



# The Open Data Opportunity

Krishna Sood

Senior IP Counsel, Microsoft

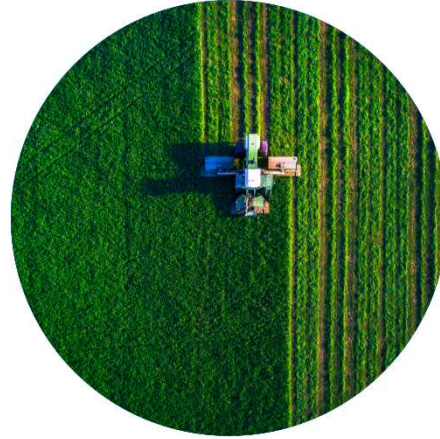
March 17, 2021

Licensing Executive Society –  
Silicon Valley Chapter

# The availability of data is critical to achieve AI innovation



Healthcare



Agriculture



Accessibility



Earth



# The Open Data Opportunity

## Background

Open data and its role in driving AI advancements

## Challenges

Overcoming limited datasets with data sharing

## How to make progress

Seizing the open data opportunity

# The CDDI is advancing biomedical research through data science



## New Cascadia Data Discovery Initiative accelerates health innovation

Jul 12, 2019 | [John Kahan - Chief Data Analytics Officer](#)



[The Cascadia Innovation Corridor](#) is home to some of the world's leading technology, research and medical organizations. In December of last year Microsoft and [Fred Hutchinson Cancer Research Center](#) together started a new chapter for our region when Microsoft President Brad Smith [announced a \\$4 million challenge gift](#) focused on accelerating cancer research. Today, in collaboration with Fred Hutch, we're proud to welcome four key players as part of the Cascadia Data Discovery Initiative (CDDI): [BC Cancer](#), [University of British Columbia](#), [University of Washington eScience Institute](#) and the [Knight Cancer Institute at Oregon Health & Science University](#).

# How to make progress

Improving the legal framework for data sharing



Advancing privacy preserving technologies

Building modern software tools to enable data sharing

# How to make progress

Improving the legal framework for data sharing



Advancing privacy preserving technologies

Building modern software tools to enable data sharing

# Advancing standard legal terms for data sharing

**Open Use of Data Agreement v0.1**  
Annotated-Discussion-DRAFT-20190722

This is the Open Use of Data Agreement, Version 0.1 (the "OUDA"). Copyright © 2019 Microsoft Corporation. All rights reserved. This document is published for informational purposes only. It is not intended to be used as a legal instrument. It is subject to change without notice. It is not intended to create a contract. It is not intended to be used as a legal instrument. It is subject to change without notice. It is not intended to create a contract.

**Computational Use of Data Agreement v0.1**  
Annotated-Discussion-DRAFT-20190722

This is the Computational Use of Data Agreement, Version 0.1 (the "CUDA"). Copyright © 2019 Microsoft Corporation. All rights reserved. This document is published for informational purposes only. It is not intended to be used as a legal instrument. It is subject to change without notice. It is not intended to create a contract. It is not intended to be used as a legal instrument. It is subject to change without notice. It is not intended to create a contract.

**Data Use Agreement for Open AI Model Development v0.1**  
Annotated-Discussion-DRAFT-20190722

This AI Data Sharing Agreement ("Agreement") is entered into between (a) ["Data Provider"] and (b) ["Data User"] ("the Parties"). Data user and data provider may also be referred to individually as "a party" or collectively as "the parties".

**DATA USE AGREEMENT FOR OPEN AI MODEL DEVELOPMENT**

This AI Data Sharing Agreement ("Agreement") is entered into between (a) ["Data Provider"] and (b) ["Data User"] ("the Parties"). Data user and data provider may also be referred to individually as "a party" or collectively as "the parties".

- Defined Terms**
  - "Algorithm" means the machine learning algorithm described in attachment A, including associated parameters and associated weights, if present.
 

**Comment:** This definition encompasses that the AI Model may be an entirely untrained machine learning algorithm, without any associated parameters or weights, or that it may be a partially trained algorithm. Attachment A should be used to identify and describe the particular AI Model that will be trained by Data User.
  - "OPTIONAL: NDIA" means a non-disclosure agreement governing the exchange of confidential information between the parties.
  - "Open Source License" means a license that meets all of the requirements of the "The Open Source Definition" as published by the Open Source Initiative at <https://opensource.org/licenses>.
 

**Comment:** The algorithm published by the Open Source License is flexible to, and generally accepted by, the open source community. Many of the most common open source licenses in use today, such as the MIT license, would satisfy this definition.
  - "Personal Data" means any information relating to an identified or identifiable natural person and any other information that constitutes personal data as personal information under any applicable law. An identifiable natural person is one who can be identified, directly or indirectly, in particular by referencing an identifier such as a name, an identification number, location data, an online identifier, or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural, or social identity of that natural person.



## Removing barriers to data innovation

*Empowering people and organizations to share and use data more effectively*

Sharing data can help address some of society's biggest challenges and help individuals and organizations be more innovative, efficient, and productive. Today, however, it is often difficult to determine how best to share data from both a legal and technical perspective. To help make it easier for individuals and organizations to share data, we're offering a set of draft agreements for consideration by the community, each designed to address a specific data sharing challenge. We look forward to improving these agreements with feedback and taking additional steps to help make it easier for individuals and organizations to share data with confidence.



microsoft / Open Use of Data Agreement · Issues

Code · Issues · Pull requests · Projects · Security · Insights

Open Use of Data Agreement - Removing Barriers to Data Innovation <https://news.microsoft.com/datainnovation...>

1 comment · 1 branch · 1 message · 1 contributor · @ 00:10

Search results	New pull request	Create new file	Upload file	Find file	Open in marketplace
microsoft/openuseofdata	initial commit	initial commit	initial commit	initial commit	initial commit
ghprflow	initial commit	initial commit	initial commit	initial commit	initial commit
UCIDE	initial commit	initial commit	initial commit	initial commit	initial commit
UCUDA-1_unannotated_discussion-draft.md	initial commit	initial commit	initial commit	initial commit	initial commit
README.md	initial commit	initial commit	initial commit	initial commit	initial commit
build-deps.txt	initial commit	initial commit	initial commit	initial commit	initial commit

microsoft / Computational-Use-of-Data-Agreement · Issues

Code · Issues · Pull requests · Projects · Security · Insights

Computational Use of Data Agreement - Removing Barriers to Data Innovation <https://news.microsoft.com/datainnovation...>

1 comment · 1 branch · 1 message · 1 contributor · @ 00:10

Search results	New pull request	Create new file	Upload file	Find file	Open in marketplace
microsoft/computational-use-of-data-agreement	initial commit	initial commit	initial commit	initial commit	initial commit
ghprflow	initial commit	initial commit	initial commit	initial commit	initial commit
UCIDE	initial commit	initial commit	initial commit	initial commit	initial commit
UCUDA-1_unannotated_discussion-draft.md	initial commit	initial commit	initial commit	initial commit	initial commit
README.md	initial commit	initial commit	initial commit	initial commit	initial commit
build-deps.txt	initial commit	initial commit	initial commit	initial commit	initial commit

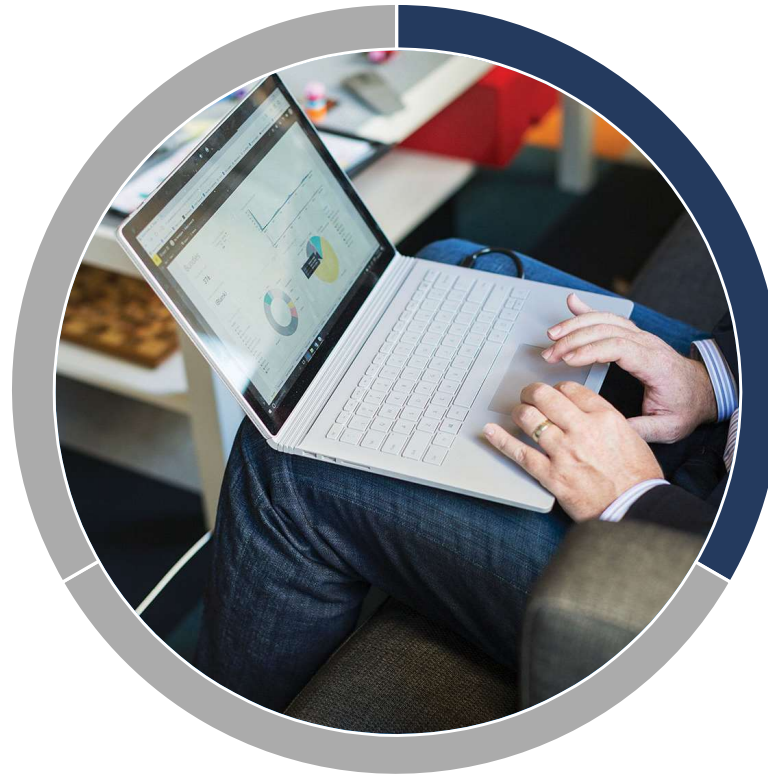
Model data use agreements

[news.microsoft.com/datainnovation](https://news.microsoft.com/datainnovation)

GitHub repos for comment

# How to make progress

Improving the legal framework for data sharing



Advancing privacy preserving technologies

Building modern software tools to enable data sharing



# Related technology developments

## Differential Privacy

Introducing "noise" to protect individuals' privacy

Life



Art



## Confidential Computing

Protecting data in use

Top data breach threats mitigated

Code protected and verified by customer

Data fully in customer control

Data and code opaque to the cloud platform



## Azure Data Share

Cross-organization big data collaboration



Easily share data



Govern your data shares



Expand analytical datasets

## Azure API for FHIR

Health data interoperability enabled by FHIR



# Confidential computing controls data through its lifecycle

## Existing



### At rest

Encrypt inactive data when stored in blob storage, database, etc.

#### Examples include:

Azure Storage Service Encryption for Data at Rest

SQL Server Transparent Database Encryption (TDE)



### In transit

Encrypt data that is flowing between untrusted public or private networks

#### Examples include:

HTTPS

TLS

## New



### In use

Protect/Encrypt data that is in use during computation

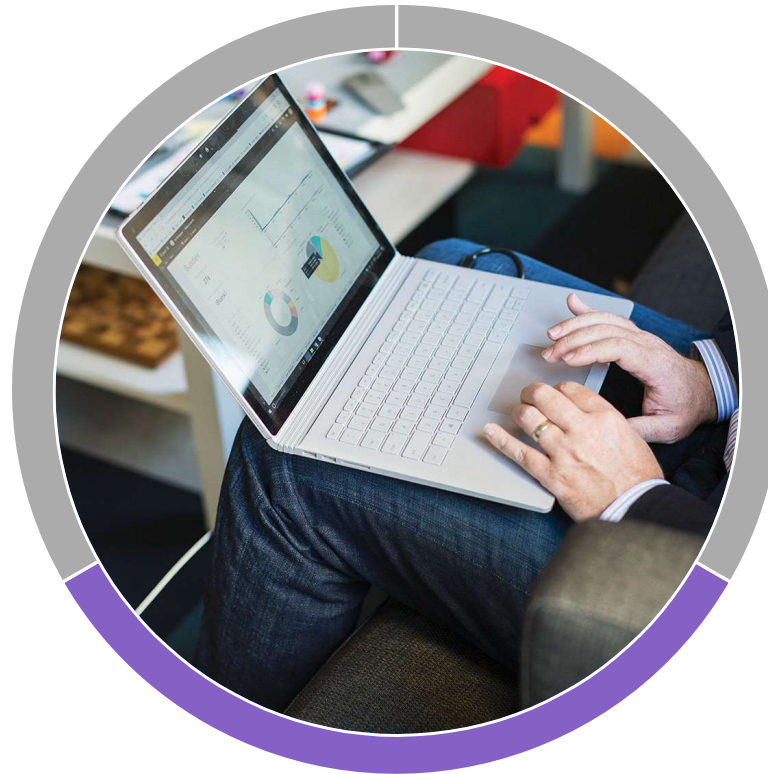
#### Examples include:

Trusted Execution Environments such as Intel SGX and VBS

Homomorphic encryption

# How to make progress

Improving the legal framework for data sharing



Advancing privacy preserving technologies

Building modern software tools to enable data sharing

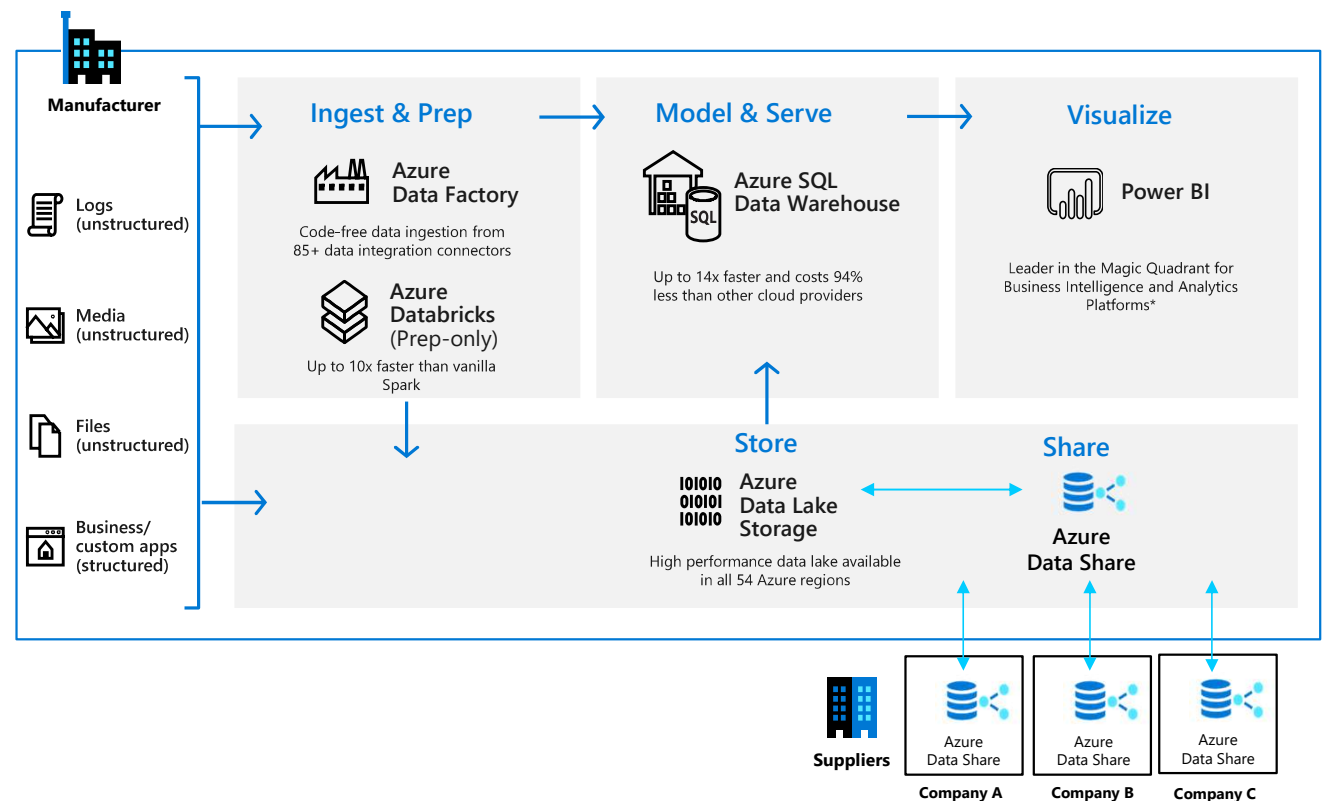
# Azure data sharing tool

A simple service for big data sharing, helping organizations to:

**Enhance** insights with data from partners and customers

**Form** industry-specific consortium to pool data among members

**Innovate** custom solutions and expand markets via new service capabilities



# Key Takeaways

- **Map the data ecosystem** to understand value flows and incentives for a data sharing arrangement
- **The value may lie in the outcomes** of sharing data, not in the data itself
- **No one size fits all:** the business should ensure policy and technical stacks evolve synergistically to support different data governance models to facilitate responsible data sharing and innovation

# The Open Data Opportunity

To learn more about the data use agreements and join the conversation, visit: <https://news.microsoft.com/opendata/>





Contact: Krishna Sood, Senior IP Counsel  
[linkedin.com/in/krishnasood](https://www.linkedin.com/in/krishnasood)