

---

# CADRE Data Access Policy

## FULL POLICY CONTENTS

Scope  
Policy Statement  
Reason for Policy  
Procedures  
Definitions  
Sanctions

### ADDITIONAL DETAILS

Additional Contacts  
Forms  
History

**Effective:** 4/22/2019  
**Last Updated:** 4/22/2019  
**Responsible Office:**  
*Indiana University Network Science Institute*

**Responsible Administrator**  
*Santo Fortunato, Director, IUNI*

**Policy Contact:**  
*Matthew Hutchinson maaahutch@indiana.edu*

---

## Scope

All agents of the Universities participating in the CADRE project who wish to access the WoS and/or MAG datasets stored on IUNI's data enclave.

---

## Policy Statement

IUNI has entered into an agreement with Clarivate Analytics to receive a copy of the 'Web of Science' (WoS) dataset and Microsoft Research to receive the 'Microsoft Academic Graph' (MAG). The WoS data set contains over sixty-nine million records documenting scholarly research from the end of the nineteenth century up to and including 2017. The MAG dataset includes over 200 million publications with the earliest dating from the early 19<sup>th</sup> century. IUNI is making this data accessible while simultaneously honoring agreements with Clarivate Analytics and Microsoft Research. To this end, the data will be stored on a discrete enclave on a dedicated within the Carbonate computing cluster. This document describes the rules users must follow in order to access the WoS and MAG datasets.

- The data is accessible via three distinct modes all of which are available on the enclave:
  - Raw XML data (WoS) or TSV (MAG)
  - A relational PostgreSQL database created by parsing the raw XML and TSV data
  - A browser based GUI running within the server

---

## Reason for Policy

IUNI wants to make the WoS and MAG data accessible to eligible individuals while adhering to the terms and conditions of the agreement with Clarivate Analytics and Microsoft Research.

The WoS data set is a proprietary data set purchased from Clarivate Analytics and the protection of this data is a key priority for the institute. A violation may result in a termination of the contract with Clarivate Analytics; resulting in significant challenges for research using publication data throughout the University.

The MAG data is openly available and is provided to IUNI by Microsoft Research. The dataset is licensed [ODC-By](#) and as such can be used freely as long as the appropriate citation is included.

The goal of the data access policy is to maximize accessibility while simultaneously ensuring adequate controls over the data and remaining in compliance with the Clarivate Analytics' licensing agreement.

---

## Procedures

- 1.1.** The data enclave will be accessible for all eligible individuals who are affiliated with a participating institution. Except where there is a compelling reason not to do so, it is the intention of IUNI to approve all applications from individuals and who complete the actions required to comply with this policy.
- 1.2.** The Indiana University Office of the Vice-President and General Counsel has final authority regarding interpretation of the Clarivate license agreement and whether the proposed use of the data comport with the agreement.
- 1.3.** Analysis and research can be conducted within the enclave using applications already installed on the Carbonate cluster. Additional applications requested by users can be installed on the enclave at the Research Data Steward's discretion (as required by [DM-02](#)). A request for new software should be submitted to the IUNI Data Manager for consideration. The Data Manager will work with the Research Data Steward and UITS to determine whether the request can be honored. All persons submitting a request will receive a response regarding the disposition of the request in a timely manner.

- 1.4.** Analysis and research can also be conducted on the indexed data in the PostgreSQL database by either submitting direct SQL queries or by using the browser based GUI from inside the enclave.
  - 1.5.** Analysis and research requiring more computational resources than the current Data Enclave provides, can be performed using the University’s advanced computational systems (e.g. Carbonate, Karst, Big Red II, Mason, etc).
  - 1.6.** The server on which the data is stored open to the internet and users are free to export data as necessary. However, when working with the WoS data, users should be aware of the volume of data they remove from the server. Users may export as much data as they wish but should consider that the more data they export, the greater the potential for that proprietary data may be lost or accessed inappropriately.
- 
- 2.1.** All users are expected to notify the IUNI Data Manager about any use of the data or derived results in publications or funding proposal submissions and include a citation for the IUNI, Clarivate Analytics and/or Microsoft Research data. Format for the citation will be posted to the IUNI WoS data website.
  - 2.2.** An individual or group that uses the XML version of the WoS data as it was purchased from Clarivate Analytics should include the following acknowledgement in the published work:

*“This work uses Web of Science data by Clarivate Analytics provided by the Indiana University Network Science Institute.”*
  - 2.3.** An individual or group that uses the WoS database in the enclave should include the following acknowledgement in the published work:

*“This work uses Web of Science data by Clarivate Analytics provided by the Indiana University Network Science Institute and the Cyberinfrastructure for Network Science Center at Indiana University.”*
  - 2.4.** An individual or group that uses the TSV version of the MAG data as Microsoft Research provides or the database on the IUNI enclave, it should include the following acknowledgement in the published work:

*“This work uses Microsoft Academic Graph data by Microsoft Research provided by the Indiana University Network Science Institute.”*
  - 2.5.** Any work that uses with any data from the MAG must include the following citation:

*Arnab Sinha, Zhihong Shen, Yang Song, Hao Ma, Darrin Eide, Bo-June (Paul) Hsu, and Kuansan Wang. 2015. An Overview of Microsoft Academic Service (MA) and Applications. In Proceedings of the 24th International Conference on World Wide Web (WWW '15 Companion). ACM, New York, NY, USA, 243-246. DOI=<http://dx.doi.org/10.1145/2740908.2742839>*

---

## Definitions

- **IUNI:** The Indiana University Network Science Institute ([iuni.iu.edu](http://iuni.iu.edu))
- **Eligible Individual:** An individual who is an employee or student of an institution participating in the CADRE project.
- **Web of Science or WoS:** The dataset and all future updates that have been provided to IUNI by Clarivate Analytics.
- **Microsoft Academic Graph or MAG:** The dataset and all future updates that have been provided to IUNI by Microsoft Research.
- **Data Enclave or ‘The enclave’:** A dedicated data intensive node in the Karst network that stores the WoS and MAG dataset and is maintained by UITS on behalf of IUNI.
- **Data Manager:** An employee of IUNI who conducts day-to-day administration of the WoS and MAG dataset and supports researchers in accessing the data. The Data Manager serves as the single point of contact for all issues regarding the enclave and the WoS data (See <https://datamgmt.iu.edu/governance/structure.php> for more details).

---

## Sanctions

Indiana University will handle reports of misuse and abuse of information and information technology resources in accordance with existing policies and procedures issued by appropriate authorities. Depending on the individual and circumstances involved this could include the offices of Human Resources, Vice Provost or Vice Chancellor of Faculties (or campus equivalent), Dean of Students (or campus equivalent), Office of the General Counsel, and/or appropriate law enforcement agencies. See policy [IT-02, Misuse and Abuse of Information Technology Resources for more detail](#).

Failure to comply with Indiana University information technology policies may result in sanctions relating to the individual's use of information technology resources (such as suspension or termination of access, or removal of online material); the individual's employment (up to and including immediate termination of employment in accordance with applicable university policy); the individual's studies within the university (such as student discipline in accordance with applicable university policy); civil or criminal liability; or any combination of these.

## Additional Contacts

<i>Subject</i>	<i>Contact</i>	<i>Phone</i>	<i>Email</i>
Web of Science	Matthew Hutchinson	(812) 855-1404	<a href="mailto:maahutch@iu.edu">maahutch@iu.edu</a>
Microsoft Academic Graph	Matthew Hutchinson	(812) 855-1404	maahutch@iu.edu

---

## Forms

If you are interested in accessing the data enclave, please complete the application form at the following address:

---

## History

This policy replaces the ‘Web of Science - Data Access Policy’.