Datalad Gooey

Release 0.2.0+43.g5bd6b92

DataLad team

Feb 20, 2024

CONTENTS

1	What DataLad Gooey is not	3
2	Overview	5
3	Commands and API	27
In	dex	35

DataLad Gooey is a Graphical User Interface (GUI) for using DataLad, a free and open source distributed data management tool. DataLad Gooey is compatible with all major operating systems and allows access to DataLad's operations via both a simplified and complete suite.

DataLad Gooey							— C]	\times
Eile <u>D</u> ataLad <u>V</u> iew <u>U</u> tilities <u>H</u> e	lp								
Name	Туре	State	Command	Properties					
✓ □ projects	directory		Create a data	set					
✓ my-project	directory								ı.
✓ □ code	directory		Create at		Jsers\mih\Docur	nents\projects\n	ny-projec	t 🗋	
analyse_data.R	file	untracked	Dataset wit	\checkmark					
📄 clean_data.R	file	⊖ untracked	Decister in	superdataset	Jsers\mih\Docur			. —	
✓ ☐ data_clean	directory		Register in	superdataset	isers\min\Docur	nents\projects\n	ny-projec		
🖹 data_clean.csv	file	\ominus untracked	Configurati	on procedure(s)	text2git			\sim	
> data_raw	directory				Add	Update	Remo	ove	
✓ ☐ figures	directory				Add	opute	Kenn	ove	
dotplot_correlati	file	😑 untracked			text2git				
> 🗋 .datalad	directory								
CITATION.cff	file	😑 untracked							
	file	😑 untracked							
README.md	file	😑 untracked	OK SU						
.gitattributes	file	⊘ clean	OK if target	directory not empty	▼				
> in other-project	directory					OK	Ci	ancel	
Command log Error log									
run(ok): C:\Users\mih\Documents\p create(ok): C:\Users\mih\Document action summary: create (ok: 1) run (ok: 1)			AppData\Local\	datalad.org\]					
							C	Clear	

While using DataLad Gooey assumes at least some familiarity with DataLad concepts, the simplified command suite makes starting with DataLad easier via tailor-made command selections, condensed parameter specifications, and tool tips. The current core functionality supported via the simplified suite includes:

- · cloning a dataset
- creating a dataset
- creating a sibling (GIN, GitHub, WebDAV)
- dropping/getting content
- pushing data/updates to a sibling
- saving the state of a dataset
- updating from a sibling

In addition, DataLad Gooey adds support for querying and setting *credentials*, *git-annex metadata*, and *general meta-data*.

CHAPTER

WHAT DATALAD GOOEY IS NOT

DataLad Gooey has a number of cool features, but here are features that you will need to use other tools for:

- An interface to visualize revision histories of DataLad datasets. Please refer to many of the available visual Git visualization tools
- An interface for advanced Git operations such as branching, resetting, reverting, or otherwise interacting with commit history. Please refer to your favourite Git client or the command line for these operations. DataLad Gooey will detect such external operations, and will update its view accordingly.

CHAPTER

TWO

OVERVIEW

2.1 Installation

2.1.1 Installing via PyPI

You can install the latest version of datalad-gooey from PyPI. It is recommended to use a dedicated virtualenv:

```
# Create and enter a new virtual environment (optional)
python3 -m venv ~/.venvs/datalad-gooey
source ~/.venvs/datalad-gooey/bin/activate
```

```
# Install from PyPI
pip install datalad_gooey
```

Dependencies

Because this is an extension to datalad, the installation process also installs the datalad Python package, although all recursive dependencies (such as git-annex) are not automatically installed. For complete instructions on how to install datalad and git-annex, please refer to the DataLad Handbook.

2.1.2 Installing on Windows

The current version of datalad-gooey comes with two installers, a full installer and a gooey-only installer. The full installer will install datalad-gooey as well as git and git-annex. it requires admin privileges to execute successfully. The gooey-only installer will only install datalad-gooey. It can be executed as admin user or as non-admin user. If you use the gooey-only installer, git and git-annex have to be provided by other means, e.g. an administrator installs them.

Note: if the full installer is executed and the system has already a newer version of git or git-annex installed, the git or git-annex-installer can be canceled and installation of the remaining components will continue.

The installers can be downloaded here.

2.1.3 Installing on Linux

Install DataLad Gooey via PyPI while specifying the --user flag:

```
pip install --user datalad_gooey
```

Then run the following line to ensure that the application's desktop file is generated:

```
datalad gooey --postinstall
```

2.1.4 Installing on macOS

macOS application pending

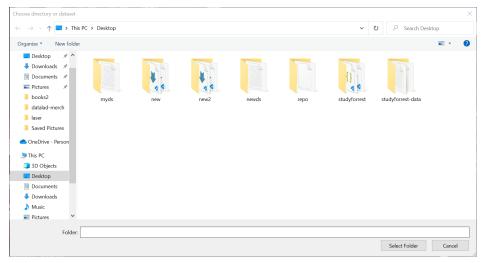
Until the macOS application is available to allow standard installation into the Applications folder, macOS users can install DataLad Gooey via PyPI.

```
# Install from PyPI
pip install datalad_gooey
```

2.2 Getting started

This section walks through the first steps to take when using DataLad Gooey and describes its main pieces and functions.

In order to start up DataLad Gooey, double-click the program's icon on your Desktop or in your Explorer if you are on Windows, find it among your Desktop files if you are using Linux, or launch it from the Terminal if you are on Mac¹. An initial dialog lets you select a base directory. Navigate to any directory you would like to open up, select it, and click Select Folder.



¹ Regardless of the operating system you are using, if you used pip to install datalad gooey you can also start it up from the command line, running datalad gooey. The optional --path argument lets you specify the root directory.

If you Cancel this dialog, the DataLad Gooey will open your home directory. The root directory can be changed at any later point using the File \rightarrow Set base directory submenu from the top task bar, or from the right-click context menus of directories in the file tree.

DataLad Gooey	- 🗆 X
Select a base directory for DataLad	×
\leftarrow \rightarrow \checkmark \uparrow \blacksquare « Desktop » science	✓ ♥ Search science
Organize • New folder	≣≣ - ?
Desktop 🖈 ^ Name	✓ Date modified
 Downloads project_gamma Documents project_beta 	14/10/2022 16:47 14/10/2022 16:46
Pictures project_alpha books2	14/10/2022 16:46
datalad-merch	
laser Saved Pictures	
OneDrive - Person	
a This PC	
🗊 3D Objects	
📃 Desktop	
	>
Folder: project_alpha	
	Select Folder Cancel

2.2.1 Initial configuration

If you are using DataLad and Git for the first time on your computer, make sure that the first thing you do is to set your Git identity by clicking Utilities \rightarrow Set author identity in the tab at the top of the application. Add your name and email address in the resulting dialog. This information will be used to include author information to the commands you will be running, and is required for many commands.

Datalad Gooey, Release 0.2.0+43.g5bd6b92

ame	Туре		State	^	С	ommand	Metadata	History	Properties
studyforrest	dataset				G	et content			
> 🗋 .datalad	directory					et content			
> 🗋 code	directory					Get conten	t in dataset at	C:\Users\a	dina\Desktop
> src	directory					Get file con	tent	\checkmark	
> 📄 stimuli	directory							Not set	
> 🗋 sub-01	directory					Limit to		NOT SET	
> sub-02	Set (Git) autho	r identity					Add	Upda
> sub-03	Jet (c	sity dutilo	ridentity					<u>)</u>	
> sub-04	Many Da	ataLad op	perations i	requi	re a d	configured a	uthor identity	/.	
>sub-05							the user's Git		
	9					ciate datase n or entity.	t	OK	Cancel
	Г				Jerso	nor entity.		_	
Command log Er	Name	Elena Pis	copia						
	E-mail	elena@e	xample.ne	et				_	
Get any dataset con									
This command only						Save	Cancel		
dataset from some s						Save	Guncer		
D									
By default this comm across potential sub									
the directory are obt	tained. Red	cursion in	ito subdat	asets	s is su	pported to	o. If		
enabled, relevant su	bdatasets	are deteo	ted and i	nstall	ed ir	order to			
fulfill a request.									
Known data location	is for each	requeste	ed file are	evalu	uated	and data a	re		
obtained from some							iration		
and possibly assigned	ed remote	priorities). unless a	spec	cific s	ource is			

2.2.2 Application Overview

The User Interface

In general, the DataLad Gooey interface has three main sections: A tree view on the upper left, pane on the upper right containing Command, Metadata, History``and ``Properties tabs, and the different log views at the bottom.

DataLad Gooey								_		\times
File DataLad View Utilities Help										
Name	Туре	State	^	Command	Metadata	History	Properties			
 istudyforrest i.datalad code 	dataset directory directory			Create a GitH	5					^
 Code src stimuli 	directory directory				ng for dataset itory name on		C:\Users\adina	\Desktop	\studytoi	
 sub-01 ses-localizer 	directory directory			Make GitHu Sibling nam	ub repo privat [,] ne	e	github			
✓ ☐ func Ø sub-01_ses-local		⊘ clean		Access prot	tocol		https		>	~
isub-01_ses-local isub-01_ses-local		⊘ clean⊘ clean	~				OK		Cancel	
Command log Error log Help										
										~
									Clear	

The tree view should display files and directories on your computer, starting from the root directory picked at start up. Double-clicking directories allows you to expand their contents, and to navigate into directory hierarchies. Double-clicking files will attempt to open them in your systems configured default application. You will notice that the Type and State annotations next to the file and directory names reveal details about files and directories: You can distinguish directories and DataLad datasets and files. Within datasets, files are either annexed-file's or file's, depending on how these files are tracked in the dataset. The State property indicates the version-state of files, datasets, or directories: A new file, for example, would be annotated with an untracked state, a directory with a newly added unsaved change would be modified, and neatly saved content would be annotated with a clean tag.

In addition to the information in the tree view, the Properties and History tab will load information for any selected directory or file. The History tab displays past commits associated with the file or directory, and the Properties tab displays known information such as the annex key or file size of annexed files.

Running a DataLad command

There are two ways of running DataLad command: either through the Dataset menu at the top, or by right-clicking on files, directories, or datasets in the tree view. The latter option might be simpler to use, as it only allows commands suitable for the item that was right-clicked on, and prefills many parameter specifications. The screenshot below shows the right-click context menu of a dataset, which has more available commands than directories or files.

D DataLad Gooey File DataLad View Utilities Help	- 🗆 X
Name Type Command Image: Studyforre Image: Studyforre Image: Studyforre Image: Image: Studyforre Image: Studyforre Image: Studyforre Image: Image: Image: Studyforre Image: Studyforre Image: Studyforre Image: Image: Image: Studyforre Image: Studyforre Image: Studyforre Image: Image	Metadata History Properties Clone a dataset Create a dataset Create a GIN sibling Create a GitHub sibling Create a GitHub sibling Create a WebDav sibling Drop content Get content Push data/updates to a sibling Save the state in a dataset Update from a sibling OK Cancel
	Clear

Once a DataLad command is selected, the Command tab contains relevant parameters and configurations for it. The parameters will differ for each command, but hovering over their submenus or editors will show useful hints what to set them to, and the Help tab below displays the commands detailed documentation. Little system-specific icons in the command panel can help you identify required input, or parameters that were wrongly specified (see detailed screenshot on the right).

ime	Туре	State	 Command 	Metadata History	Properties		
studyforrest dataset datalad directory			Create a Git	Hub cibling			
			Create a Git	Hub sibiling			
> 🗋 code		Create sib	ling for dataset at	C:\Users\adina\Desktop\stud		dyfor	
> 🗋 src		New repo	sitory name on Github (!)	Not set			
> 📋 stimuli							
✓ □ sub-01 directory				lub repo private			
 ses-localizer 	directory		Sibling na	me	github		
✓ ☐ func	directory		Access pr	otocol	https		
sub-01_ses-loc		⊘ clean					>
sub-01_ses-loc		⊘ clean				_	
sub-01_ses-loc	al annexed-file	⊘ clean			OK	Ca	ancel
ommand log Error log Hel	2						
						C	lear
						C	lear
Command Metadat	a History	Prop	perties			C	lear
metadat	a History	Prop	perties			C	lear
metadat				:top\studyforrest		c	(lear
Create a GitHub sibling	set at	C:\User	rs\adina\Desk	top\studyforrest			(lear
Create a GitHub sibling	set at on Github (!)	C:\User	rs\adina\Desk	top\studyforrest			ilear
Treate a GitHub sibling Create sibling for data New repository name	set at on Github (!)	C:\User	rs\adina\Desk t	.top\studyforrest		C	llear
Create a GitHub sibling Create sibling for data New repository name Make GitHub repo pri	set at on Github (!)	C:\User	rs\adina\Desk t	top\studyforrest			ilear

Once input validation passes on all parameters, the OK button will become functional and will execute the command. The Command log will continuously update you on the state of running and finished commands, displaying, where available, progress bars, result reports, or command summaries.

Command log Error log Help	
Running: clone(<i>source='///', path=</i> 'C:/Users/adina/Desktop/super')	^
<pre>install(ok): C:\Users\adina\Desktop\super (dataset) -> Done: clone(source='///', path='C:/Users/adina/Desktop/super')</pre>	
	~ Clear

During command execution, a small hammer symbol lets you interrupt ongoing commands. If you accidentally started to get hundreds of Gigabytes worth of files instead of only one directory, clicking this button will stop the command.

		Clear
Started `clone`	100%	🎀 clone

Note: The stop button will stop the execution at the next possible chance, which is after the next part of the command (such as the retrieval of the next file) has finished. This means it is not an immediate stop like a CTRL-C press in the command line, but a safe interruption. This also means that stopping might take longer than you expect, depending on the command that gets interrupted.

Should a command fail, a detailed traceback with details about the failure will be send to the Error log tab right next to the Command log. You can use the information from this tab to investigate and fix problems.

Navigation

The interface can be navigated via mouse clicks, or, on most operating systems, via keyboard shortcuts as well. Low lines under specific letters of menus or submenus identify the shortcut². Accessing the shortcut to a menu requires pressing Alt and the respective letter: Alt + f for example will open the File menu. Pressing further letters shortcuts to submenu actions: Alt + f + q will shortcut to Quit and close the application, while Alt + d + g will open a get command in the Command panel.

In addition, path parameters (such as the dataset parameter) can be filled via drag and drop from your system's native file browser.

The View Menu

The View menu contains two submenus that allow you to alter the appearance of the interface. Whenever you change the appearance of the interface, you need to close and reopen the program in order to let the change take effect.

The Theme submenu lets you switch between a light, dark, and system theme, shown below in order.

 $^{^2}$ Windows users may not automatically see underlined letters. To make them visible, press the Alt key. Mac users won't see underlined letters as it would violate the guidelines of macOS graphical user interface aqua.

sme	_ ^			Command Metadata His	story Properties
studyforrest	Type dataset	State	0		noy riopinis
 studyforrest datalad 	directory			Get content	
> Code	directory directory			Get content in dataset	at a\Desktop\studyforr
	directory			Get file content	
> sub-01	directory			Limit to	Not set
> sub-02	directory directory				Add Update
	directory				Add Update
>	directory directory				OK Cancel
				-	
mmand log Error	log Help				
Get any dataset cor					
This command only	operates on da	itaset conten	t. To o	btain a new independent	
By default this com across potential sul	mand operates odatasets, i.e. if	recursively w a directory is	ithin a provid	dataset, but not led, all files in is supported too. If	
the directory are of enabled, relevant s	stained. Recursio	on into subda	tasets	is supported too. If	
fulfill a request.					
Known data locatio	ns for each requ	ested file an	evalu	ated and data are it-annex configuration ific source is	
obtained from som and possibly assign	e available locat ed remote prior	tion (accordir ities), unless	ng to g a spec	it-annex configuration ific source is	
enorified					
_					
Datal ad Groney					
					- 0
DataLad View Ut	nities Help				
				Command Metadata	
studyforrest	dataset				
datalad code	directory directory				
> 📄 src					
 stimuli sub-01 	directory directory				
> sub-02					
> isub-03					
> sub-04	directory directory				
5 🖿 mis-06	directions				OK Cancel
mmand log Error	log Help				
emmand log Error Get any dataset co		tories/subda	tasets)		1
Get any dataset co	ntent (files/direc				1
Get any dataset co This command only dataset from some	ntent (files/direc / operates on da source use the	itaset conten 'clone' comn	t. To o hand.		
Get any dataset co This command only dataset from some	ntent (files/direc / operates on da source use the	itaset conten 'clone' comn	t. To o hand.		
Get any dataset co This command only dataset from some	ntent (files/direc / operates on da source use the	itaset conten 'clone' comn	t. To o hand.		
Get any dataset co This command only dataset from some By default this com across potential su the directory are of enabled, relevant s fulfill a request.	ntent (files/direc v operates on da source use the mand operates bdatasets, i.e. if stained. Recursi ubdatasets are o	itaset conten 'clone' comm recursively w a directory is on into subdi letected and	t. To o hand. ithin a provio atasets install	btain a new independent dataset, but not Jed, all filles in is supported too. If ed in order to	
Get any dataset co This command only dataset from some By default this com across potential su the directory are of enabled, relevant s fulfill a request.	ntent (files/direc v operates on da source use the mand operates bdatasets, i.e. if stained. Recursi ubdatasets are o	itaset conten 'clone' comm recursively w a directory is on into subdi letected and	t. To o hand. ithin a provio atasets install	btain a new independent dataset, but not Jed, all filles in is supported too. If ed in order to	
Get any dataset co This command only dataset from some By default this com across potential su the directory are of enabled, relevant s fulfill a request.	ntent (files/direc v operates on da source use the mand operates bdatasets, i.e. if stained. Recursi ubdatasets are o	itaset conten 'clone' comm recursively w a directory is on into subdi letected and	t. To o hand. ithin a provio atasets install	btain a new independent dataset, but not Jed, all filles in is supported too. If ed in order to	
Get any dataset co This command only dataset from some By default this com across potential su the directory are of enabled, relevant s fulfill a request.	ntent (files/direc v operates on da source use the mand operates bdatasets, i.e. if stained. Recursi ubdatasets are o	itaset conten 'clone' comm recursively w a directory is on into subdi letected and	t. To o hand. ithin a provio atasets install		
Get any dataset co This command only dataset from some By default this com across potential su the directory are of enabled, relevant s fulfill a request.	ntent (files/direc v operates on da source use the mand operates bdatasets, i.e. if stained. Recursi ubdatasets are o	itaset conten 'clone' comm recursively w a directory is on into subdi letected and	t. To o hand. ithin a provio atasets install	btain a new independent dataset, but not Jed, all filles in is supported too. If ed in order to	
Get any dataset co This command only dataset from some By default this com across potential su the directory are of enabled, relevant s fulfill a request.	ntent (files/direc v operates on da source use the mand operates bdatasets, i.e. if stained. Recursi ubdatasets are o	itaset conten 'clone' comm recursively w a directory is on into subdi letected and	t. To o hand. ithin a provic stasets install	btain a new independent dataset, but not Jed, all filles in is supported too. If ed in order to	
Get any dataset cor This command only dataset from some get default this com across potential su the directory are of enabled, relevant s fulfill a request. Known data locatic obtained from som and possibly assign remerified	ntent (files/direc v operates on da source use the mand operates bdatasets, i.e. if stained. Recursi ubdatasets are o	itaset conten 'clone' comm recursively w a directory is on into subdi letected and	t. To o hand. ithin a provic stasets install	btain a new independent dataset, but not Jed, all filles in is supported too. If ed in order to	
Get any dataset co This command only dataset from some By default this com across potential su the directory are of enabled, relevant 5 fulfill a request. Known data locatic obtained from som and possibly assign remeified	ntent (files/direc y operates on di source use the mand operates bdatasets i.e. if stained. Recursis bdatasets are c ubdatasets are c ubdatasets are c ns for each requ envirable foca remote prior	itaset conten clone' comm recursively w a directory is on into subdi detected and jetected and uested file an tion (accordii rities), unless	t. To o hand. ithin a provic stasets install	btain a new independent dataset, but not Jed, all filles in is supported too. If ed in order to	
Get any dataset co Get any dataset co dataset from some By default this commander across potential so a maked, relevant 3 datalität a request. Known data locatic bobained from som and possibly assign mer find	ntent (tiles/direc v operates on di source use the mand operates bdatasets, i.e. if stained. Recursi ubdatasets are o vis for each requ evailable local versificate ender evailable local Ublittes Ho	taset conten 'clone' comm recursively w a directory is on into subd fetected and letected and uested file an tion (accordin tities), unless	t. To o hand. ithin a provic stasets install	brain a new independent dataset, but not led all files in is supported too too if ed in order to and and data are and and data are and and data are it source is	
Get any dataset co This command only dataset from some by default this com across potentials across potentials the directory are of the directory ar	ntent (files/direc source use the source use the mand operates datasets, i.e. if atalind, Recurst atalind, Recurst atalind, Recurst atalind, Recurst ataliand, Recurst atalian	itaset conten clone' comm recursively w a directory is on into subdi detected and jetected and uested file an tion (accordii rities), unless	t. To o hand. ithin a provic stasets install	dataset, but not dataset, but not led, all files in eller, all files in eller out of eller out of another to asted and data are in ennes configuration file source is Command Metadata	
Get any dataset co This command only dataset from some by default this com across potentials across potentials the directory are of the directory ar	ntent (files/direc source use the source use the mand operates datasets, i.e. if atalind, Recurst atalind, Recurst atalind, Recurst atalind, Recurst ataliand, Recurst atalian	taset conten 'clone' comm recursively w a directory is on into subd fetected and letected and uested file an tion (accordin tities), unless	t. To o hand. ithin a provic stasets install	brain a new independent dataset, but not led all files in is supported too too if ed in order to and and data are and and data are and and data are it source is	
Get any dataset co This command only dataset from some by default this com across potentials the directory are of enabled, relevant a huffill a request. Known data locatic dataset of the comment DataLast Goocy DataLast View me j (studyforrest >) dataBad	ntent (files/direc operates on de source use the mand operates bdatasets, ice if stained. Recursi ubdatasets are o exvisible location exvisible location exvisible location exvisible location exvisible location exvisible location directory directory	taset conten 'clone' comm recursively w a directory is on into subd fetected and letected and uested file an tion (accordin tities), unless	t. To o hand. ithin a provic stasets install	dataset, but not dataset, but not led, all files in el supported too til el node too asted and data are il nome configuration file source is Command Metadata Get content	History Properties
Get any dataset co This command only dataset from some by default hits command only dataset from some arroso predential so arroso predential so arroso predential so dataset discount on based of the source of the source of the source of DataLad Cocey DataLad View me Distudyforrest Distud	ntent (files/direc source use the source use the mand operates bubdatasets, ice if stained. Recursi ubdatasets are of available local result of the source of the source of the the the source of the the source of the source of the source of the the source of the source of the source of the the source of the source of the source of the the source of the source of the source of the the source of the source of the source of the source of the the source of the source of the source of the source of the the source of the source of the source of the source of the the source of the source of the source of the source of the the source of the source of the source of the source of the source of the the source of the source of	taset conten 'clone' comm recursively w a directory is on into subd fetected and letected and uested file an tion (accordin tities), unless	t. To o hand. ithin a provic stasets install	dataset, but not dataset, but not led, all files in el supported too til el node too asted and data are il nome configuration file source is Command Metadata Get content	History Properties
Get any dataset co This command only dataset from some by default his common arross potential su arross potential su methods are used. Known data locatic datalet a constri- batalet down Datalet View The second of the second Dataset View The second of the second The second of the second Dataset View The second of the second The second of the second The second of the second The second of the second of the second of the second The second of the second o	ntent (files/direct operates on de source use the mand operates obtatasets, i.e. if stained. Recursi ubdatasets are of evaniable local evaniable local evaniable local evaniable local terms of the type dataset directory directory directory	taset conten 'clone' comm recursively w a directory is on into subd fetected and letected and uested file an tion (accordin tities), unless	t. To o hand. ithin a provic stasets install	dataset, but not detaset, but not bet, at files in joo at of norder to at phones configuration file source is	History Properties t C\Users\adina\Desktop
Get any dataset co This command only dataset from some by default his common arross potential su arross potential su methods are used. Known data locatic datalet a constri- batalet down Datalet View The second of the second Dataset View The second of the second The second of the second Dataset View The second of the second The second of the second The second of the second The second of the second of the second of the second The second of the second o	ntent (files/direct operates on de source use the mand operates obtatasets, i.e. if stained. Recursi ubdatasets are of evaniable local evaniable local evaniable local evaniable local terms of the type dataset directory directory directory	taset conten 'clone' comm recursively w a directory is on into subd fetected and letected and uested file an tion (accordin tities), unless	t. To o hand. ithin a provic stasets install	dialactic han not dialactic han not dialactic han not dialactic han not dialactic han not di a adolen no is supported too if di source di orden to di en oder to di en oder to di contre di Get content Get content Get content Get content di adalacti a	History Properties t C\Users\adina\Desktop Not set Add Upda
Get any dataset co This command only dataset from some by default his common arross potential su arross potential su arross potential substanced from some difficult arequest by default his costation data location data lo	ntent (files/direct operates on de source use the mand operates obtatasets, i.e. if stained. Recursi ubdatasets are of evaniable local evaniable local evaniable local evaniable local terms of the type dataset directory directory directory	taset conten 'clone' comm recursively w a directory is on into subd fetected and letected and uested file an tion (accordin tities), unless	t. To o hand. ithin a provic stasets install	dialactic han not dialactic han not dialactic han not dialactic han not dialactic han not di a adolen no is supported too if di source di orden to di en oder to di en oder to di contre di Get content Get content Get content Get content di adalacti a	History Properties t C\Users\adina\Desktop
Get any dataset co This command only dataset from some by default his common arross potential su arross potential su arross potential substanced from some difficult arequest by default his costation data location data lo	ntent (files/direct or operates on di source use the mend-operates batasets i.e. if shined. Records batasets are of e available loca e available e a	taset conten 'clone' comm recursively w a directory is on into subd fetected and letected and uested file an tion (accordin tities), unless	t. To o hand. ithin a provic stasets install	dataset, but not dataset, but not bel, al filter in our if all and the in our of all normal our our our pis-protec configuration filte source is Get content Get content Limit to	History Properties t C\Users\adina\Desktop Not set Add Upda
Get any dataset co this command out dataset from some by default this command out arrows potential so arrows potential so arrows potential by default this down dut a location dataset discovery dataset discovery dataset discovery dataset discovery dataset Dataset discovery dataset discovery dataset discovery dataset discovery dataset discovery dataset discovery dataset discovery dataset discovery dataset discovery dataset discovery dataset discovery dataset discovery discovery dataset discovery dataset discovery dataset discovery dataset discovery discov	tent (files/direct or operates on de source use the mand operates datasets, i.e. if trained, Recursi biolisaset are e available loca e available loca directory directory directory directory directory directory directory directory directory	taset conten 'clone' comm recursively w a directory is on into subd fetected and letected and uested file an tion (accordin tities), unless	t. To o hand. ithin a provio atasets install	didact hannot didact hannot is supported too if or order too de order to source is outport and data are province configuration de outre is command Get content Get content Get for content Umit to Also get subdatasets	History Properties t C\Users\adina\Desktop Not set Add Upda
Get any dataset co This command only dialest from some by default this convert the directory are of energy and the directory are of energy the directory are of energy the directory are of energy and possibly assign Datalact Gromeyr Datalact System Datalact System Databaset View The System S	Intern (files/direct or operates on du source use the mand operates) bublates to end trained. Recurso bublatesets are of trained. Recurso bublatesets are of trained. Recurso trained. Recurso trained. Recurso trained. Recurso directory	ataset content content of the observation of the ob	t. To o hand. ithin a provio atasets install	didact hannot didact hannot is supported too if or order too de order to source is outport and data are province configuration de outre is command Get content Get content Get for content Umit to Also get subdatasets	History Properties C\Users\adina\Desktop C Not set Add Upda
Get any dataset co this command out dataset from some by default this command out a composition of the dataset from some by default this composition fulfill a request. Known data location dataset discovery DataLad Groups me Distudiof from some dataset DataLad Groups me Distudiof	Intern (files/direct or operates on du source use the mand operates) bublates to end trained. Recurso bublatesets are of trained. Recurso bublatesets are of trained. Recurso trained. Recurso trained. Recurso trained. Recurso directory	ataset content content of the observation of the ob	t. To o hand. ithin a provio atasets install	didact hannot didact hannot is supported too if or order too de order to source is outport and data are province configuration de outre is command Get content Get content Get for content Umit to Also get subdatasets	History Properties C\Users\adina\Desktop C Not set Add Upda
Get any dataset co This command only dialest from some by default this convert the directory are of energy and the directory are of energy the directory are of energy the directory are of energy and possibly assign Datalact Gromeyr Datalact System Datalact System Databaset View The System S	tent (files/direct roperates on di source use the mand operates balances, je. if alande, je. euroris- balances, je. euroris- balances, je. euroris- balances, je. euroris- balances, je. euroris- balances, je. euroris- balances, je. euroris- devinable (e.e. versione) directory	ataset content content of the constraints of the co	t. To o hand. ithin a provid stasets install e eval. ig to c a spec	didact hannot didact hannot is supported too if or order too de order to source is outport and data are province configuration de outre is command Get content Get content Get for content Umit to Also get subdatasets	History Properties C\Users\adina\Desktop C Not set Add Upda
Get any dataset co This command only discretifying services and only the directory are al- nonso potentially the directory are al- nonso potentially the directory are al- nonso potentially the directory are al- shall a request. Norman discretifying Dataset View me () (varier files/cline/ varier and operates on di- outrain and the second operates and distances in the second operates and distances are de- not for an operation of the second operation operation of the second operation of the second operation of the second operation operation of the second operation of the second operation operation of the second operation operation operation operation of the second operation ope	atalet content content of the second of the	t. To o hand. ithin a provide statement of the statement	dialact: bannot dialact: bannot dialact: bannot dialact: bannot dialactione is supported too (f another and the another another and data are given and the another discourse (is discourse (is discourse) det ties content limit to Also get subdatasets c	History Properties C\Users\adina\Desktop C Not set Add Upda
Get any dataset co This command only distributions the distribution of the distribution of the distribution the distribution of the distribution of the distribution the distribution of the distribution of the distribution the distribution of the distribution of the distribution of the distribution of the distribution of the distribution of the distribution Dataset of the distribution of the distr	terrer (Bes/Ghreen verband) outputs and operates outputs and operates outputs of perates outputs ou	atalet content content of the conten	t. To o hand. ithin a provide the second se	Landard a new independent dislate: han not dislate: han	History Properties C\Users\adina\Desktop C Not set Add Upda
Get any dataset co This command only distributions the distribution of the distribution of the distribution the distribution of the distribution of the distribution the distribution of the distribution of the distribution the distribution of the distribution of the distribution of the distribution of the distribution of the distribution of the distribution Dataset of the distribution of the distr	terrer (Bes/Ghreen verband) outputs and operates outputs and operates outputs of perates outputs ou	atalet content content of the conten	t. To o hand. ithin a provide the second se	Landard a new independent dislate: han not dislate: han	History Properties C\Users\adina\Desktop C Not set Add Upda
Get any dataset co This command only distributions the distribution of the distribution of the distribution the distribution of the distribution of the distribution the distribution of the distribution of the distribution the distribution of the distribution of the distribution of the distribution of the distribution of the distribution of the distribution Dataset of the distribution of the distr	terrer (Bes/Ghreen verband) outputs and operates outputs and operates outputs of perates outputs ou	atalet content content of the conten	t. To o hand. ithin a provide the second se	Landard a new independent dislate: han not dislate: han	History Properties C\Users\adina\Desktop C Not set Add Upda
Get any dataset co This command only distance from some by default this concerness of the some the distance from some any distance of the some some some some some some some some some Dataset of the some some some some some some some Dataset of the some some some some some some some som	terrer (Bes/Ghreen verband) outputs and operates outputs and operates outputs of perates outputs ou	atalet content content of the conten	t. To o hand. ithin a provide the second se	dialact: bannot dialact: bannot dialact: bannot dialact: bannot dialactione is supported too (f another and the another another and data are given and the another discourse (is discourse (is discourse) det ties content limit to Also get subdatasets c	History Properties C\Users\adina\Desktop C Not set Add Upda
Get any dataset co This command only dataset from some by default this com by default this com the differences. In the differences that is a second to the differences and possibly assign Datalact Over Dataset View The Dataset Vi	terr (Beddinese Inter- dietektor) (Beddinese Inter- dietektor) (Beddinese Inter- ter (Beddinese Inter- dietektor) (Beddinese Inter- Beddinese Inter- Bed	statet content control of the contro	t. To o nand. ithin a providuation of the second of the s	dalaes but not dalaes but not dalaes but not de asterne to en order to en order to en order to de note to de note to de code to	History Properties C\Users\adina\Desktop C Not set Add Upda
Get any dataset co This command only distance from some by default this concerness of the some the distance from some any distance of the some some some some some some some some some Dataset of the some some some some some some some Dataset of the some some some some some some some som	there files/clines reported to a file of the second	statet contentions of the second seco	t. To o nand. ithin a provide installed re evaluation of the second of	dadaet balan oot dada	History Properties C\Users\adina\Desktop C Not set Add Upda

The Suite submenu lets you switch between suites that alter the command selection. The two suites you will always be able to select between is a "simplified" command set, reduced to the most essential commands and parameters, and a "complete" command set, which is a development preview. DataLad extensions can add additional suites when you install them. Please note that we recommend the "simplified" command suite to users, as the complete suite can contain experimental implementations.

The Utilities and Help Menu

The Utilities menu has a several useful functions. One is a convenience version checker that can tell you whether there is a newer DataLad version available.

Note: The Check for new version feature requires a network connection.

The Manage credentials submenu opens an interface to the Credential manager. The section *Credential management with DataLad Gooey* has a detailed overview. The Set author identity submenu lets you set and change your Git identity and was already covered above.

The Help menu contains a range of actions to find additional information or help. Report a problem contains links for filing issues and getting in touch with the developers. Diagnostic infos will create a report about the details of your installation and system that you can copy-paste into such issues.

2.3 Central Concepts (Glossary)

Working with DataLad is easier if you know a few technical terms that are regularly referred to in the documentation and the GUI interface. This glossary provides short definitions, and links relevant additional documentation, where available.

annex

git-annex concept: a different word for the internal location in a dataset that *annexed-file*'s are version controlled in.

annexed-file

Files managed by *git-annex* are annotated as "annexed-file". Annexed files have access to additional commands in their context menus such as *get* and *drop*.

branch

Git concept: A lightweight, independent history streak of your dataset. Branches can contain less, more, or changed files compared to other branches, and one can merge the changes a branch contains into another branch. DataLad Gooey only views the currently checked out branch in your dataset, and does not support Git commands that expose branching functionality.

clone

The datalad clone command retrieves a copy of a *Git* repository or *DataLad dataset* from a local or remote path or URL. In Git-terminology, all "installed" datasets are clones.

commit

Git concept: Adding selected changes of a file or dataset to the repository, and thus making these changes part of the revision history of the repository. The datalad save command creates a commit in the selected dataset. Commits should always have an informative *commit message*.

commit message

Git concept: A concise summary of changes you should attach to a **datalad save** command. This summary will show up in your *DataLad dataset* history.

DataLad dataset

A DataLad dataset is a Git repository that may or may not have a data annex that is used to manage data referenced in a dataset. In practice, most DataLad datasets will come with an annex.

DataLad extension

Python packages that equip DataLad with specialized commands. The section *extensions_intro* <*http://handbook.datalad.org/en/latest/r.html?extensions>_* of the DataLad Handbook gives and overview of available extensions and links to Handbook chapters that contain demonstrations.

DataLad subdataset

A DataLad dataset contained within a different DataLad dataset (the parent or DataLad superdataset).

DataLad superdataset

A DataLad dataset that contains one or more levels of other DataLad datasets (DataLad subdataset).

drop

The datalad drop command drops file content of annexed files. It is the antagonist to get.

GIN

A web-based repository store for data management that you can use to host and share datasets. Find out more about GIN here.

Git

A version control system to track changes made to small-sized files over time. You can find out more about git in this (free) book or these interactive Git tutorials on *GitHub*.

git-annex

A distributed file synchronization system, enabling sharing and synchronizing collections of large files. It allows managing files with *Git*, without checking the file content into Git.

git-annex branch

A *branch* that exists in your dataset, if the dataset contains an annex. The git-annex branch is completely unconnected to any other branch in your dataset, and contains different types of log files. Its contents are used for git-annex's internal tracking of the dataset and its annexed contents. DataLad Gooey provides support for adding git annex metadata, but does not otherwise support operations on dataset branches

GitHub

GitHub is an online platform where one can store and share version controlled projects using Git (and thus also DataLad project). See`GitHub.com https://github.com/>`_.

GitLab

An online platform to host and share software projects version controlled with *Git*, similar to *GitHub*. See Gitlab.com.

get

The datalad get command gets file content of annexed files. It is the antagonist to *drop*.

https

Hypertext Transfer Protocol Secure; A protocol for file transfer over a network.

pip

A Python package manager. Short for "Pip installs Python". pip install <package name> searches the Python package index PyPi for a package and installs it while resolving any potential dependencies.

remote

Git-terminology: A repository (and thus also *DataLad dataset*) that a given repository tracks. A *sibling* is DataLad's equivalent to a remote.

SSH

Secure shell (SSH) is a network protocol to link one machine (computer), the *client*, to a different local or remote machine, the *server*.

SSH key

An SSH key is an access credential in the SSH protocol that can be used to login from one system to remote servers and services, such as from your private computer to an SSH server, without supplying your username or password at each visit. To use an SSH key for authentication, you need to generate a key pair on the system you would like to use to access a remote system or service (most likely, your computer). The pair consists of a *private* and a *public* key. The public key is shared with the remote server, and the private key is used to authenticate your machine whenever you want to access the remote server or service. Services such as *GitHub*, *GitLab*, and

GIN use SSH keys and the SSH protocol to ease access to repositories. This tutorial by GitHub is a detailed step-by-step instruction to generate and use SSH keys for authentication.

sibling

DataLad concept: A dataset clone that a given *DataLad dataset* knows about. Changes can be retrieved and pushed between a dataset and its sibling. It is the equivalent of a *remote* in Git.

version control

Processes and tools to keep track of changes to documents or other collections of information.

2.4 Walk-through: Dataset hosting on GIN

In this walkthrough, we will use DataLad Gooey to create a dataset, save it contents, and publish it to GIN (G-Node Infrastructure).

2.4.1 Prerequisites

In order to use GIN for hosting and sharing your datasets, you need to:

- Register a GIN account;
- add a personal access token to your GIN account (for creation of repositories with DataLad);
- add an SSH key to your GIN account (for uploading annexed contents).

Follow the instructions on GIN to do so.

If you need to generate an SSH key pair and want to stay in the world of graphical interfaces, we recommend using PuTTYgen for this purpose. Your private key needs to be placed in the *.ssh* folder in your home directory for it to be picked up correctly.

2.4.2 Create a dataset

Let's assume that we are starting with an existing folder which already has some content, but is not yet a DataLad dataset. Let's open the DataLad Gooey and set a base directory to our folder, or its parent directory.

Our first operation is to to create a DataLad dataset. For this, right-click your folder and select *Directory commands* \rightarrow *Create a dataset*. This will populate the Command tab on the right with options for the selected command. The first value (*Create at*) is already populated, since we used right-click to issue the command. We leave *Dataset with file annex* checked (default), and *Register in superdataset* not set (default). In this example we want to configure our dataset to annex binary, but not text files. To do so, select *text2git* from the list of *Configuration procedure(s)* and click *Add*. Finally, check the *OK if target directory not empty* to enforce dataset creation out of a non-empty folder. With the options selected, click *OK*.

DataLad Gooey ile <u>D</u> ataLad <u>V</u> iew <u>U</u> tilities <u>H</u> e	elp					-	- 🗆	×
Name projects my-project code analyse_data.R	Type directory directory directory file	State	Command Create a data Create at		Jsers\mih\Docur	nents\projects\m	ny-project	
 clean_data.R data_clean data_clean.csv data_raw 	file directory file directory	 untracked untracked 	-	h file annex superdataset on procedure(s)	✓ Jsers\mih\Docur text2git Add	nents\projects\m Update	ny-project	~
 figures dotplot_correlati datalad CITATION.cff LICENSE README.md gitattributes other-project 	directory file directory file file file directory	 untracked untracked untracked untracked untracked clean 	OK if target	directory not empty	text2git	ОК	Cance	2
Command log Error log run(ok): C:\Users\mih\Documents\p create(ok): C:\Users\mih\Document action summary: create (ok: 1) run (ok: 1)			AppData\Local\	datalad.org\]			Clear	▲

2.4.3 Save the contents

Right-click the newly created dataset, and select *Dataset commands* \rightarrow *Save the state in a dataset*. Parameters required for the Save command should appear in the Command tab. Fill in the *Description of change* (this is the commit message associated with the save). Leave all other fields default (note: *Do not put files in annex* is greyed out, not checked, i.e. it has no value). Here, we are saving all files at once, but if we wanted we could limit the save operation to selected files, or trigger it by clicking on a specific file. Once ready, click *OK*.

Note that after this operation, "untracked" files changed their state to "clean". Different from files, the dataset state is still "untracked", because it is not registered in any superdataset. Because we used the *text2git* configuration, only the PNG file changed its type to "annexed-file" in the screenshot below.

DataLad Gooey			- 🗆 X
File DataLad View Utilities Help Name	Type ^ directory dataset directory directory	State	Command Properties Save the state in a dataset Save changes in dataset at Description of change Add all content that we have right now
 analyse_data.R clean_data.R data_clean data_clean.csv data_raw figures dotplot_correlations.png .gitattributes CITATION.cff LICENSE README.md other-project 	file file directory file directory directory annexed-file file file file file file directory	 ⊘ clean 	Limit to Not set Add Update Remove Include changes in subdatasets Do not put files in annex Tag for saved dataset state Amend last saved state
Command log Error log add(ok): data_raw\datapackage.json (file) add(ok): figures\dotplot_correlations.png save(ok): . (dataset) action summary: add (ok: 9) save (ok: 1)			OK Cancel

2.4.4 Create a GIN sibling

Creating a GIN sibling will create a new repository on GIN, and configure your dataset with its address. To perform this action, right-click your dataset, and select *Dataset commands* \rightarrow *Create a GIN sibling*. Fill in the *New repository name on Gin* (and, optionally, check the *Make GIN repository private*). You can leave all other options default.

In the *Name of the credential to be used* field, you can pick previously used credentials. If no value is given, and no previous credentials exist, the credentials will be save with website name (*gin.g-node.org*) by default.

Click OK.

At this point, a pop-up window will appear and you will be asked for a token. Paste the access token generated from GIN website, and click OK.

me	Туре	State	Command	Properties	
projects	directory		Create a GIN	cibling	
✓ Ď my-project	dataset	⊖ untracked	Create a Gill	sibilitig	
> 🗋 .datalad	directory		Create sibli	ng for dataset at	\Users\mih\Documents\projects\my-project
✓ □ code	directory		Now reposi	itory name on Gin	my-project
analyse_data.R	file	🖉 clean			
j clean_data.R	file	🕢 clean	Make GIN repository private Sibling name Get/push protocol Name of credential to be used		
✓ ☐ data_clean	directory				gin
data_clean.csv	file	⊘ clean			https-ssh
> 🗋 data_raw	directory				
✓ ☐ figures	directory	-	Name of cr	edential to be used	
dotplot_correlations.png	annexed-file	⊘ clean	If the sibling exists already		error
] .gitattributes	file	⊘ clean			
CITATION.cff	file	⊘ clean			
	file	⊘ clean			
E README.md	file	⊘ clean			
> 🗋 other-project	directory				OK Cance
mmand log Error log					
n access token is required for https://gin reate_sibling_gin(ok): [sibling repository '					eate a token
onfigure-sibling(ok): . (sibling) ction summary:	gin created at i	https://gini.g houe.or	g/msz/my projec		
configure-sibling (ok: 1)					
create_sibling_gin (ok: 1)					

2.4.5 Push to the GIN sibling

Right-click *Dataset commands* \rightarrow *Push data/updates to a sibling*. The only thing you need to select is the value of *To dataset sibling* - this will be the sibling name from the step above. Leave other options default, and click *OK*.

me	Туре	State	Command Properti	es	
projects	directory		Push data/updates to a	cibling	
✓ Ď my-project	dataset	\ominus untracked	Fush data/updates to a		
> 🗋 .datalad	directory		Push from dataset at	C:\Users\mih\Documents\projects\my-project	
✓ ☐ code	directory		To dataset sibling	gin	
analyse_data.R	file	🖉 clean		-	
📄 clean_data.R	file	🖉 clean	What to push	auto-if-wanted	N
✓ ☐ data_clean	directory		Limit to	Not set	
■ data_clean.csv	file	🖉 clean			
> 🗋 data_raw	directory			Add Update Ren	nove
✓ ☐ figures	directory	-	Recursive		
dotplot_correlations.png		⊘ clean	Force operation	none	~
.gitattributes	file	⊘ clean			
CITATION.cff	file	⊘ clean			
	file	⊘ clean			
= README md	file	🖉 clean			
README.md				ОК	Cancel
Other-project	directory			UK	Cancer
other-project ommand log Error log copy(ok): figures\dotplot_correlations.pn publish(ok): . (dataset) [refs/heads/git-an publish(ok): . (dataset) [refs/heads/maste	g (file) [to gin] nex->gin:refs/he			UK	Cancer
 > other-project > other-project > pmmand log Error log > opy(ok): figures\dotplot_correlations.png > ublish(ok): . (dataset) [refs/heads/git-an 	g (file) [to gin] nex->gin:refs/he			UK	Cancer

2.4.6 Retrieve the data from GIN

Finally we can confirm that our dataset can be obtained from GIN (possibly by other users who have access). Let's try making a clone in the same parent directory as our dataset, but under a different name. To do this, from the menu select $DataLad \rightarrow Clone \ a \ dataset$. In the *Clone from* field, enter the dataset URL taken from GIN. Here, you can use either the HTTPS (for public repositories) or SSH (for private repositories) URL. Note that when using HTTPS, you need to remove the .git from the URL ending for proper interaction with GIN. Next, click the directory selection icon to the right of *Clone into* field, and use the directory picker to create and select a new directory named, for example, cloned-dataset. Afterwards, click *OK*.

To obtain the annexed contents in the cloned dataset, right click it in the file tree, select *Dataset commands* \rightarrow *Get content*, and click *OK*. With other options kept default, this will download all annexed content in the dataset; if you wanted to obtain contents selectively, you could use the *Limit to* option. Alternatively, you could right-click individual files, and use *File commands* \rightarrow *Get content*.

2.5 Credential management with DataLad Gooey

DataLad can store credentials and use them to authenticate against a variety of services or infrastructures. Among them are, for example, the credentials required to push datasets to siblings on Gin, GitHub, or similar services. DataLad Gooey's credential manager makes setting, querying, and removing credentials much easier than performing the actions via the command line. It is a front-end for the modernized datalad credential command from datalad-next.

Access to this tooling is provided via the Utilities -> Manage credential menu in the tab bar.

DataLad Gooey								—		×
File DataLad View	Utilities He	lp								
Name	Туре	^ State	e	Î.	Command N	letadata H	istory Properti	es.		
 b studyforre datala 	Credent	ial manager					×			
> Code	Select kno		Select existing credential				~			
 ▶ □ stimuli ▼ □ sub-01 	Name	Enter a creden	tial name							
> set	Secret		for the credential				Show			
>	Prope		verify (optional)				match			
> sub-04 > sub-05		Name		Value					Cancel	
Command log Er	1									
				A	dd property	Remove p	property			
	Delete					Reset fo	orm Save			
									Cle	ar

All credentials are identified via a unique name, and should at least contain a secret such as a token, a password, or a key. The Select known credential drop-down menu will list all known secrets for query or modification. This drop-down menu can be helpful to discover credentials, update them after their secrets changed, or remove credentials that became obsolete via the Delete button at the bottom of the form.

In order to create a new credential, for example a token to authenticate against *Gin*, enter a name of your choice (here it is gin) and a corresponding secret in the Credential tab underneath. As the Secret field is confidential, the characters you type or paste will be hidden, but the show button on the right can be ticked to view it in clear text. To prevent typos, the secret has to be repeated and match the previous entry. The Save button at the bottom will store the secret in your systems credential store.

DataLad Gooey File DataLad View	Utilities Help				- 🗆 X
Name Name Studyforre Studyforre Adatala Code Struct		State ger ential Select existing credential		Aetadata History Propert	es
Command log Er	1 Nam	e	Value Add property	Remove property	Cancel
	Delete			Reset form Save	Clear

Credentials can have properties denoting additional details about them useful for queries or required for authentication to particular services. Once saved, the gin credential gained an automatic property, the last-modified property. You can add additional arbitrary properties or alter or remove existing ones in the central Properties tab.

	Туре	Sta	te	Comn	nand Metadata H	istory Properties	\$
 studyforre datala 	Creden	tial manager				×	
 Code 	Select kno	own credential	gin			~	
> src	Credent	tial					
 ▶ □ stimuli ▼ □ sub-0⁻ 	Name	gin				⊗	
> 🗋 se:	Secret	•••••			8	Show	
> ses					8	match!	
>	⊂ Prope				W	match:	
> 🗋 sub-04	Prope						
> sub-05		Name		Value			Cancel
	1	last-modified	2022-10-14T08:27:47.232018				
mmand log Er							
				Add pro	operty Remove	property	

Once saved, the credential will be available in all commands with a credential parameter via a drop-down menu for you to select. This ensures that you can precisely select which credential is used in every operation. Especially if you have several accounts on one and the same service or several authentication methods with a different set of permissions, such as one for your private account and one for your organizations account, this comes in handy.

In the screenshot below, the newly created credential gin is used in the parametrization of a create-sibling-gin command.

					—
DataLad View Uti	ilities Help				
lame	Туре	State	Command Metadata History Pro	operties	
 studyforrest 	dataset	_	Create a GIN sibling		
> 🗋 .datalad	directory				
> 🗋 code	directory		New repository name on Gin	data-backup	^
> 📄 src	directory		Make GIN repository private		
> 📄 stimuli	directory		Sibling name	ain	
✓ ☐ sub-01	directory	1	Sibling name	gin	
> 🗋 ses-l			Get/push protocol	https-ssh	~
> ses			Name of credential to be used	ain	~
> sub-02	directory		Name of credential to be used	gin	
> sub-03	directory		If the sibling exists already	error	▼ ▼
> sub-04	directory				OK Cancel
> sub-05	directory				Cancer
		Þ	-		
control to shared da host DataLad datase This command uses default target, but c parameter.	bling on a GIN sit tructure) is a free ased repository s ata. GIN is built c ets, including the s the main GIN in other deploymen	e data manage store and prov on Git and git- eir data conter istance at http its can be used	ement system. It is a ides fine-grained access annex, and can natively nt! us://gin.g-node.org as the d via the 'api'		
Create a dataset sib GIN (G-Node infrast GitHub-like, web-ba control to shared da host DataLad datase This command uses default target, but o parameter. An SSH key, propert	bling on a GIN sit tructure) is a free ased repository s ata. GIN is built c ets, including the s the main GIN in other deploymen ly registered at t	e data manage store and prov on Git and git- eir data conter istance at http its can be used he GIN instance	ement system. It is a ides fine-grained access annex, and can natively nt! s://gin.g-node.org as the		

2.6 Setting git-annex metadata

git-annex has its own concept and implementation of file metadata, and DataLad Gooey provides a graphical user interface for querying, removing, and setting metadata. This functionality is exposed via the Metadata tab or right-click context menue on any annexed file.

DataLad Gooey	-		\times
File DataLad View Utilities Help			
Name Image: Ima	Type Command Metadata History Properties datas direc file For val\2017_06_09_How_Digital_Memory_Is_Shaping_Our_Future_Abby_Smith direc anne: anne: anne: Add field Command Metadata For val\2017_06_09_How_Digital_Memory_Is_Shaping_Our_Future_Abby_Smith Command Com	+	3
 2017_06_09_Talking_wi 2017_06_09_The_Red p 2017_07_03_Transformi 	anne:	Cancel	
Command log Error log Help Git-annex metadata editor Git-annex supports assigning metadat More precisely, such metadata are ass the content identifier. Therefore, mult content will also report the same meta git-annex metadata is modified, the m	sociated with the underlying annex-key, tiple files that have the exact same tadata record. When a file with		~

Each annexed file's content¹ can have any number of metadata *fields* attached to it to describe it, and each field can have any number of *values*. This metadata is stored internally, in the *git-annex branch*.

The Annex metadata editor will show all existing fields and values for a selected file, and let users add additional fields and values as well as deleting them. A subsequent Save will write the provided annex metadata to the file.

2.6.1 Input validation

git-annex's metadata is highly flexible and allows arbitrary content, but it is impossible to have two fields of the same name, or two identical values belonging to one field. This is validated automatically after each entry, and only valid metadata can be saved. Icons attached to conflicting fields will let you know which metadata elements need fixing.

me	Type ^ Con	nmand	Metadata	History F	Properties		
longnow-podcasts		nnex meta	data				
 .datalad .gitattributes 	direc Fo	or 2017_0	06_09How_C	igital_Memory	/_ls_Shaping_Our_Future	eAbby_Smith_Rumsey.mp3	i.
 .gitattributes Long_Now_Conversations 		eld			Value		
2017_06_09_How_Digit		emid		8 🕥 🗎	227e1dcbdffa4c98348	ac 🔇 🕋 +	
2017_06_09_Pace_Laye 2017_06_09_Proof_Th		alidfield		⊗ √ 🗎	duplicate value	8 🗸 🍿	
2017_06_09Seveneves.	. anne:				duplicate value	❸ 💼 +	
2017_06_09Talking_wi 2017_06_09The_Red_P.		Add field					
2017_07_03Transformi.	. anne: 🗸					Save Cancel	
	>				L.		
ommand log Error log Help							
Git-annex metadata editor							,
Git-annex supports assigning metada More precisely, such metadata are as							1
he content identifier. Therefore, mul	tiple files that hav	e the exact	t same				
content will also report the same met git-annex metadata is modified, the r							
jit-annex metadata is modified, the r	netadata of the pl	evious ver	sion will				`

¹ Note that metadata is attached to file content, not file names, i.e. the git-annex key corresponding to the content of a file, not to a particular filename on a particular git branch. This means that all files with the same key share the same metadata, which is stored in the *git-annex branch*. If a file is modified, the metadata of the previous version will be copied to the new key when git-annex adds the modified file.

2.7 Setting Metadata

Coming soon - stay tuned!

CHAPTER

THREE

COMMANDS AND API

3.1 Command line reference

This module reference extends the manual with an overview of the available functionality built into datalad gooey. Please note that apart from the main command, the commands this module provides are internal helpers used within DataLad Gooey for fast annotations and file tree overviews. Unlike the DataLad core package or its extensions, the commands provided by DataLad Gooey are thus not intended to be used directly.

3.1.1 datalad gooey

Synopsis

datalad gooey [-h] [-p PATH] [--postinstall] [--version]

Description

DataLad GUI

Long description of arbitrary volume.

Examples

Launch the DataLad Graphical User Interface (GUI, a.k.a Gooey) at the specified location .:

% datalad gooey --path 'path/to/root/explorer/directory'

Options

-h, --help, --help-np

show this help message. --help-np forcefully disables the use of a pager for displaying the help message

-p PATH, --path PATH

The root location from which the Gooey file explorer will be launched (default is current working directory).

--postinstall

Perform post-installation tasks.

--version

show the module and its version which provides the command

Authors

datalad is developed by The DataLad Team and Contributors <team@datalad.org>.

3.1.2 datalad gooey-askpass

Synopsis

datalad gooey-askpass [-h] [--version]

Description

Internal helper for datalad-gooey

Options

-h, --help, --help-np

show this help message. --help-np forcefully disables the use of a pager for displaying the help message

--version

show the module and its version which provides the command

Authors

datalad is developed by The DataLad Team and Contributors <team@datalad.org>.

3.1.3 datalad gooey-Isdir

Synopsis

datalad gooey-lsdir [-h] [--version] path

Description

Internal helper for datalad-gooey

Options

path

-h, --help, --help-np

show this help message. --help-np forcefully disables the use of a pager for displaying the help message

--version

show the module and its version which provides the command

Authors

datalad is developed by The DataLad Team and Contributors <team@datalad.org>.

3.1.4 datalad gooey-status-light

Synopsis

datalad gooey-status-light [-h] [-d DATASET] [--version] path

Description

Internal helper for datalad-gooey

Options

path

-h, --help, --help-np

show this help message. --help-np forcefully disables the use of a pager for displaying the help message

-d DATASET, --dataset DATASET

specify the dataset to query. If no dataset is given, an attempt is made to identify the dataset based on the current working directory. Constraints: Value must be a Dataset or a valid identifier of a Dataset (e.g. a path) or value must be NONE

--version

show the module and its version which provides the command

Authors

datalad is developed by The DataLad Team and Contributors <team@datalad.org>.

3.2 Python module reference

This module reference extends the manual with an overview of the available functionality built into datalad gooey. Please note that apart from the main command, the commands this module provides are internal helpers used within DataLad Gooey for fast annotations and file tree overviews. Unlike the DataLad core package or its extensions, the commands provided by DataLad Gooey are thus not intended to be used directly.

gooey([path, postinstall])	DataLad GUI
gooey_askpass()	Internal helper for datalad-gooey
gooey_lsdir(path)	Internal helper for datalad-gooey
<pre>gooey_status_light([dataset, path])</pre>	Internal helper for datalad-gooey

3.2.1 datalad.api.gooey

datalad.api.gooey(path: str = None, postinstall: bool = False)

DataLad GUI

Long description of arbitrary volume.

Examples

Launch the DataLad Graphical User Interface (GUI, a.k.a Gooey) at the specified location .:

```
> gooey(path='path/to/root/explorer/directory')
```

Parameters

- **path** -- The root location from which the Gooey file explorer will be launched (default is current working directory). [Default: None]
- **postinstall** (*bool*, *optional*) -- Perform post-installation tasks. [Default: False]

- **on_failure** (*{'ignore', 'continue', 'stop'}, optional*) -- behavior to perform on failure: 'ignore' any failure is reported, but does not cause an exception; 'continue' if any failure occurs an exception will be raised at the end, but processing other actions will continue for as long as possible; 'stop': processing will stop on first failure and an exception is raised. A failure is any result with status 'impossible' or 'error'. Raised exception is an IncompleteResultsError that carries the result dictionaries of the failures in its *failed* attribute. [Default: 'continue']
- **result_filter** (*callable or None, optional*) -- if given, each to-be-returned status dictionary is passed to this callable, and is only returned if the callable's return value does not evaluate to False or a ValueError exception is raised. If the given callable supports ***kwargs* it will additionally be passed the keyword arguments of the original API call. [Default: None]
- **result_renderer** -- select rendering mode command results. 'tailored' enables a command- specific rendering style that is typically tailored to human consumption, if there is one for a specific command, or otherwise falls back on the the 'generic' result renderer; 'generic' renders each result in one line with key info like action, status, path, and an optional message); 'json' a complete JSON line serialization of the full result record; 'json_pp' like 'json', but pretty-printed spanning multiple lines; 'disabled' turns off result rendering entirely; '<template>' reports any value(s) of any result properties in any format indicated by the template (e.g. '{path}', compare with JSON output for all key-value choices). The template syntax follows the Python "format() language". It is possible to report individual dictionary values, e.g. '{metadata[name]}'. If a 2nd-level key contains a colon, e.g. 'music:Genre', ':' must be substituted by '#' in the template, like so: '{metadata[music#Genre]}'. [Default: 'tailored']
- **result_xfm** ({'datasets', 'successdatasets-or-none', 'paths', 'relpaths', 'metadata'} or callable or None, optional) -- if given, each to-be-returned result status dictionary is passed to this callable, and its return value becomes the result instead. This is different from *result_filter*, as it can perform arbitrary transformation of the result value. This is mostly useful for top- level command invocations that need to provide the results in a particular format. Instead of a callable, a label for a pre-crafted result transformation can be given. [Default: None]
- **return_type** ({'generator', 'list', 'item-or-list'}, optional) -- return value behavior switch. If 'item-or-list' a single value is returned instead of a one-item return value list, or a list in case of multiple return values. *None* is return in case of an empty list. [Default: 'list']

3.2.2 datalad.api.gooey_askpass

datalad.api.gooey_askpass()

Internal helper for datalad-gooey

3.2.3 datalad.api.gooey_lsdir

datalad.api.gooey_lsdir(path: Path)

Internal helper for datalad-gooey

Parameters

- path --
- **on_failure** ({'ignore', 'continue', 'stop'}, optional) -- behavior to perform on failure: 'ignore' any failure is reported, but does not cause an exception; 'continue' if any

failure occurs an exception will be raised at the end, but processing other actions will continue for as long as possible; 'stop': processing will stop on first failure and an exception is raised. A failure is any result with status 'impossible' or 'error'. Raised exception is an IncompleteResultsError that carries the result dictionaries of the failures in its *failed* attribute. [Default: 'continue']

- **result_filter** (*callable or None*, *optional*) -- if given, each to-be-returned status dictionary is passed to this callable, and is only returned if the callable's return value does not evaluate to False or a ValueError exception is raised. If the given callable supports ***kwargs* it will additionally be passed the keyword arguments of the original API call. [Default: None]
- **result_renderer** -- select rendering mode command results. 'tailored' enables a command- specific rendering style that is typically tailored to human consumption, if there is one for a specific command, or otherwise falls back on the the 'generic' result renderer; 'generic' renders each result in one line with key info like action, status, path, and an optional message); 'json' a complete JSON line serialization of the full result record; 'json_pp' like 'json', but pretty-printed spanning multiple lines; 'disabled' turns off result rendering entirely; '<template>' reports any value(s) of any result properties in any format indicated by the template (e.g. '{path}', compare with JSON output for all key-value choices). The template syntax follows the Python "format() language". It is possible to report individual dictionary values, e.g. '{metadata[name]}'. If a 2nd-level key contains a colon, e.g. 'music:Genre', ':' must be substituted by '#' in the template, like so: '{metadata[music#Genre]}'. [Default: 'tailored']
- **result_xfm** ({'datasets', 'successdatasets-or-none', 'paths', 'relpaths', 'metadata'} or callable or None, optional) -- if given, each to-be-returned result status dictionary is passed to this callable, and its return value becomes the result instead. This is different from *result_filter*, as it can perform arbitrary transformation of the result value. This is mostly useful for top- level command invocations that need to provide the results in a particular format. Instead of a callable, a label for a pre-crafted result transformation can be given. [Default: None]
- **return_type** ({'generator', 'list', 'item-or-list'}, optional) -- return value behavior switch. If 'item-or-list' a single value is returned instead of a one-item return value list, or a list in case of multiple return values. *None* is return in case of an empty list. [Default: 'list']

3.2.4 datalad.api.gooey_status_light

datalad.api.gooey_status_light(dataset=None, path: Path = None)

Internal helper for datalad-gooey

Parameters

- **dataset** (*Dataset or None, optional*) -- specify the dataset to query. If no dataset is given, an attempt is made to identify the dataset based on the current working directory. [Default: None]
- **path** -- [Default: None]
- **on_failure** ({*'ignore'*, *'continue'*, *'stop'*}, *optional*) -- behavior to perform on failure: 'ignore' any failure is reported, but does not cause an exception; 'continue' if any failure occurs an exception will be raised at the end, but processing other actions will continue for as long as possible; 'stop': processing will stop on first failure and an exception is raised. A failure is any result with status 'impossible' or 'error'. Raised exception is an IncompleteResultsError that carries the result dictionaries of the failures in its *failed* attribute. [Default: 'continue']

- **result_filter** (*callable or None, optional*) -- if given, each to-be-returned status dictionary is passed to this callable, and is only returned if the callable's return value does not evaluate to False or a ValueError exception is raised. If the given callable supports ***kwargs* it will additionally be passed the keyword arguments of the original API call. [Default: None]
- **result_renderer** -- select rendering mode command results. 'tailored' enables a command- specific rendering style that is typically tailored to human consumption, if there is one for a specific command, or otherwise falls back on the the 'generic' result renderer; 'generic' renders each result in one line with key info like action, status, path, and an optional message); 'json' a complete JSON line serialization of the full result record; 'json_pp' like 'json', but pretty-printed spanning multiple lines; 'disabled' turns off result rendering entirely; '<template>' reports any value(s) of any result properties in any format indicated by the template (e.g. '{path}', compare with JSON output for all key-value choices). The template syntax follows the Python "format() language". It is possible to report individual dictionary values, e.g. '{metadata[name]}'. If a 2nd-level key contains a colon, e.g. 'music:Genre', ':' must be substituted by '#' in the template, like so: '{metadata[music#Genre]}'. [Default: 'tailored']
- **result_xfm** ({'datasets', 'successdatasets-or-none', 'paths', 'relpaths', 'metadata'} or callable or None, optional) -- if given, each to-be-returned result status dictionary is passed to this callable, and its return value becomes the result instead. This is different from *result_filter*, as it can perform arbitrary transformation of the result value. This is mostly useful for top- level command invocations that need to provide the results in a particular format. Instead of a callable, a label for a pre-crafted result transformation can be given. [Default: None]
- **return_type** ({'generator', 'list', 'item-or-list'}, optional) -- return value behavior switch. If 'item-or-list' a single value is returned instead of a one-item return value list, or a list in case of multiple return values. *None* is return in case of an empty list. [Default: 'list']

INDEX

А

annex, 13 annexed-file, 13

В

branch, 13

С

clone, 13 commit, 13 commit message, 13

D

DataLad dataset, 13 DataLad extension, 13 DataLad subdataset, 14 DataLad superdataset, 14 drop, 14

G

get, 14 GIN, 14 Git, 14 git-annex, 14 git-annex branch, 14 GitHub, 14 GitLab, 14 gooey() (in module datalad.api), 30 gooey_askpass() (in module datalad.api), 31 gooey_lsdir() (in module datalad.api), 31 gooey_status_light() (in module datalad.api), 32

Η

https, 14

Ρ

pip, **14**

R

remote, 14

S

sibling, 15 SSH, 14 SSH key, 14

V

version control, 15