Understanding NCBI Cloud Users' Needs

Chae Eun Park – Civic Digital Fellow

U.S. National Library of Medicine
National Center for Biotechnology Information

Cloud Savvy

The early adopters, cloud power users, technically proficient, work with huge datasets, bioinformaticians

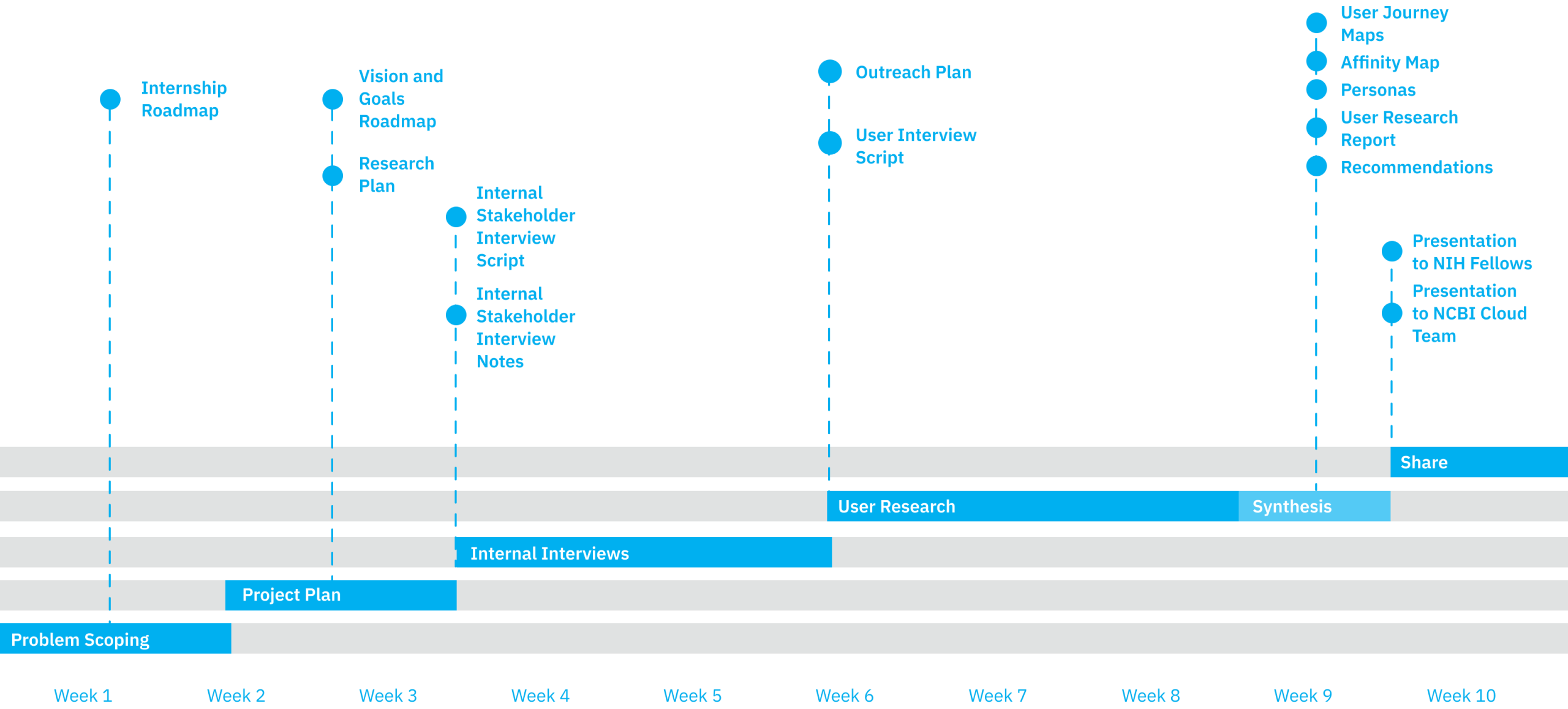
Cloud Curious

Majority segment, interested in cloud but lack technical expertise to be onboarded readily, made up of biologists, researchers

The Vision

Empathizing with New Users

Understand the needs of cloud curious users to facilitate their adoption of our cloud data/tools.


Internship Roadmap
Vision and Goals Roadmap
Research Plan
Internal Stakeholder Interview Script
Internal Stakeholder Interview Notes
Outreach Plan
User Interview Script
User Journey Maps
Affinity Map
Personas
User Research Report
Recommendations
Presentation to NIH Fellows
Presentation to NCBI Cloud Team
Share
User Research
Synthesis
Internal Interviews
Project Plan
Problem Scoping
Week 1
Week 2
Week 3
Week 4
Week 5
Week 6
Week 7
Week 8
Week 9
Week 10

Interviewing Internal Stakeholders

What Do We Know?

Conducted 5 interviews with internal members of the team

Research Goals

1. Understand what's been done and how team perceives direction and progress
2. Compare user needs perceived by internal stakeholders and reconcile those with needs expressed by real world cloud curious users

Interviewing End Users

What Do We Want to Know?

Conducted 6 interviews with NCBI end users who attended past cloud related NCBI events.

They ranged from computational biologists, students to librarians.

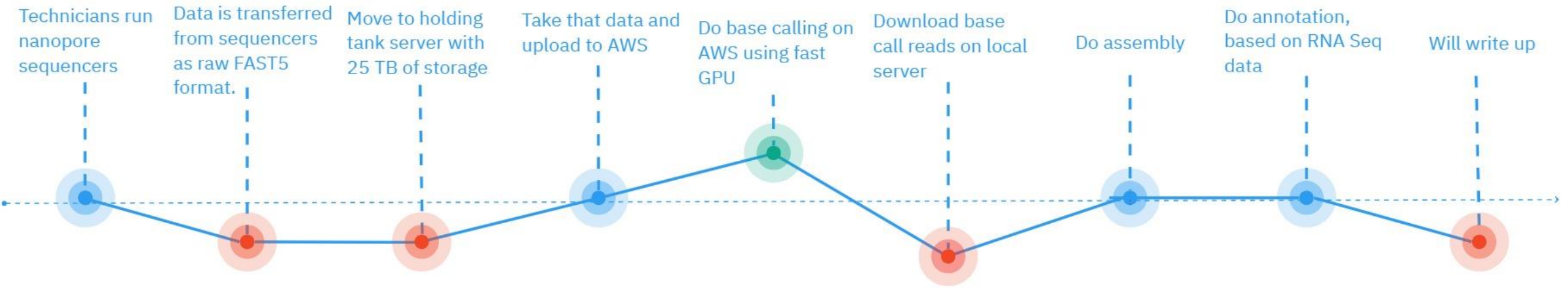
Research Goals

1. Glean insights into motivations for Cloud Curious users about why they are interested in learning more about working in the cloud.
2. Identify research needs of Cloud Curious users
3. Identify pain points of Cloud Curious users within their current workflow.
4. Identify blockers/obstacles of Cloud Curious users in adopting Cloud services into their workflow

Understanding Process

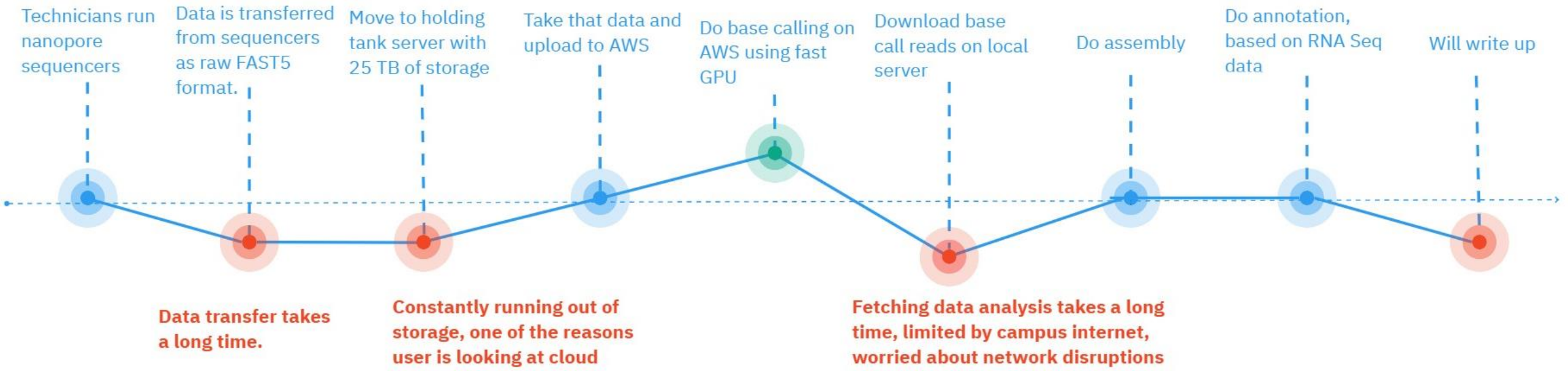
Current Workflow

I asked each subject to walk me through their current workflow and mapped their experience. Below is one of those workflows.



Understanding Pain Points

Areas of Opportunity



Subjects

- Systems Biology Grad Student
- Computational Biologist
- Technical Librarian
- Entomology PhD Candidate
- Bioinformatics Master's Student
- Research Chemist

300+

Quotes

26

Insights

Cloud Curious

Cloud Novice

Goals

1. Keep up with industry technology trends
2. Integrate cloud into my workflow by replacing aspects of my workflow where I am under severe local limitations
   - I'm interested in the compute resources available to me on the cloud
   - I'm interested in data storage on the cloud

Role

Studying bioinformatics in higher education (Master's or PhD)

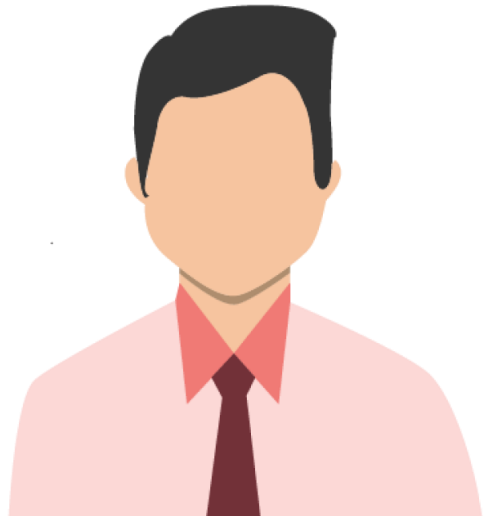
Working in research lab relating to bioinformatics or proteomics

Working within a team, will exchange helpful information between teammates

> **I feel like moving forward, it's all going to be cloud based technologies because there's so much data being generated that it's not going to be feasible just to store data physically.**

> **I knew even though I wasn't going to use it at that particular time, I will probably use it in the future, so it's better to really find out and even if I don't understand it completely, to have a bigger,clearer picture of it would be great.**

> **I'm an outside user. I would definitely not call myself a power user. It's something that I can use.**



Experience

I have utilized the cloud a couple times or occasionally utilize it for my work.

I am used to using local servers or a HPC if it's available to me.

I frequently use NCBI resources (SRA, BLAST, GEO Datasets, GenBank) in my research projects.

Explicit Needs

Work in an integrated workflow that combines local and cloud based tools

Have robust, consistent, and quick troubleshooting support

Know how to allocate resources for my specific workflows

Implicit Needs

Feel confident that community is increasingly adopting cloud

Receive validation that I am implementing what I know correctly

Feel that system is being honest with me before trusting system with my data

Frustrations

Constantly running into storage issues with my data

Downloading and transferring data takes a long time

COVID has exacerbated this frustration and made me reevaluate the severity to which local limitations interfere with my work

Friction when using on premises resources

Don't know best combination for allocating cloud compute resources to increase effectiveness and shave off cost

NIH U.S. National Library of Medicine
National Center for Biotechnology Information

NCBI

Cloud Curious

Cloud Naive

Goals

1. Keep up with industry technology trends
   - Can be prepared to answer potential questions regarding new tech
2. Support other groups of users who are running into a lot of frustrations due to local limitations
3. Learn more about what the cloud has to offer

Role

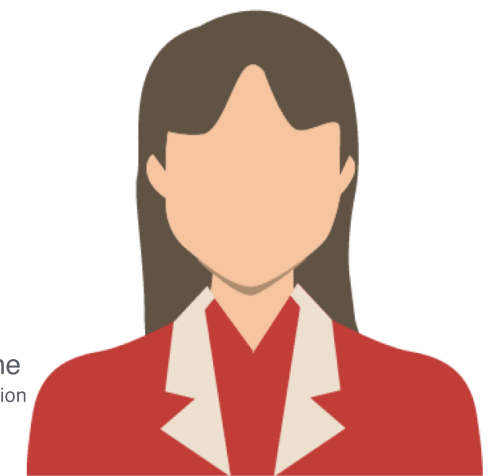
Technical librarian, product manager (housed within a bigger bioinformatics, biology department/company/institution)

Supporting institutional bioinformatics programming and research

Integrated into multiple teams

> **If you gave me a need I would not know what cost that would incur, which is a concern to me.**

> **I never spun up an AWS instance, but at a class I was in, they recommended we play around, and we had a trial but they said make sure you don't do too much or else we'll have to pay and that whole concept honestly just scared me away.**



Experience

I have never utilized the cloud before.

I have knowledge on NCBI resources (SRA, BLAST, GEO Datasets, GenBank, ClinVar) and maybe I have used NCBI resources before (not frequently)

Explicit Needs

Train in a safe space with no real world consequences to practice using cloud tools

Work in an integrated workflow that combines local and cloud based tools

Have robust, consistent, and quick troubleshooting support and feedback

Know who is responsible for incurred costs

Implicit Needs

Feel supported at each step fo the process within cloud workflow

Feel comfortable navigating the cloud's self-service model.

Avoid feeling overwhelmed by potential use cases when first onboarding onto the cloud.

Feel confident that the community is increasingly adopting the cloud.

Frustrations

Desire to implement some cloud infrastructure is blocked by general lack of technical expertise

Intimidated by the complexity of the cloud self-service model (cost is not transparent to me)

Afraid of unintentionally making a mistake that will affect cost or affect data

Do not know who is expected to pay for the cloud services within the context of my institution or organization

Recommendation

Speak Their Language

Provide cost documentation based on user segment - Define these segments by adding short descriptors or short example use cases so that users themselves can identify what segment they fall under and they can **refer to the cost documentation that best fits their comfort level with the cloud**

Cloud Savvy → Probably won't need much help navigating cost as they have used the cloud extensively

Cloud Novice → Give helpful estimates based on common use cases from their current workflow

Cloud Naive → Provide very simplified options where customization of resources is minimized, create a "default" for users.

Why?

1. **Lack of clarity behind cloud self-service model**
2. **Bridge gap between commercial cloud and end users**
3. **User perception of their own cloud proficiency is highly relative**

Recommendation

Centralization

Provide **NCBI in the Cloud related documentation, resources, and updates in a central hub, make this central hub** highly visible and accessible from the NCBI homepage

Connect tool-specific cloud initiatives spaces to centralized space

Why?

1. **User overlap between NCBI tools**
2. **Decrease cognitive effort on user**
3. **Gives more confidence to user that NCBI will invest in cloud long-term**

Recommendation

Educational Outreach - Virtual Sandbox

Provide **a space for users to get used to working in the cloud**, familiarize themselves with the process, and explore the tools available. **This space should be separate from real world work and should be free of consequences such as cost**, perhaps playing with dummy datasets and working with temporary accounts.

Users should be able to pull up and review directions to workshops and activities at any time during activities or sessions.

Why?

1. **Users explicitly want this**
2. **Create trust between system and user**
3. **Alleviate user anxieties regarding data, bills**

NCBI in the Cloud

Chaeeun's Work

Everything After Chaeeun

Improved outreach and education initiatives, implementing changes to help the Cloud Curious onboard comfortably

Thank you.

This research was supported by the Intramural Research Program of the NIH, National Library of Medicine.

NCBI Team/NIH

Ravinder Eskandary

Lydia Fleischmann

Carl Leubsdorf

Guangfeng Song

Yuriy Skripchenko

Rana Morris

Core STRIDES Team

BLAST Team

Jessica Mazerik

Coding It Forward

Rachel Dodell

Chris Kuang

Ariana Soto

My Fellow Fellows

Mentors

Eric Chiu

Neille Ilel