

# PsychNotebook

Inhaltliche und technische Anforderungen an eine JupyterHub-basierte Umgebung für die Studienplanung, Datenanalyse und Befunddokumentation in der Psychologie



**Jupyter...?**

Quelle: NASA/JPL/University of Arizona  
<https://photojournal.jpl.nasa.gov/catalog/PIA02873>

# Jupyter Notebook / Lab

- **ausführbare** “Notebook”-Dokumente, die sowohl Quellcode als auch Text- und Graphikelemente enthalten
- Nutzung **unterschiedlicher Programmiersprachen** möglich (z.B. Python, R, Julia, Java, PHP etc.)
- einfacher Export in statische Formate (HTML, PDF) und Weitergabe von Notebookdokumenten

# Jupyter Notebook bzw. JupyterLab

The screenshot displays a Jupyter Notebook environment with the following components:

- Header:** "PsychNotebook PREVIEW OA\_in\_PSYNDEX Last Checkpoint: 23.02.2018 (autosaved)" with "Logout" and "Control Panel" buttons.
- Menu:** File, Edit, View, Insert, Cell, Kernel, Widgets, Help, Snippets.
- Toolbar:** Includes icons for file operations, execution, and output viewing.
- Code Cell:**

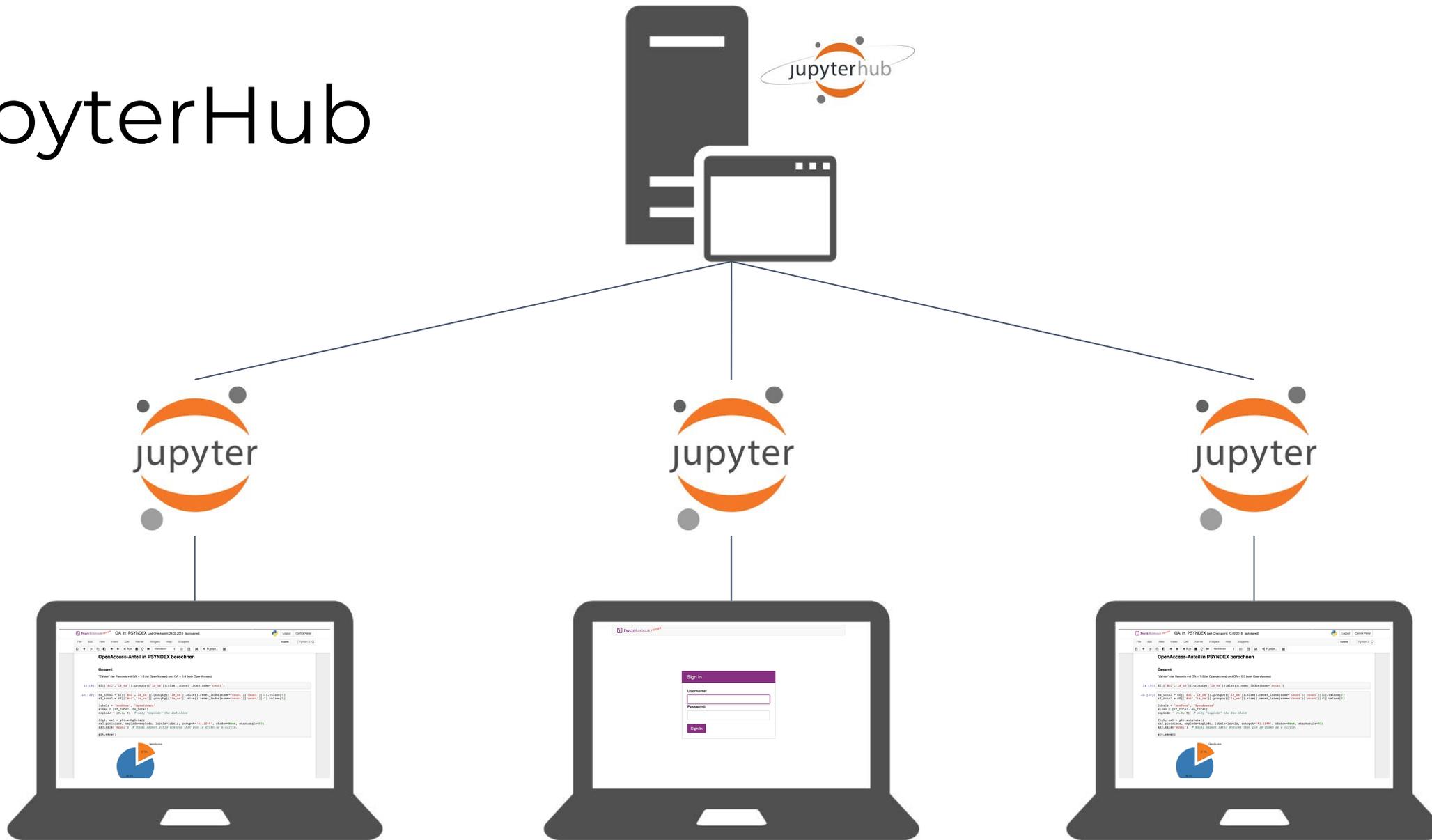
```
df[['doi', 'is_oa']].groupby(['is_oa']).size().reset_index(name='count')
oa_total = df[['doi', 'is_oa']].groupby(['is_oa']).size().reset_index(name='oa_total')
nf_total = df[['doi', 'is_oa']].groupby(['is_oa']).size().reset_index(name='nf_total')
labels = 'nonFree', 'OpenAccess'
sizes = [nf_total, oa_total]
explode = (0.3, 0) # only "explode" the 2nd slice
fig1, ax1 = plt.subplots()
ax1.pie(sizes, explode=explode, labels=labels, autopct='%1.1f%%', shadow=True, radius=100)
ax1.axis('equal') # Equal aspect ratio ensures that pie is drawn as a circle.
plt.show()
```
- Figure 1:** A pie chart titled "OpenAccess-Anteil in PSYNDEX berechnen" showing 82.5% for "nonFree" (blue) and 17.5% for "OpenAccess" (orange).
- Code Cell:**

```
N = len(publishers)
ind = np.arange(N) # the x locations for the groups
width = 0.75 # the width of the bars: can also be len(x) sequence

p1 = plt.bar(ind, nonFree, width, color='d62728')
p2 = plt.bar(ind, openAccess, width, color='00aa00', bottom=nonFree)
plt.ylabel('Anzahl')
plt.title('OpenAccess in PSYNDEX nach Verlagen (nur DOI-Check)')
xtick_pos = (ind + width*0.5)
plt.xticks(xtick_pos, publishers, rotation='vertical')
plt.yticks(np.arange(0, 14000, 1000))
plt.legend((p1[0], p2[0]), ('Non-Free', 'OpenAccess'), loc='upper right')
plt.rcParams["figure.figsize"] = [20,9]
plt.grid(color='gray', linestyle='-', linewidth=0.1)
```
- Figure 2:** A stacked bar chart titled "OpenAccess in PSYNDEX nach Verlagen (nur DOI-Check)". The y-axis is "Anzahl" (0 to 13000). The x-axis lists publishers. Red bars represent "Non-Free" and green bars represent "OpenAccess".
- Table of Contents:** A sidebar on the right with a "# Table of Contents" link.
- Raw NBConvert Format:** A sidebar showing metadata: 

```
{ "toc": "true" }
```

# JupyterHub



# PsychNotebook

- interner Demonstrator am ZPID basierend auf JupyterHub
  - Authentifizierung gegen ActiveDirectory
  - Testen verschiedener Kernels (v.a. Python und R)
  - Anpassungen an der Oberfläche, z.B. Einbindung von Snippets
  - “Sharing” über Directory mit Schreib-/Leserechten für Gruppe

# Use Cases

- Bereitstellung einer **einfach** zu bedienenden Umgebung in der Lehre
- kollaboratives Arbeiten von verteilten Forschergruppen
- Notebooks als Supplements für Zeitschriftenartikel, um Analysen reproduzieren zu können
- direkte Verknüpfung mit Forschungsdaten- bzw. Coderepositories

# Inhaltliche Anforderungen

## **Komponente zu Planung des Studiendesign, z.B.**

- Berechnung des optimalen Stichprobenumfangs
- Planung der Stichprobenzusammensetzung

## **Komponente zur Datenanalyse**

Primär- und Sekundäranalyse psychologischer Forschungsdaten

# Nutzer-Anforderungen

- Oberfläche menügeführt (SPSS-like)
- einfache Erweiterbarkeit durch Nachinstallation von Paketen

# Authentifizierung

- **interne Nutzung** (z.B. gegen ActiveDirectory)
- **externe Nutzung** (ORCID, Google, GitHub, Facebook etc.)

# Sicherheit

- **Sicherheit der Benutzerdaten** (Daten und Code) vor unberechtigtem Zugriff
- **Serversicherheit** (und Sicherheit der Infrastruktur) vs. Ausführbarkeit beliebigen Codes durch den Nutzer
- **Backup**konzept

# Schnittstellen

- einfache Einbindung externer Datenquellen und Systeme (z.B. Fachzeitschriften, Repositories)
- Einbindung von Datenquellen, die **Nutzungsbeschränkungen** unterliegen (z.B. scientific use, Genehmigung durch Datengeber erforderlich etc.)
- Export von Notebooks in Publikationsformate

# Skalierung

- Speicherplatz
- Rechenleistung

# Vielen Dank

Peter Weiland, <[pw@leibniz-psychology.org](mailto:pw@leibniz-psychology.org)>