



Data Basics, Management and Sharing

Steve Slevinski

Agenda

Why Data Management?

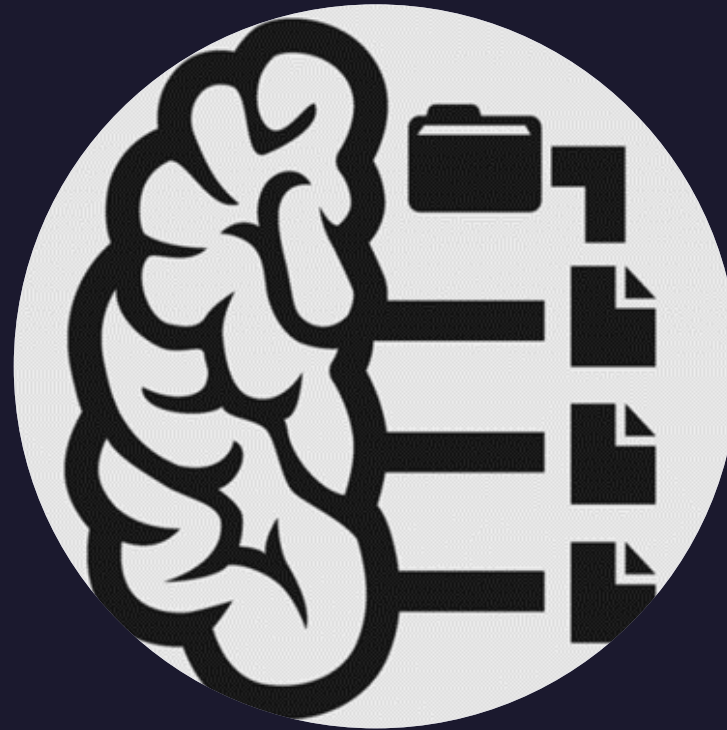
Brain Imaging Data Structure (BIDS)

NIMH Data Archive (NDA)

Globus data transfers

Flywheel

Data Basics





Why Data Management?

- Make data more accessible and understandable
- Reduce data redundancy
- Produce high quality data
- Improve data security
- Increase reproducibility and transparency
- Facilitate scientific discovery and enquiry
- Expand research recognition and attribution

Brain Imaging Data Structure (BIDS)

<https://bids.neuroimaging.io/>

BIDS Version History

v1.7.0 (2022-02-15)

v1.6.0 (2021-04-22)

v1.5.0 (2021-02-23)

v1.4.1 (2020-10-13)

v1.4.0 (2020-06-11)

v1.3.0 (2020-04-14)

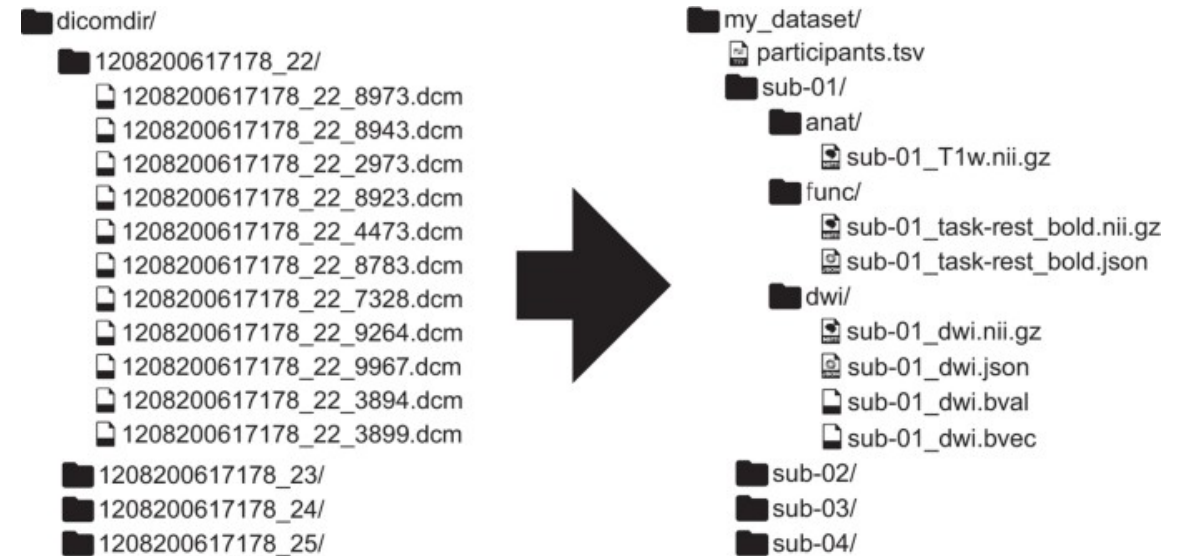
v1.2.2 (2020-02-12)

v1.2.1 (2019-08-14)

v1.2.0 (2019-03-04)

v1.1.2 (2019-01-10)

Scanner to BIDS Organization



DICOM (.dcm) Digital Imaging and Communications in Medicine
NIFTI (.nii) Neuroimaging Informatics Technology Initiative

Benefits of BIDS

- Easier to use
- Reduces error
- Optimized analysis
- Automated tools



<https://xkcd.com/2054>



Features of BIDS



Modular data



Standard folder hierarchy



Meaningful file names for
humans and code

Modular Data



Source Data

Beginning input



Raw Data

Standard data formats



Derivatives

Output by pipeline

Standard Folder Hierarchy

```
dataset/  
├── sourcedata  
│   ├── input_stream_1  
│   └── input_stream_2  
├── rawdata  
│   ├── dataset_description.json  
│   ├── participants.tsv  
│   ├── participants.json  
│   ├── phenotype  
│   │   ├── <measurement_tool_name>.tsv  
│   │   └── <measurement_tool_name>.json  
│   ├── sub-01  
│   └── sub-02  
└── derivatives  
    ├── pipeline_1  
    └── pipeline_2
```

Subject Only

Subject and Session

```
rawdata/  
├── sub-01  
│   ├── anat  
│   ├── dwi  
│   ├── fmap  
│   └── func  
└── sub-02  
    ├── anat  
    ├── dwi  
    ├── fmap  
    └── func
```

```
rawdata  
├── sub-01  
│   ├── ses-01  
│   │   ├── anat  
│   │   ├── dwi  
│   │   └── func  
│   └── ses-02  
│       ├── anat  
│       ├── dwi  
│       └── func  
└── sub-02
```

Meaningful File Names for Humans and Code

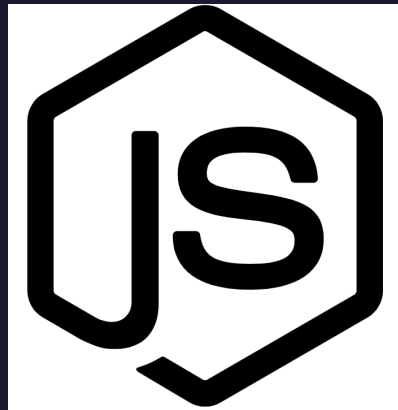
- **Template:** sub-<label>[_ses-<label>][_acq-<label>][_run-<index>]_<suffix>.nii[.gz]
- **File Name:** sub-10622_ses-20201020133843_acq-SL50_Tl rho.nii.gz
- A file name is a combination of key-value pairs, a suffix and an extension
- Dashes - separate key and value
- Underscores _ separate key-value pairs and separates the suffix
- Square brackets [] indicate that a key-value pair is optional
- <label> is an alphanumeric string
- <index> is a positive integer
- <suffix> is an alphanumeric string

BIDS Validator

Validate Dataset

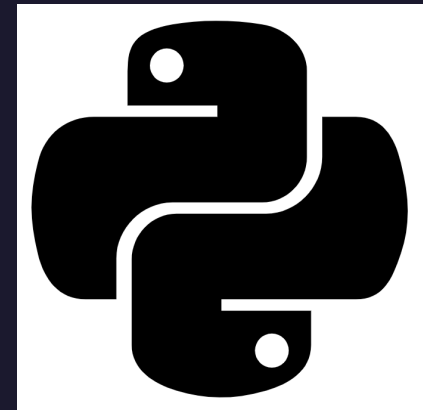


Chrome Browser



Node JS

Validate File Names



Python

<https://github.com/bids-standard/bids-validator>

BIDS Validator in the Browser



Select a **BIDS dataset** to validate

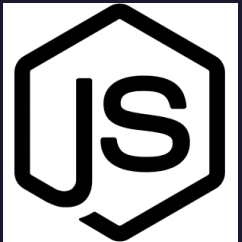
No file chosen

Options: Ignore Warnings Ignore NIFTI Headers Skip Subject Filename Consistency Check

Note: Selecting a dataset only performs validation. Files are never uploaded.

<https://bids-standard.github.io/bids-validator/>

BIDS Validator with Node JS



```

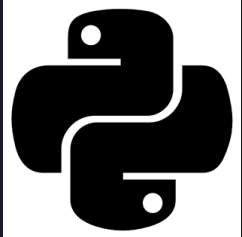
# install latest node and npm versions
curl -o- https://raw.githubusercontent.com/nvm-sh/nvm/v0.38.0/install.sh | bash
nvm install --lts
nvm install-latest-npm

# install bids-validator
npm install -g bids-validator

# validate dataset
bids-validator rawdata/
```

BIDS Validator with Python

Install with “pip install --user bids_validator”



```
from bids_validator import BIDSValidator
validator = BIDSValidator()
filepaths = ["/sub-01/anat/sub-01_rec-CSD_T1w.nii.gz", "/sub-01/anat/sub-01_acq-23_rec-CSD_T1w.exe"]
for filepath in filepaths:
    print(validator.is_bids(filepath)) # will print True, and then False
```

https://github.com/bids-standard/bids-validator/tree/master/bids-validator/bids_validator/rules

NDA

Welcome to the NIMH Data Archive



The National Institute of Mental Health Data Archive (NDA) makes available human subjects data collected from hundreds of research projects across many scientific domains. NDA provides infrastructure for sharing research data, tools, methods, and analyses enabling collaborative science and discovery. De-identified human subjects data, harmonized to a common standard, are available to qualified researchers. Summary data are available to all.

The NDA mission is to accelerate scientific research and discovery through data sharing, data harmonization, and the reporting of research results.

<https://nda.nih.gov/>

NDA Webinars And Tutorials

New Grantee Orientation

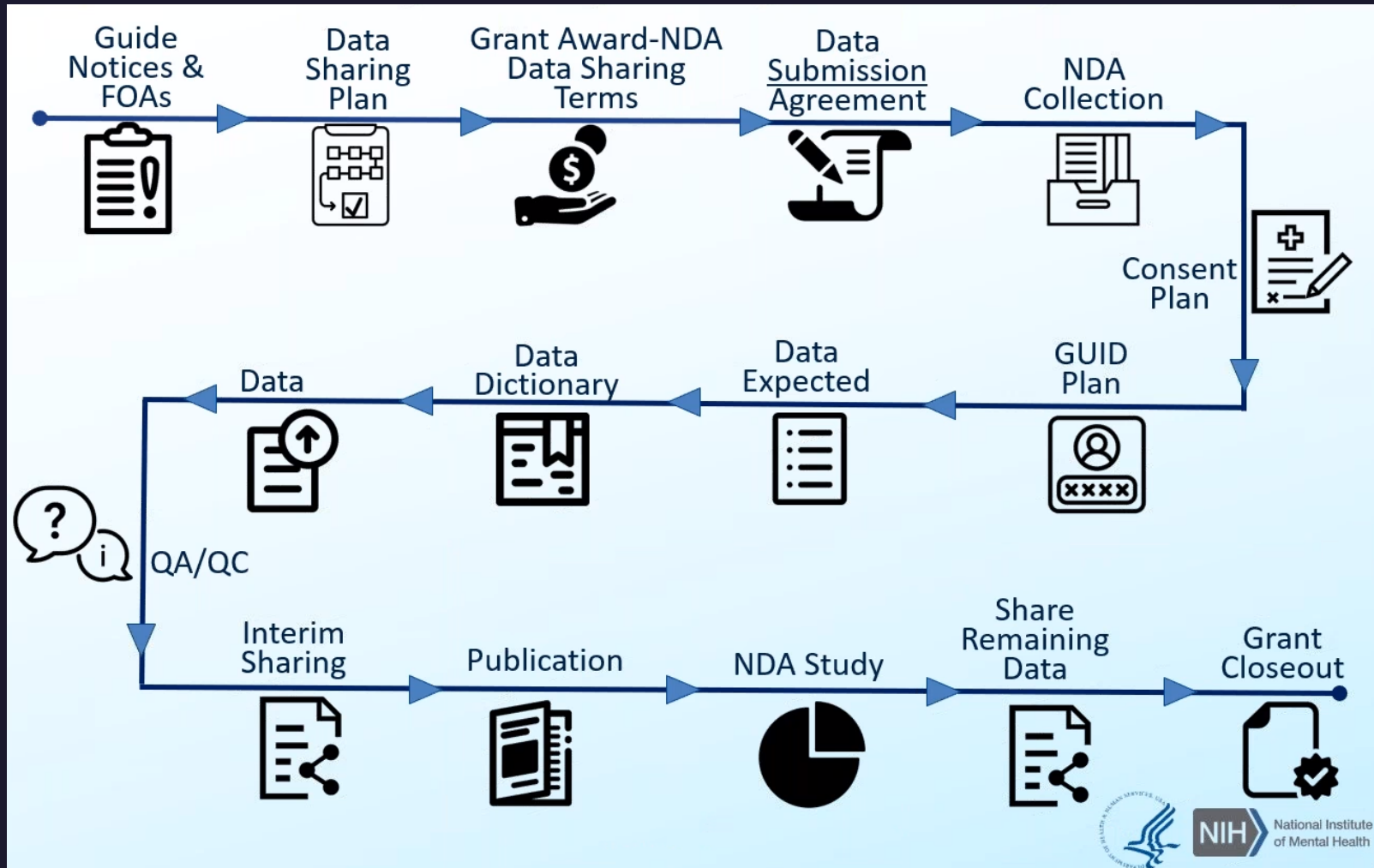
Data Harmonization

Data Validation & Submission

Data Access

<https://nda.nih.gov/webinars-and-tutorials>

Data Sharing Process Overview



NDA Collection

Growth and development of Striatal-Cerebellum circuitry in subjects at risk for Huntington's Disease #3409

State: Shared

General

Experiments (1)

Shared Data

Publications (9)

Data Expected (14)

Submissions

Associated Studies (0)

Permissions

- Experiments for fMRI
- Data Expected for data to submit
- Permissions for users: query, submission, and admin

Global Unique Identifier

Example GUID - NDARAB533LMW

- Download GUID Tool from NDA
- Enter information from birth certificate
- Generate GUID
- Batch mode available



<https://nda.nih.gov/s/guid/nda-guid.html>

NDA Data Dictionary

3,937 Results Found [Detail View](#) | [Table View](#)

SELECT	TITLE	SHORT NAME
<input type="checkbox"/>	Axis 1 Diagnoses	a1d01
<input type="checkbox"/>	Alternative Aspects of Personality Disorders and Social Functioning	aa01
<input type="checkbox"/>	Adapted ADOS Module 1	aados_m101
<input type="checkbox"/>	Adapted ADOS Module 2	aados_m201
<input type="checkbox"/>	Attitudes and Expectations Form	aaef01

<https://nda.nih.gov/general-query.html>

NDA Data Structure

ELEMENT NAME	DATA TYPE	SIZE	REQUIRED	CONDITION	DESCRIPTION
subjectkey	GUID		Required		The NDAR Global Unique Identifier (GUID) for research subject
src_subject_id	String	20	Required		Subject ID how it's defined in lab/project
interview_date	Date		Required		Date on which the interview/genetic test/sampling/imaging/biospecimen was completed. MM/DD/YYYY
interview_age	Integer		Required		Age in months at the time of the interview/test/sampling/imaging.
sex	String	20	Required		Sex of subject at birth









Additional columns of value range, notes, and aliases

Required column can be Required, Recommended, or Conditional

Recommended rows must be submitted if available

Conditional rows specify a Boolean test using another element value

NDA Data Expected

Data Expected 	Targeted Enrollment 	Initial Submission 	Subjects Submitted 
Child Behavior Checklist (CBCL) 	400	07/15/2021	41
Adult Behavior Checklist (ABCL) 	400	07/15/2021	37
Research Subject and Pedigree 	400	07/15/2021	95
Picture Sequence Memory 	400	07/15/2021	67



Format Image Data with bids2nda

- Download main.py and rename to bids2nda.py
- Requires code modifications for errors and issues
- Creates image03 data structure and zipped metadata files
- Ready to validate and upload

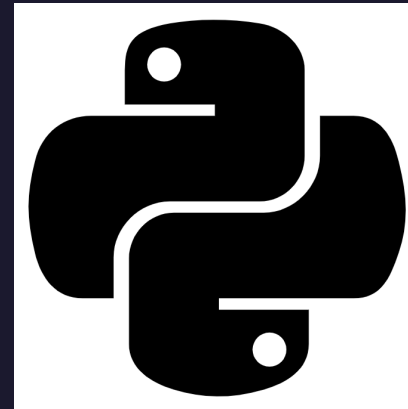


<https://github.com/bids-standard/bids2nda/blob/master/bids2nda/main.py>

Validation and Upload Tool



Chrome Browser



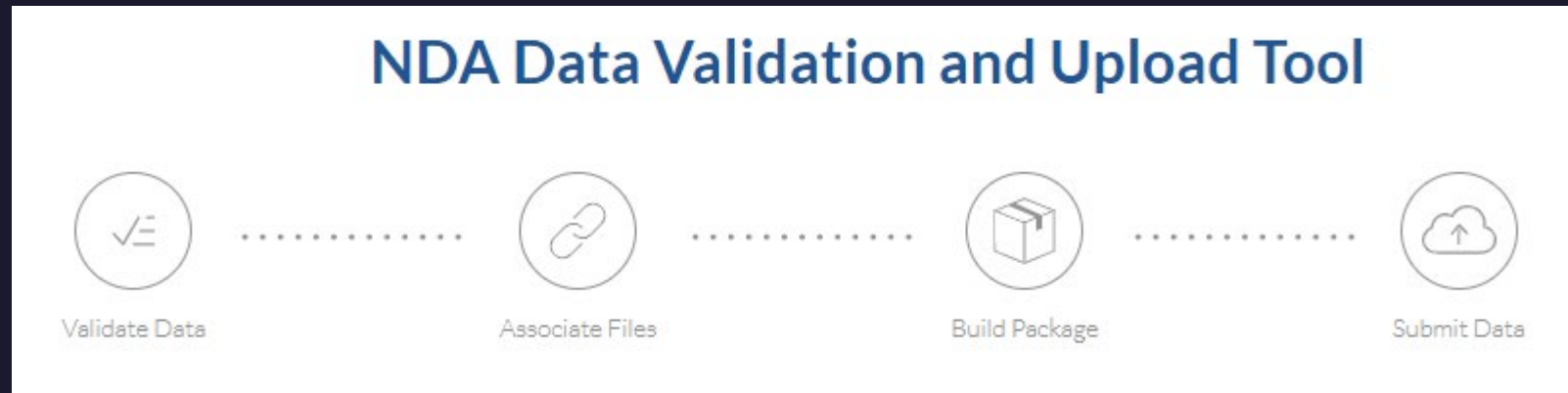
Python

<https://nda.nih.gov/tools/nda-tools.html#vt>

Validation and Upload Tool



Chrome
Browser

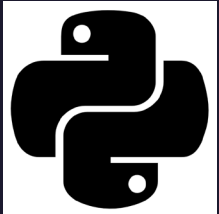


RESUME INCOMPLETE SUBMISSION

CREATE NEW SUBMISSION

<https://nda.nih.gov/vt/>

Validation and Upload Tool



Python

```
# install validation and upload tool
pip install --user nda-tools

# validate datasheet
vtcmd psm.csv

# upload datasheet after passing validation
vtcmd psm.csv -l /dir/associated/files -b -c 3409 -u username -d "Description" -t "Title"
```

<https://nda.nih.gov/tools/nda-tools.html#vt>

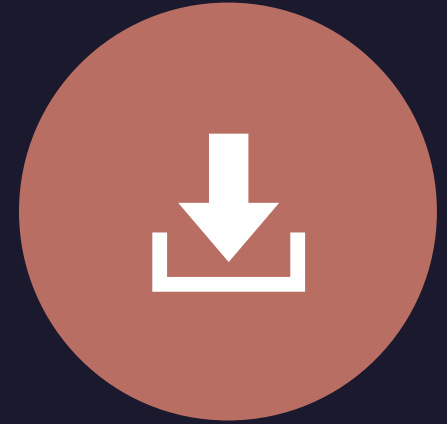
NDA Data Access



Request Access




Search Data



Download Data



NDA Request Access

- Make an account with NDA
 - Go to Data Permissions
 - Complete Data Use Agreement
 - Send to DSP workflow for signature
 - Upload signed Data Use Agreement to NDA
 - Once approved, acknowledge permissions on NDA site
- 

General Search

FILTER BY SITES

[Select All](#) | [Clear All](#)

- NIMH Data Archive
- Adolescent Brain Cognitive Development Study
- CCF
- Osteoarthritis Initiative
- NIAAADA
- Accelerating Medicines Partnership - Schizophrenia (AMP SCZ) Data Repository

Filter Data Types

CONTENT (15)

NDA STUDIES (37)

PUBLICATIONS (1166)

COLLECTIONS (601)

DATA STRUCTURES (118)

DATA ELEMENTS (80)

EXPERIMENTS (133)

2150 search results for

fmri

<https://nda.nih.gov/search>

Data Search

Anyone

- Summary
- General information

Approved Access

- Detailed subject level data
- Build data packages

Search by

- Featured Datasets
- Data from Labs
- Data from Paper
- Data Dictionary
- Experiments
- Global Unique Identifier (GUID)



<https://nda.nih.gov/general-query.html?q=query=collections>

Download Data

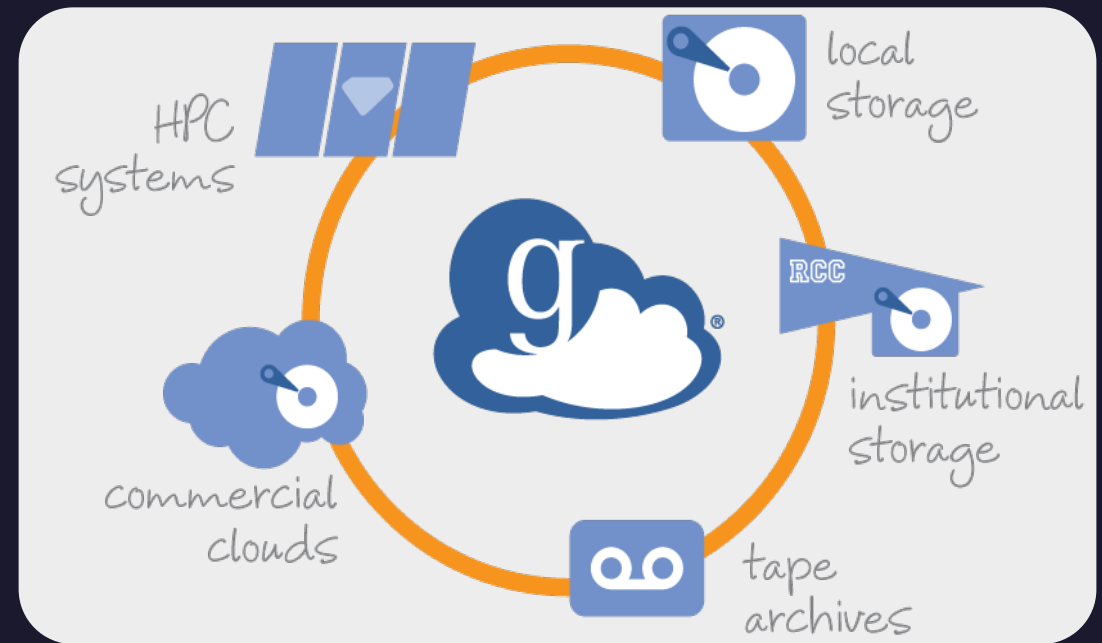
- Create data package
- Create miNDAR from package
- Oracle database
 - Requires Oracle SQL Developer
- Amazon Web Services
- Granular download
- Download Manager (Electron app with 200GB limit)
- nda-tools (Python with security token)



Globus

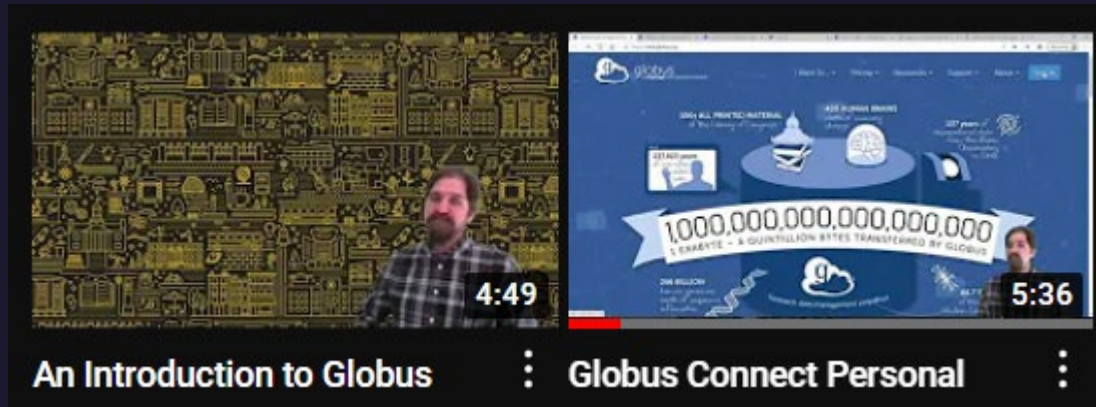
<https://www.globus.org>

Large Scale Data Transfer



Collection: uiowadata#data

Access University of Iowa data



<https://www.youtube.com/channel/UCeSfxCh8iohP-9Bg0TOhDJA>

File Manager

Collection

Path

select all

NAME ▾

- Dedicated
- home
- hpchomes
- nfsscratch
- scratch
- Shared

globus connect software

globus connect **personal**



For researchers and
other end users



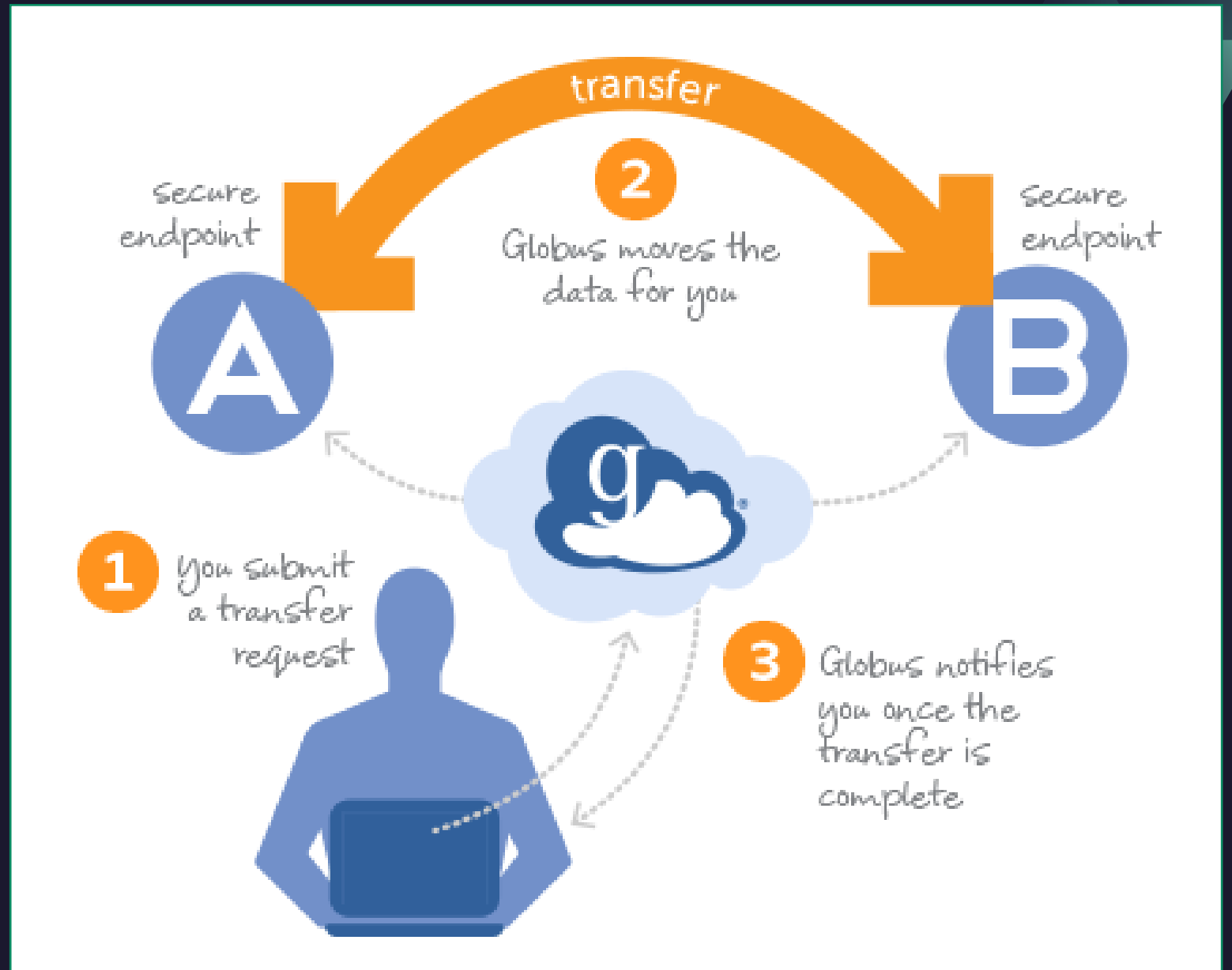
globus connect **server**

For network and
system administrators



Transfer

- Move data from endpoint to endpoint
- Endpoints on Globus connect server
- Endpoints on Globus connect personal



Sharing

- Authenticated University Login
- Share authenticated security rights
 - Globus interacts as you
- Share read and/or write access
 - Let others interact as you
- Share with users and groups
 - Authenticated with Globus

Log in to use Globus Web App

Use your existing organizational login

e.g., university, national lab, facility, project

University of Iowa

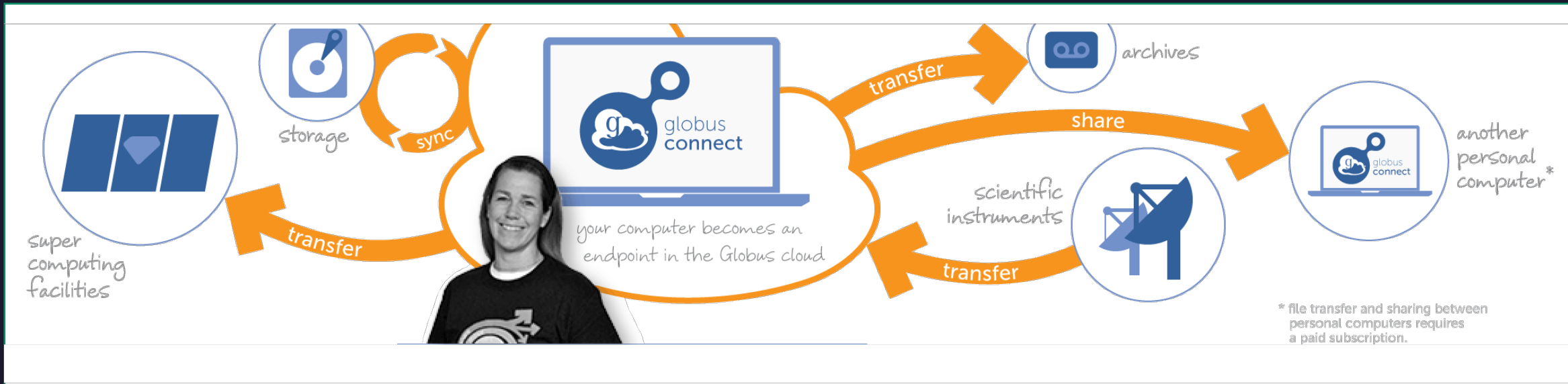
By selecting Continue, you agree to Globus [terms of service](#) and [privacy policy](#).

Continue

Know your data classification

Public, Internal, Restricted, Critical

<https://its.uiowa.edu/support/article/11090>



globus connect personal

Collection Search

Collection

- Steve Slevinski's MacBook
Owner: sslevinski@uiowa.edu
macOS 12
- Steve Slevinski's Desktop
Owner: sslevinski@uiowa.edu
Windows 11

File Manager

Collection: uiowadata#data Steve Slevinski's Desktop

Path: /Dedicated/inc_database/ /C/Users/slevi/Desktop/

Start Transfer & Timer Options Start

NAME	LAST MODIFIED	SIZE
admin	4/14/2022, 10:...	
agreements	12/29/2020, 2:...	
archive	3/3/2021, 11:51	
bigdata	7/30/2021, 8:5	

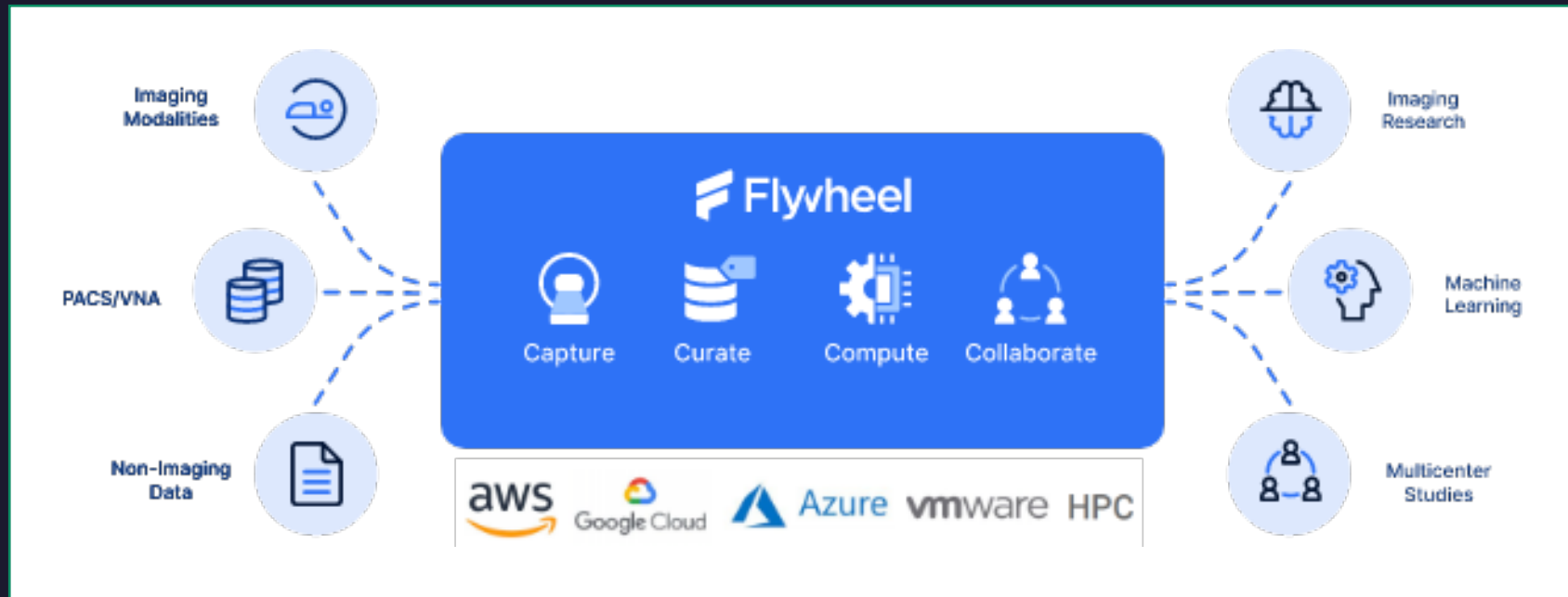
NAME	LAST MODIFIED	SIZE
Globus	1/14/2022, 5:...	

- Share
- Transfer or Sync to...
- New Folder
- Rename
- Delete Selected
- Download

Flywheel

<https://flywheel.io>

Platform for imaging research



Data Basics

JSON

ASCII and Unicode

Bash variables, scripts, and arguments

Regular Expressions for text

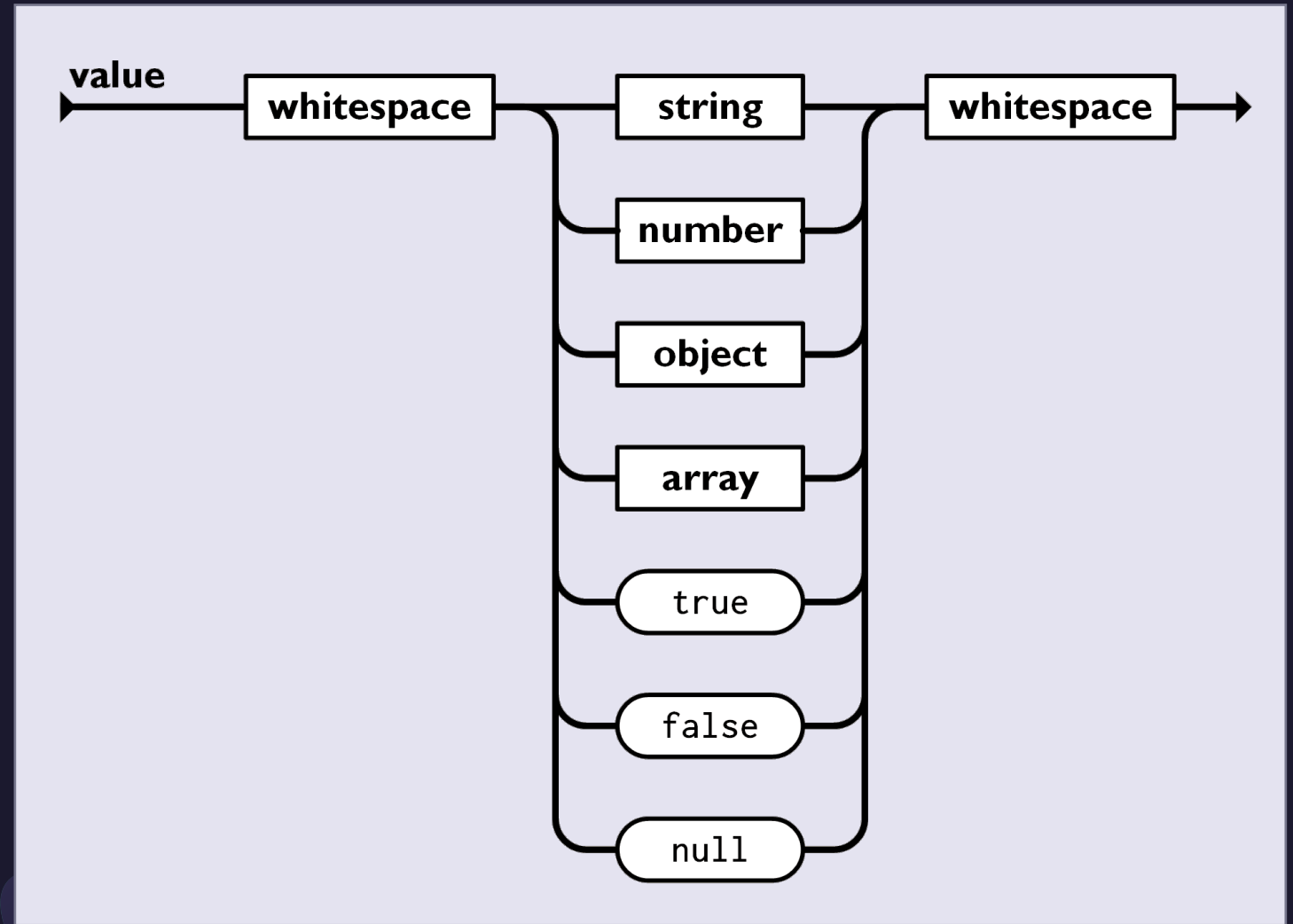
Globbering for files



JSON

A lightweight data-interchange format

<https://jsonbeautify.com>



<https://www.json.org>

JSON

A lightweight data-interchange format

<https://jsonbeautify.com>

```
string = "hello world"
```

```
number = 0.2
```

```
array = [1,2,3]
```

```
object = {  
  "string": "hello",  
  "number": 7  
  "array": [1,2,3]  
}
```

<https://www.json.org>

ASCII

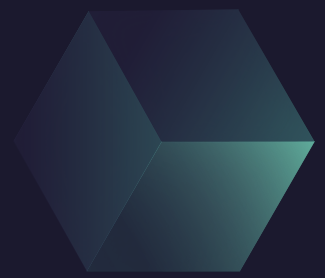
The universal 7-bit character set

- 7-bits have 128 different combinations
- Int: 0 to 127
- Hex: 00 to 7F

- 8-bits have 256 different combinations
- Int: 0 to 255
- Hex: 00 to FF

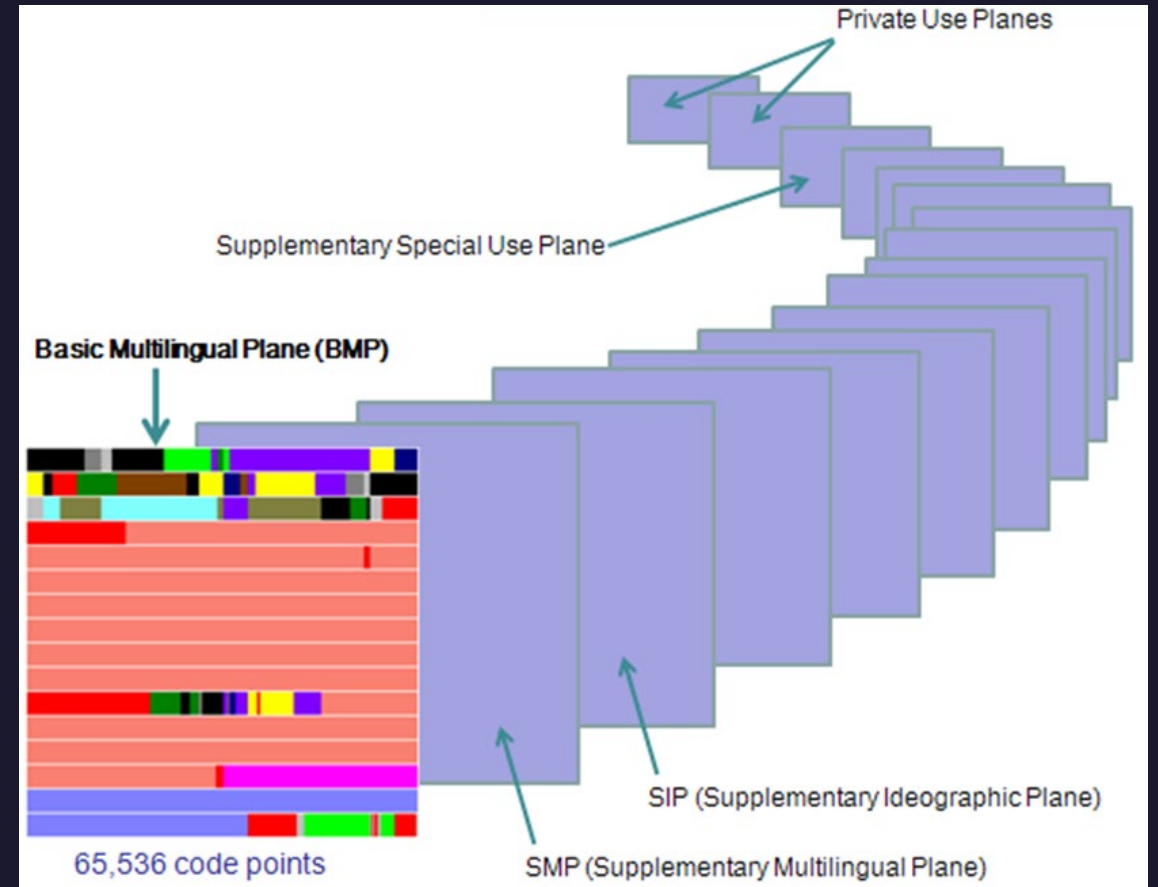
	0	1	2	3	4	5	6	7
0	NUL	DLE	space	0	@	P	`	p
1	SOH	DC1 XON	!	1	A	Q	a	q
2	STX	DC2	"	2	B	R	b	r
3	ETX	DC3 XOFF	#	3	C	S	c	s
4	EOT	DC4	\$	4	D	T	d	t
5	ENQ	NAK	%	5	E	U	e	u
6	ACK	SYN	&	6	F	V	f	v
7	BEL	ETB	'	7	G	W	g	w
8	BS	CAN	(8	H	X	h	x
9	HT	EM)	9	I	Y	i	y
A	LF	SUB	*	:	J	Z	j	z
B	VT	ESC	+	;	K	[k	{
C	FF	FS	,	<	L	\	l	
D	CR	GS	-	=	M]	m	}
E	SO	RS	.	>	N	^	n	~
F	SI	US	/	?	O	_	o	del

Unicode



The international 32-bit character set

- Divided into multiple 16-bit planes
- 16-bits have 65,536 different combinations
- Hex: 0000 to FFFF
- Each character can be referenced with a plane followed by a 16-bit value.
- U+007F is plane zero and character 007F.
- This is ASCII character DEL.
- U+1D800 is plane one with character D800.
- This character is not available in ASCII.



UTF-8 encoding form: ASCII superset

First code point	Last code point	Byte 1	Byte 2	Byte 3	Byte 4
U+0000	U+007F	0xxxxxxx			
U+0080	U+07FF	110xxxxx	10xxxxxx		
U+0800	U+FFFF	1110xxxx	10xxxxxx	10xxxxxx	
U+10000	U+10FFFF	11110xxx	10xxxxxx	10xxxxxx	10xxxxxx

<https://www.urlencoder.io/>

"ASCII-Double-Quotes"
“Unicode-Fancy-Quotes”

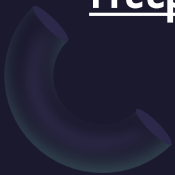


UTF-16 encoding: an unfortunate reality

High	Low	DC00	DC01	...	DFFF
D800		010000	010001	...	0103FF
D801		010400	010401	...	0107FF
⋮		⋮	⋮	⋮	⋮
DBFF		10FC00	10FC01	...	10FFFF

<https://niddapp.com/json-encoder.html>

"ASCII-Double-Quotes"
“Unicode-Fancy-Quotes”



Bash

Bash cheat sheet

<https://devhints.io/bash>

Bash quick start

<https://learnxinyminutes.com/docs/bash/>

Bash reference

<https://www.gnu.org/software/bash/manual/>



Variables

Command	Result
test1=hello	Variable test1 = "hello"
test2="hello world"	Variable test2 = "hello world"
readonly test3="value"	Variable test3 = "value" and can not be change
echo \$test2	Prints "hello world" with \$1 = hello, \$2 = world
echo "\$test2"	Prints "hello world" with \$1 = "hello world"
echo \${test1}	Prints "hello"
echo \$test1_world	Prints nothing
echo \${test1}_world	Prints "hello_world"

Variable parameter expansion

Command	Result
echo \${test:-alternate}	Prints “alternate” if test is not set
echo \${test:=alternate}	Updates test and prints if test is not set
echo \${test:+alternate}	Prints “alternate” if test is set, otherwise prints nothing
echo \${test:1}	Prints test starting after first character
echo \${test:3:2}	Prints two characters starting after the third character
echo \${#test}	Prints length of test
echo \${test/pattern/string}	Prints test after substituting first matching pattern with string
echo \${test//pattern/string}	Prints test after substituting all matching patterns with string
echo \${!test}	Prints values of variable referenced by value of test

<https://wiki.bash-hackers.org/syntax/pe>

Array Variables

Command	Result
Declare -a ARRAYNAME	Explicitly create an empty array variable
ARRAYNAME=(A B "C D")	Create array variable with three values
ARRAYNAME[3]="E F G"	Set value of array variable at index
ARRAYNAME+=(H)	Add an element to end of array variable
ARRAYNAME+=(I "J K L")	Add two elements to end of array variable
echo \$ARRAYNAME	Prints first element of array variable
echo \${ARRAYNAME[0]}	Prints first element of array variable
echo \${ARRAYNAME[*]}	Prints all elements of array variable
echo \${#ARRAYNAME[@]}	Prints length of array
for ELEMENT in "\${ARRAYNAME[@]"; do echo \$ELEMENT done	Prints all elements of array variable

Scripts and input arguments

```
1 #!/bin/bash
2
3 echo "Script is $0"
4 echo "First arg is ${1:-unset}"
5 echo "Second arg is ${2:-unset}"
```

Command	Output
./script.sh	Script is ./script.sh First arg is unset Second arg is unset
./script.sh hello world	Script is ./script.sh First arg is hello Second arg is world
./script.sh "hello world"	Script is ./script.sh First arg is hello world Second arg is unset



Positional Arguments versus flags

Example	Explanation
<code>echo hello world</code>	\$1 is “hello”, \$2 is “world”
<code>echo “hello world”</code>	\$1 is “hello world”
<code>cp current.txt new.txt</code>	\$1 is “current.txt”, \$2 is “new.txt”
<code>cp “old file” “new file”</code>	\$1 is “old file”, \$2 is “new file”
<code>ls -ltr</code>	Turn on flags “-l” for long, “-t” for time order, and “-r” for reverse
<code>ls -l -t -r</code>	Turn on flags “-l” for long, “-t” for time order, and “-r” for reverse
<code>ls --ignore README.txt</code>	Flag “--ignore” will list files other than README.txt

For flag arguments use `getopts` command.

Regular Expression for text

Characters	Description	Example
.	Match any character	. matches A, B, C, or any single character
*	Match preceding literal 0 or more times	ABC* matches AB, ABC, ABCC, ...
+	Match preceding literal 1 or more times	ABC+ matches ABC, ABCC, ABCCC, ...
?	Match preceding literal 0 or 1 times	ABC? matches AB or ABC
{#}	Match preceding literal “#” times	AB{2} matches ABB
[]	Match any single literal from a list	[ABC] matches A, B or C
[-]	Match any single literal from a range	[A-C] matches A, B, or C
()	Creates a group for matching	A(BC)+ matches ABC, ABCBC, ABCBCBC, ...
()	Matches one of several alternatives	(AB BC CD) will match AB, BC, or CD
(?:)	Creates a non-capturing group	A(?:BC) will match ABC as one group
^	Match beginning of line	^A.* will match any line starting with A
\$	Match ending of line	.*A\$ will match any line ending with A

Globbing for files

Characters	Description	Example
?	Match any single character	ls test.???
*	Match any string of characters, except /	ls test.*
**	Match any string of characters, including /	ls **/test.txt
[]	Match any characters from a list	ls [ac-z]*
[!]	Match any characters not in a list	ls [!ac-z]*
{ }	Use multiple matching patterns	ls {*.txt,*.zip}
+()	Match one of several alternatives	ls *+(txt zip)

Use “set -f” to turn off globbing and “set +f” to globbing back on.

Use “shopt -s globstar” to turn on ** if not enabled.

Globs versus Regular Expressions

Characters	Meaning in Globs	Meaning in Regular Expressions
*	Match zero or more characters	Match zero or more of the preceding literal
?	Match any single character	Match zero or one of the preceding literal
.	Literal “.” character	Match any single character

Globs for files	Regular Expressions for text
<code>names*</code> matches <code>names.txt</code> and <code>names.zip</code>	<code>names*</code> matches <code>name</code> , <code>names</code> , and <code>names</code>



Thank You

Steve Slevinski
Data Manager

steve-slevinski@uiowa.edu

<https://neuroimaging.uiowa.edu>

