

Collaboration Network Data Dictionary

About

This data is provided to meet the data availability requirements for a scientific publication. Data is derived from NIH's internal datasets and from Digital Science's Dimensions Platform. Data from Digital Science is based on metadata as of March 2024 from Digital Science's Dimensions platform, available at <https://app.dimensions.ai/>. Access was granted under license agreement with the National Cancer Institute. Researchers interested in exploring the data further should visit the Dimensions platform website.

Available Data Files

Data File	Row Counts	Last update
nodes.csv	10,330	3/13/2024
agg_edges.csv	262,907	3/13/2024

Dataset Development Notes

Network Inclusion Criteria

NIH PIs qualified for network inclusion if they were awarded one of the following between Fiscal Year (FY) 2017 and FY 2022: Research Project Grants (RPGS): R00, R01, R03, R15, R16, R21, R33, R34, R35, R36, R37, R50, R56, R61, RC1, RC2, RC3, RC4, RF1, RL1, RL2, RL9, P01, P42, PM1, PN1, RM1, UA5, UC1, UC2, UC3, UC4, UC7, UF1, UG3, UH2, UH3, UH5, UM1, UM2, U01, U19, U24, U34, U3R, DP1, DP2, DP3, DP4, DP5; Research Center Grants: P50, U10, U19, U54, U2C; and Cancer Center Support Grants: P30. Project and Subproject PIs were included. PIs who received 21st Century Cures Act in FY 2017 - FY 2023 were included. Some 21st Cures Act recipients have SBIR/STTR awards (R41, R42, R43, R44).

Project Counting

NIH projects included in the network were counted by base project number per fiscal year (FY). This means multiple awarded applications in the same FY for the same base project number were counted once.

Identifying 21st Century Cures Act Funding

21st Century Cures Act NIH grants are those that received any funding from the 21st Century Cures Act - Beau Biden Cancer Moonshot and were identified via Assistance Listing (number 93.353). Publications acknowledging 21st Century Cures Act NIH grants were identified through automated linkages in the Dimensions database. Publications associated with 21st Century Cures Act-supported NIH intramural projects were manually collected.

Withheld Values in the Data

Some NIH PPIDs, Application IDs, and base project numbers were withheld if they were not available in public NIH data sources (RePORTER, EXPORTER) at the time of data release (March 2025). If a column may contain withheld values, this is noted in the column description.

nodes.csv Column Name	nodes.csv Column Description	Includes Withheld Values
ppid	NIH PPID. Any PPIDs that were not publicly available at the time of data release were withheld. Withheld values have the format 99999999X.	Y
researcher_ids	Semicolon-delimited list of Dimensions researcher IDs matched to the PI. This field will be empty if no researcher IDs were matched to the PI at the time of dataset development.	
n_pubs	Total number of publications (article and proceeding) published between 2012 and 2023 linked to Dimensions researcher IDs matched to the PI.	
n_nci_funded_pubs	Number of NCI-supported publications published between 2012 and 2023 linked to Dimensions researcher IDs matched to the PI.	
n_nci_funded_projs	Number of funded projects with any NCI funding between FY2012 and 2023 associated with the PI in the NIH application tables.	
nci_role	One of "project_pi", "research_subproject_pi", or "supplement_pi". The role of the PI derived from all role(s) the PI had on their qualifying award(s), using the priority order "project pi" > "research subproject pi" > "supplement_pi". Each PI will only have one role.	
qual_nci_appl_ids	Semicolon-delimited list of NIH Application IDs between FY2017 and FY2023 that qualified the PI to be in the dataset. Any application IDs that were not publicly available at the time of data release were withheld. Withheld values have the format WITHHELDAPPLX.	Y
qual_nci_projs	Semicolon-delimited list of projects between FY2017 and FY2023 by base project number that qualified the PI to be in the dataset. Any NIH project numbers that were not publicly available at the time of data release were withheld. Withheld values have the format WITHHELDPROJX.	Y

nodes.csv Column Name	nodes.csv Column Description	Includes Withheld Values
collab_nci_funded_only	A boolean flag indicating if the PI collaborated with others on NCI funded publications/projects in the time window 2012-2023. For publications, this is calendar year (CY) 2012 - 2023, for projects this is fiscal year (FY) 2012 - 2023. This field will be empty if the PI did not collaborate with other NCI PIs in the network.	
moonshot_pi	A y/n boolean flag indicating if the PI had 21st Century Cures Act Funding in fiscal year (FY) 2017 - 2023.	
min_moonshot_fy	The earliest fiscal year (FY) the PI received 21st Century Cures Act funding. This field will be empty if the PI did not receive 21st Century Cures Act funding.	
moonshot_projects	The 21st Century Cures Act funded projects the PI is associated with, with the fiscal year (FY) of the project. This field will be empty if the PI did not receive 21st Century Cures Act funding.	
med_authors_overall	Median number of authors on the PI's publications published between 2012 and 2023. PI publications are identified based on linkages to a Dimensions researcher IDs matched to the PI. This field may be empty if no publication linkages were available at the time of dataset development. This field may be empty if no publication linkages with author counts were available at the time of dataset development.	
avg_authors_overall	Average number of authors on the PI's publications published between 2012 and 2023. PI publications are identified based on linkages to a Dimensions researcher IDs matched to the PI. This field may be empty if no publication linkages with author counts were available at the time of dataset development.	

nodes.csv Column Name	nodes.csv Column Description	Includes Withheld Values
nci_pub_avg_authors	Average number of authors on the PI's NCI-supported publications published between 2012 and 2023. PI publications are identified based on linkages to a Dimensions researcher IDs matched to the PI. This field may be empty if no publication linkages with author counts were available at the time of dataset development.	
msfunded_pub_avg_authors	Average number of authors on the PI's 21st Century Cures Act supported publications published between 2012 and 2023. PI publications are identified based on linkages to a Dimensions researcher IDs matched to the PI. This field will be empty if the PI did not receive 21st Century Cures Act funding in 2017-2023 or if they did not collaborate with another PI who received 21st Century Cures Act funding on a publication.	
collab_types	One of "Publication", "Project", or "Both" indicating the collaborative events that the PI participated within network, in the time window 2012-2023. This field will be empty of the PI did not collaborate with another NCI PI in the network in 2012-2023.	
n_collab_pubs	Number of publications where the NCI PI collaborated with another NCI PI in the network in the time window 2012-2023. This field will be empty of the PI did not collaborate with another NCI PI in the network in 2012-2023.	
n_collab_projs	Number of projects where the NCI PI collaborated with another NCI PI in the network in the time window 2012-2023. This field will be empty of the PI did not collaborate with another NCI PI in the network in 2012-2023.	
n_network_collaborators	Overall number of NCI collaborators (either co-authors or co-PIs) within network. This field will be empty of the PI did not collaborate with another NCI PI in the network in 2012-2023.	

nodes.csv Column Name	nodes.csv Column Description	Includes Withheld Values
n_network_coauthors	Number of NCI co-authors on collaborating publications within network. This field will be empty of the PI did not collaborate with another NCI PI in the network in 2012-2023.	
n_network_copis	Number of NCI co-PIs on collaborating projects within network. This field will be empty of the PI did not collaborate with another NCI PI in the network in 2012-2023.	
n_ms_network_collaborators	Overall number of collaborators (either co-authors or co-PIs) that are were also supported by the 21st Century Cures Act in 2017-2023. This field will be empty if the PI did not receive 21st Century Cures Act funding in 2017-2023.	
n_ms_network_coauthors	Number of co-authors on collaborating publications who also received 21st Century Cures Act funding. This field will be empty if the PI did not receive 21st Century Cures Act funding in 2017-2023.	
n_ms_network_copis	Number of co-PIs on collaborating projects who also received 21st Century Cures Act funding. This field will be empty if the PI did not receive 21st Century Cures Act funding in 2017-2023.	
n_collab_nci_docs	Number of NCI funded publications and projects where the NCI PI collaborated with another NCI PI in the network in the time window 2012-2023. This field will be empty of the PI did not collaborate with another NCI PI in the network in 2012-2023.	
n_collab_nci_funded_pubs	Number of NCI funded publications where the NCI PI collaborated with another NCI PI in the network in the time window 2012-2023. This field will be empty of the PI did not collaborate with another NCI PI in the network in 2012-2023.	
n_collab_nci_funded_projs	Number of NCI funded projects where the NCI PI collaborated with another NCI PI in the network in the time window 2012-2023. This field will be empty of the PI did not collaborate with another NCI PI in the network in 2012-2023.	

nodes.csv Column Name	nodes.csv Column Description	Includes Withheld Values
n_collab_ms_docs	Number of 21st Century Cures Act-supported publications and projects where the PI collaborated with another PI who received 21st Century Cures Act funding. This field will be empty if the PI did not receive 21st Century Cures Act funding in 2017-2023 or if they did not collaborate with another PI who received 21st Century Cures Act funding.	
n_collab_msfunded_pubs	Number of 21st Century Cures Act-supported publications where the PI collaborated on with another 21st Century Cures Act PI. This field will be empty if the PI did not receive 21st Century Cures Act funding in 2017-2023 or if they did not collaborate with another PI who received 21st Century Cures Act funding.	
n_collab_msfunded_projs	Number of projects with 21st Century Cures Act funding (excluding Type 3 applications and P30s) where the PI collaborated on with another PI who had 21st Century Cures Act funding. This field will be empty if the PI did not receive 21st Century Cures Act funding in 2017-2023 or if they did not collaborate with another PI who received 21st Century Cures Act funding.	
collab_ms_funded_only	A y/n boolean flag indicating if the PI collaborated with others on 21st Century Cures Act funded publications/projects in the time window 2012-2023. This field will be empty if the PI did not receive 21st Century Cures Act funding in 2017-2023.	
collab_post_ms	A y/n boolean flag indicating if the PI collaborated with other 21st Century Cures Act funded PIs after receiving 21st Century Cures Act funding. This field will be empty if the PI did not receive 21st Century Cures Act funding in 2017-2023.	

agg_edges.csv Column Name	agg_edges.csv Column Description	Includes Withheld Values
source	NIH PPID for the source node in the collaboration pair. Any PPIDs that were not publicly available at the time of data release were withheld. Withheld values have the format 99999999X.	Y
target	NIH PPID for the target node in the collaboration pair. Any PPIDs that were not publicly available at the time of data release were withheld. Withheld values have the format 99999999X.	Y
source_role	One of "project pi", "supplement pi", or "research subproject pi".	
target_role	One of "project pi", "supplement pi", or "research subproject pi".	
source_moonshot_pi	A y/n boolean flag indicating if the source node had 21st Century Cures Act funding in 2017-2023.	
target_moonshot_pi	A y/n boolean flag indicating if the target node had 21st Century Cures Act funding in 2017-2023.	
ms_collab	A boolean flag (y/n) indicating if both the source and target nodes had 21st Century Cures Act funding in 2017-2023.	
source_min_moonshot_fy	The earliest 21st Century Cures Act funding year of the source node. This field is empty if the source did not have 21st Century Cures Act funding in 2017-2023.	
target_min_moonshot_fy	The earliest 21st Century Cures Act funding year of the target node. This field is empty if the target did not have 21st Century Cures Act funding in 2017-2023.	
pair_min_moonshot_fy	The earliest 21st Century Cures Act funding year of the collaboration pair. This field is empty if the source and the target both did not have 21st Century Cures Act funding in 2017-2023.	
n_tot_collabs	The total number of documents (publications/projects) the pair has collaborated on within network in the time window 2012-2023.	
n_pub_collabs	The number of publications the pair has collaborated on within network in the time window 2012-2023.	
n_proj_collabs	The number of projects the pair has collaborated on within network in the time window 2012-2023.	
n_nci_pub_collabs	The number of NCI-supported publications the pair has collaborated on within network in the time window 2012-2023.	
n_nci_proj_collabs	The number of NCI-supported projects, excluding Type 3s, that the pair has collaborated on within network in the time window 2012-2023.	

agg_edges.csv Column Name	agg_edges.csv Column Description	Includes Withheld Values
n_ms_pub_collabs	The number of 21st Century Cures Act-supported publications the pair has collaborated on.	
n_ms_proj_collabs_excl_type3	The number of projects the pair has collaborated on that are linked to one or more 21st Century Cures Act-funded applications, excluding Type 3s.	
first_collab_year	The earliest collaboration year between the pair within network in the time window 2012-2023.	
first_collab_type	One of "Publication", "Project", or "Multiple", indicating the type of document(s) with the earliest year between the pair, within network in the time window 2012-2023.	
collab_type_all	One of "Publication", "Project", or "Multiple", indicating overall collaboration types between the pair within network in the time window 2012-2023.	
earliest_pub_collab_year	The earliest year of collaboration on a publication between the two PIs within network in the time window 2012-2023. This field will be empty if the PIs did not collaborate on a publication in 2012-2023.	
latest_pub_collab_year	The latest year of collaboration on a publication between the two collaborators. This field will be empty if the PIs did not collaborate on a publication in 2012-2023.	
earliest_proj_collab_year	The earliest year of collaboration on a project between the two PIs within network in the time window 2012-2023. This field will be empty if the PIs did not collaborate on a project in 2012-2023.	
latest_proj_collab_year	The latest year of collaboration on a project between the two collaborators. This field will be empty if the PIs did not collaborate on a project in 2012-2023.	
nci_proj_collab_funding	One of "NCI Funded" or "Not NCI Funded" to indicate the overall funding source of the collaborative pair for projects within network in the time window 2012-2023.	
nci_pub_collab_funding	One of "NCI Funded" or "Not NCI Funded" to indicate the overall funding source of the collaborative pair for publications within network in the time window 2012-2023. Note that "NCI Funded" can include other funding sources.	
nci_collab_funding_overall	One of "NCI Funded" or "Not NCI Funded" to indicate the overall funding source of the collaborative pair for projects and publications within network in the time window 2012-2023. Note that "NCI Funded" can include other funding sources.	

agg_edges.csv Column Name	agg_edges.csv Column Description	Includes Withheld Values
ms_collab_period	One of "Pre-Moonshot", "Post-Moonshot", or "Pre&Post-Moonshot". If none of the projects or publications between a collaborative pair occurred in a year \geq either one of the collaborators earliest 21st Century Cures Act funding fiscal year (FY), then all collaboration occurred prior to 21st Century Cures Act funding (Pre-Moonshot). If some projects and/or publications occurred in a year \geq the earliest 21st Century Cures Act funding FY of either collaborator and others did not, then this would be a mix (Pre&Post-Moonshot). If all the projects and/or publications occurred in year \geq the earliest 21st Century Cures Act funding FY of either collaborator, then all collaboration occurred after 21st Century Cures Act funding (Post-Moonshot). This field will be empty if the PI did not receive 21st Century Cures Act funding in 2017-2023 and did not collaborate with another PI who received 21st Century Cures Act funding in 2017-2023.	
ms_proj_collab_funding	One of "Moonshot Funded" or "Not Moonshot Funded" to indicate the overall funding source of the collaborative pair for projects within network in the time window 2012-2023. Projects were determined to be "Moonshot Funded" by presence of a 21st Century Cures Act-funded application (excluding Type 3 applications). This field will be empty if the PI did not receive 21st Century Cures Act funding in 2017-2023 and did not collaborate with another PI who received 21st Century Cures Act funding in 2017-2023.	
ms_pub_collab_funding	One of "Moonshot Funded" or "Not Moonshot Funded" to indicate the overall funding source of the collaborative pair for publications within network in the time window 2012-2023. Note that "Moonshot Funded" can include other funding sources. Publications were matched to a mapping table of supported publications. This field will be empty if the PI did not receive 21st Century Cures Act funding in 2017-2023 and did not collaborate with another PI who received 21st Century Cures Act funding in 2017-2023.	

agg_edges.csv Column Name	agg_edges.csv Column Description	Includes Withheld Values
ms_collab_funding_overall	One of "Moonshot Funded" or "Not Moonshot Funded" to indicate the overall funding source of the collaborative pair for projects and publications within network in the time window 2012-2023. Note that "Moonshot Funded" can include other funding sources. Projects were determined to be "Moonshot Funded" by presence of a 21st Century Cures Act-funded application (excluding Type 3 applications) while publications were matched to a mapping table of supported publications. This field will be empty if the PI did not receive 21st Century Cures Act funding in 2017-2023 and did not collaborate with another PI who received 21st Century Cures Act funding in 2017-2023.	
n_post_ms_single_proj	The number of projects between the collaboration pair that occurred \geq the earliest 21st Century Cures Act funding fiscal year of either of the collaborators.	
n_post_ms_single_pub	The number of publications between the collaboration pair that occurred \geq the earliest 21st Century Cures Act funding fiscal year of either of the collaborators.	
msfunded_pub_activities	Semicolon-delimited list of the activity and IC codes of 21st Century Cures Act projects that supported publications shared by the two collaborators.	
msfunded_proj_activities	Semicolon-delimited list of the activity and IC codes of 21st Century Cures Act projects shared by the two collaborators.	
non_msfunded_pub_activities	Semicolon-delimited list of NIH the activity and IC codes of non-21st Century Cures Act projects that supported publications shared by the two collaborators. This field will be empty if there were no NIH projects supporting publications shared by the two PIs at the time of dataset development.	
non_msfunded_proj_activities	Semicolon-delimited list of the NIH activity and IC codes of non-21st Century Cures Act projects shared by the two collaborators. This field will be empty if there were no NIH projects shared by the two PIs at the time of dataset development.	