Board of Scientific Counselors
NCHS Strategic Planning

Brian C. Moyer, Ph.D., NCHS Director

January 27, 2021

Strategic Planning Process

Case Studies and Beyond

Remote Data Access

Internal Communications

Disclosure Risk Assessment

NCHS: State of Data Science

Board of Scientific Counselors

1/27/2021

Chief Data Scientist

Travis Hoppe, PhD

What is "data science"?

NIST: *Data science is the empirical synthesis of actionable knowledge from raw data through the complete data lifecycle process*

What is "data science"?

NIST: *Data science is the empirical synthesis of actionable knowledge from raw data through the complete data lifecycle process*

What does data science mean for NCHS?

Scale
Sustainability
Automation
Subject matter expertise
Collaboration
Culture change

How do we get there?

Three programs

Strategic planning: Data Science Case Study
Guide NCHS over next 5-10 years

Data Modernization Initiative (DMI)
Management of CDC-wide effort to modernize

Innovation group
Survey, inspire, and encourage Data Science projects

Staff-run Innovation Breakout Sessions

AI & Machine Learning

Data Modernization

Engagement

Privacy & Disclosure

API implementation

Consolidated Statistical Platform (CSP)

AI: Artificial Intelligence
CSP: Consolidated Statistical Platform
API: Application Protocol Interface

NCHS Data Science Inventory

Hi Travis, when you submit this form, the owner will be able to see your name and email address.
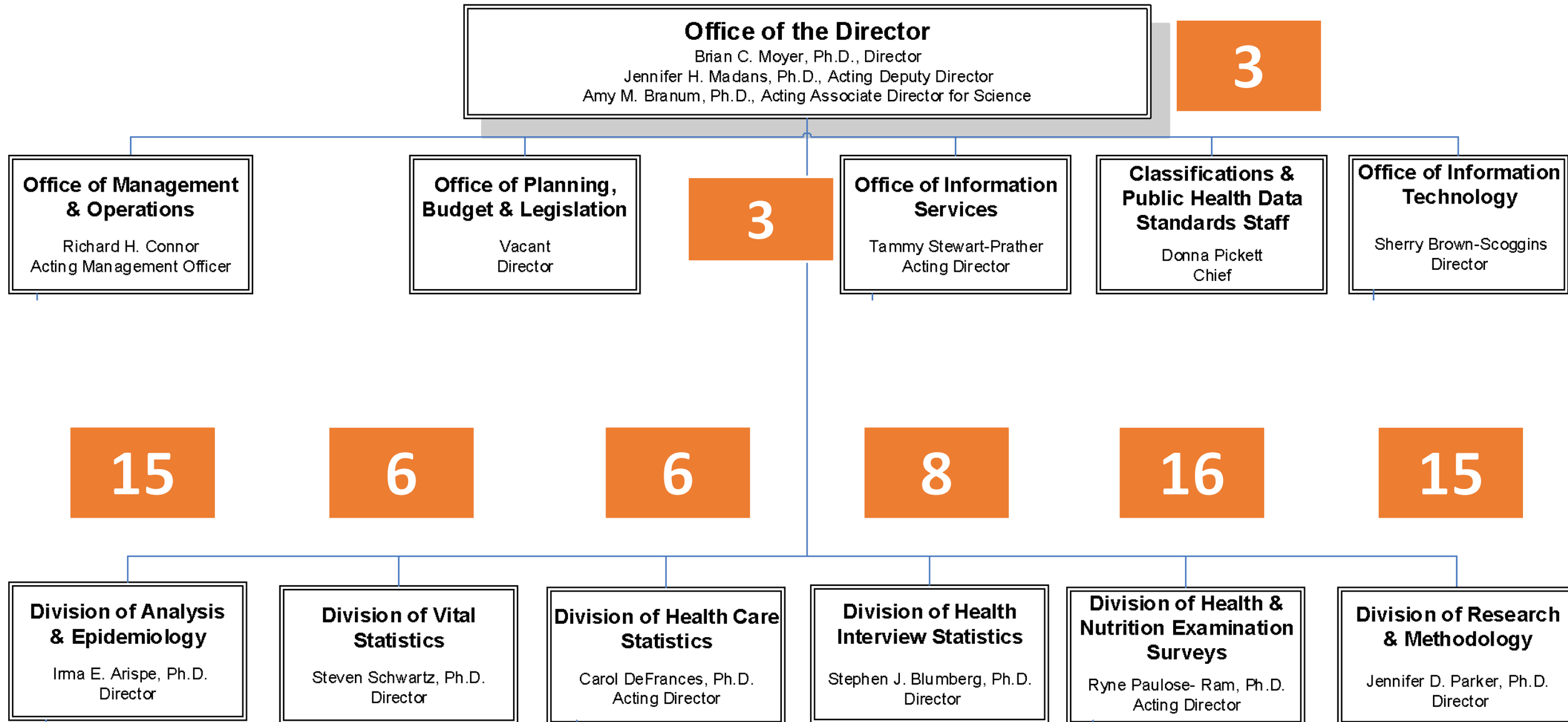
The purpose of this form is to gather information on NCHS data science projects and activities. This information will be summarized to facilitate conversation about potential partnerships and collaborations. For questions about form content, please reach out to Chad Heilig (cqh9@cdc.gov). For questions about form logistics, please reach out to Jessly Joy (oys4@cdc.gov). Please complete one form entry for each data science related project or program within your

NCHS 2020 Innovation survey

- Broaden perspective on the range of activities within the Center
- Learn about concrete projects underway
- Identify new cross-center collaborations
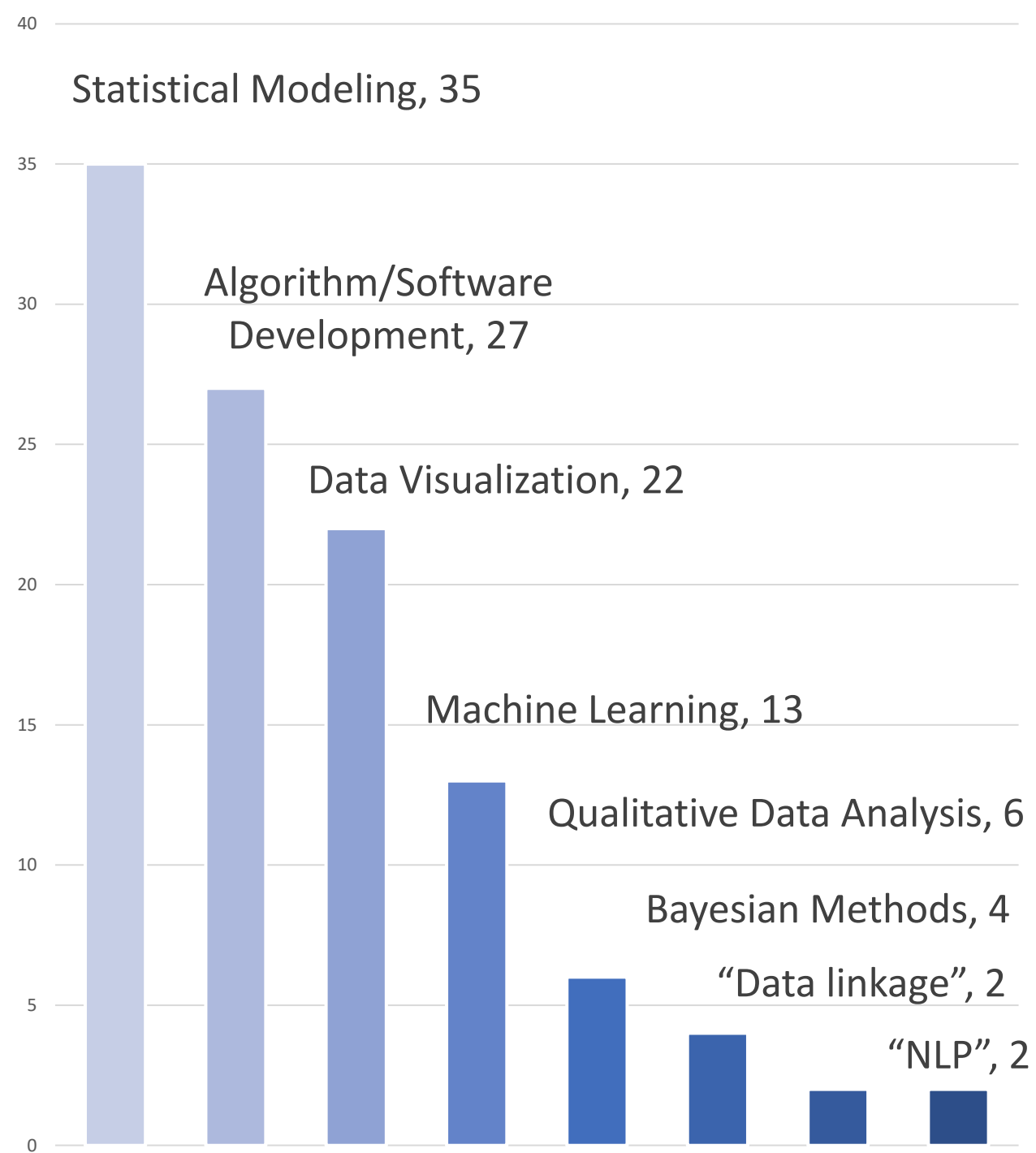- Identify ongoing and prospective R&D within the Center

Responses per division (72 total)

Office of the Director
Brian C. Moyer, Ph.D., Director
Jennifer H. Madans, Ph.D., Acting Deputy Director
Amy M. Branum, Ph.D., Acting Associate Director for Science

3

Office of Management & Operations

Richard H. Connor
Acting Management Officer

Office of Planning, Budget & Legislation

Vacant
Director

3

Office of Information Services

Tammy Stewart-Prather
Acting Director

Classifications & Public Health Data Standards Staff

Donna Pickett
Chief

Office of Information Technology

Sherry Brown-Scoggins
Director

15

6

6

8

16

15

Division of Analysis & Epidemiology

Irma E. Arispe, Ph.D.
Director

Division of Vital Statistics

Steven Schwartz, Ph.D.
Director

Division of Health Care Statistics

Carol DeFrances, Ph.D.
Acting Director

Division of Health Interview Statistics

Stephen J. Blumberg, Ph.D.
Director

Division of Health & Nutrition Examination Surveys

Ryne Paulose- Ram, Ph.D.
Acting Director

Division of Research & Methodology

Jennifer D. Parker, Ph.D.
Director

Data Analytics

multi-response question,
single responses omitted

Data Science is already active in the Center

Statistical Modeling, 35

Algorithm/Software Development, 27

Data Visualization, 22

Machine Learning, 13

Qualitative Data Analysis, 6

Bayesian Methods, 4

"Data linkage", 2

"NLP", 2

Activity Type

multi-response question,
single responses omitted

Viz, linkage, privacy,
pipelines, and more!
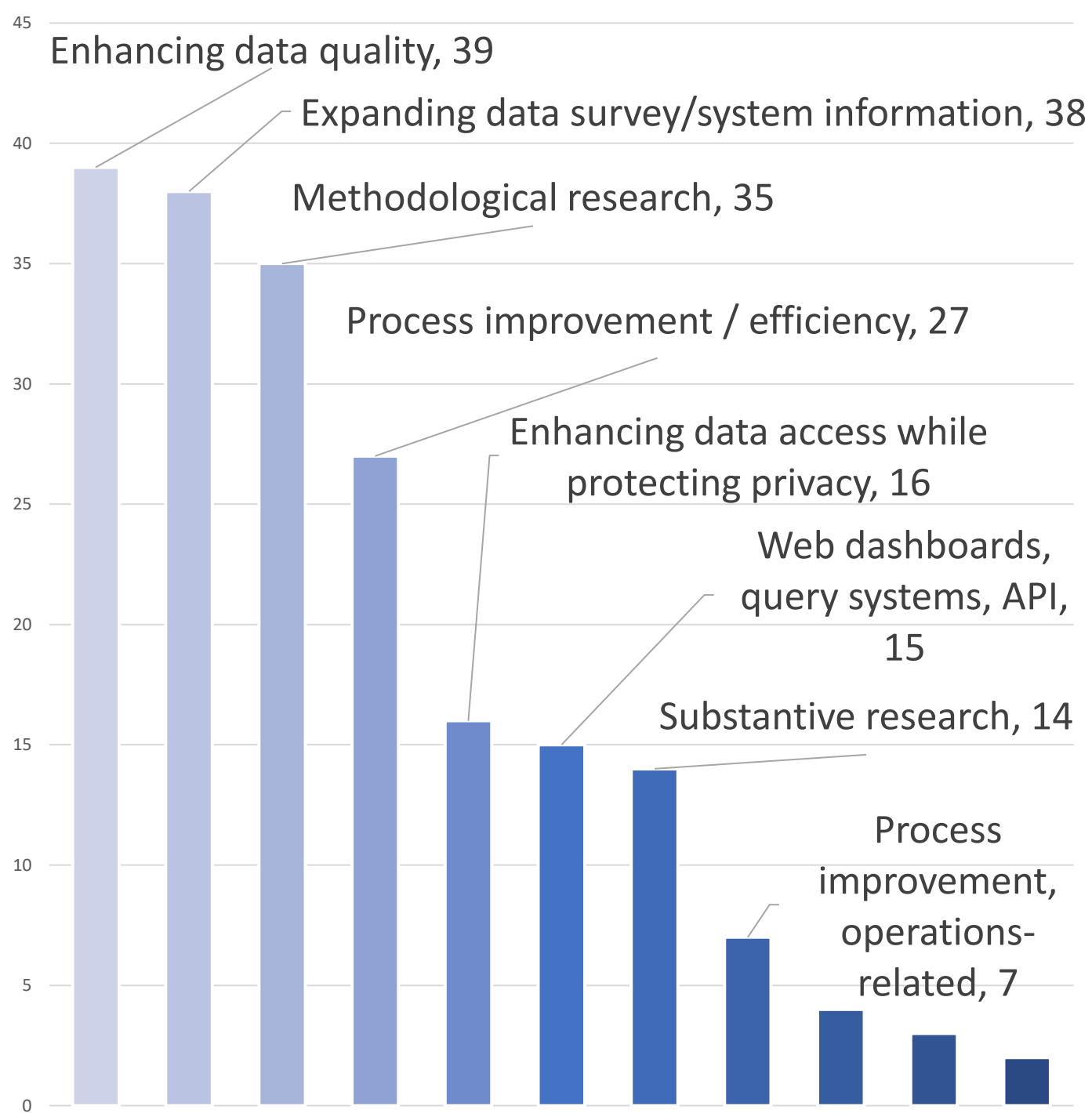
Enhancing data quality, 39
Expanding data survey/system information, 38
Methodological research, 35
Process improvement / efficiency, 27
Enhancing data access while protecting privacy, 16
Web dashboards, query systems, API, 15
Substantive research, 14
Process improvement, operations-related, 7
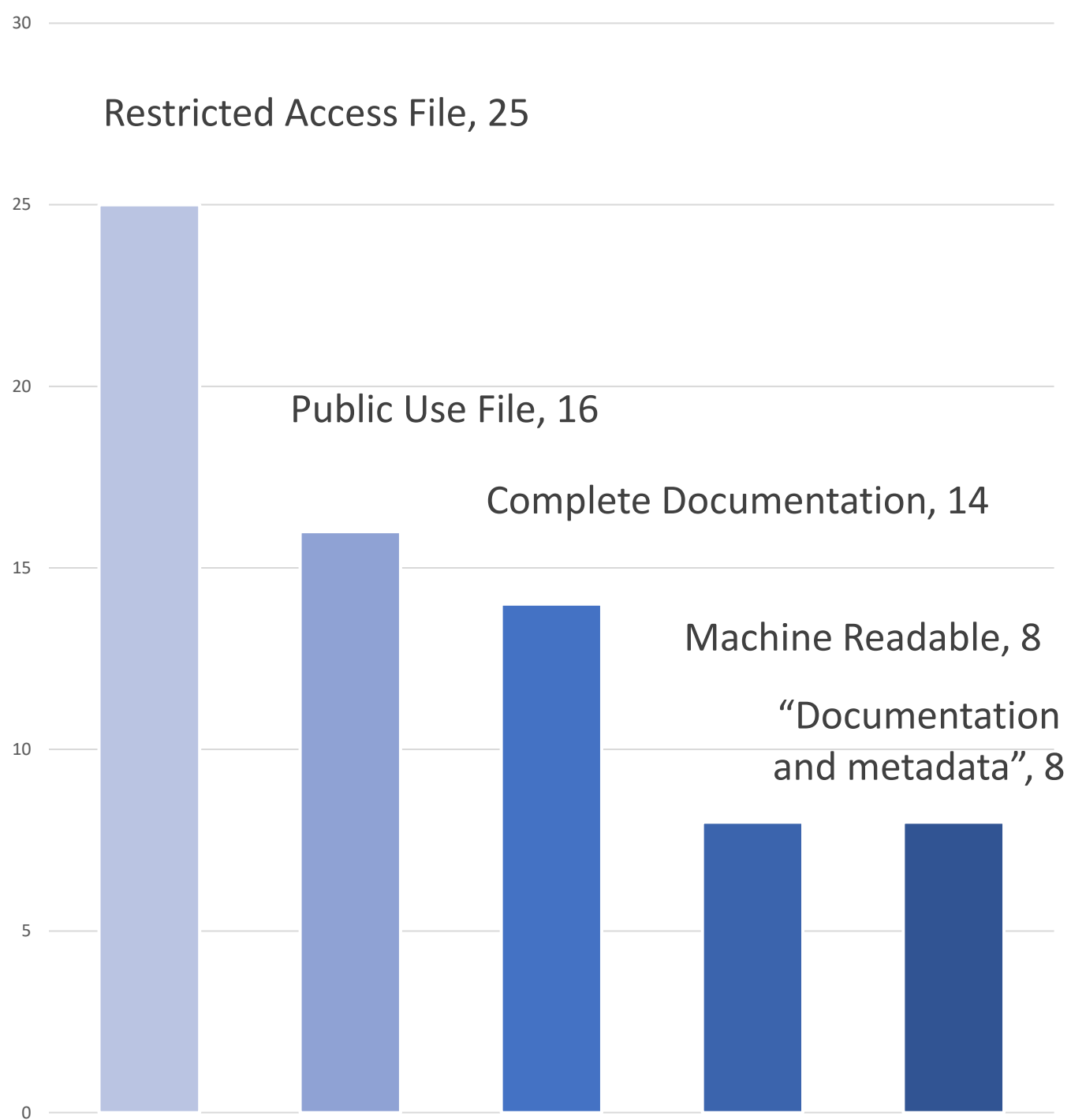
Active highlights

Health, US Redesign

Topic based taxonomy to organize data, report key findings across NCHS major data systems, and a development of a robust API.

Small Area Estimation for Cancer Risk

Hierarchical Bayesian model to produce county-level estimates of cancer screening rates, current/former smoking prevalence, and hysterectomy prevalence. Combines information from the National Health Interview Survey (NHIS) and the Behavioral Risk Factors Surveillance Survey (BRFSS).

Health, United States -2018-

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES
Centers for Disease Control and Prevention
National Center for Health Statistics

Identification of Opioid-Involved Health Outcomes

Make use of all available structured and unstructured data from the National Hospital Care Survey (NHCS), the National Death Index (NDI), and the National Vital Statistics System restricted mortality data (NVSS-M) to identify specific opioids (e.g., fentanyl and heroin) involved in outcomes such as drug-related hospital visits and drug poisoning deaths.

Natural Language Processing to classify drowning deaths

Ascertain places of drowning using literal text fields from death certificate records received from the National Vital Statistics System, Mortality (NVSS-M) using modern Natural Language Processing (NLP).

Nowcasting and Geospatial Analyses

Tools and analytic methods from NVSS (Vital Statistics) data to provide more accurate estimates of COVID-19 mortality at the state and county-level in near-real time, mitigating the impact of reporting delays.

Enhancing the identification of COVID-19 deaths

Identify COVID-19 deaths from unstructured text to enhance COVID-19 surveillance using the provisional mortality data.

Expanding beyond traditional statistics

New areas of growth

- Predictive analytics: small area estimations, nowcasting, …
- Classification tasks: EHR, RANDS, …
- Remote Access: public APIs, data enclaves, …
- Named entity recognition: opioids, ICD-10 coding, …
- Differential privacy: synthetic data, federated learning, …
- New datasets: DXA, PAM, Retinal imaging, …
- Bibliometric analysis: NCHS publications, reports, and linked data, and social media engagement

NCHS: State of Data Science

Jennifer Parker

Director

Division of Research and Methodology

Three programs

Strategic planning: Data Science Case Study
Guide NCHS over next 5-10 years

Data Modernization Initiative (DMI)
Management of CDC-wide effort to modernize

Innovation group
Survey, inspire, and encourage Data Science projects

Strategic planning: Data Science Case Study

- Identify connections and opportunities of Data Science Innovations and possibilities with HHS priorities, agency goals, division requirements and the varied needs of our NCHS data users and stake-holders

- Promote grassroots ideas, NCHS-wide engagement and staff development

- Investigate barriers and solutions

- How can we manage risks?
  - New methods may depart from usual processes and be less transparent
  - Outputs may be sensitive to assumptions and require revisions or updates