Enhancing the impact infrastructure Working at scale, and working together

Neil Thakur, PhD
Office of Extramural Research
Thakurn@od.nih.gov
NIH

HRA
March 12, 2018

How do we fund great science?

- Impact
- Efficiency

Impact: Are we funding quality or reputation?

Artificial scarcity and Influence = Reputation

- Scarcity: limited positions imperfectly assigned (universities, journals, academic honors)
- Influence: JIF, citations, social media, press

Integrity, rigor and change = Quality

- Reproducibility
  - Shared data
  - Documented, validated and appropriate methods
- Breakthroughs (Patents?)
- Improved outcomes

Efficiency: using time and resources effectively

Complementary funding- we are part of a whole

- Not duplicative with other funders
- Not competitive

Selection through efficient due diligence, not red tape

- Effective peer-reviewer selection
- Thorough understanding of portfolio

Impact Infrastructure

The data model to measure impact and work efficiently

- People and their career trajectories
- Funding
- Research Products
- All the interconnections between and across people, funding, and products

Status quo: Fragmented systems, poor results

- **Duplicate data and wasted effort:** Researchers have to curate and combine data that is scattered across public and private sources- ORCID, SCOPUS, PubMed, RPPRs, Vivo, Trellis, etc., and must do this in multiple times in multiple systems.
- **Poor tracking and measurement**– Funders can't track their impacts on researcher careers, especially across different funders.
- **Inefficient research networks** – Researchers and associated groups do not use modern technology for networking and hiring (e.g., finding mentors, collaborators, employees, reviewers, etc.)
- **Bad incentives**- Current measures of research productivity do not adequately incentivize openness, rigor and impact. Current fragmentation in research and career data and reporting makes it difficult to implement new measures.

Goals for a better impact infrastructure

- Track funder impact
- Encourage development of better productivity measures and incentives
- Support collaboration, networking and expert locator services
- Maintain researcher control and privacy
- Reduce researcher burden

Solving at scale: Design or adoption challenge?

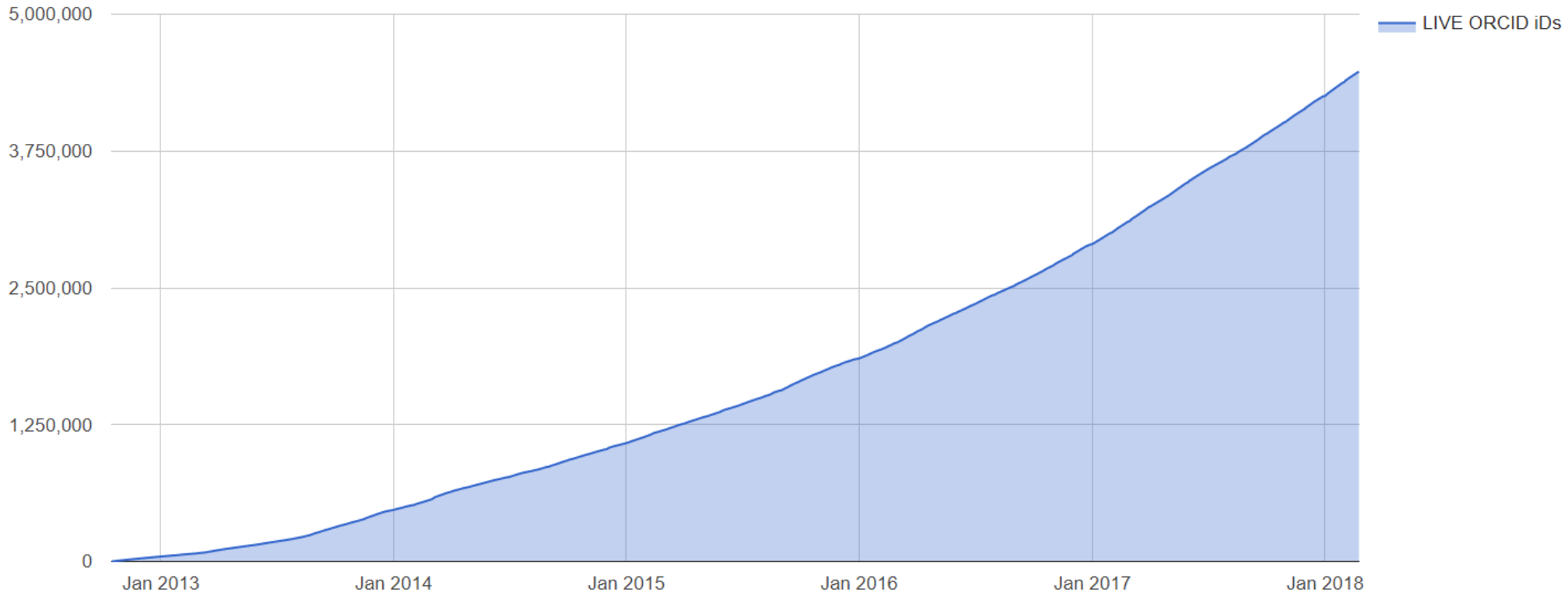
Do funders have the leverage to address many larger goals?

- Funders are small in scale
- Research funding is only one of many incentives and systems

Funder systems are not the burden, so silos are not the answer

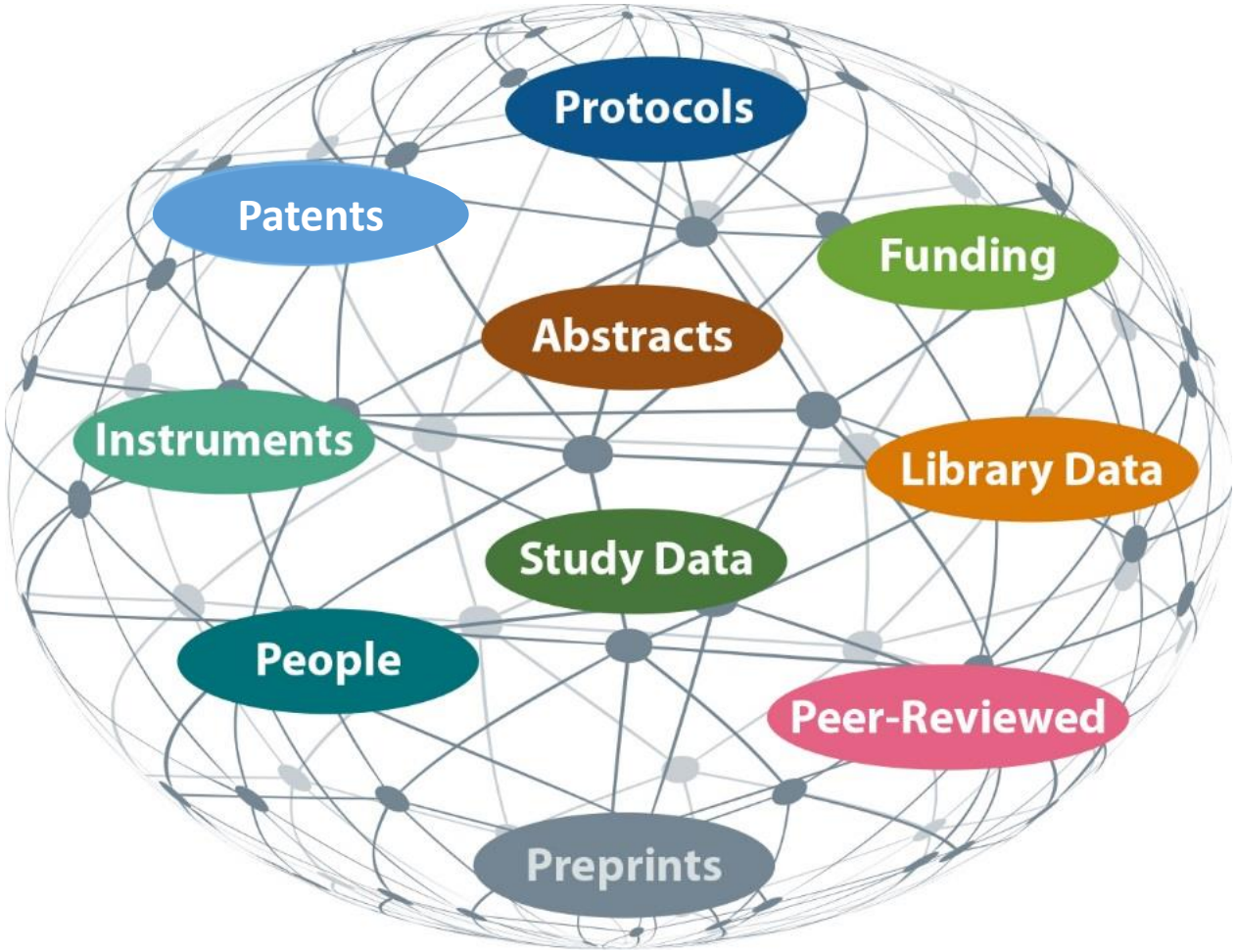
- FDP experience with profile data: Fragmentation, burden, inefficiency

How many users does your system have?



https://support.orcid.org/knowledgebase/articles/150557
https://orcid.org/statistics.

Create a comprehensive research impact infrastructure with unique identifiers



Link

- Products (RRID, DOI, ORCID)
- Funding (DOIs?)
- People (ORCID)
- Institutions (?)

Enable

- Burden reduction
- Impact analysis
- Metrics
- Innovation and economic growth

Persistent identifiers

ORCID

- A persistent unique identifier for researchers
- Helps track and validate people/product associations
- Over 4M users, supported by thousands of journals

Digital Object Identifiers (DOIs)

- Developed as a universal, persistent, overlay identifier
- Used as a universal article identifier supported by multiple publisher data systems
- Infrastructure for metadata, validation, citation tracking
- 63M articles, 11M books and book chapters, agency level funding identifiers, data sets, reviews, etc.

DOIs for funding (grants, contracts, etc)

Utilize the publications tracking infrastructure to track grants

- Better tracking of people across their careers and funding agencies
- More accurate identification of research products
- More robust data to identify potential reviewers and assess conflicts of interest
- Validation for grant /product associations

As an overlay, a universal funding number system for all funding agencies

- Provide a ‘common denominator’ funding identifier format to harmonize NIH’s grants system and contract system, and harmonize with other funders
- An inexpensive way for funding agencies to develop unique identifiers for their funding. Requires permanent location for funding information

https://www.crossref.org/community/funders/

ORBIT: ORCID Reducing Burden and Improving Transparency

ORCID will enhance their data model and 3rd party service integrations to:

- broaden connections to research and career data usually reported on CVs
- link researchers to funding and professional activities with verified and structured data
- serve as an open hub for other systems
- will also explore institutional identifiers

Goals

- **Reduce researcher burden** of applying for funds and maintaining multiple profiles
- **Track impact** of research and professional development through transparently-curated open data
- **Support collaboration and networking services** to build efficient and equitable markets for reviewers, collaborators, mentors, etc.
- **Maintain researcher control** of their own data and how it is used across platforms
- **Encourage development of better productivity measures and incentives**

https://orcid.org/content/orbit-project

Use Case: Better Measures

ORBIT aggregates by person…

- Products (DOIs, Etc)
- Funding (DOIs?)
- Institutions (institutional identifiers?)

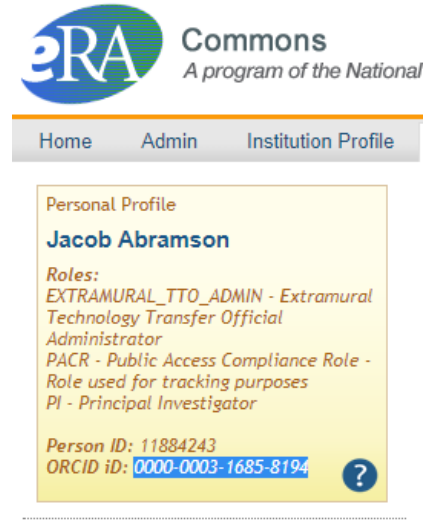
- **Product level metrics can be retained in metadata**
  - Relative citation ratio (https://www.ncbi.nlm.nih.gov/pubmed/27599104)
  - Openness? (licenses?)
  - Rigor? (badges?)

Product level metrics can aggregate to…

- Person level measures
- Award level measures
- Funding initiative level measures
- Institution level measures

ORCID integration with NIH systems

ORCID provides investigators with persistent digital identifiers and helps them track their research products



Phase 1: integration with SciENcv

- Link to ORCID in SciENcv and download ORCID citations into biosketches

Phase 2 (current): Allow ORCIDs in eRA profiles

- Facilitate data exchange, funding/ORCID linkages

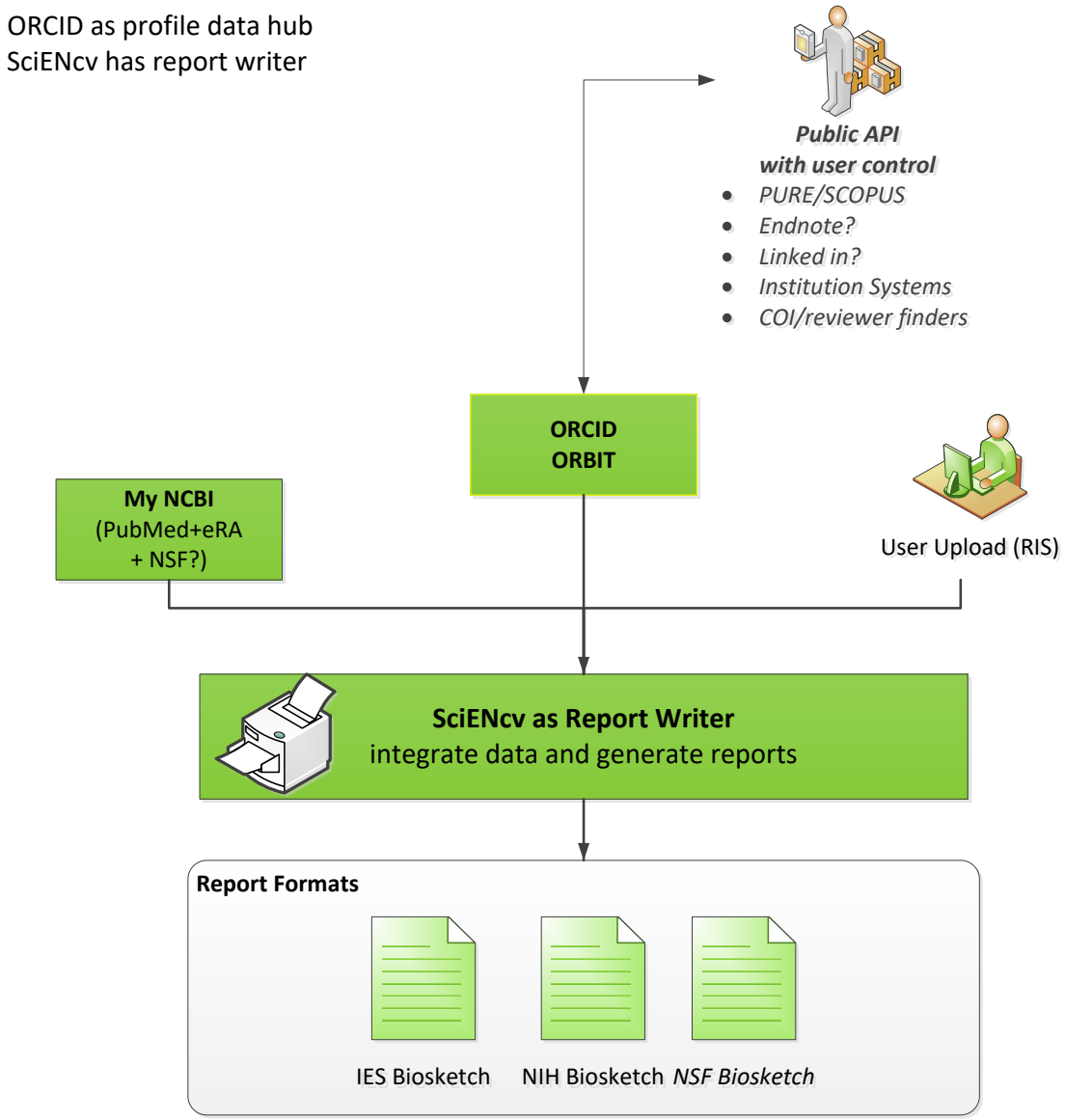
Phase 3 (future): Expand ORCID data model and integration with eRA

- Use ORCID data to automate other forms like Other Support, RPPR?
- Upload NIH data (funding, products, profile data) into ORCID?
- Use ORCID as a hub and interchange for all profile data, reducing burden for federal and private profile systems?

Additional use cases and information

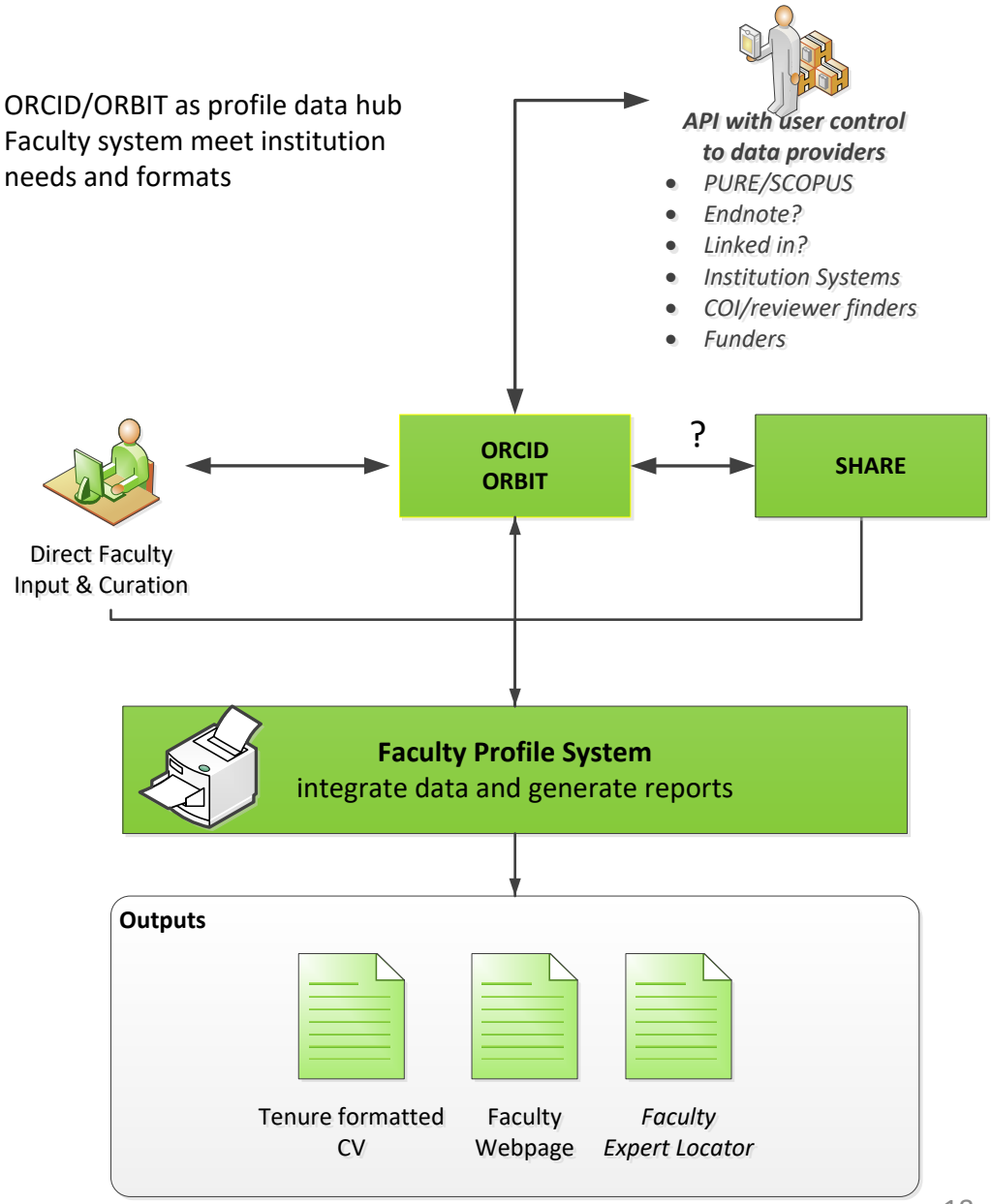
Use Case: Application Forms

- ORCID/ORBIT data hub
- SciENcv writes creates biosketches for NIH, NSF, ED
- User approval for data linkage
- Reduced burden, validation, structured data
- Scaling: eRA as 1/10th users of ORCID



Use case: Better university data

- ORCID/ORBIT integrates data streams for linked accounts
- Users can manage their data in the system they prefer
- Primary source of burden for PI profiles



SciENcv = Science Experts Network Curriculum Vitae

Vision- *Let investigators harvest their data from multiple systems to support funding applications, reporting and collaboration with less burden and complexity*

Goals

- **Reduce burden** of applying for federal funds and maintaining federal profiles
- **Track impact** of federal investments in science and scientist careers through scientist-curated data
- **Support collaboration and networking services** to find reviewers, collaborators, mentors, etc.

Products to date

- NIH biosketches, NSF biosketch, Ed IES biosketch
- Embedded XML
- Integration with ORCID, Fastlane, PubMed and eRA
- Bulk upload of citations from reference manager software
- Internal refinements: user testing, adopting agile software principles