Managing Data to Manage Evidence:
Social and Technical Challenges

Irene Pasquetto, Ph.D. - Research Associate @ the UCLA Information Science Department

Labs locations
Data collection 2014-2017

Irene Pasquetto, Ph.D. - Research Associate @ the UCLA Information Science Department

"REUSABLE" RESEARCH DATA ?

Irene Pasquetto, Ph.D. - Research Associate @ the UCLA Information Science Department

DATA REUSE:
EXTRACTING "NEW KNOWLEDGE" FROM "OLD DATA"

Irene Pasquetto, Ph.D. - Research Associate @ the UCLA Information Science Department

Data reuse empowers data-driven research

"Hiding within those mounds of data is knowledge that could change the life of a patient, or change the world"

Atul Butte
Stanford School of Medicine

RE, C., & LEN, W. (2013) Biology's Dry Future. Science Magazine

Irene Pasquetto, Ph.D. - Research Associate @ the UCLA Information Science Department

PART I:
WHY WE WANT DATA REUSE, AND HOW WE WANT IT.

Irene Pasquetto, Ph.D. - Research Associate @ the UCLA Information Science Department

DATA REUSE=
Return On Investment (ROI)
for
OLD AND **NEW** DATA
COLLECTIONS

Irene Pasquetto, Ph.D. - Research Associate @ the UCLA Information Science Department

NCBI
National Center for
Biotechnology Information

All Databases ⌄

NCBI Home

Resource List (A-Z)

All Resources

Chemicals & Bioassays

Data & Software

DNA & RNA

Domains & Structures

Genes & Expression

Genetics & Medicine

Genomes & Maps

Homology

Literature

Proteins

Sequence Analysis

Taxonomy

Training & Tutorials

Variation

Welcome to NCBI

The National Center for Biotechnology Information advances science and health by providing access to biomedical and genomic information.

About the NCBI | Mission | Organization | NCBI News & Blog

Submit

Deposit data or manuscripts into NCBI databases

Download

Transfer NCBI data to your computer
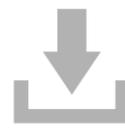
Learn

Find help documents, attend a class or watch a tutorial

Develop

Use NCBI APIs and code libraries to build applications

Analyze

Identify an NCBI tool for your data analysis task

Research

Explore NCBI research and collaborative projects

National Center for Biotechnology Information (NCBI), https://www.ncbi.nlm.nih.gov/

From
HYPOTHESIS-DRIVEN RESEARCH…

- Funding of pre-determined research questions;
- Research design guides data collection;
- Raw data are rarely shared

To
HYPOTHESIS-FREE DATASETS

- Research data as "fungible" goods;
- Datasets can be repurposed in an infinite number of studies;
- Raw data are publicly available

SPECIALIZED DATASETS Vs. REUSABLE DATA

Leonelli, S. (2013). Why the current insistence on open access to scientific data? Big data, knowledge production, and the political economy of contemporary biology. Bulletin of Science, Technology & Society, 33(1-2), 6-11.

Mirowski, P. (2018). The future (s) of open science. *Social studies of science*, *48*(2), 171-203.

The future of health begins with you

All of Us
RESEARCH PROGRAM   ABOUT   HOW TO JOIN   NEWS & EVENTS   COMMUNITY   LOG IN   Search
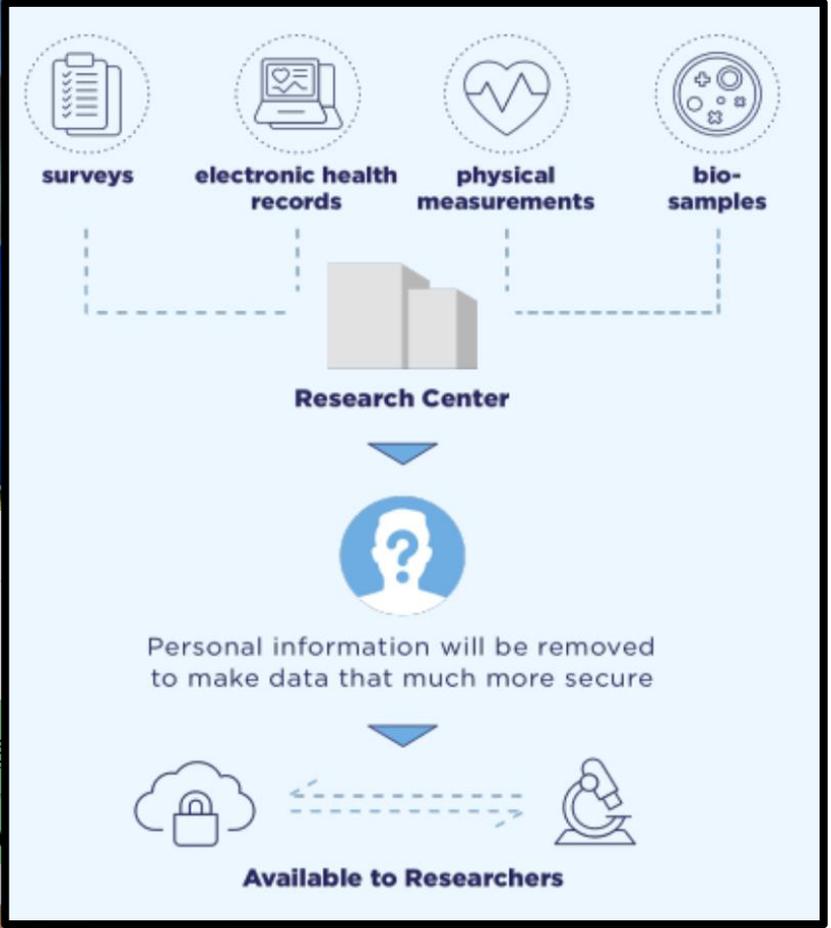
Why People Are Joining

"For this to be a hit, it takes all of us."
—BJ

"We all want our kids to live healthier lives than we did."
—Steffinie

"You nee
types of
in
—A

surveys   electronic health records   physical measurements   bio-samples

Research Center

Personal information will be removed to make data that much more secure

Available to Researchers

WHY WE CARE ABOUT DATA REUSE

- INNOVATION. Reuse increases innovation enabling integrated, systemic, and data-driven approaches to science;

- EFFICIENT SCIENCE. Increases ROI on old and new collections of data.

Irene Pasquetto, Ph.D. - Research Associate @ the UCLA Information Science Department

Open Data Policies

- European Union

- U.S. Federal research policy

- Research Councils of the UK

- Australian Research Council

- Individual countries, funding agencies, journals, universities

National Institutes of Health
Turning Discovery Into Health

Irene Pasquetto, PhD – Research Associate @ the UCLA Information Science Department

12

Policy RECommendations for Open Access to Research Data in Europe

PART II:
WHO IS REUSING RESEARCH DATA? HOW? WHEN?

Irene Pasquetto, Ph.D. - Research Associate @ the UCLA Information Science Department

Long tail of data reuse

Volume of data (y-axis) vs. # frequency of reuse (x-axis)
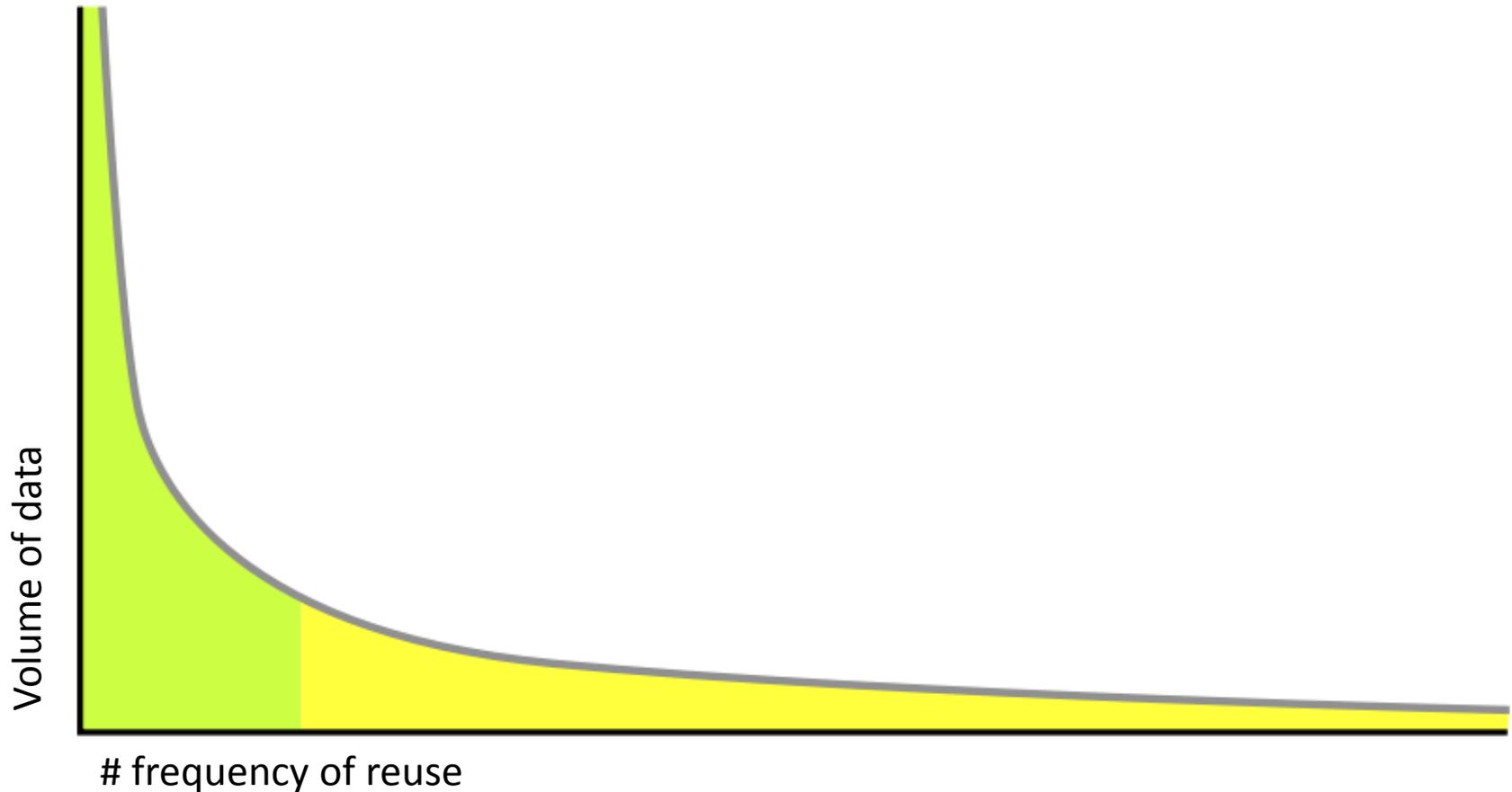
Wallis, J. C., Rolando, E., & Borgman, C. L. (2013). If we share data, will anyone use them? Data sharing and reuse in the long tail of science and technology. *PloS one*, *8*(7), e67332.

Pasquetto, I. V., Randles, B. M., & Borgman, C. L. (2017). On the reuse of scientific data.

THE IDEALS:
A RESEARCH DATA 'LIFE CYCLE'...
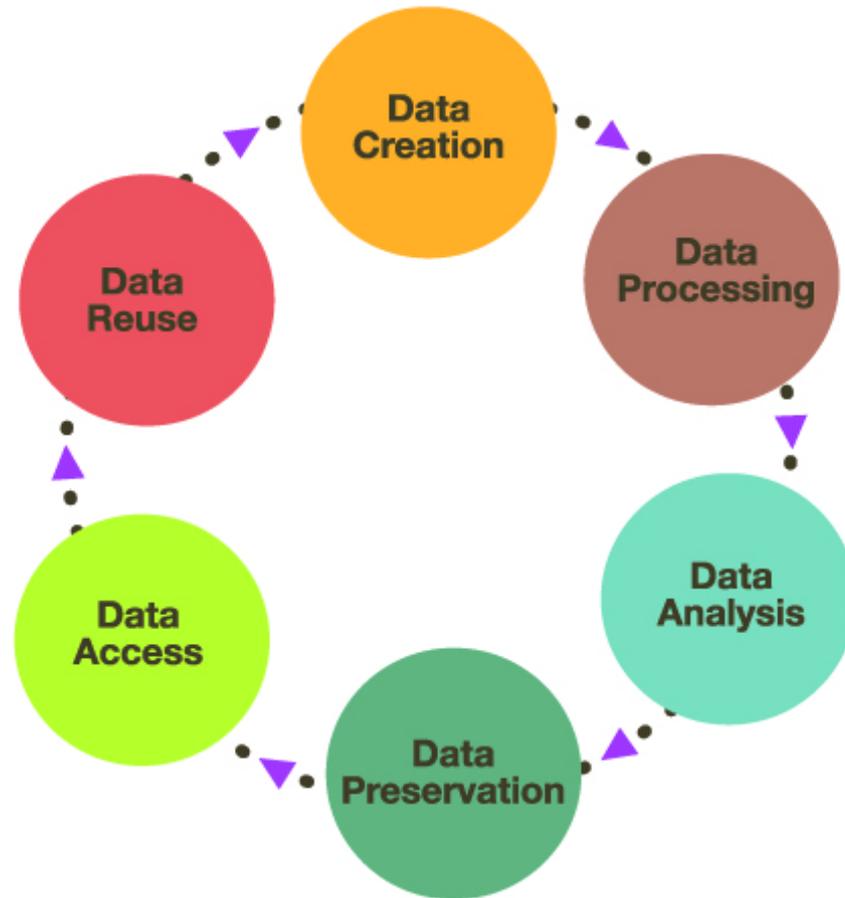
Ray, J. M. (Ed.). (2014). *Research data management: Practical strategies for information professionals*. Purdue University Press.
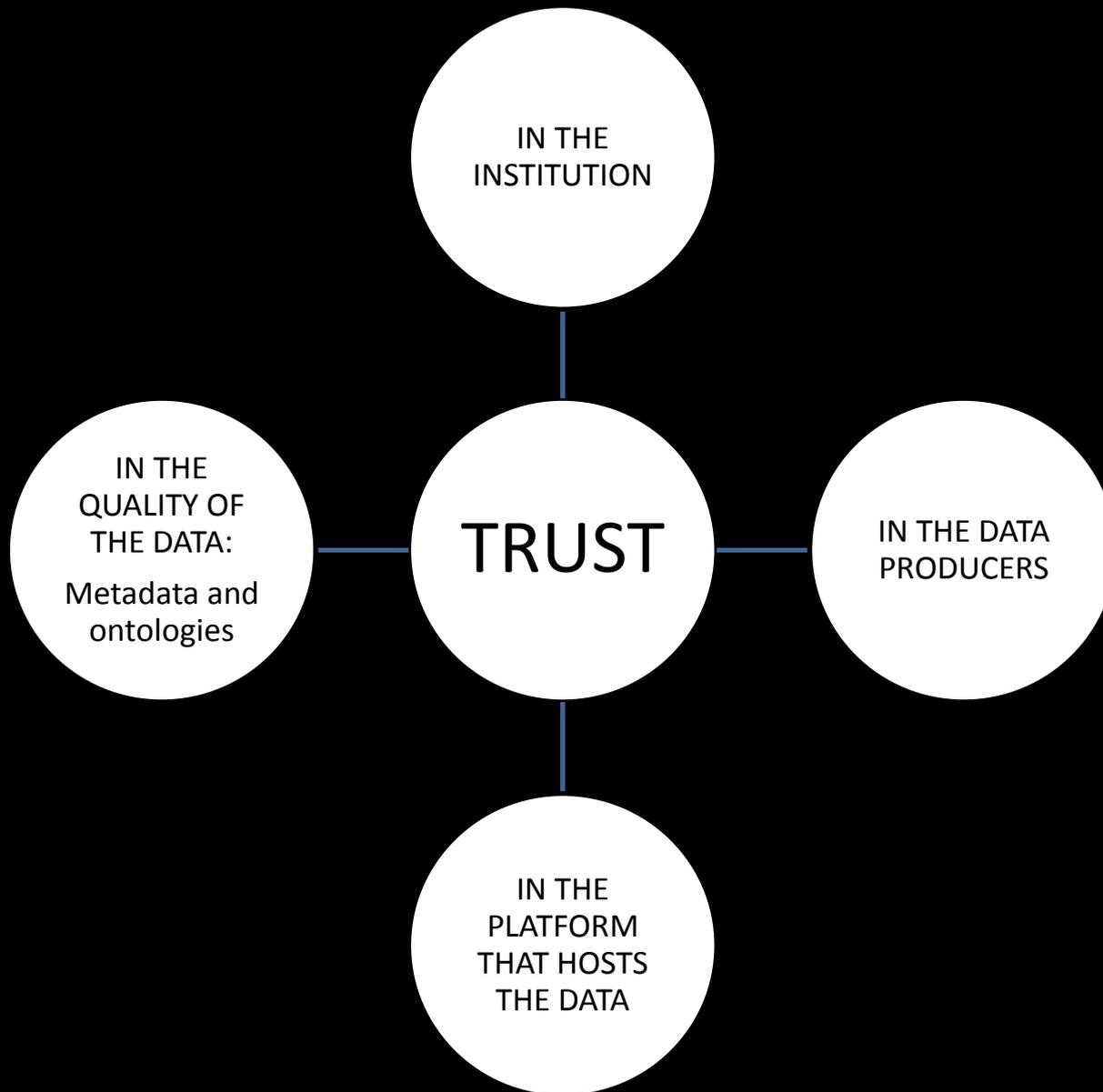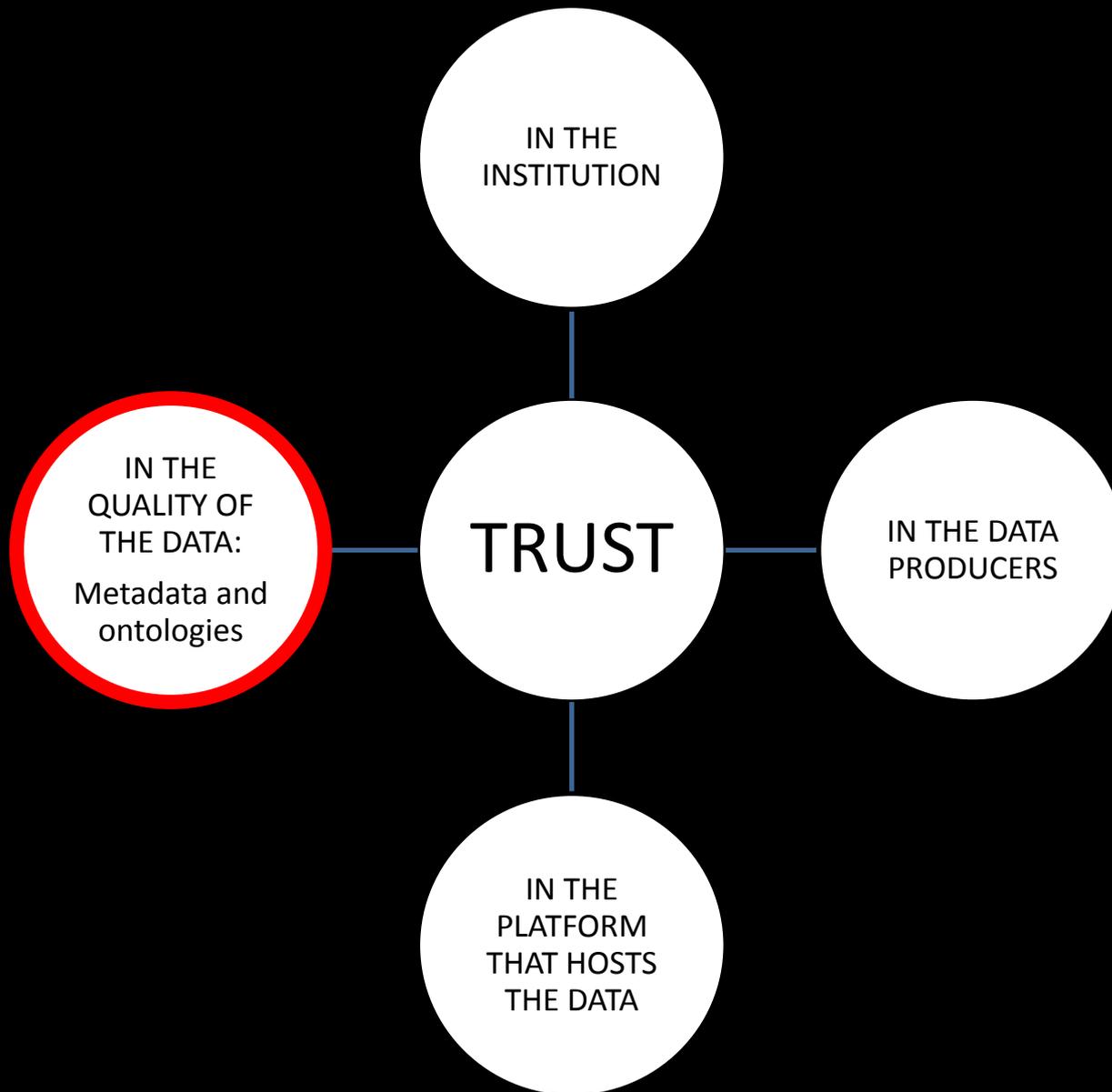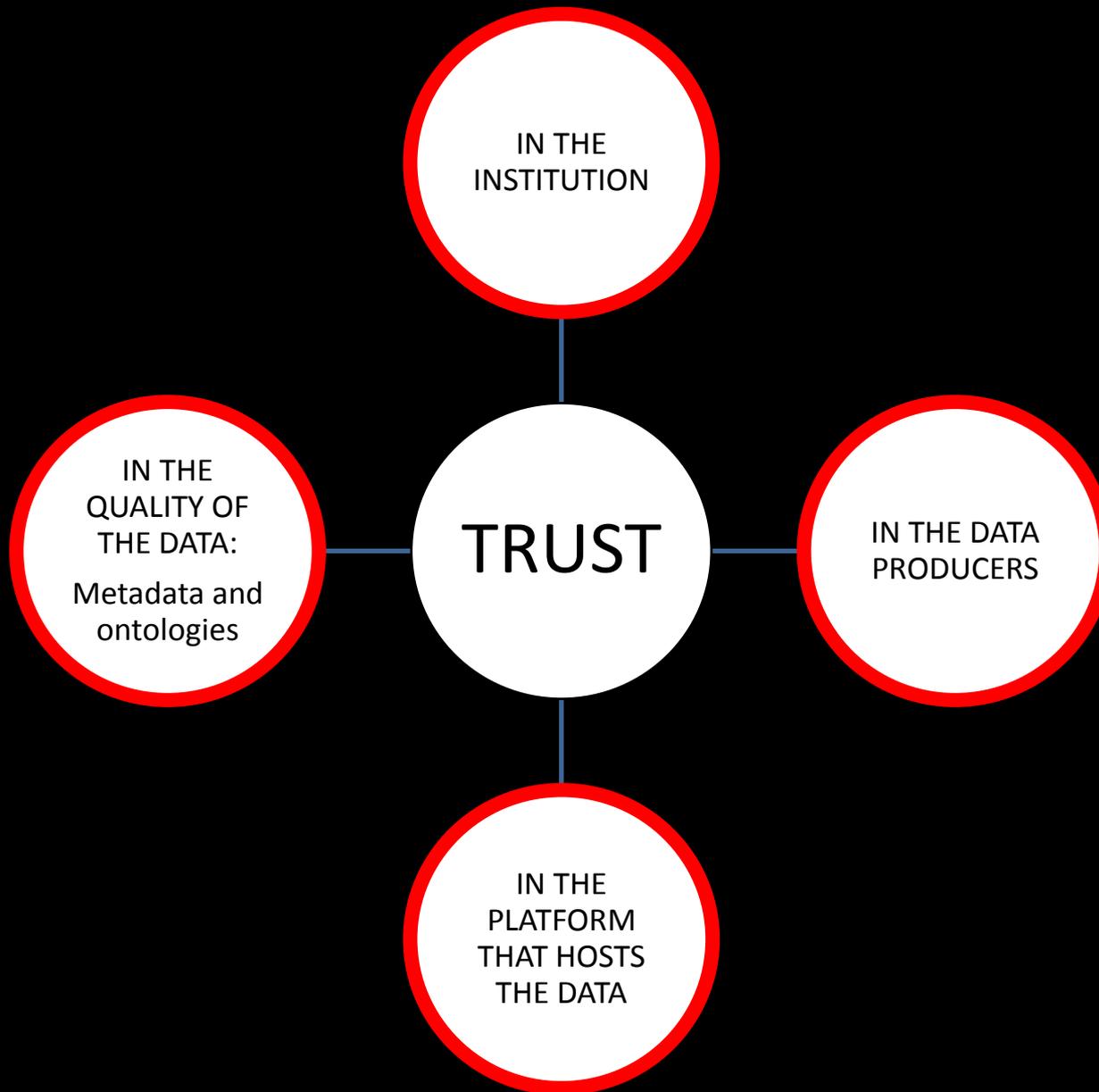
…AND DATA STEWARDSHIP PRINCIPLES.
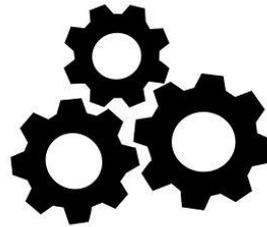
Wilkinson, et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, *3*, http://dx.doi.org/10.1038/sdata.2016.18

Irene Pasquetto, Ph.D. - Research Associate @ the UCLA Information Science Department

Point A

Point B

@ the UCLA Information Science Department

Irene Pasquetto, Ph.D. - Research Associate @ the UCLA Information Science Department

Irene Pasquetto, Ph.D. - Research Associate @ the UCLA Information Science Department

Irene Pasquetto, Ph.D. - Research Associate @ the UCLA Information Science Department

Access to Specialized Knowledge and Complimentary Skills

Irene Pasquetto, Ph.D. - Research Associate
@ the UCLA Information Science
Department
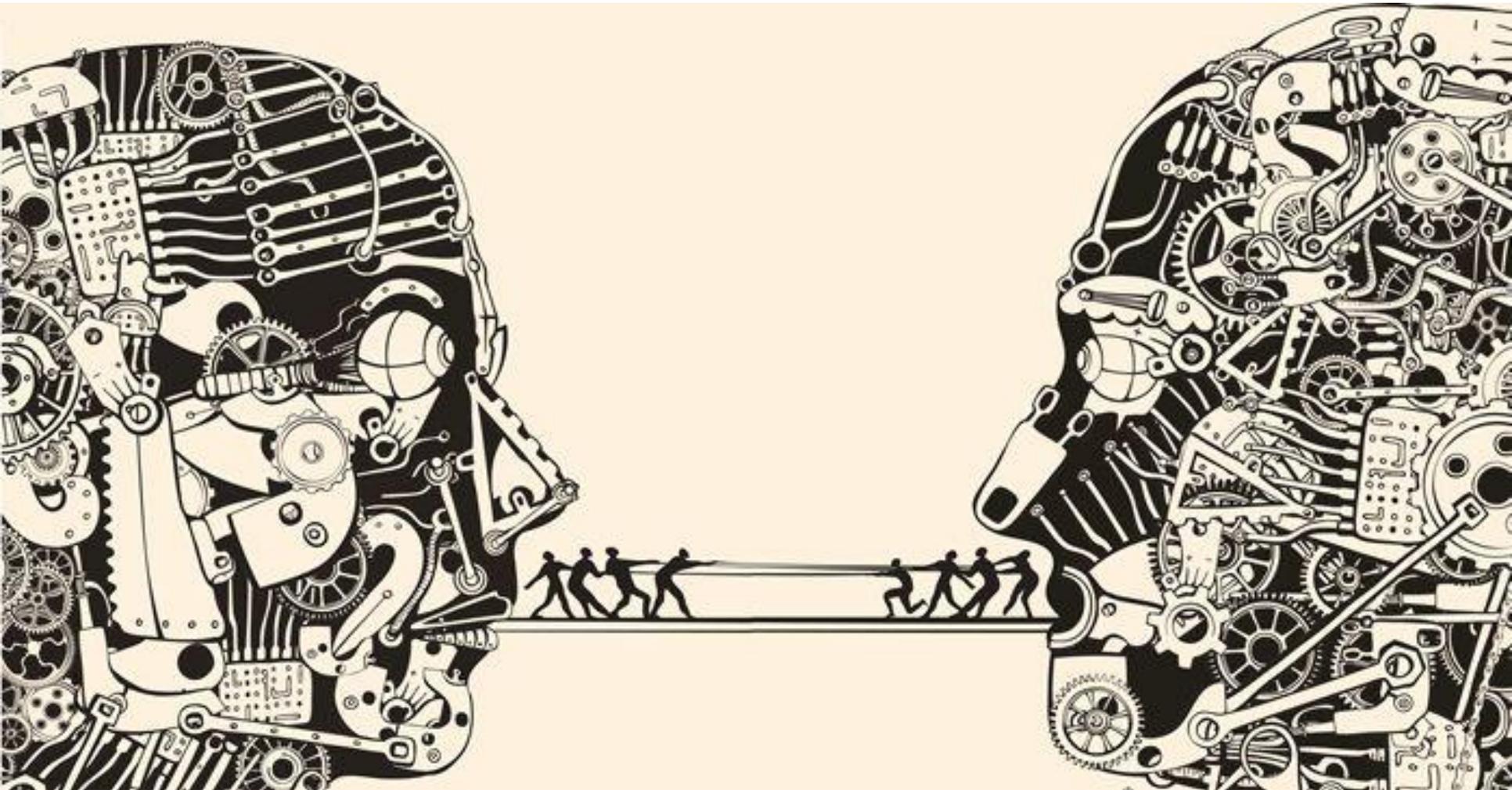
Takeaways

- Reusing research data can be beneficial for innovation and can increase ROI on science;
- But…it is not an easy thing to do!
- Research data are not fungible commodities…(datasets are not like oil!)
- Metadata and data models are necessary but not sufficient for reuse;
- Other factors that impact reuse are:
  - Trust in evidence, repository reputation, and interpersonal exchange of tacit and specialized knowledge

Reusing research data is a social AND a technical challenge, it needs much more than data curation and management to be implemented.

Reuse requires trust & collaboration among science stakeholders, as much as technical infrastructures.