National Institutes of Health
Center for Information Technology
Cloud Services

Reply

Reply all

Forward

CIVIC
DIGITAL
FELLOWSHIP

Summary

- Analyzed cost and usage data
- Automated billing report generation and distribution



Weber, Nick (NIH/CIT) [E] 8:59 AM

Our previous approach was extremely cumbersome and not scalable. You've made a lasting impact on STRIDES in terms of our ability to sustainably provide cost and usage information to our customers. Thank you!

The beauty of your code is that prep time is low and the email control table gives us the recipients for every report.

9:52 AM

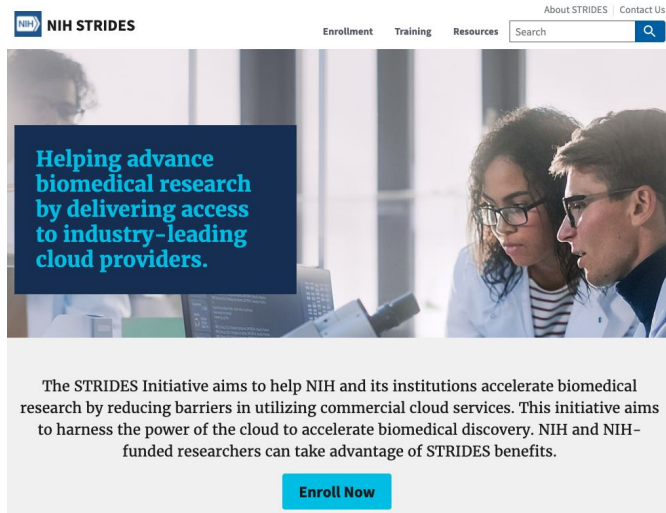
the R code is ideal for me cause im not doing 50hrs worth of work plus my job lol



1



Thanks!



The screenshot shows the NIH STRIDES website. At the top, there is a navigation bar with the NIH logo, the text "NIH STRIDES", and links for "Enrollment", "Training", and "Resources". On the right side of the navigation bar, there are links for "About STRIDES" and "Contact Us", a search bar with the placeholder text "Search", and a magnifying glass icon. Below the navigation bar is a large banner image featuring two researchers, a woman and a man, looking at a laptop. Overlaid on the left side of the banner is a dark blue box with white text that reads: "Helping advance biomedical research by delivering access to industry-leading cloud providers." Below the banner, there is a paragraph of text: "The STRIDES Initiative aims to help NIH and its institutions accelerate biomedical research by reducing barriers in utilizing commercial cloud services. This initiative aims to harness the power of the cloud to accelerate biomedical discovery. NIH and NIH-funded researchers can take advantage of STRIDES benefits." At the bottom of the banner area, there is a blue button with white text that says "Enroll Now".

NIH STRIDES

About STRIDES | Contact Us

Enrollment Training Resources Search

Helping advance biomedical research by delivering access to industry-leading cloud providers.

The STRIDES Initiative aims to help NIH and its institutions accelerate biomedical research by reducing barriers in utilizing commercial cloud services. This initiative aims to harness the power of the cloud to accelerate biomedical discovery. NIH and NIH-funded researchers can take advantage of STRIDES benefits.

Enroll Now

Questions?