



USC Alzheimer's Disease Research Center (USC ADRC) Data Sharing and Publication Policy

Introduction

Sharing USC ADRC data with the general scientific community is an objective established from the onset in the Request for Applications and is supported by the USC ADRC Executive Committee and representatives of the National Institute on Aging.

This document outlines the DPC recommendations for sharing data from the USC ADRC as well as policies for publication and publication credits for those who use USC ADRC data. The intent of these policies is to facilitate the sharing of data with all interested investigators, to encourage academic productivity, and to provide a mechanism for tracking and archiving data requests, intended analyses and publications related to and resulting from USC ADRC data.

Data Requests

The expectation of both the USC ADRC Executive Committee and the NIA has been that the USC ADRC de-identified data would be made available to the general scientific community. The DPC will review the applications of each investigator requesting data and make a judgment based on their affiliation with a scientific or educational institution, and on the basis of the reason for the request. It is anticipated that most requests will be legitimate and can be approved rapidly, and that only a few will require clarification. The DPC does not feel it is appropriate to monitor IRB compliance by the applicant, however requestors will be asked to provide IRB numbers and expiration dates.

Therefore, we recommend full, open access of all de-identified USC ADRC imaging and clinical data to individuals who register with the USC ADRC and agree to the conditions in the "USC ADRC Data Use Agreement" and who undergo limited screening by the DPC. These data fall under the definition of a "Limited Data Set" as defined in 45 CFR Part 164.514.

Data Request Application and Tracking

A web-based, semi-automated Data Sharing Application and Tracking system has been instituted to manage the data requests. Users will complete a web-based application form and a Data Use Agreement. When the application is received and approved by the DPC, the applicant



will be given permission to access the data using the on-line interactive query interface. Applicants will be required to provide the following information:

- Name
- Address and Contact information
- Academic Affiliation (if any) or Institution/Company Name
- Proposed analyses with a named lead investigator on each analysis
- Certification to each point of the Data Use Agreement.

Applicants will receive annual requests to update the application information. These requests will also solicit responses to the following queries about manuscripts:

- Title of each manuscript in development
- Lead (or senior) author of each manuscript in development and contact information for that author
- Status of each manuscript in development (In development, In submission, In press/published)
- Citation of each published manuscript
- Upload a file of each published manuscript

Non-compliance with the required updates will jeopardize further access to USC ADRC data. The purpose of listing each proposed analysis with a lead investigator is to allow and encourage investigators considering similar projects to collaborate and to identify an individual who is taking responsibility for the use and reporting of requested data.

Publication Policy

The USC ADRC DPC will not attempt to evaluate the scientific merit of proposed analyses, nor evaluate proposed abstracts. The DPC will be charged with the administrative review of manuscripts only.

Regardless of what subset of data is used in analyses or publications, as part of the Data Use Agreement, investigators will be asked to agree to a Publication Policy as follows:

- All manuscripts will acknowledge the methods of USC ADRC data gathering using language recommended by the USC ADRC (see Data Use Agreement).
- All manuscripts will acknowledge the support of NIA using the NIH grant number for the USC ADRC, as well as naming the financial sponsors of the USC ADRC (see Data Use Agreement).



- Manuscripts will be submitted to the USC ADRC DPC for administrative review prior to submission. The purpose of this review will be to make sure that all of the above requirements are met.
- Full citation of all published manuscripts will be provided to the USC ADRC upon acceptance of manuscripts.
- A copy of the manuscript will be provided to the USC ADRC upon publication of the manuscript. If permitted by the journal, a version of the article or link to the article will be available online at the USC ADRC website.

Tracking of Data Use and Publications

If USC ADRC data are used in publications, authors will be required to:

1. Add USC ADRC acknowledgement (see Data Use Agreement)
2. Include recommended language describing USC ADRC methods and data gathering
3. Cite USC ADRC sponsors (see Data Use Agreement)
4. Provide manuscript to USC ADRC DPC for review 2 weeks prior to planned submission (see Data Use Agreement)
5. Apprise USC ADRC of acceptance or rejection
6. Provide manuscript citation to USC ADRC upon acceptance
7. Provide URL to published work, if possible.

Unusual Situations

It is expected that most users of USC ADRC data will follow these guidelines in good faith and that most analyses will be of reasonable quality. The DPC does not intend to review manuscripts for scientific quality, preferring to let the peer-review process sort out quality. It is recognized that USC ADRC data may even be used to support publications with conflicting results. However, we do anticipate the possibility of some unusual circumstances.

1. **Egregiously Poor Manuscripts.** If a review of a proposed manuscript reveals that it is egregiously poor in terms of language, writing or sensible substance, the DPC can recommend to the authors that it not be submitted without significant revision. If the authors choose to submit the manuscript anyway, the DPC will have three options: 1) request that the acknowledgement credit be withheld; 2) request that the authors publish a statement to the effect that while USC ADRC data were used, the USC ADRC did not find this manuscript of sufficient merit to warrant submission for publication, or; 3) revocation of privileges to use USC ADRC data in the future.
2. **Failure to Follow Data Use Agreement.** If users inadvertently violate the Data Use Agreement, it is likely that they will self-correct as infractions are discovered. If users willfully violate the Data Use Agreement, the sole sanction available to the USC ADRC will be to revoke access to the USC ADRC data.
3. **Fraudulent Use of Data.** Open access using only email registration raises the possibility that individuals could obtain access to data under false pretenses. Should the DPC discover an attempt to publish data obtained fraudulently, the data user will be sanctioned through NIH communication with them or their academic supervisors. As soon as the DPC becomes aware of any breach of the Data Use Agreement, immediate steps will be taken to cure the breach or end the violation. This may include discontinuing the user's data access and/or reporting the violation.



4. Journal Objections. It is possible that journal editors may object to the acknowledgement format with the caveat that those USC ADRC names credited have gathered data but not participated in the analysis and writing of the paper. The DPC can authorize exceptions to the acknowledgement policy or can negotiate with journal editors as needed on a case-by-case basis.

Committee Structure and Funding

The USC ADRC Data and Publications Committee was formed by the USC ADRC Executive Committee and will make recommendations to the Executive Committee regarding these and other related policies. Data and publications issues will be adjudicated by the DPC, however, if investigators disagree with the decisions of the DPC, they can appeal to the EC for final arbitration.



Alzheimer's Disease Research Center (USC ADRC)
DATA USE AGREEMENT

I request access to data collected by the University of Southern California Alzheimer's Disease Research Center (USC ADRC) for the purpose of scientific investigation, teaching or the planning of clinical research studies.

I agree to the following terms:

1. I will receive access to de-identified data and will not attempt to establish the identity of /or attempt to contact any of the USC ADRC participants.
2. I will not attempt to make direct contact with USC ADRC PIs or staff at sites concerning the specific results of individual subjects.
3. I will not further disclose these data beyond the uses outlined in this agreement and my data use application. **I understand that redistribution of individual participant-level data in any manner is prohibited.** Reports and publication of summary (not participant-level) data are allowed.
4. I will not utilize AI tools, especially (but not exclusively) those with public-facing interfaces that do not offer guarantees regarding the containment of data inputs for USC ADRC data, recognizing that the use of AI tools, including generative and analytical models, poses a **potential risk of inadvertent data sharing** due to the nature of how these tools process information and may lead to data being sent, saved, viewed, or used in unforeseen ways by parties not covered by the DUA, which is a direct violation of this agreement. See Appendix A for additional AI tool concerns, restrictions, and guidance.
5. I will not disclose any participant-level raw or derived datasets beyond the uses outlined in this agreement. Derived datasets containing participant-level data that I create and that I anticipate will benefit the scientific community, will be submitted to the USC ADRC study PIs (Drs. Helena Chui and Arthur Toga) using the format described in *Derived Data Submission Form* (available in the USC ADRC repository). The study PI will determine whether my data may be distributed through the USC ADRC-LONI and/or other NIA-designated data repository. I understand derived datasets should only be shared through the USC ADRC-LONI and/or other NIA-designated data repository.
6. I will require anyone on my team (i.e. in my lab or company) who utilizes these data to comply with this Data Use Agreement. I understand that I am not allowed to distribute USC ADRC data outside of my team, only the USC ADRC study website operated by LONI-IDA can distribute data.
7. I will accurately provide the requested information for persons who will use these data and the analyses that are planned using these data.
8. I will respond promptly and accurately to annual requests to update information about my use of USC ADRC data.
9. I will comply with any rules and regulations imposed by my institution and its institutional review board in requesting these data.

If I publish abstracts using data from USC ADRC, I agree to the following:

1. I will cite USC ADRC as the source of data and the USC ADRC funding sources in the abstract as space allows.
2. For abstracts, you are *not* required to cite USC ADRC in the authorship line.

If I publish manuscripts using data from USC ADRC, I agree to the following:

1. On the by-line of the manuscript, after the named authors, I will include USC ADRC as an author by using the phrase "for the University of Southern California Alzheimer's Disease Research Center*" with the asterisk referring to the following statement and list of names:

*Data used in preparation of this article were obtained from the University of Southern California Alzheimer's Disease Research Center (USC ADRC) database (ida.loni.usc.edu). As such, the investigators within the USC ADRC contributed to the design and implementation of USC ADRC and/or provided data but did not participate in analysis or writing of this report. A complete listing of USC ADRC investigators can be found at: <https://adrc.usc.edu/about/leadership/>

2. I will include language similar to the following in the Methods section of my manuscripts in order to accurately acknowledge data gathering by the USC ADRC personnel. Depending upon the length



and focus of the article, it may be appropriate to include more or less than the example below. However, inclusion of some variation of the language shown below is mandatory.

Data used in the preparation of this article were obtained from the University of Southern California Alzheimer's Disease Research Center (USC ADRC) database (ida.loni.usc.edu). Founded in 1984 as one of the first five Alzheimer's Disease Research Centers designated by the National Institute on Aging, the University of Southern California Alzheimer's Disease Research Center (USC ADRC) has stood at the forefront of ADRC research for decades. USC's pioneering spirit is bolstered by its historic contributions, notably establishing the first School of Gerontology in the United States in 1975. The USC ADRC is led by Principal Investigators Helena Chui, MD and Arthur Toga, PhD. For up-to-date information, see <https://adrc.usc.edu/>

I will acknowledge funding by the USC ADRC in the support acknowledgement section of the manuscript using language similar to the following:

Data collection and sharing for the University of Southern California Alzheimer's Disease Research Center (USC ADRC) is funded by the National Institute on Aging (National Institutes of Health Grant 5P30AG066530-05). The USC ADRC data are contributed by the NIA-funded projects: 5U19NS120384 (PI Charles DeCarli, MD), R01AG058162 (PI Vasilis Marmarelis, MS, PhD), P01AG052350 (PI Arthur Toga, PhD), 2P30AG066530 (PI Helena Chui, MD, PI Arthur Toga, PhD), R01AG055770-03S1 (PI Hussein Yassine, MD), P30AG066530-01S1 (PI Arthur Toga, PhD), 5U24AG067418-05 (PI William J. Jagust, MD).

3. I will submit all manuscripts to the USC ADRC Data and Publications Committee (DPC) **prior** to submitting them to a journal. This review will not be a scientific review, but it is intended to ensure that the items above are correctly implemented. The DPC will maintain confidentiality of the manuscript and will complete its review within 2 weeks. The DPC can be reached by logging into your LONI-IDA account: 1) click "My Account" in the upper right corner 2) click the "Update" button in the Project line for USC ADRC; 3) On the USC ADRC User Account Update page, select "Publication Update" tab; 4) Enter and upload your publication information; 5) Email your manuscript to the DPC at: adrc@ini.usc.edu Note: in the event that a journal requires a Conflict of Interest form because USC ADRC is listed on the author line of the manuscript, the DPC will provide one on request.
4. I will ensure that Investigators who utilize USC ADRC data use appropriate administrative, physical, and technical safeguards to prevent the use or disclosure of the data other than as provided for by this Agreement.
5. I will report any use or disclosure of the data not provided for by this Agreement of which I become aware within 15 days of becoming aware of such use or disclosure.

IMPORTANT NOTE: *It is the policy of the University of Southern California Alzheimer's Disease Research Center to make analyzed data available to investigators as quickly as possible. However, data analysis for this project is expected to take years as methods for analysis of these datasets evolve. Therefore, I understand that any processed data that I download might be preliminary and that results may change as new methods of analysis are implemented. I will familiarize myself with the analysis methods so that I am aware of the limitations of these data prior to using them for scientific purposes.*

Finally, because "preliminary data" will be posted on the database, in the event that I download data from the USC ADRC database for the purposes of analysis and future presentation/publication in the form of abstracts and/or manuscripts, I will note the version (download date) of the data I download and provide this information in any Methods section (if publishing), and I will check the database to determine if updated data has been provided prior to submission of any material for publication.

USC ADRC maintains the right to modify terms of this agreement and may do so by posting notice of such modifications on this page: <https://adrc.usc.edu/> Any modification made is effective immediately upon posting the modification (unless otherwise stated). You should visit this page periodically to review the current use agreement terms.

APPENDIX A

USC ADRC DUA: Use and Requirements of AI tools

Prohibited use of AI Tools: This USC ADRC Data Use Agreement explicitly prohibits any individual who has obtained USC ADRC data from sharing raw data, specifically participant-specific data, with any other individual or entity, in any forum, in any medium. For a variety of reasons, including the important need to protect the privacy of research participants, all data must be directly released to requesting individuals from our USC ADRC LONI website.

The use of AI tools, including generative and analytical models, pose a **genuine risk of inadvertent data sharing** due to the nature of how AI tools process, store, and regenerate information. AI tools, including but not limited to those with public-facing interfaces, such as OpenAI's ChatGPT, usually do not offer guarantees regarding the containment of data inputs. Any data input into a prompt or output generated from a prompt could result with proprietary data infringed, sent, saved, viewed, misinterpreted, or used in unforeseen public ways. **As a result, the use of such AI tools which may release USC ADRC participant level data to others, is in direct violation of USC ADRC Data Use Agreement policies.**

However, an important exception to the above stated prohibition would be AI tools, including generative and analytical models, which **explicitly prevent sharing data** with others. Such AI tools may be developed and internally trained by individual research groups or academic institutions, and strictly used for the individual research group or academic institution provided that "sharing" data beyond the tool is strictly prohibited. Therefore, USC ADRC data may be analyzed using AI tools which provide guarantees that all data is safeguarded, contained, and will not be released to others.

Note that this amendment does not impose any new restrictions on the use of USC ADRC data but is intended for clarification on item #3 in the DUA (above): "I will not further disclose these data beyond the uses outlined in this agreement and my data use application and understand that redistribution of data in any manner is prohibited."

As of the writing of this Appendix the term "Artificial Intelligence" or "AI" means a broad range of technologies and tools that can generate and regenerate content in various mediums including text, images, videos, and audio, based on the content or data input to generate output in any medium response. The recent commercialization of these AI tools represents a significant challenge in maintaining the privacy of participants in the USC ADRC study, and of participants in human subjects' research in general. Uploading USC ADRC data to a third-party platform that is not explicitly in compliance with the prohibition and restrictions on redistribution of USC ADRC data set forth herein is a clear violation of this Agreement.

USC ADRC requires investigators to adhere to the following terms when deciding whether a project utilizing these tools is in compliance with the terms of this Agreement:

- Computational resources
 - The application of any statistical or other analytic methods to USC ADRC data - whether branded as 'AI' and 'AI Tool' or otherwise - should not be performed using computational or AI resources owned by third parties who engage in the long-term retention of user content that is either shared with other parties or used in the training of public-facing models, such activity could include the contents of the USC ADRC data set. This includes the use of third-party platforms for inference (e.g. the inclusion of USC ADRC data as part of an input or 'prompt' for a generative language model), training (including both the training of foundations model and the fine-tuning of pre-trained models via transfer learning or some other process), prediction, or any other task that could result] in USC ADRC data being retained, shared, regenerated or otherwise by entities that are not in compliance with the DUA.
 - Some commercial entities may offer an 'opt-out' clause in the terms of use for their models, allowing users to decline the right of the entity to retain the data for training and other purposes. This does not constitute sufficient protection, and under no circumstances should this be viewed as an acceptable safeguard for sharing USC ADRC data with these third parties.
 - Investigators making use of remote computational resources such as Microsoft Azure, Google Cloud, Amazon Web Services, Runpod, or other computational services for the purposes of training AI models must take care to review the terms of use for the service to ensure that their workflow is secure.
- Public release of models trained using USC ADRC data
 - The use of USC ADRC data in training models is both permitted and encouraged, provided that the process is carried out in compliance with the DUA as outlined above.



However, investigators who intend on making the weights of a trained model publicly available should consider whether the model could be used to easily reconstruct parts of the USC ADRC data set. USC ADRC prohibits release of participant level data, even if it is sufficiently processed or altered and lack USC ADRC subject codes. See Term (3) above **"I understand that redistribution of individual participant-level data in any manner is prohibited"**. All participant level data generated by USC ADRC can only be released by the LONI-IDA website. Individual investigators who have generated participant level data based on USC ADRC data can request for their data to be shared by contacting Dr. Toga at adrc@ini.usc.edu

- This precaution is particularly important in the case of large language models and other highly overparameterized generative models. Numerous studies have demonstrated that it is possible to extract training examples from these large and complex models, and investigators should keep this fact in mind.

I understand that failure to abide by these guidelines will result in the termination of my privileges to access USC ADRC data. Furthermore, USC ADRC reserves the right to pursue damages for actions which violate USC ADRC guidelines.

Electronic Signature of User