Data stewardship in the information age: Science respecting formal confidentiality protections

John M. Abowd
Chief Scientist and Associate Director for Research and Methodology
U.S. Census Bureau

2022 IRS-Tax Policy Center Annual Conference on Tax Administration

Keynote Address Thursday, June 16, 2022, 12:40-1:15PM EDT

United States® Census Bureau

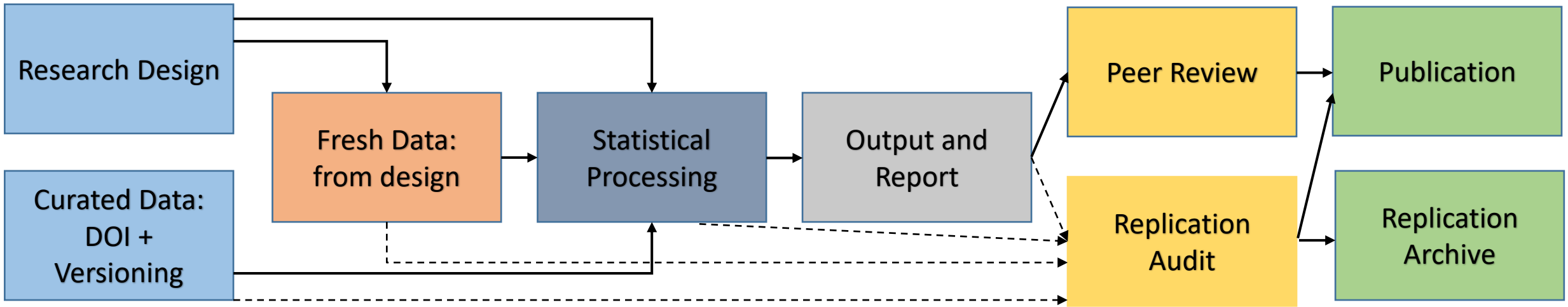
The views expressed in this talk are my own and not those of the U.S. Census Bureau.

Basics

- **Science:** research destined for peer-reviewed outlets in any discipline.
- **Transparency:** ability of the peer reviewers to understand the research based on the details in the paper and supporting appendices including digital materials.
- **Replication:** ability of the publishing journal to confirm that the inputs and workflow described in the research produce the outputs in that research.
- **Formal confidentiality (or privacy):** ability of the steward of the research inputs to quantify and limit the information leakage used to produce the outputs reported in the paper and supporting appendices including digital materials.
- **Inference validity:** ability of the reader to correctly assess the uncertainty of the output reported in the research paper in support of statistical reasoning about the conclusions.

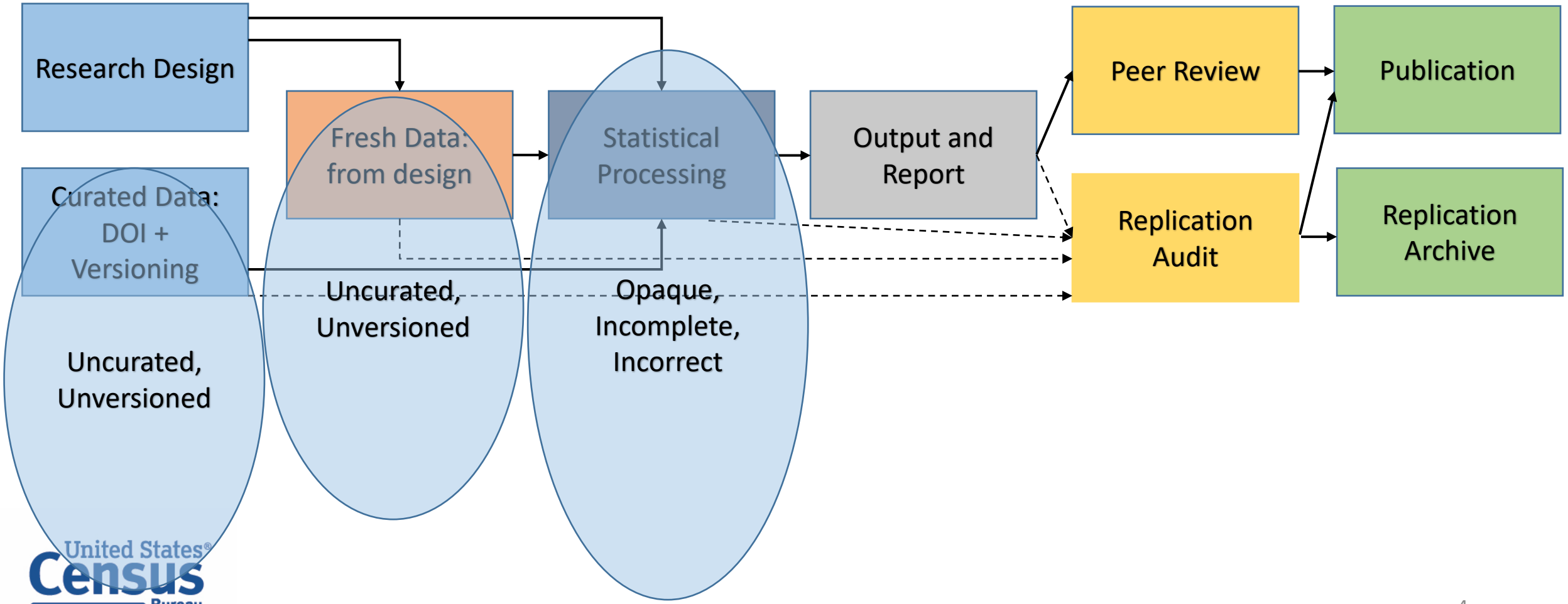
United States® Census Bureau

Ideal Workflow



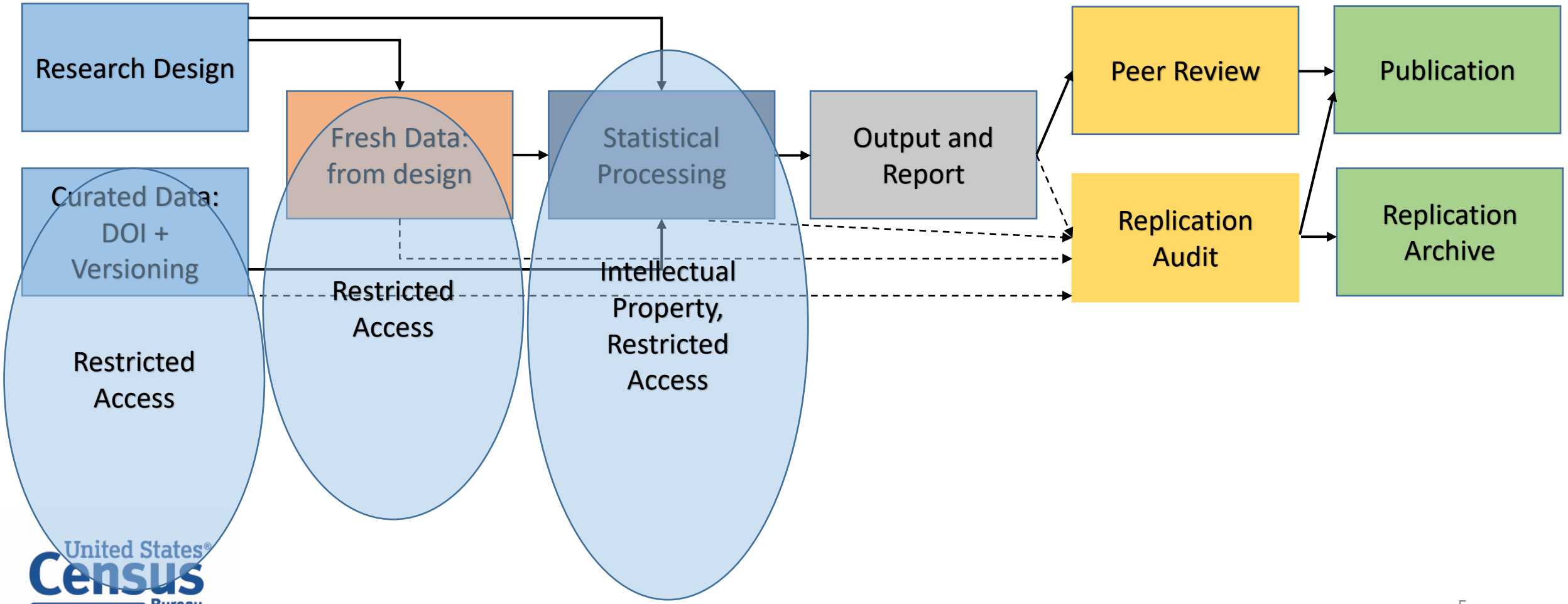
United States® Census Bureau

Challenges to Transparency



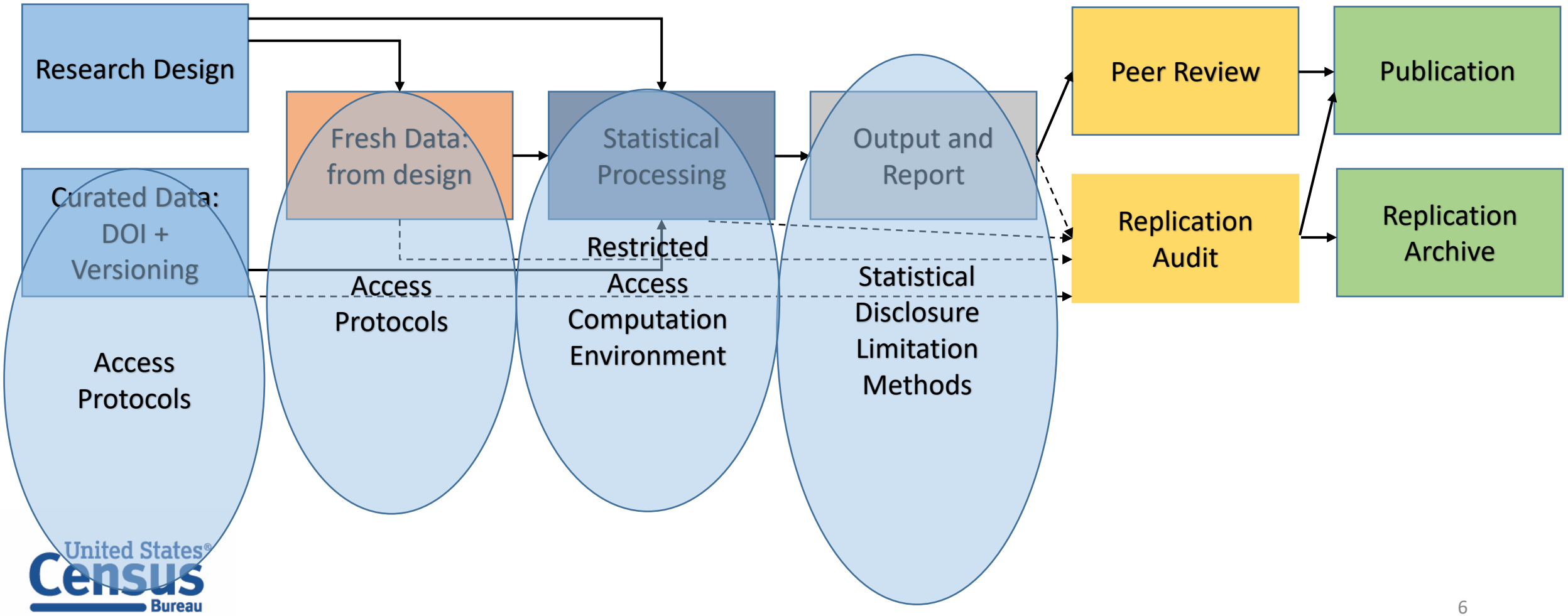
United States® Census Bureau

Challenges to Replication



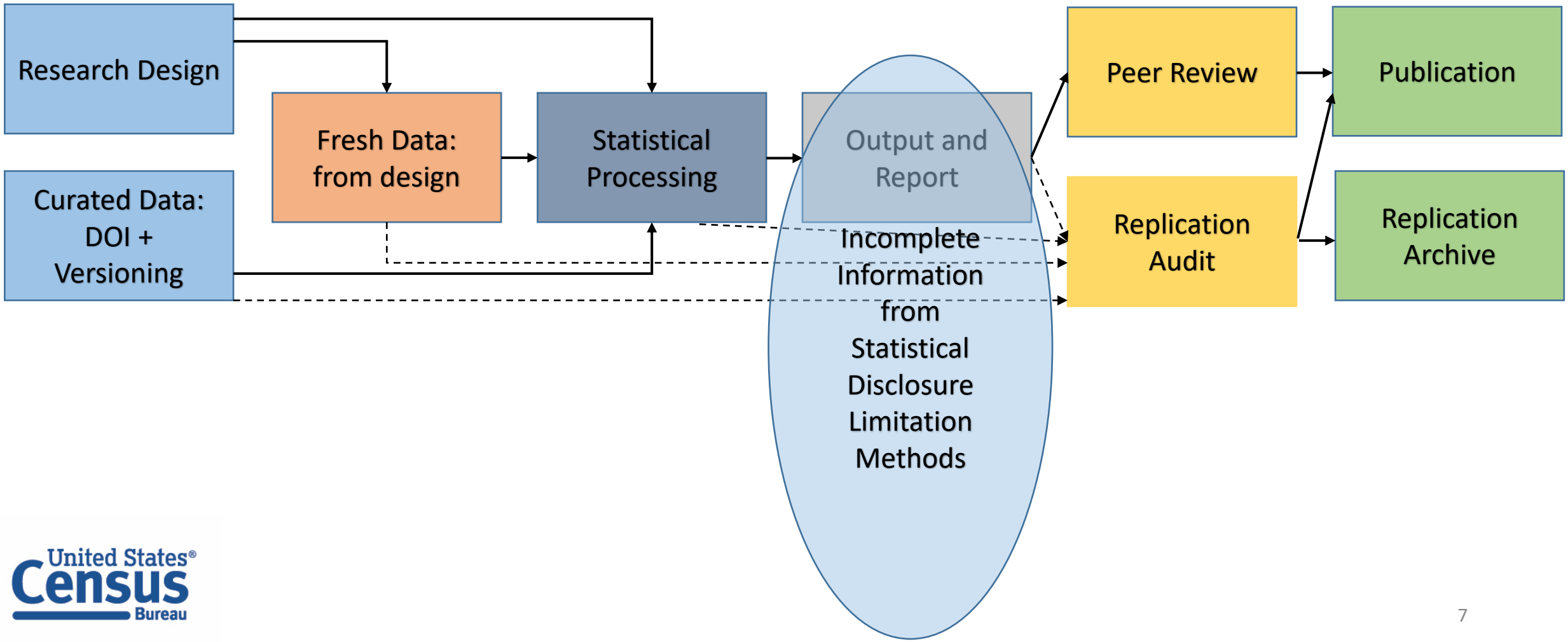
United States® Census Bureau

Challenges to Privacy and Confidentiality



United States® Census Bureau

Challenges to Inference Validity



United States® Census Bureau

Policies to Promote Peer Review and Transparency

Data Stewardship Policy 001

"The Census Bureau recognizes that Title 13 benefits are not fully realized until the research has been reviewed and published. The integrity of research done under Census Bureau auspices depends upon the confidence of the scientific community in our adherence to the principle of peer review. Every effort should be made to allow for timely response to peer review, consistent with the requirements of Title 13 and agreements with external data providers."

Data Stewardship Policy 027

"The Census Bureau will, subject to the availability of funds and resources and the feasibility of the proposed methodology, support external research designed to assess the quality of our data products and programs and recognizes the implicit value to the agency of this work. This support may come in the form of provisioned access to data through the Federal Statistical Research Data Center (FSRDC) network for related external projects, joint statistical partnerships with other agencies, internal project support from outside researchers, and peer review by federal grant administering agencies."

United States® Census Bureau

Policies to Promote Peer Review and Transparency

American Economic Association

"It is the policy of the American Economic Association to publish papers only if the data and code used in the analysis are clearly and precisely documented and access to the data and code is non-exclusive to the authors.

Authors of accepted papers that contain empirical work, simulations, or experimental work must provide, prior to acceptance, information about the data, programs, and other details of the computations sufficient to permit replication, as well as information about access to data and programs."

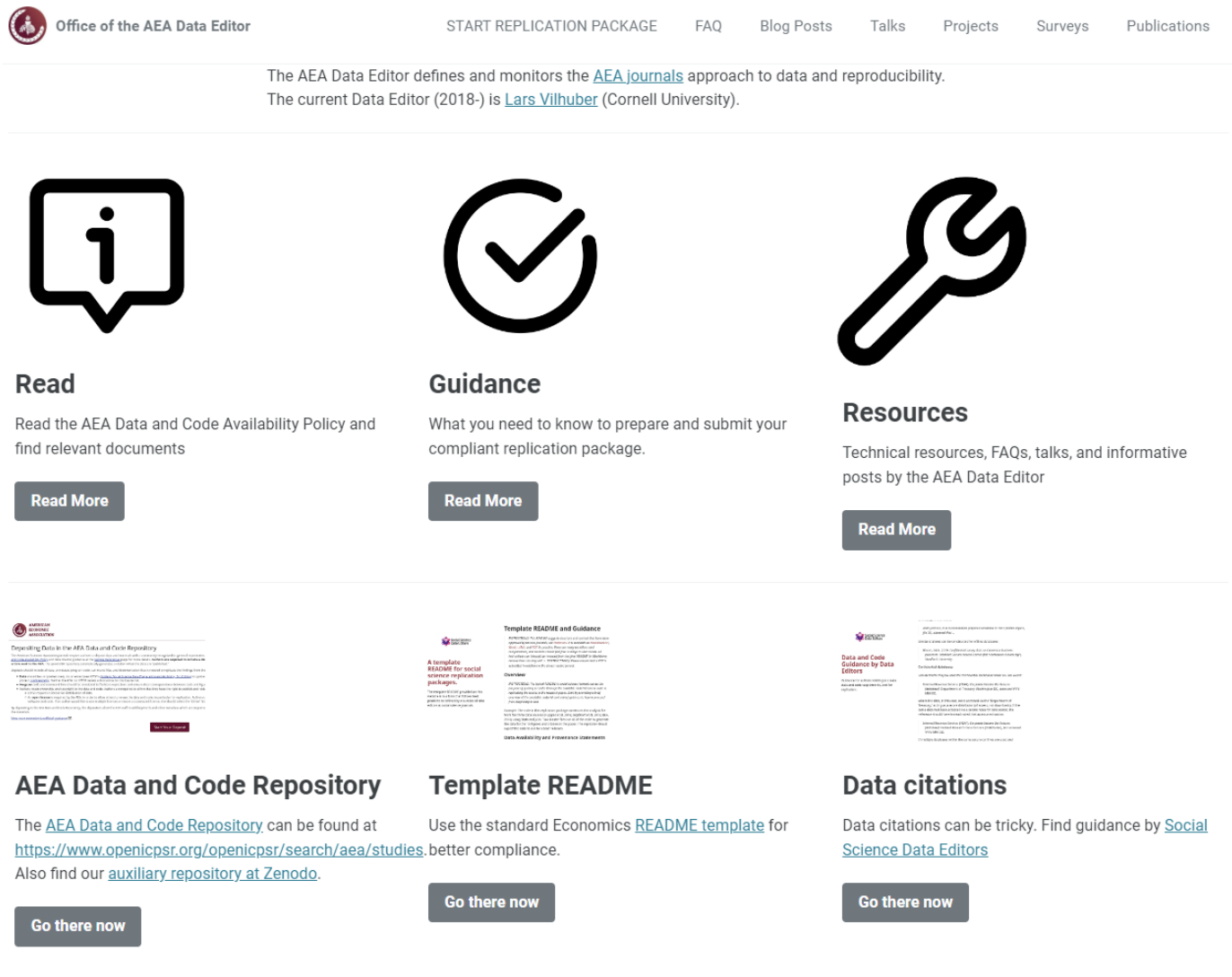
AAPOR Transparency Initiative

"AAPOR's Transparency Initiative is designed to promote methodological disclosure through a proactive, educational approach that assists survey organizations in developing simple and efficient means for routinely disclosing the research methods associated with their publicly-released studies. The Transparency Initiative is an approach to the goal of an open science of survey research by acknowledging those organizations that pledge to practice transparency in their reporting of survey-based research findings. In doing so, AAPOR makes no judgment about the approach, quality or rigor of the methods being disclosed."

United States® Census Bureau

Tools to Promote Peer Review and Transparency

[AEA Data Editor]



United States® Census Bureau

Tools to Respect Firewalls for Audits

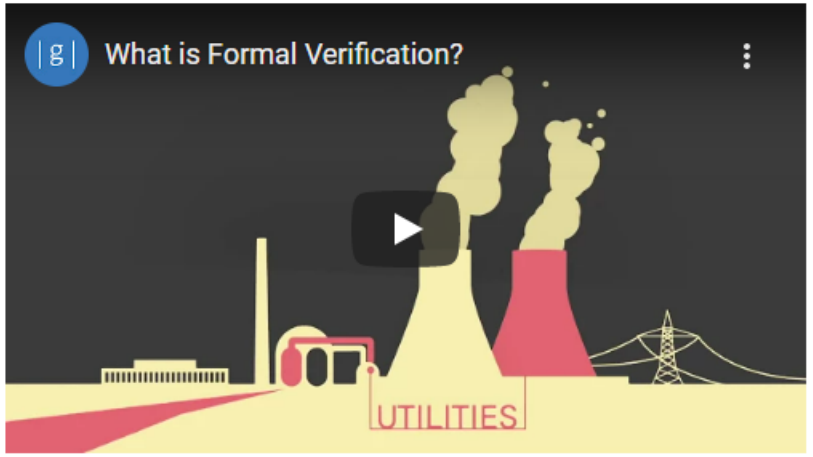
[Galois Software Correctness]

Software Correctness

Our software correctness tools guarantee that your systems do exactly what you want, and no more.

In today's complex, mission-critical environments, hidden defects and security gaps in software are an unaffordable liability. Traditional test-based validation techniques aren't sufficient to provide the high-confidence assurance guarantees that are required. Developers and evaluators need the ability to provide rigorous evidence of software correctness that supports the creation of enhanced functionality for demanding environments.

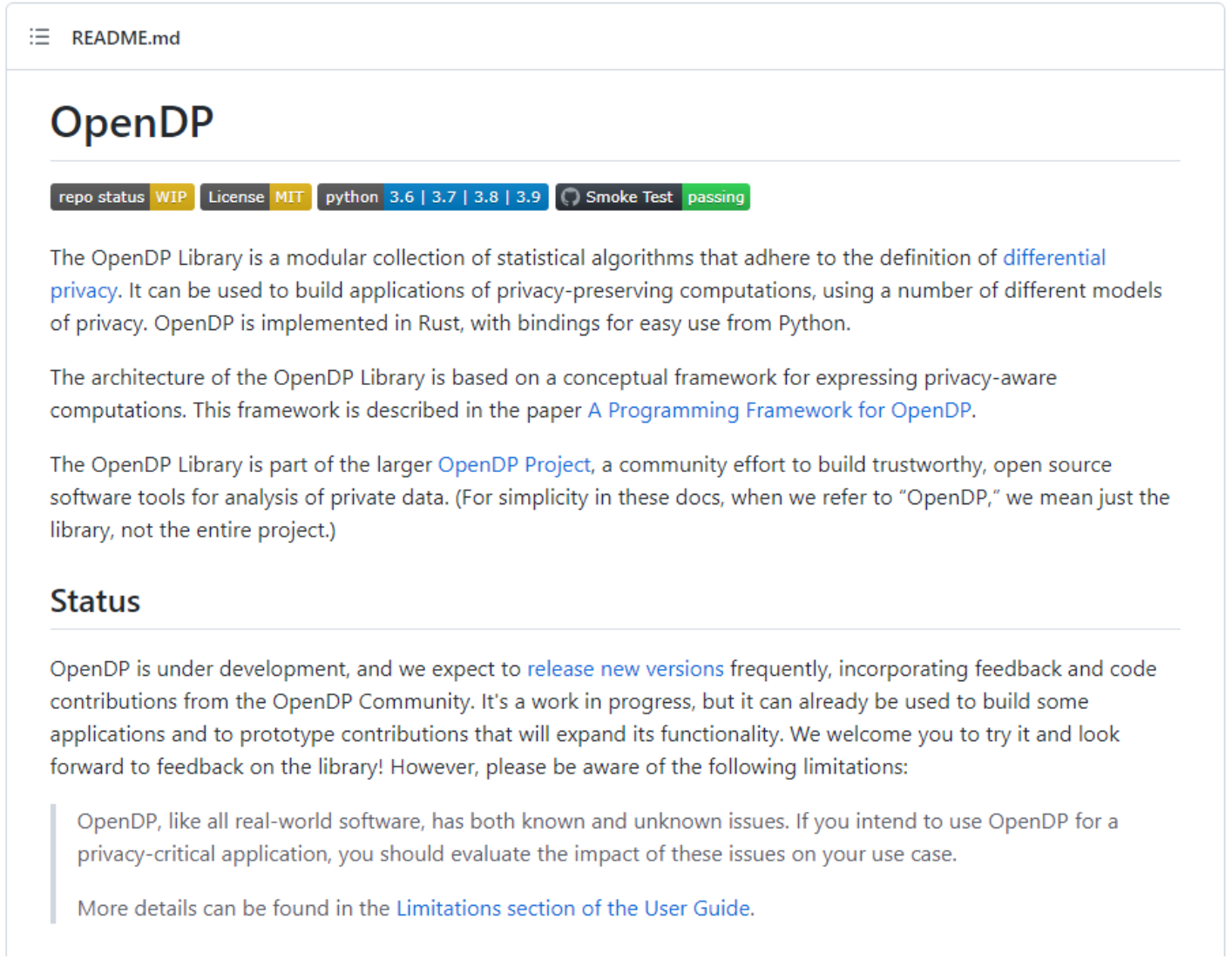
The Galois software correctness portfolio includes capabilities in program understanding, code analysis, and software provenance. And to bring these technologies to bear on complex software systems, we also offer frameworks for modeling and assessing trust relationships between system components.



United States® Census Bureau

Tools to Promote Confidentiality and Inference Validity

[OpenDP]



README.md

OpenDP

repo status WIP | License MIT | python 3.6 | 3.7 | 3.8 | 3.9 | Smoke Test passing

The OpenDP Library is a modular collection of statistical algorithms that adhere to the definition of differential privacy. It can be used to build applications of privacy-preserving computations, using a number of different models of privacy. OpenDP is implemented in Rust, with bindings for easy use from Python.

The architecture of the OpenDP Library is based on a conceptual framework for expressing privacy-aware computations. This framework is described in the paper A Programming Framework for OpenDP.

The OpenDP Library is part of the larger OpenDP Project, a community effort to build trustworthy, open source software tools for analysis of private data. (For simplicity in these docs, when we refer to "OpenDP," we mean just the library, not the entire project.)

Status

OpenDP is under development, and we expect to release new versions frequently, incorporating feedback and code contributions from the OpenDP Community. It's a work in progress, but it can already be used to build some applications and to prototype contributions that will expand its functionality. We welcome you to try it and look forward to feedback on the library! However, please be aware of the following limitations:

> OpenDP, like all real-world software, has both known and unknown issues. If you intend to use OpenDP for a privacy-critical application, you should evaluate the impact of these issues on your use case.
>
> More details can be found in the Limitations section of the User Guide.

United States® Census Bureau

Questions?

John.Maron.Abowd@census.gov

United States® Census Bureau