



Introduction to SODA

Bhavesh Patel, Ph.D.
Associate Research Professor



CALIFORNIA MEDICAL
INNOVATIONS INSTITUTE

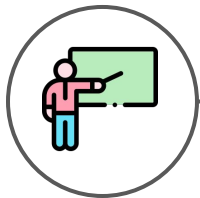


FAIR DATA
INNOVATIONS HUB

About This Presentation



<20 min + Q&A



Learn about SODA &
how it can help you

SODA = Software to Organize Data Automatically



Background

SPARC Guidelines

SPARC imposes **strict**
data preparation and submission **guidelines**

- ❑ Organize data into a standard folder structure
- ❑ Follow file and folder naming conventions
- ❑ Include specific metadata files
- ❑ Upload dataset on the Pennsieve platform
- ❑ etc.

Problem

The guidelines can be **time consuming and difficult** to implement



I experienced this firsthand in 2017 while managing data from our group's SPARC-funded project

Solution

Automating the SPARC data preparation and submission process

2017

Started developing
Python scripts for
our group's data



2018

Presented the idea
at the SPARC
Hackathon



2019

Received funding
from SPARC
→ Birth of SODA





About SODA

Overview of SODA

SODA is a **desktop software** that simplifies preparation and submission of datasets according to the SPARC guidelines



Cross-platform



Free

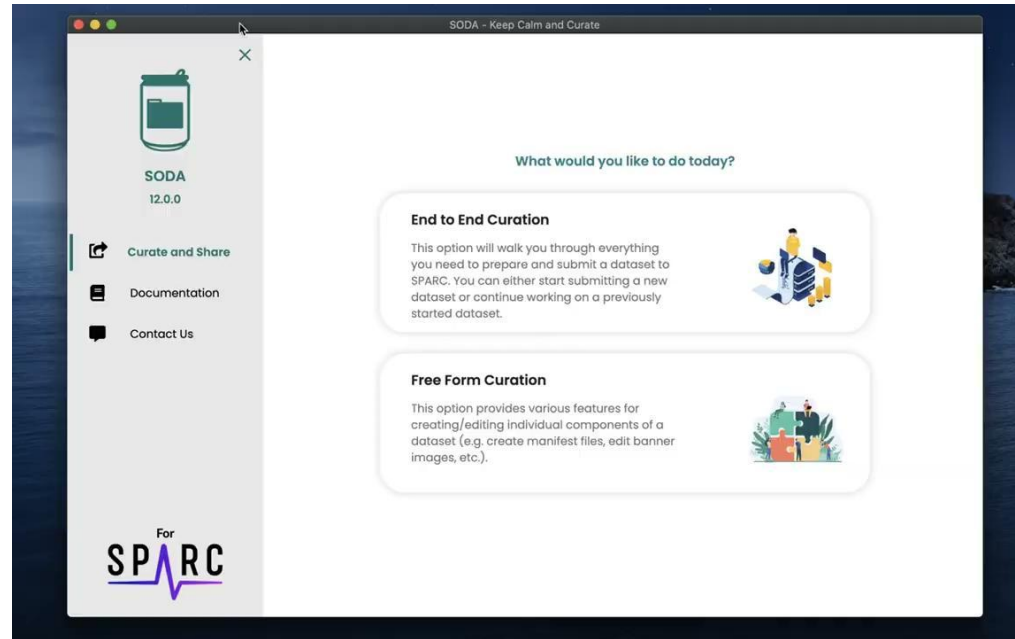


Open source



Web style frontend
Python backend

Overview of SODA



([Link to video](#))

Like a tax filing tool but for filing SPARC datasets

User Testing (10 SPARC Investigators, 2020)



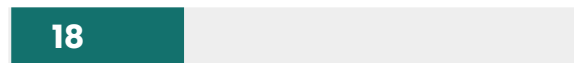
Without SODA



**Prepare and submit
a sample dataset**



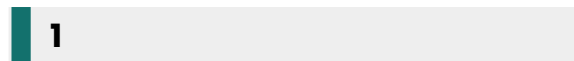
With SODA



Average Ease
of use (/5)



Total number
of errors



Usage Tracking (Last 2 Years)

SODA has been **widely used** by SPARC 1 investigators




Installed on 680 + unique computers



16 TB + data processed



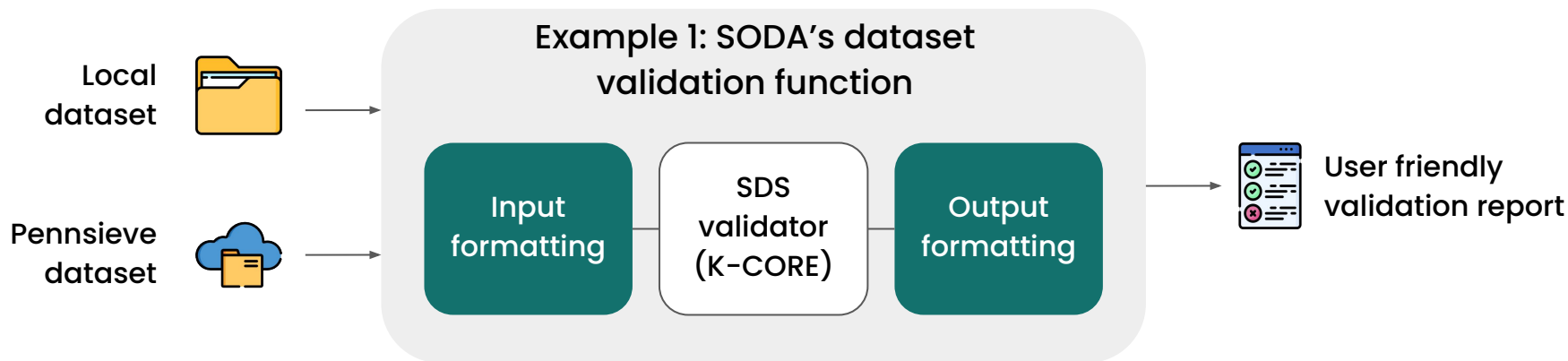
230k + files processed



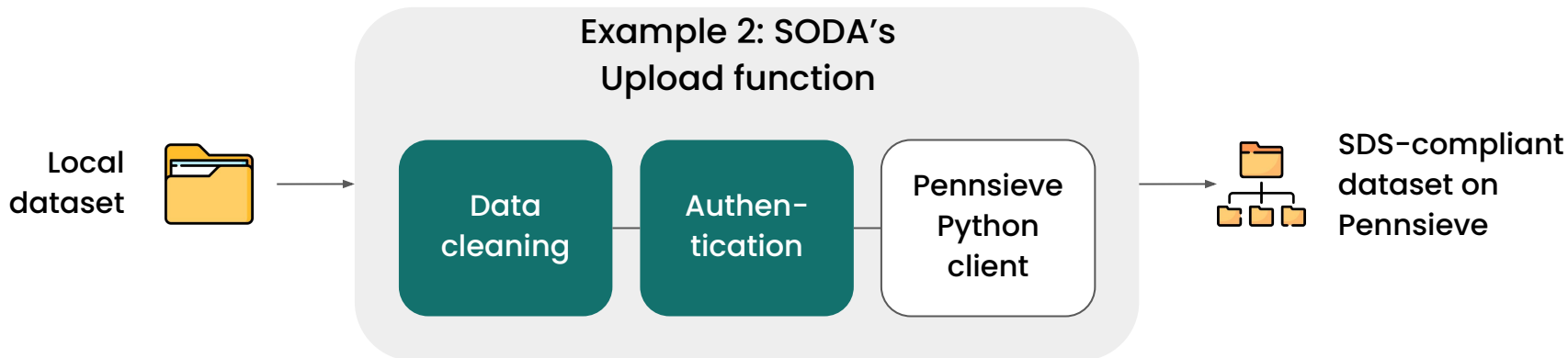
How SODA Can Be Useful to You (SPARC 2 teams)

Python Backend Code

SODA's thoroughly **tested and validated backend Python functions** can be reused in your automated workflows



Python Backend Code



- Many more functions: create metadata, upload banner image, etc.
- Use, copy, modify, as needed (MIT license)

Don't reinvent the wheels!

SODA Python Package?

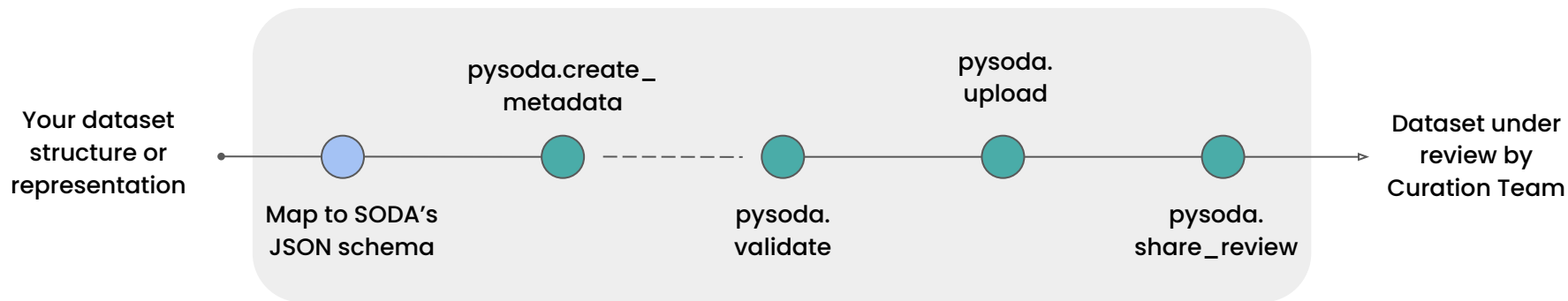
We plan to develop a **SODA Python package** if there is interest for easy integration in your workflows



pysoda

SODA Python Package?

Potential data preparation and submission workflow of a SPARC 2 team:



Why a SODA Python Package?



Prevent siloed development



Build easier to maintain workflows



Provide support to non-SPARC data submitters

Focus more on the science, less on the data submission

Tell us if you are interested in a SODA Python package!



bpatel@calmi2.org



bit.ly/gitsoda

Come chat with me at the SODA poster at lunch time
(12:25 PM – 1:15 PM)

Find these slides and
all resources here

