


Ok, we shared the data.

Can we reuse it?

What does it take to reuse it?

Pasquetto, I. V., Borgman, C. L., & Wofford, M. F. (2019). Uses and reuses of scientific data: The data creators' advantage. Harvard Data Science Review, 1(2).



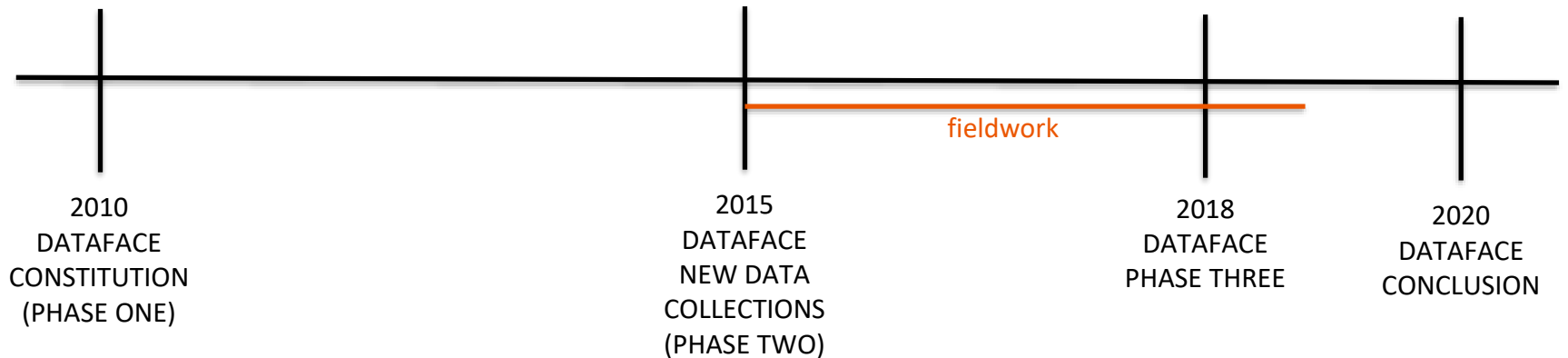
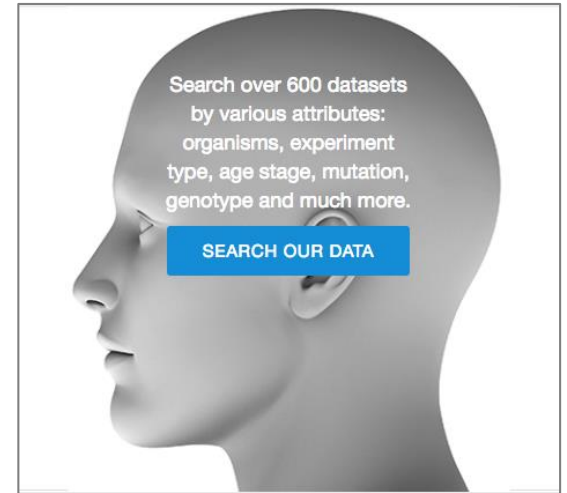
Reusing Science Data and Software: Not easy.
Trust in the **data** (metadata, ontologies, annotation etc.)

Trust in the **systems** (repositories, software etc.)

Trust in the “**people**,” interpersonal trust (in the data creators, in the curators etc.)



The “DataFace Consortium”



RQ1: Where do scientists find reusable data?

RQ2: How do scientists reuse others' data?

RQ3: How do scientists interpret others' data?



Key findings

- Data reuse as a common practice, essential to conduct research;
- Most of others' data were reused for comparison, control, and reference;
- Secondary analyses on others' data:
 - Rare;
 - **Scientists would reuse others' data for hypothesis testing exclusively in the context of a collaborative effort with the data creators.**

WHY?

COMPARATIVE DATA REUSE ← INTEGRATIVE DATA REUSE

GOAL

'Ground truthing': calibrate, compare, confirm

Secondary analysis: identify patterns, correlations, causal relationships

EXAMPLE

Instrument calibration, sequence annotation, review summary-level data

Meta-analyses, novel statistical analyses

FREQUENCY

Easy, frequent, routine practice

Difficult, rare, emergent practice

INTERPRETATION

Limited expertise (interactional expertise, 'knowledge that')

Specialized expertise (contributory expertise, 'knowledge how,' tacit knowledge)

- **The specialized expertise needed to analyze others' data needs time and resources to be mastered by reusers;**
 - **Data creators already own this expertise, while data reusers often do not;**
 - **Reusing data in collaboration with data creators is faster than mastering new expertise.**
-



The “Data Creators’ Advantage”

>> those who create data have intimate and tacit knowledge that can be used as barter to form collaborations for mutual advantage.



Conclusions

- Data can be reused for many different ends
- High quality, highly curated data does not equal reuse
- Data reuse requires **TRUST & EXPERTISE** in data creators >> community building