The integrated *iSearch* suite of tools: delivering powerful analytics to NIH staff and the public

Office of Portfolio Analysis
DPCPSI/OD/NIH
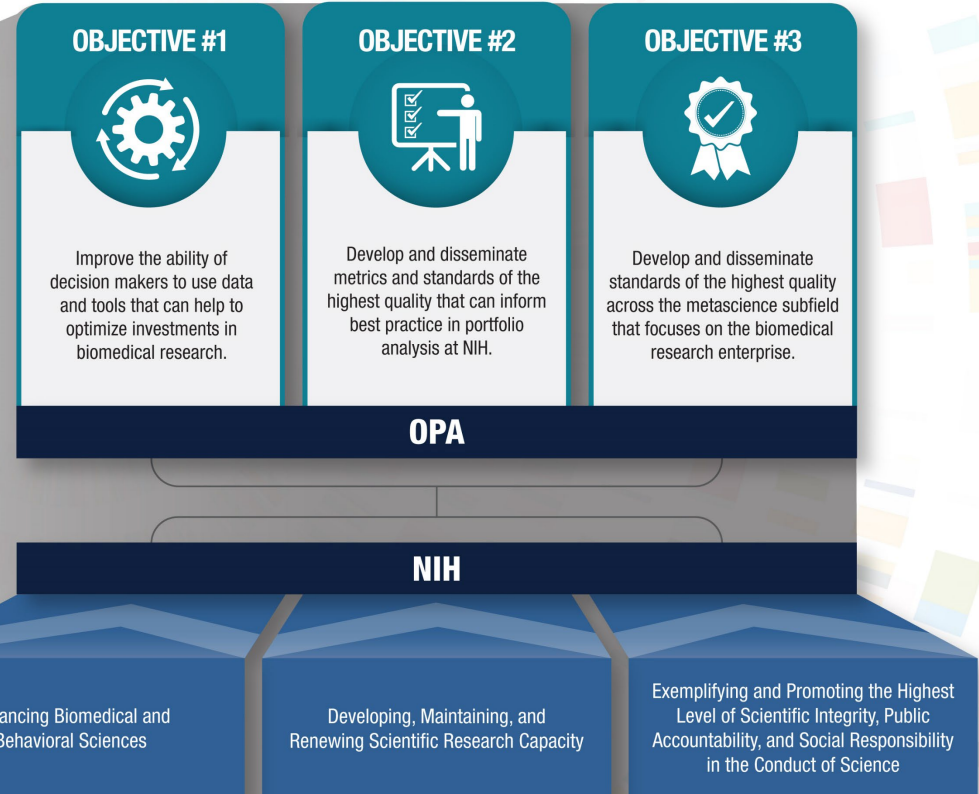
January 28, 2022

National Institutes of Health
Office of Portfolio Analysis

OFFICE OF PORTFOLIO ANALYSIS
STRATEGIC PLAN, FISCAL YEARS 2021–2025

OVERARCHING GOAL | To accelerate biomedical research by providing access to improved methods of data-driven decision making
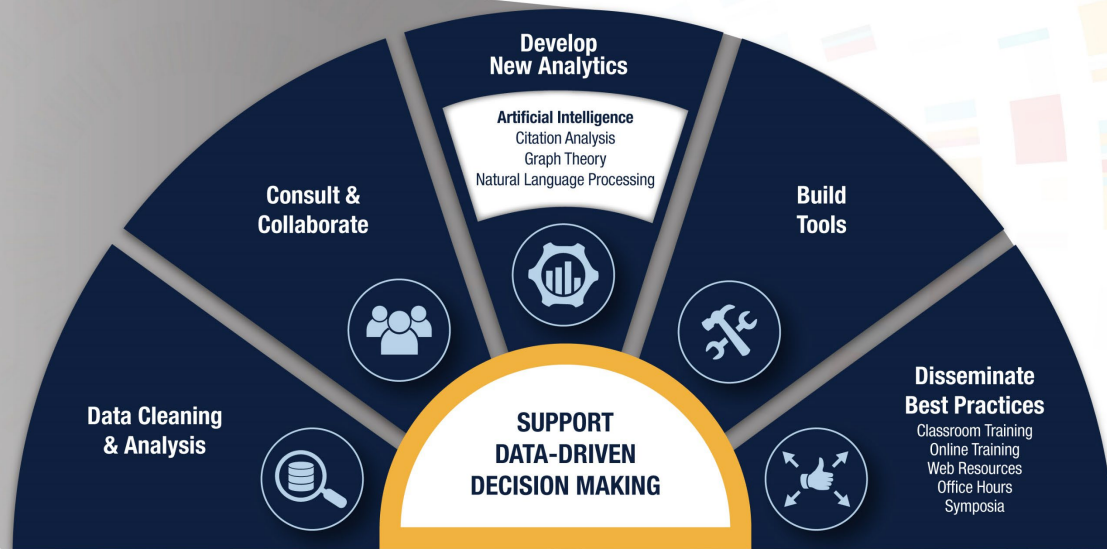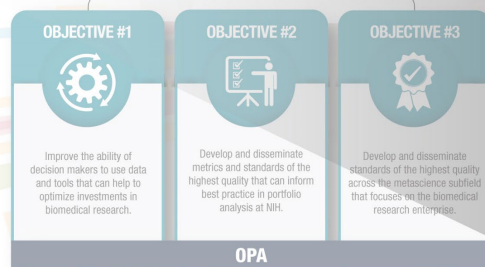
Develop New Analytics
Artificial Intelligence
Citation Analysis
Graph Theory
Natural Language Processing

Consult & Collaborate

Build Tools

Data Cleaning & Analysis

Disseminate Best Practices
Classroom Training
Online Training
Web Resources
Office Hours
Symposia

SUPPORT DATA-DRIVEN DECISION MAKING

OBJECTIVE #1

Improve the ability of decision makers to use data and tools that can help to optimize investments in biomedical research.

OBJECTIVE #2

Develop and disseminate metrics and standards of the highest quality that can inform best practice in portfolio analysis at NIH.

OBJECTIVE #3

Develop and disseminate standards of the highest quality across the metascience subfield that focuses on the biomedical research enterprise.

OPA

NIH

Advancing Biomedical and Behavioral Sciences

Developing, Maintaining, and Renewing Scientific Research Capacity

Exemplifying and Promoting the Highest Level of Scientific Integrity, Public Accountability, and Social Responsibility in the Conduct of Science

OFFICE OF PORTFOLIO ANALYSIS
STRATEGIC PLAN, FISCAL YEARS 2021–2025

OVERARCHING GOAL | To accelerate biomedical research by providing access to improved methods of data-driven decision making

Develop New Analytics

Artificial Intelligence
Citation Analysis
Graph Theory
Natural Language Processing

Consult & Collaborate

Build Tools

Data Cleaning & Analysis

Disseminate Best Practices
Classroom Training
Online Training
Web Resources
Office Hours
Symposia

SUPPORT DATA-DRIVEN DECISION MAKING

OBJECTIVE #1
Improve the ability of decision makers to use data and tools that can help to optimize investments in biomedical research.

OBJECTIVE #2
Develop and disseminate metrics and standards of the highest quality that can inform best practice in portfolio analysis at NIH.

OBJECTIVE #3
Develop and disseminate standards of the highest quality across the metascience subfield that focuses on the biomedical research enterprise.

OPA

NIH

Advancing Biomedical and Behavioral Sciences

Developing, Maintaining, and Renewing Scientific Research Capacity

Exemplifying and Promoting the Highest Level of Scientific Integrity, Public Accountability, and Social Responsibility in the Conduct of Science

The complete OPA strategic plan can be found on our website:
https://dpcpsi.nih.gov/opa/strategicplan

New *iSearch*

A comprehensive portfolio analysis platform
for NIH / HHS and public users

iSearch 3.0

- NIH and other HHS users
- Login required
- All grants data
- Higher export limits
- Additional features, e.g., user preferences
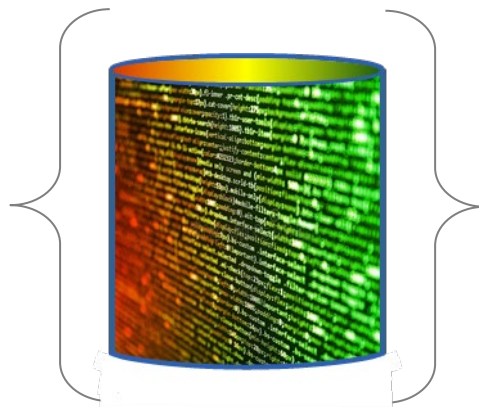
iSearch Analytics 1.0

- Public users
- Login optional
- Funded grants data only
- Some limits on anonymous users, e.g., export limits

Goal of today's presentation

Seek feedback about new *iSearch*
and our development processes

01 *iSearch* **Vision**

02 **Requirements gathering & design**

03 **Workstreams in progress**

04 **Summary**

01

iSearch Vision

iSearch Vision

An intuitive, user-centric analytics platform

Streamlined

Integrates key functions currently available across a range of OPA tools into one comprehensive toolkit

Comprehensive

Delivers comprehensive portfolio analysis for a wide range of users, from data scientists to the casual trend observer

Configurable

Enables tailored views and visualizations by user, organization, location, or topic

Scalable

Scales to capacity for high-demand usage

Integrate with existing NIH tools

Uniquely positioned to support metascience analytics
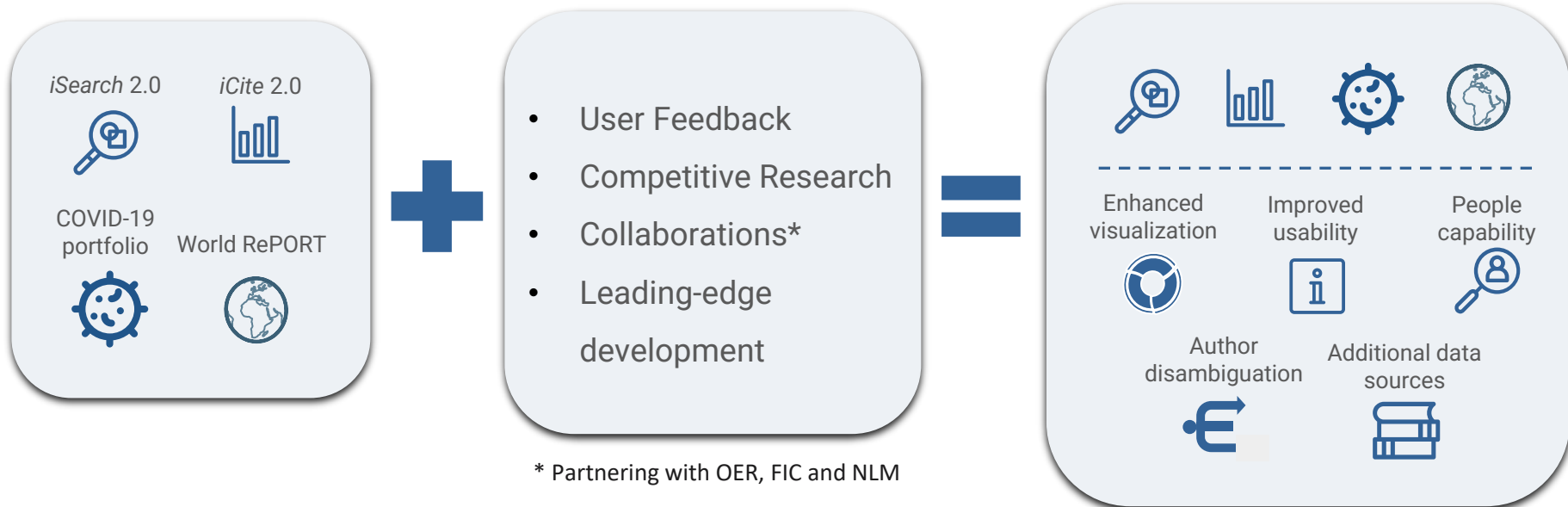
iSearch 3.0 / iSearch Analytics

iSearch 2.0   *iCite* 2.0

COVID-19 portfolio   World RePORT

+

- User Feedback
- Competitive Research
- Collaborations*
- Leading-edge development

* Partnering with OER, FIC and NLM

=

Enhanced visualization   Improved usability   People capability

Author disambiguation   Additional data sources

Reimagine topic visualization

Current visualization

Current *iSearch* visualization
(Lingo3G/Foamtree)

NIGMS FY2020: 4882 grant applications

Reimagine topic visualization

Current and planned visualization

Current *iSearch* visualization
(Lingo3G/Foamtree)

NIGMS FY2020: 4882 grant applications

AI-labeled universal word2vec clusters
(uClusters)
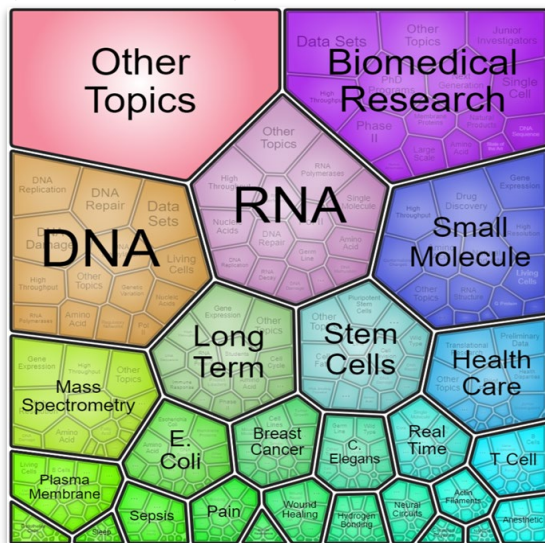
NIGMS FY2020: 4882 grant applications

Enhance all records with person disambiguation

Doing accurate person-level searches is necessary to avoid major errors in capturing a scientist's body of work

Author and publication features

Author / publication neural network

Hierarchical agglomerative clustering

Disambiguated person cluster enrichment
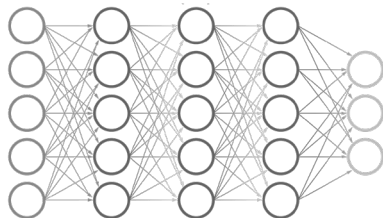
Author / publication entries
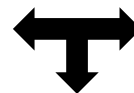
NIH applicant records

A recent OPA preprint describes this disambiguation method:
The effect of mentee and mentor gender on scientific productivity of applicants for NIH training fellowships
https://www.biorxiv.org/content/10.1101/2021.02.02.429450v1

National Institutes of Health
Office of Portfolio Analysis

Enhance all records with person disambiguation

Accurate person-level searches will enable users to avoid major errors in identifying a scientist's body of work

Real-world example

Is the author "AK Stewart" on these papers the same person or two (or three) different people?

	RCR	APT
Improvement in Overall Survival With Carfilzomib, Lenalidomide, and Dexamethasone in Patients With Relapsed or Refractory Multiple Myeloma. Siegel DS, Dimopoulos MA, Ludwig H, Facon T, Goldschmidt H, Jakubowiak A, San-Miguel J, Obreja M, Blaedel J, Stewart AK. J Clin Oncol. 2018 Mar 10;36(8):728-734. doi: 10.1200/JCO.2017.76.5032. Epub 2018 Jan 17. PMID: 29341834 Clinical Trial.	10.0	95%
Comparison of cases captured in the national cancer data base with those in population-based central cancer registries. Lerro CC, Robbins AS, Phillips JL, Stewart AK. Ann Surg Oncol. 2013 Jun;20(6):1759-65. doi: 10.1245/s10434-013-2901-1. Epub 2013 Mar 9. PMID: 23475400	7.0	95%
Metabolomics-Guided Discovery of Microginin Peptides from Cultures of the Cyanobacterium Microcystis aeruginosa. Stewart AK, Ravindra R, Van Wagoner RM, Wright JLC. J Nat Prod. 2018 Feb 23;81(2):349-355. doi: 10.1021/acs.jnatprod.7b00829. Epub 2018 Feb 6. PMID: 29405714	1.1	5%

Enhance all records with person disambiguation

Accurate person-level searches will enable users
to avoid major errors in identifying a scientist's body of work

		RCR	APT

Real-world example

Alexander Keith Stewart

Improvement in Overall Survival With Carfilzomib, Lenalidomide, and Dexamethasone in Patients With Relapsed or Refractory Multiple Myeloma.
Siegel DS, Dimopoulos MA, Ludwig H, Facon T, Goldschmidt H, Jakubowiak A, San-Miguel J, Obreja M, Blaedel J, Stewart AK.
J Clin Oncol. 2018 Mar 10;36(8):728-734. doi: 10.1200/JCO.2017.76.5032. Epub 2018 Jan 17.
PMID: 29341834   Clinical Trial.

10.0   95%

Andrew Kenneth Stewart

Comparison of cases captured in the national cancer data base with those in population-based central cancer registries.
Lerro CC, Robbins AS, Phillips JL, Stewart AK.
Ann Surg Oncol. 2013 Jun;20(6):1759-65. doi: 10.1245/s10434-013-2901-1. Epub 2013 Mar 9.
PMID: 23475400

7.0   95%

Allison Kathleen Stewart

Metabolomics-Guided Discovery of Microginin Peptides from Cultures of the Cyanobacterium Microcystis aeruginosa.
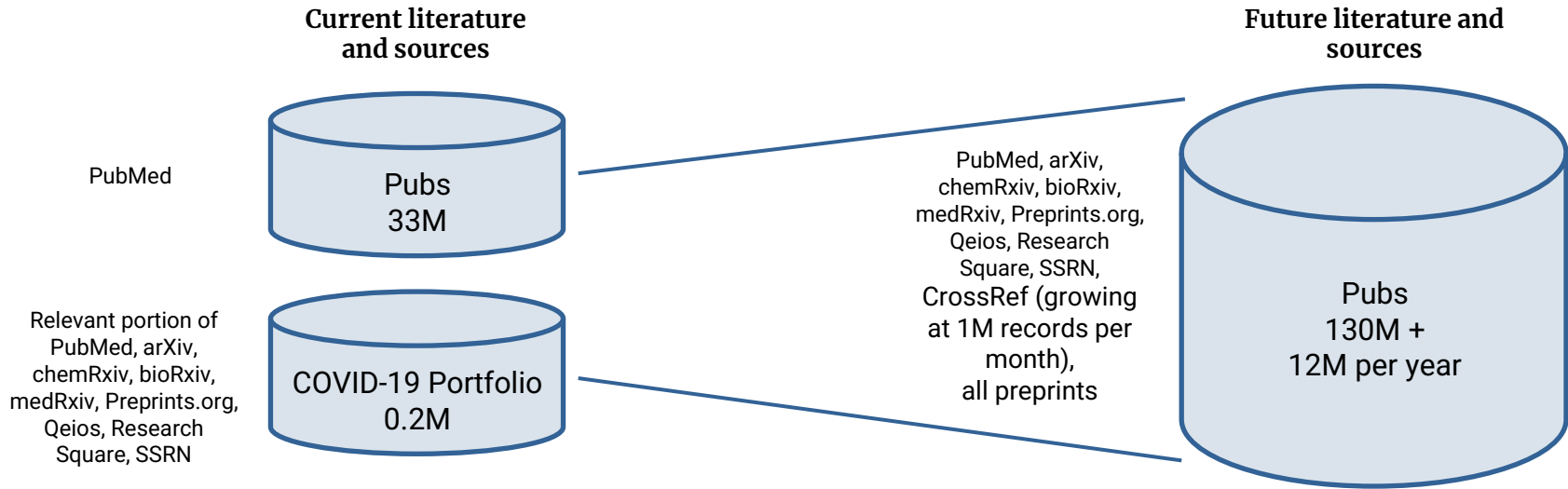Stewart AK, Ravindra R, Van Wagoner RM, Wright JLC.
J Nat Prod. 2018 Feb 23;81(2):349-355. doi: 10.1021/acs.jnatprod.7b00829. Epub 2018 Feb 6.
PMID: 29405714

1.1   5%

National Institutes of Health
Office of Portfolio Analysis

Greatly expand literature coverage

Coverage beyond PubMed to include
a broad range of scientific fields

Current literature and sources

Future literature and sources

PubMed

Pubs
33M

Relevant portion of PubMed, arXiv, chemRxiv, bioRxiv, medRxiv, Preprints.org, Qeios, Research Square, SSRN

COVID-19 Portfolio
0.2M

PubMed, arXiv, chemRxiv, bioRxiv, medRxiv, Preprints.org, Qeios, Research Square, SSRN, CrossRef (growing at 1M records per month),
all preprints

Pubs
130M +
12M per year

02

Requirements gathering & design

National Institutes of Health
Office of Portfolio Analysis

UX design process

1 Discovery & Research

2 Synthesize & Build Design Strategy

3 Placement & Layout Design

4 Execute & Validate

Discovery

Requirements gathering

Research

Competitive analysis of existing relevant tools to assess their strengths and weaknesses

Empathize

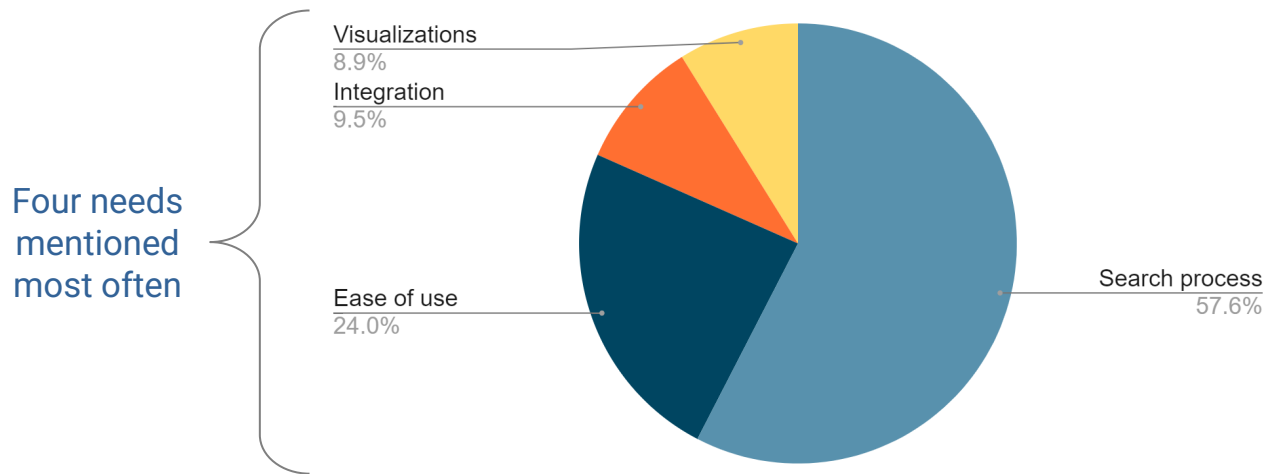Empathize with users to begin to understand their needs, motivations, and expectations

Interviews

User interviews to identify issues with functionality and areas of improvement

Research summary

What do NIH users care about most?

NIH surveys (March 2021 through June 2021)

Visualizations
8.9%

Integration
9.5%

Four needs
mentioned
most often

Ease of use
24.0%

Search process
57.6%

Total responses: 763

Research summary

Who are the prospective public users?

How can we best meet their analytical needs?

September 2021 Request for Information (RFI) public responses

A total of **6850** people responded to the RFI

1360 agreed to being contacted for follow-up

Respondent categories

Member of the public

Industry

Health professional

Government

Advocacy group

Non-profit

Other

Academia

Category	Percentage
Academia	89.2%
Industry	3.58%
Government	3.25%
Non-profit	2.34%
Health professional	0.72%
Other	0.53%
Member of the public	0.35%
Advocacy group	0.07%
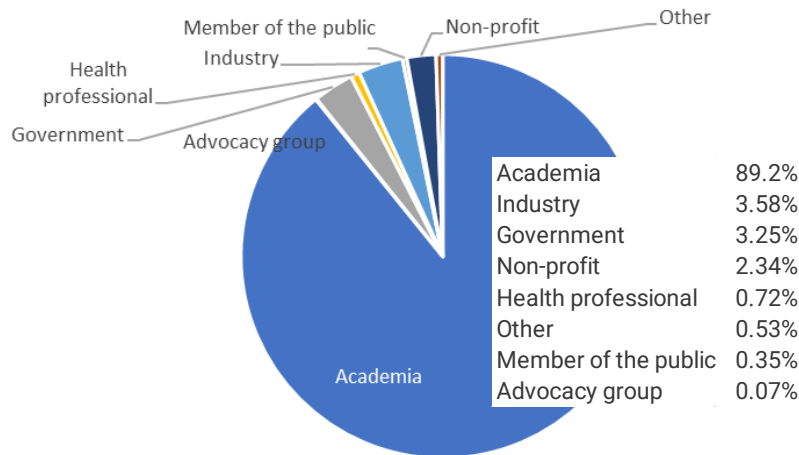
Interviews & User testing

- NIH Staff
- Staff of other government agencies
- Public RFI respondents who agreed to be contacted

NIH National Institutes of Health
Office of Portfolio Analysis
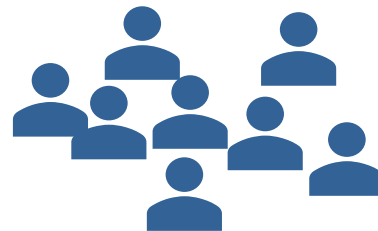
Leveraging requirements gathering

Optimizing support for analytics

Include strengths of existing tools

- Advanced search features
- Summary view of abstracts
- "Similar articles" feature
- Alerts for specific topics
- Centralized and accessible information
- "Cited by" and "Citing" functionality

Satisfy common needs

- Flexible search process
- Integration across databases
- Easy to use
- Quick results
- Visualizations
- Reliable

iSearch

Avoid weaknesses of existing tools

- Paywalls blocking access to resources
- Lack of name disambiguation
- Lack of ability to construct complex queries
- Lack of (or suboptimal) data visualization
- Incomplete coverage of the literature
- Inflexible/lack of filters

iSearch user personas

Based on goal pathways

1

EXPLORERS

Discover engaging information through minimal interactions

2

SEARCHERS

Learn information on the topic they came to research

3

INTERMEDIATE USERS

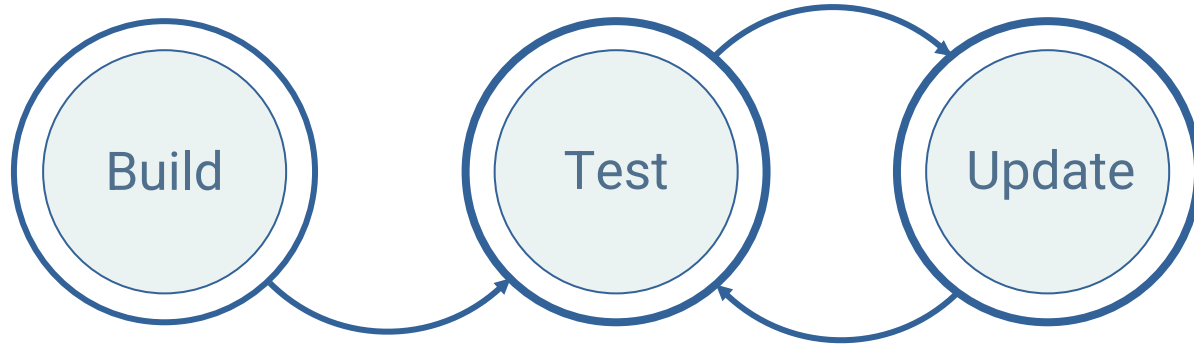Identify/analyze a collection of documents they are interested in

4

POWER USERS

Do a deep dive on specific portfolio topics of interest

Level of engagement

Execution and validation

Using usability testing

03

Workstreams in progress

iSearch architecture

Structure and benefits

User feedback

Fast results

Reliable

Trustworthy

Intuitive

Implementation

Cloud-based

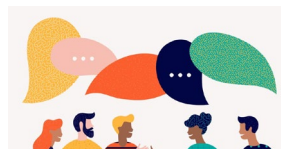Highly reliable and available

Automatically scales with demand

Incorporates lead-edge technology
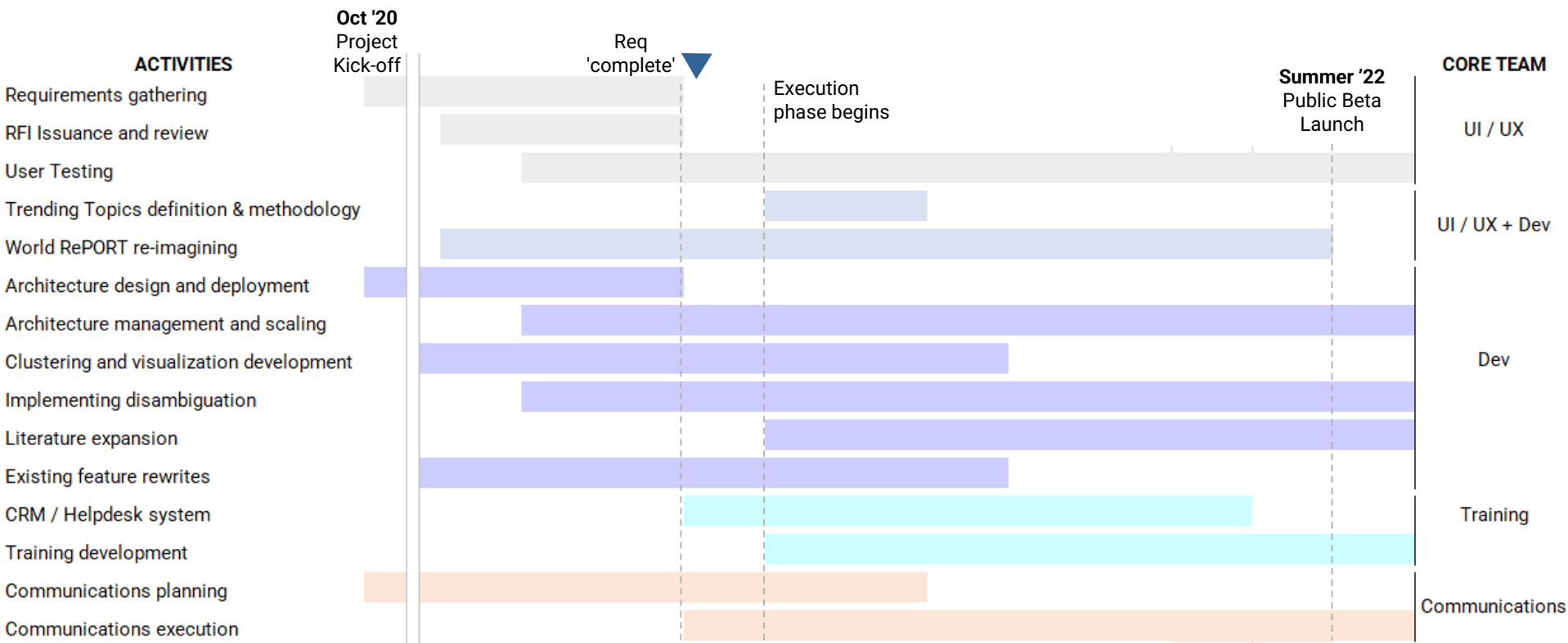
Communications Plan

Goals

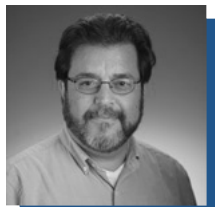Raise awareness of *iSearch* to new users

Foster engagement with new and existing users of *iSearch*

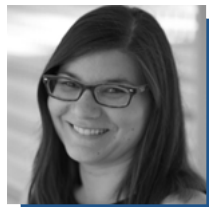| Audience	Key Messages	Delivery methods	
Explorers Searchers	Big picture: hook audience with intriguing visualizations of data and share potential uses of the tool	❖ Engaging questions ❖ Example scenarios ❖ Infographics	❖ Shareable social media images ❖ Videos
Intermediate Users Power Users	Technical details: share how *iSearch* 3.0 is different from other industry systems (e.g., QVR, RePORTER, Dimensions)	❖ Email campaign ❖ 1-pager summary ❖ User guide update	❖ Seminar with Q&A ❖ White paper
All personas	Announce new tool and where to access it	❖ Press release via NIH ❖ IC newsletters, e.g. Fogarty ❖ OPA website	❖ *iSearch* Landing page ❖ Scientific communities

NIH National Institutes of Health
Office of Portfolio Analysis

Major milestones

ACTIVITIES		CORE TEAM
Requirements gathering		
RFI Issuance and review		UI / UX
User Testing		
Trending Topics definition & methodology		UI / UX + Dev
World RePORT re-imagining		
Architecture design and deployment		
Architecture management and scaling		
Clustering and visualization development		
Implementing disambiguation		Dev
Literature expansion		
Existing feature rewrites		
CRM / Helpdesk system		Training
Training development		
Communications planning		Communications
Communications execution		

Oct '20 Project Kick-off

Req 'complete'

Execution phase begins

Summer '22 Public Beta Launch

iSearch team

George Santangelo, Ph.D.
Director

Rebecca Meseroll, Ph.D.
Special Advisor to the Director

Paula Fearon, Ph.D.
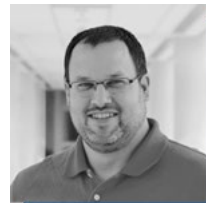**Project Manager**

Krista Callender, M.B.A.
IT Project Manager

Emily Lampe
UX/UI Designer

Anthony Cheu
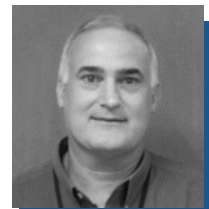**UX Researcher / BA**

Matt Davis, M.S.
Software Engineer

Will Millman
Software Engineer

Payam Meyer, M.S.
Software Engineer

Karl Doughty
Software Engineer

Chuck Lynch, Ph.D.
COR/ Cloud Specialist

Matt Perkins
Training Director and Policy Analyst

Shannon Davis
Technical Product Manager

Winnie Wong, Ph.D.
Communications Officer

Michael Cheetham
Senior Science Policy Analyst

NIH National Institutes of Health
Office of Portfolio Analysis

04

Summary

Summary

The journey to new *iSearch*

1

User feedback

Surveys, RFI and interviews captured input from current and prospective users

2

Standard processes

Industry accepted processes leveraged to ensure an efficient and accurate delivery of features

3

Iterative development

Basic designs and architecture are in place as the foundation for a sophisticated yet intuitive analytics platform