Programmatic Access to PDG Data

Juerg Beringer
Physics Division
Lawrence Berkeley National Laboratory

For the PDG Collaboration

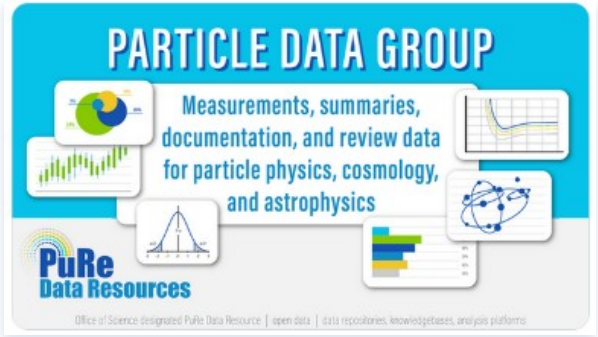
HADRON 2023 Conference
Genova, June 5 – 9, 2023

Outline

- **About PDG and the *Review of Particle Physics***
- **New tools for accessing PDG data in machine-readable format**
  - REST API
  - Python API
  - Database files
- **Status and Challenges**
- **Conclusions**

PDG in a Nutshell

- **The Particle Data Group provides a comprehensive summary of particle physics and related areas (cosmology, astrophysics) in a single publication, the *Review of Particle Physics***
  - International collaboration with 236 scientific authors from 171 institutions and 26 countries
  - Coordination team at LBNL provides overall scientific leadership, central coordination, and technical expertise & infrastructure for whole collaboration
  - Funded primarily by the DOE Office of High Energy Physics with important contributions from MEXT (Japan), INFN, JPS, and CERN
  - Recognized by DOE Office of Science as authoritative data resource in 2021 ("PuRe Data Resource")



- ***Review of Particle Physics* consists of**
  - **Summary Tables**
  - **Particle Listings**
  - **120 review articles** covering a wide range of topics (Standard Model, searches, cosmology, experimental methods, mathematical tools, atomic and nuclear properties, ...)

- **Updated yearly (on the web) and published in a journal every two years**

Summary Tables and Particle Listings

- **Summary Tables give PDG world averages (or best limits)**
  - Some provided by specialized averaging groups (HFLAV, ...)
- **Particle Listings detail how averages/fits were obtained**
  - Including **curated, annotated** lists of published results
- **PDG scans published literature for relevant results**
- **For each measurement, the PDG entry is carefully put together by two experts and finally checked by the corresponding experiment**

Summary Table

$D^*(2007)^0$ $I(J^P) = \frac{1}{2}(1^-)$

I, *J*, *P* need confirmation.

Mass $m = 2006.85 \pm 0.05$ MeV (S = 1.1)

$m_{D^{*0}} - m_{D^0} = 142.014 \pm 0.030$ MeV (S = 1.5)

Full width $\Gamma < 2.1$ MeV, CL = 90%

$\overline{D}^*(2007)^0$ modes are charge conjugates of modes below.

$D^*(2007)^0$ DECAY MODES	Fraction (Γ_i/Γ)	p (MeV/c)
$D^0\pi^0$		
$D^0\gamma$		
$D^0 e^+ e^-$		

Particle Listing

$m_{D^*(2007)^0} - m_{D^0}$

The fit includes $D^\pm$, D^0, $D_s^\pm$, $D^{*\pm}$, D^{*0}, $D_s^{*\pm}$, $D_1(2420)^0$, $D_2^*(2460)^0$, and $D_{s1}(2536)^\pm$ mass and mass difference measurements.

VALUE (MeV)	*EVTS*	*DOCUMENT ID*		*TECN*	*COMMENT*
142.014±0.030 OUR FIT Error includes scale factor of 1.5.					
142.016±0.030 OUR AVERAGE Error includes scale factor of 1.5.					
142.007±0.015±0.014	10k	[1] TOMARADZE	15	CLEO	$e^+e^- \to$ hadrons
142.2 ±0.3 ±0.2	145	ALBRECHT	95F	ARG	$e^+e^- \to$ hadrons
142.12 ±0.05 ±0.05	1176	BORTOLETTO	92B	CLE2	$e^+e^- \to$ hadrons
• • • We do not use the following data for averages, fits, limits, etc. • • •					
142.2 ±2.0		SADROZINSKI	80	CBAL	$D^{*0} \to D^0\pi^0$
142.7 ±1.7		[2] GOLDHABER	77	MRK1	e^+e^-

[1] Obtained by analyzing CLEO-c data but not authored by the CLEO Collaboration . This value comes from the average of the results for two decay modes, $D^0 \to K^-\pi^+$ and $D^0 \to K^-\pi^+\pi^-\pi^+$.

[2] From simultaneous fit to $D^*(2010)^+$, $D^*(2007)^0$, D^+, and D^0.

66 Years of PDG

- **PDG started in 1957 with journal article by Rosenfeld and Gell-Mann**
  - Followed by wallet card

Barkas and Rosenfeld UCRL-8030 Table I

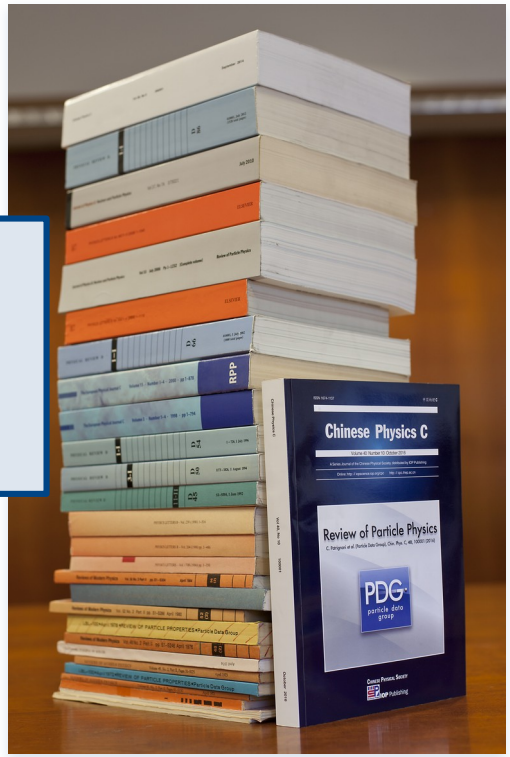
Masses and mean lives of elementary particles; November, 1957
(The antiparticles are assumed to have the same spins, masses, and mean lives as the particles listed)

	Particle	Spin	Mass (Errors represent standard deviation) (Mev)		Mass difference (Mev)		Mean life (sec)		
Photon	γ	1	0				stable		
Leptons	ν	$\frac{1}{2}$	0				stable		
	e^-	$\frac{1}{2}$	0.510976	(a)			stable		
	μ^-	$\frac{1}{2}$	105.70 ±0.06	(a)			(2.22 ±0.02) ×10^{-6}		
Mesons	π^+	0	139.63 ±0.06	(a)	4.6 (a)		(2.56 ±0.05) ×10^{-8}		
	π^0	0	135.04 ±0.16	(a)			<4 ×10^{-15}		
	K^+	0	494.0 ±0.2	(g)	0.4±1.8		(1.224±0.013)×10^{-8}		
	K^0	0	494.4 ±1.8	(i)		K_1:	(0.95 ±0.08) ×10^{-10}		
						K_2:	(4< τ <13) ×10^{-8}		
Baryons	p	$\frac{1}{2}$	938.213 ±0.01	(a)			stable		0.0
	n	$\frac{1}{2}$	939.506 ±0.01	(a)			(1.04 ±0.13) ×10^{+3}	(a)	0.96 ×10^{-3}
	Λ	$\frac{1}{2}$	1115.2 ±0.14	(j)			(2.77 ±0.15) ×10^{-10}	(k)	0.36 ×10^{10}
	Σ^+	$\frac{1}{2}$	1189.4 ±0.25	(l)	7.1 ± 0.4		(0.83 $^{+.06}_{-.05}$) ×10^{-10}	(m)	1.21 ×10^{10}
	Σ^-	$\frac{1}{2}$	1196.5 ±0.5	(n)			(1.67 ±0.17) ×10^{-10}	(o)	0.60 ×10^{10}
	Σ^0	$\frac{1}{2}$	1190.5 $^{+0.9}_{-1.4}$	(p)	6.0 $^{+1.4}_{-0.9}$		(<0.1) ×10^{-10} theoretically ~10^{-19}	(b)	>10 ×10^{10} theoretically ~10^{19}
	Ξ^-	?	1320.4 ±2.2	(q)			(4.6< τ <200) ×10^{-10}	(f)	(>0.005, <0.2) ×10^{10}
	Ξ^0	?	?				?		

Table IV

For much of PDG's history, focus has (naturally) been on printed materials

- **PDG Book and Booklet as staples of particle physics**



Today's *Review of Particle Physics*

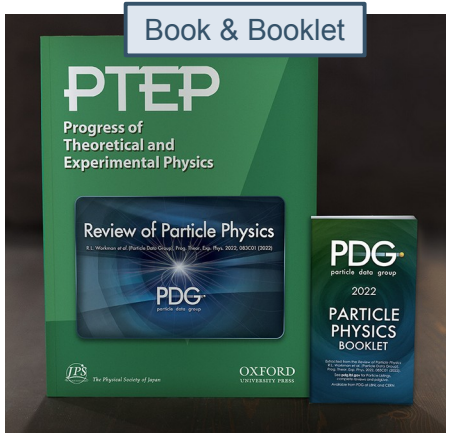
2,270 page open-access journal article (2022 ed.)

- 120 review articles
- 47k measurements from 13k publications
- One of/the most cited publication in HEP

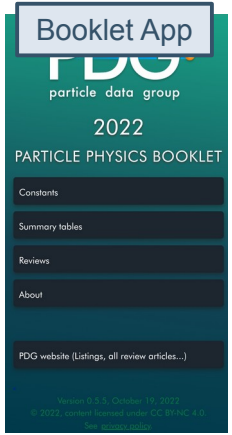
PDG website

pdgLive

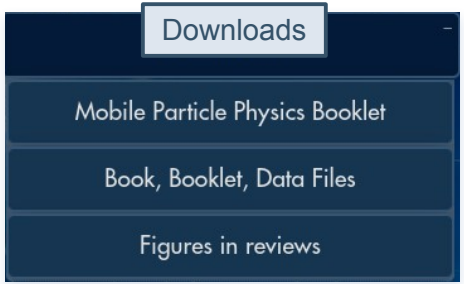
Book & Booklet



Booklet App

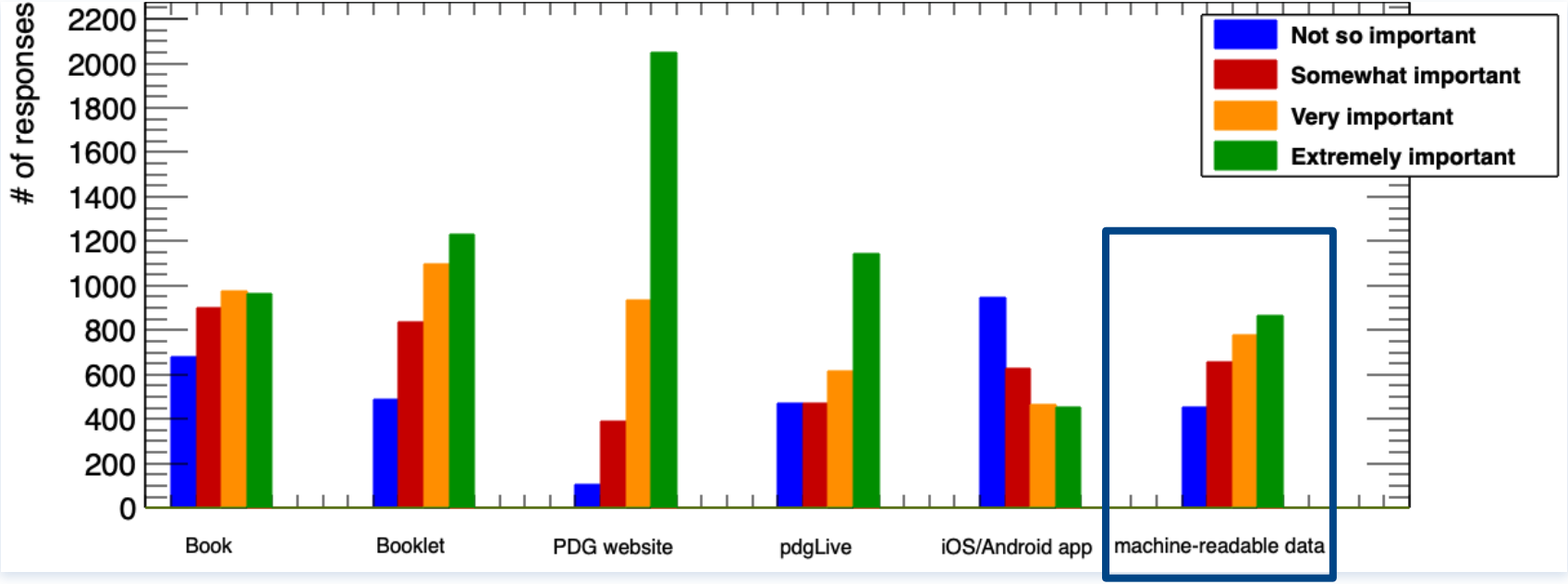


Downloads



What Does Community Use and Need?

- **Comprehensive user survey in 2022**
  - Widely advertised (online, PDG address list, advertised by many collaborations)
  - 4,231 responses, over 2,500 comments (135 pages)



Which Data Should be Made Machine-Readable?

- **Some PDG data has been available as fixed-format text files for very long time**
  - Particle charges, masses, and widths (available since at least 1994!)
  - Cross-section data and related quantities
  - Data from interactive version of Atomic and Nuclear Properties of Materials
- **Current focus is on making core PDG data fully available in modern way**
  - Summary Tables (**including branching fractions**)
  - Particle Listings, including
    - Measurements and limits
    - Comments, footnotes, references
    - Fit information

FAIR principles

- Findable (digital identifiers)
- Accessible (standard protocol)
- Interoperable (shared language for knowledge representations)
- Reusable (metadata, license)

- **Eventually also**
  - Data from tables in certain PDG review articles (Collider Parameters, Physical Constants, ...)
  - Data from figures in PDG review articles
  - ...

Some Use Cases

- **Partly or fully automated update of PDG data in HEP software for simulation, data analysis, or reconstruction**
  - Example: update of EvtGen decay tables
- **Simplify making calculations or figures based on PDG data**
  - What particles decay into a certain combination of particles?
- **Download PDG data to do custom averages/fits**
  - What if the PDG average did (not) include a certain measurement?
- **Enable machine-learning applications of PDG data**
  - Automatic document classification based on PDG Identifiers and their metadata
- **Meta-analysis of particle physics data**
  - Can the impact of blind analyses on statistical properties of averages (avoiding bias, ...) be quantified?
- **Full-featured pdgLive-like smartphone app that works offline**
- **…**

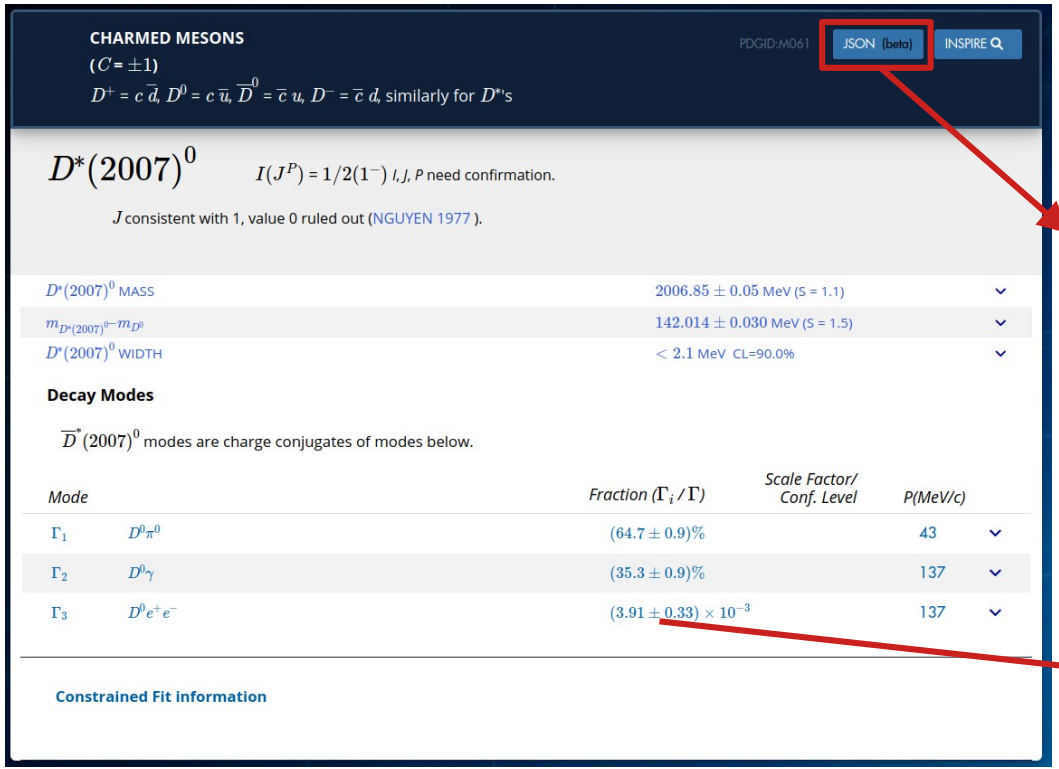
New PDG API with Three Tools

Developing three closely related tools, aimed at different use cases

- **REST API**
  - Download JSON data directly from pdgLive
  - Can also be used in scripts/programs
  - Intended for incidental, rate-limited use
- **Python API**
  - High-level API for programmatic access to PDG data
  - Includes local data store
- **Database files**
  - SQLite files with part of or whole PDG dataset
  - Aimed primarily at software developers

Get Data in JSON Format from pdgLive

From D*(2007)⁰ page in pdgLive



CHARMED MESONS

($C = \pm 1$)

$D^+ = c\,\overline{d}$, $D^0 = c\,\overline{u}$, $\overline{D}^0 = \overline{c}\,u$, $D^- = \overline{c}\,d$, similarly for D^*'s

PDGID:M061 JSON (beta) INSPIRE

$D^*(2007)^0$ $I(J^P) = 1/2(1^-)$ *I, J, P* need confirmation.

J consistent with 1, value 0 ruled out (NGUYEN 1977).

$D^*(2007)^0$ MASS	2006.85 ± 0.05 MeV (S = 1.1)
$m_{D^*(2007)^0} - m_{D^0}$	142.014 ± 0.030 MeV (S = 1.5)
$D^*(2007)^0$ WIDTH	< 2.1 MeV CL=90.0%

Decay Modes

$\overline{D}^*(2007)^0$ modes are charge conjugates of modes below.

Mode		Fraction (Γ_i / Γ)	Scale Factor/ Conf. Level	P(MeV/c)
Γ_1	$D^0\pi^0$	$(64.7 \pm 0.9)\%$		43
Γ_2	$D^0\gamma$	$(35.3 \pm 0.9)\%$		137
Γ_3	$D^0 e^+ e^-$	$(3.91 \pm 0.33) \times 10^{-3}$		137

Constrained Fit information


```
status_code: 200
status_message: "OK"
request_timestamp: "2023-05-27 17:30:35 PST"
request_url: "https://pdgapi.lbl.gov/summaries/M061"
edition: "2023"
about: "For further information see https://pdg.lbl.gov/api"
pdgid: "M061"
description: "D^*(2007)0"
summaries:
  properties:
    0: {…}
    1: {…}
    2: {…}
  branching_fractions:
    0: {…}
    1: {…}
    2:
      pdgid: "M061.3"
      description: "D^*(2007)0  --> D0 e+ e-"
      mode_number: 3
      pdg_values:
        0:
          value: 0.003914359217882152
          error_positive: 0.0003346667716125103
          error_negative: 0.0003346667716125103
          value_text: "(3.91+-0.33)E-3"
          type: "OUR AVERAGE"
```

REST API

- JSON (beta) **buttons in pdgLive provide links to a REST API**
  - REST – Representational State Transfer
    - Typical architecture for web-based APIs (Application Programming Interface)
  - Provides URLs to download PDG data in machine-readable format
    - JSON files
  - Implemented at https://pdgapi.lbl.gov
- **General format: https://pdgapi.lbl.gov/PATH**

Path	Example	Description
/info	https://pdgapi.lbl.gov/info	Get metadata (edition, citation, version, license)
/summaries/PDGID	https://pdgapi.lbl.gov/summaries/S126M	Get summary data for PDG Identifiers
/summaries/PDGID/EDITION	https://pdgapi.lbl.gov/summaries/S126M/2020	Get summary data from an earlier edition

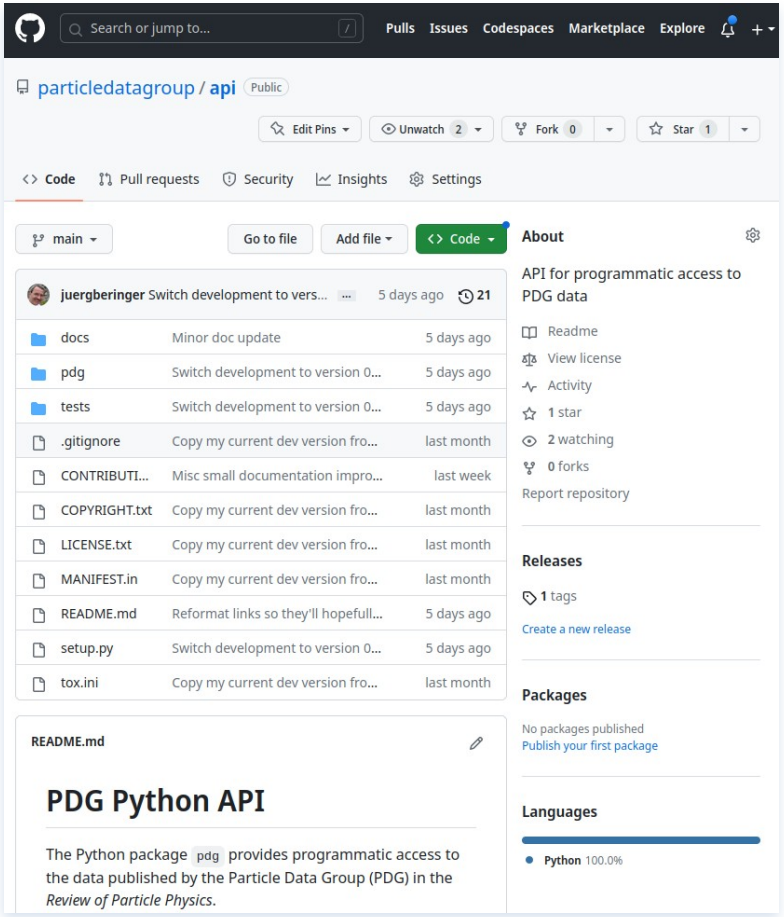
PDG Identifiers

- **Digital object identifiers**
  - Case-insensitive, alphanumeric strings
    - First 4 alphanumeric characters typically denote particle; additional characters for properties
    - STRING.NUMBER for branching fractions
  - Unambiguous way to refer to physics quantities in a program
    - Example: **M061** → $D^*(2007)^0$
    - Example: **S042.214** → $B^0 \to J/\Psi(1S)K^*(892)^0\pi^+\pi^-$
  - In most cases, NOT expected to know/use PDG Identifiers directly
    - Can use pdgLive to look up PDG Identifiers:

- **Currently about 19,000 PDG Identifiers**
  - Used by all PDG systems as well as for cross-linking with INSPIRE
- **Might be embedded into ISO 26324 DOI system**
  - E.g. 10.NNNN/S042.214 (NNNN = future unique number for PDG)
  - Benefit: improved public visibility/searchability, standardization

Python API

- **Implemented in Python package pdg**
  - Supports Python 3 (and for now also 2.7)
  - Installed like any other package
    - For example: **python -m pip install pdg**
  - Released as open-source software
    - https://github.com/particledatagroup/api
- **Comes with local data source**
  - PDG database file containing data current at time of release



Example: Particle Properties

General usage pattern

- **Import pdg package**
- **Connect to data source**
  - By default uses included database file
  - Can download/use other PDG database files (more data or different edition)
- **Access desired particle or property**
  - By Monte Carlo number
  - By particle name
  - By PDG Identifier
  - By iterating over particles/properties

```
>>> import pdg
>>> api = pdg.connect()
>>>
>>> api.get_particle_by_mcid(423)
PdgParticle('M061/2023')
>>>
>>> api.get_particle_by_name('D^*(2007)0')
PdgParticle('M061/2023')
>>>
>>> api.get('M061')
PdgParticle('M061/2023')
>>>
>>> for p in api.get_particles():
...     print(p.name)
...
gamma
g
graviton
W+
Z0
H
e-
mu-
tau-
u
```

Example: B^0 Exclusive Branching Fractions

```
Python 3.10.10 (main, Mar 01 2023, 21:10:14) [GCC] on linux
Type "help", "copyright", "credits" or "license" for more information.
>>>
>>> import pdg
>>> api = pdg.connect()
>>>
>>> for bf in api.get_particle_by_name('B0').exclusive_branching_fractions():
...     if not bf.is_limit:
...         print('%-40s %s' % (bf.description, bf.value))
...
B0 --> D- pi+                                0.0025060902816920 1
B0 --> D- rho+                               0.007574637658365 27
B0 --> D- K^*(892)+                          0.0004457772020725 39
B0 --> D- omega pi+                          0.0028
B0 --> D- K+                                 0.0002051413814562 25
B0 --> D- K+ pi+ pi-                         0.0003530128447741 61
B0 --> D- K+ Kbar^*(892)0                    0.00088
B0 --> Dbar0 pi+ pi-                         0.0008802598590904 93
```

Ambiguities and Pedantic Mode

Mass of the top quark?

t

$I(J^P) = 0(\frac{1}{2}^+)$

Charge $= \frac{2}{3}\ e$ Top $= +1$

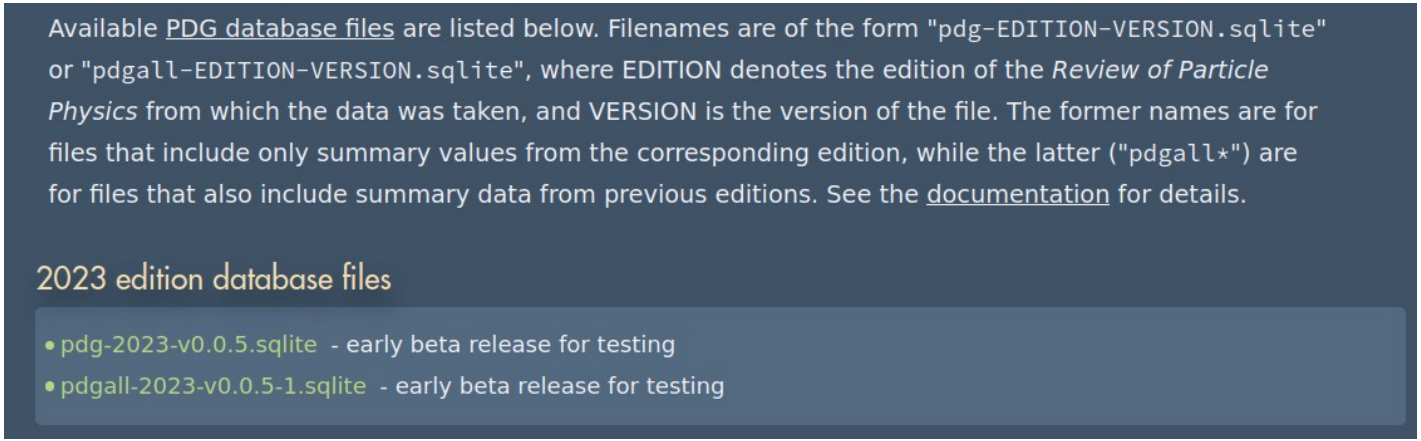
Mass (direct measurements) $m = 172.69 \pm 0.30$ GeV [a,b] (S = 1.3)
Mass (from cross-section measurements) $m = 162.5^{+2.1}_{-1.5}$ GeV [a]
Mass (Pole from cross-section measurements) $m = 172.5 \pm 0.7$ GeV

```
>>> [m.display_value_text for m in pdg.connect().get_particle_by_name('t').masses()]
['172.69+-0.30', '162.5+2.1-1.5', '172.5+-0.7']
>>>
>>> pdg.connect().get_particle_by_name('t').mass
172.687433377613
>>>
>>> pdg.connect(pedantic=True).get_particle_by_name('t').mass
Traceback (most recent call last):
  File "<stdin>", line 1, in <module>
  File "/home/jb/PDG/work/PyApiPub/api/pdg/particle.py", line 237, in mass
    best_mass_property = best(self.masses(), self.api.pedantic, '%s mass (%s)' % (self.name, self.pdgid))
  File "/home/jb/PDG/work/PyApiPub/api/pdg/utils.py", line 76, in best
    raise PdgAmbiguousValueError('Ambiguous best property%s' % for_what)
pdg.errors.PdgAmbiguousValueError: Ambiguous best property for t mass (Q007/2023)
```

Default: make assumption about “best” value

Using Other Database Files

- **Python API package includes minimal set of PDG data**
  - Summary Table data for current edition of *Review of Particle Physics*
- **Other database files are available from https://pdg.lbl.gov/api**
  - Database files with data from Particle Listings (future)
  - Database files for a different edition (future)
  - Files including historical Summary Table data (available now)



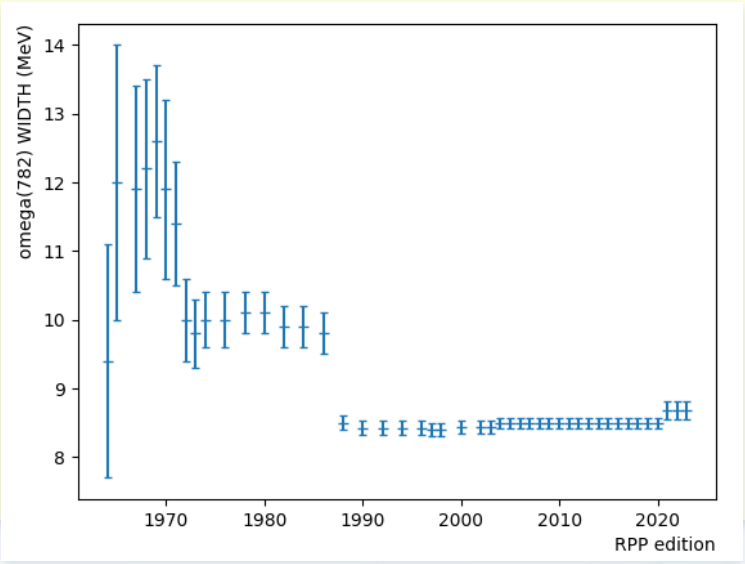
Available PDG database files are listed below. Filenames are of the form "pdg-EDITION-VERSION.sqlite" or "pdgall-EDITION-VERSION.sqlite", where EDITION denotes the edition of the *Review of Particle Physics* from which the data was taken, and VERSION is the version of the file. The former names are for files that include only summary values from the corresponding edition, while the latter ("pdgall*") are for files that also include summary data from previous editions. See the documentation for details.

2023 edition database files

- pdg-2023-v0.0.5.sqlite - early beta release for testing
- pdgall-2023-v0.0.5-1.sqlite - early beta release for testing

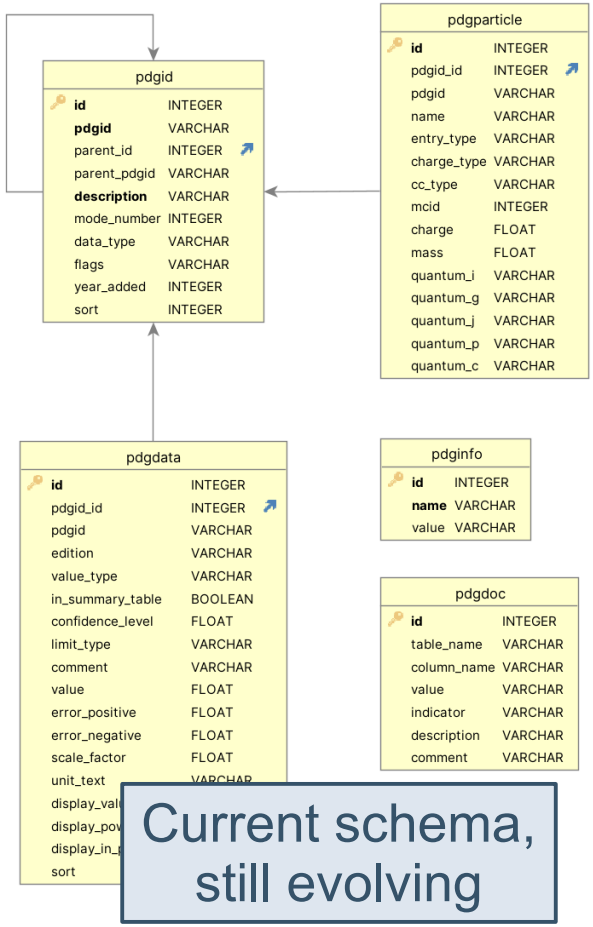
Example: Make Your Own History Plot

```
>>> pdg.connect().editions
['2023']
>>>
>>> api = pdg.connect('sqlite:///pdgall-2023-v0.0.5-1.sqlite')
>>> api.editions
['2023', '2022', '2021', '2020', '2019', '2018', '2017', '2016', '2015', '2014', '2013', '2012', '2011',
 '2010', '2009', '2008', '2007', '2006', '2005', '2004', '2003', '2002', '2000', '1998', '1997', '1996',
 '1994', '1992', '1990', '1988', '1986', '1984', '1982', '1980', '1978', '1977', '1976', '1974', '1973',
 '1972', '1971', '1970', '1969', '1968', '1967', '1966', '1965', '1964', '1963', '1961', '1958']
>>>
>>> years, values, errors = list(), list(), list()
>>>
>>> quantity = api.get('M001W')
>>>
>>> for year in range(1958, 2024):
...     try:
...         quantity.edition = year
...         values.append(quantity.value)
...         errors.append(quantity.error)
...         years.append(year)
...     except:
...         pass
```



Database Files

- **All (or subset of) PDG data in a single file in SQLite format**
  - SQLite is very widely used “relational database in a single file”
- **Aimed at software developers wishing to include PDG data in their applications without using Python API**
- **Low-level access to PDG data**
  - Queried via SQL
  - Must understand special cases in PDG data
  - Source of Python API can serve as guide



Status

- **Beta release available for testing and to get feedback**
  - Includes access to most data in PDG Summary Tables via REST API, Python API, and database files
  - In particular, access to PDG branching fraction data in machine-readable format
- **It's a beta release ...**
  - There may be bugs, and some data may be incomplete or wrong
  - There are features missing that obviously should be there (and will be added)
- **Ongoing development to**
  - Improve and extend all tools, **improve metadata**, better handle special cases
  - Add access to data in Particle Listings
  - ...

Extending Metadata

- **For most of its existence, PDG produced summaries to be read by humans**
  - Meaning of ℓ, interpretation of footnotes, indenting, and more is intuitively understood or “well known” by community
- **For programmatic access, all this needs to be modeled by metadata**
  - Must first be added to PDG production database
  - Need tools to maintain/update metadata
  - Then needs to be properly and intuitively interpreted by e.g. the Python API

B^0 DECAY MODES		Fraction (Γ_i/Γ)
$\ell^+ \nu_\ell X$	[a]	(10.33 ± 0.28) %
$e^+ \nu_e X_c$		(10.1 ± 0.4) %
$D \ell^+ \nu_\ell X$		(9.4 ± 0.9) %
$D^- \ell^+ \nu_\ell$	[a]	(2.31 ± 0.10) %
$D^- \tau^+ \nu_\tau$		(1.08 ± 0.23) %
$D^*(2010)^- \ell^+ \nu_\ell$	[a]	(5.06 ± 0.12) %
$D^*(2010)^- \tau^+ \nu_\tau$		(1.57 ± 0.09) %

[a] An ℓ indicates an e or a μ mode, not a sum over these modes.

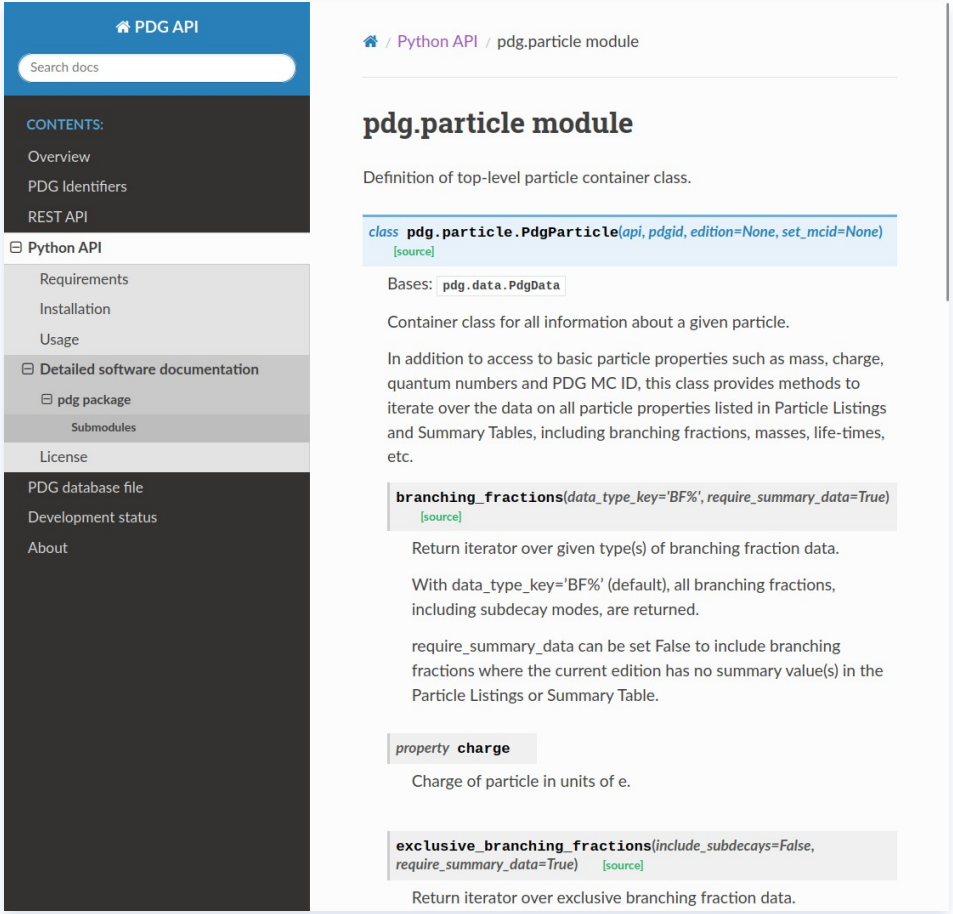
This is what requires most of the work for making PDG data available in machine-readable format

How To Get Started with Present Version

See pdg.lbl.gov/api to get started

- **Links to documentation**
- **Downloadable database files**

Please send feedback, suggestions, and bug reports go api@pdg.lbl.gov



Conclusions

- **New PDG API consists of three complementary tools to access PDG data in machine-readable format, addressing different use cases**
  - REST API with direct access to JSON data in pdgLive
  - Python API
  - Database files
- **This is a new PDG product**
  - Expected to be supported in the long-term
- **First beta release available for testing**
  - Provides, for the first time, programmatic access to PDG branching fraction data
- **Development is ongoing**
- **We look forward to your feedback**
  - Contact api@pdg.lbl.gov

Extra Slides

Complete Code for History Example

```
import numpy as np, matplotlib.pyplot as plt
import pdg

pdg.connect().editions

api = pdg.connect('sqlite:///pdgall-2023-v0.0.5-1.sqlite')
api.editions

years, values, errors = list(), list(), list()

quantity = api.get('M001W')

for year in range(1958, 2024):
    try:
        quantity.edition = year
        values.append(quantity.value)
        errors.append(quantity.error)
        years.append(year)
    except:
        pass

plt.errorbar(years, values, errors,marker='_', capsize=2, linestyle='')
plt.xlabel('RPP edition', loc='right')
plt.ylabel('%s (%s)' % (quantity.description, quantity.units), loc='top')
plt.show()
```

From Production Database to PDG Value

- **PDG production database aimed at producing and updating Particle Listings and Summary Table values**
  - In many cases does not store values directly
  - Instead stores relationships between measurements used in PDG averages and fits
  - Done using “PDG macros”
- **Tools to make PDG data available in machine-readable format must do this evaluation**
  - Much more involved than just copying values from the production database into a downloadable database file

From a LHCb paper

align with expectations from quark-helicity conservation in B decays. The ratio of branching fractions $[\mathcal{B}(B^0 \to D^{*-} D_s^{*+}) \times \mathcal{B}(D_s^{*+} \to D_s^+ \gamma)]/\mathcal{B}(B^0 \to D^{*-} D_s^+)$ is measured to be $2.045 \pm 0.022 \pm 0.071$ with world-best precision. In ad-

(Part of) entry in production database

```
pdgprod=# select node, reference_id, measurement, place from measurement where node='S042P66';
  node   | reference_id |                    measurement                    | place
---------+--------------+---------------------------------------------------+-------
 S042P66 |        61307 | br_adjust: 2.045 +-0.022 +-0.071; *, ADJUST, S074 1 | U
(1 row)
```

Entry in Particle Listings

$\Gamma(D^*(2010)^- D_s^{*+})/\Gamma(D^*(2010)^- D_s^+)$ Γ_{85}/Γ_{83}

VALUE	DOCUMENT ID		TECN	COMMENT
2.19±0.08±0.02	[1] AAIJ	21S	LHCB	pp at 13 TeV

[1] AAIJ 21S reports $[\Gamma(B^0 \to D^*(2010)^- D_s^{*+})/\Gamma(B^0 \to D^*(2010)^- D_s^+)] \times [\mathrm{B}(D_s^{*+} \to D_s^+ \gamma)] = 2.045 \pm 0.022 \pm 0.071$ which we divide by our best value $\mathrm{B}(D_s^{*+} \to D_s^+ \gamma) = (93.5 \pm 0.7) \times 10^{-2}$. Our first error is their experiment's error and our second error is the systematic error from using our best value.