New Ways to Access PDG Data

Juerg Beringer
Physics Division
Lawrence Berkeley National Laboratory

For the PDG Collaboration

ICHEP 2024 Conference
Prague, July 17 - 24, 2024

PDG and the *Review of Particle Physics*

- **The Particle Data Group (PDG) provides a comprehensive summary of particle physics and related areas from cosmology and astrophysics in a single publication, the *Review of Particle Physics***
  - International collaboration of 240 scientists, 173 institutions, 25 countries
  - Berkeley Lab provides scientific leadership, central coordination, and technical expertise & infrastructure
  - Funded primarily by the DOE Office of High Energy Physics with important contributions from MEXT (Japan), INFN, and CERN

- **The *Review of Particle Physics* consists of**
  - **Summary Tables**
  - **Particle Listings**
  - **120 individual review articles**, covering a wider range of topics
  - Available online at pdg.lbl.gov and pdgLive.lbl.gov, as well as in print (PDG Book and Booklet), as an Android/web app, and downloadable files

- **Journal publication every 2 years (updated online every year)**
  - This year's journal publication to appear in Phys. Rev. D
  - 2,400 pages

PDG particle data group

Summary Tables and Particle Listings

BERKELEY LAB

Summary Tables

- PDG world averages (or best limits)
- Masses, widths/lifetimes, branching fractions, ...

Particle Listings

- Detailed information on how PDG arrived at any given average
- Curated and annotated lists of relevant published results

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

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)	Confidence level	p (MeV/c)
$D^0\pi^0$	(64.7 ± 0.9) %		43
$D^0\gamma$	(35.3 ± 0.9) %		137
$D^0 e^+ e^-$	$(3.91 \pm 0.33) \times 10^{-3}$		137
$\mu^+\mu^-$	$< 2.5 \times 10^{-8}$	90%	998
$e^+ e^-$	$< 1.7 \times 10^{-6}$	90%	1003

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

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.

120 Review Articles

Covering wide range of topics, including

- Constants, units, atomic & nuclear properties
- Standard Model and related topics
- Astrophysics and cosmology
- Experimental methods and colliders
- Mathematical tools
- Kinematics, cross-section formulae and plots
- Particle properties
- Hypothetical particles and concepts

Sample page from PDG review
"Passage of Particles Through Matter"

As shown in Fig. 34.11, stopping powers for e^-, e^+, and heavy particles are not dramatically different. In silicon, the minimum value for electrons is 1.50 MeV cm^2/g (at $\gamma = 3.3$); for positrons, 1.46 MeV cm^2/g (at $\gamma = 3.7$), and for muons, 1.66 MeV cm^2/g (at $\gamma = 3.58$).

34.4.2 *Radiation length*

High-energy electrons predominantly lose energy in matter by bremsstrahlung, and high-energy photons by e^+e^- pair production. The characteristic amount of matter traversed for these related interactions is called the radiation length X_0, usually measured in g cm^{-2}. It is the mean distance over which a high-energy electron loses all but $1/e$ of its energy by bremsstrahlung. It is also the appropriate scale length for describing high-energy electromagnetic cascades. X_0 has been calculated and tabulated by Y.S. Tsai [42]:

$$\frac{1}{X_0} = 4\alpha r_e^2 \frac{N_A}{A}\left\{Z^2\left[L_{\rm rad} - f(Z)\right] + Z\,L'_{\rm rad}\right\}. \qquad (34.25)$$

For $A = 1$ g mol^{-1}, $4\alpha r_e^2 N_A/A = (716.408\ {\rm g\ cm^{-2}})^{-1}$. $L_{\rm rad}$ and $L'_{\rm rad}$ are given in Table 34.2. The function $f(Z)$ is an infinite sum, but for elements up to uranium can be represented to 4-place accuracy by

$$f(Z) = a^2\left[(1+a^2)^{-1} + 0.20206 - 0.0369\,a^2 + 0.0083\,a^4 - 0.002\,a^6\right], \qquad (34.26)$$

where $a = \alpha Z$ [43].

Table 34.2: Tsai's $L_{\rm rad}$ and $L'_{\rm rad}$, for use in calculating the radiation length in an element using Eq. (34.25).

Element	Z	$L_{\rm rad}$	$L'_{\rm rad}$
H	1	5.31	6.144
He	2	4.79	5.621
Li	3	4.74	5.805
Be	4	4.71	5.924
Others	>4	$\ln(184.15\,Z^{-1/3})$	$\ln(1194\,Z^{-2/3})$

The radiation length in a mixture or compound may be approximated by

$$1/X_0 = \sum w_j/X_j, \qquad (34.27)$$

where w_j and X_j are the fraction by weight and the radiation length for the jth element.

34.4.3 *Bremsstrahlung energy loss by $e^\pm$*

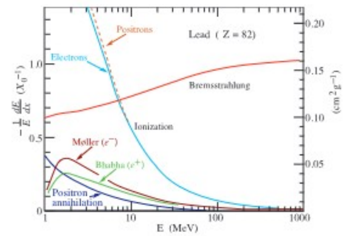


Figure 34.11: Fractional energy loss per radiation length in lead as a function of electron or positron energy. Electron (positron) scattering is considered as ionization when the energy loss per collision is below 0.255 MeV, and as Møller (Bhabha) scattering when it is above. Adapted from Fig. 3.2 from Messel and Crawford, *Electron-Photon Shower Distribution Function Tables for Lead, Copper, and Air Absorbers*, Pergamon Press, 1970. Messel and Crawford use $X_0({\rm Pb}) = 5.82\ {\rm g/cm^2}$, but we have modified the figures to reflect the value given in the Table of Atomic and Nuclear Properties of Materials ($X_0({\rm Pb}) = 6.37\ {\rm g/cm^2}$).

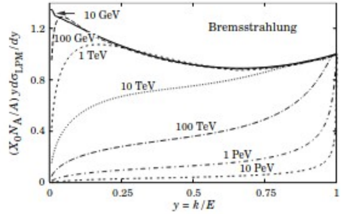


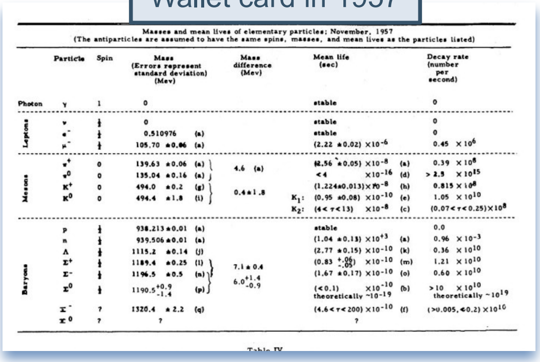
Figure 34.12: The normalized bremsstrahlung cross section $k\,d\sigma_{LPM}/dk$ in lead versus the fractional photon energy $y = k/E$. The vertical axis has units of photons per radiation length.

nearby scattering centers (the LPM effect) [44, 45] and dielectric suppression [46, 47]. These and other suppression effects in bulk media are discussed in Sec. 34.4.6.

With decreasing energy ($E \lesssim 10$ GeV) the high-y cross section drops and the curves become rounded as $y \to 1$. Curves of this familar shape can be seen in Rossi [2] (Figs. 2.11.2,3); see also the

Distribution of the *Review of Particle Physics*

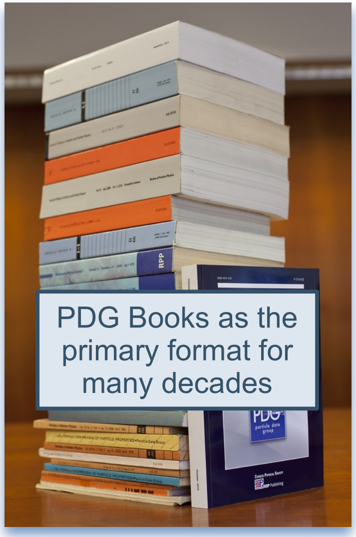
Wallet card in 1957



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)

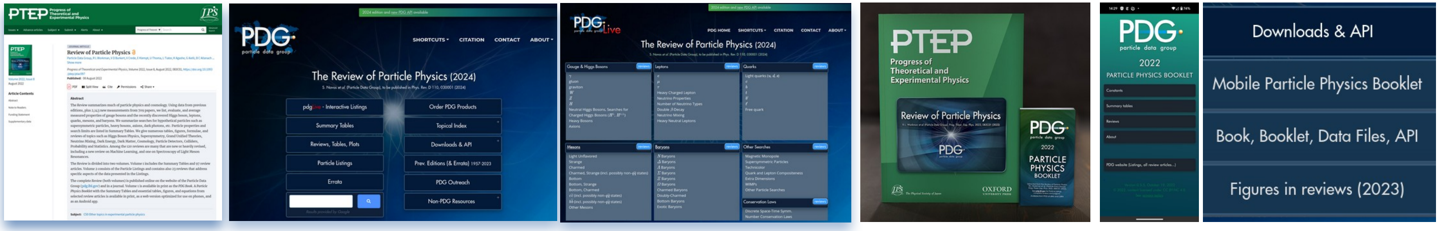
	Particle	Spin	Mass (Errors represent standard deviation) (Mev)	Mass difference (Mev)	Mean life (sec)	Decay rate (number per second)
Photon	γ	1	0		stable	0
Leptons	ν	½	0		stable	0
	e^-	½	0.510976 (a)		stable	0
	μ^-	½	105.70 ±0.06 (a)		(2.22 ±0.02) $\times 10^{-6}$	0.45 $\times 10^{6}$
Mesons	π^+	0	139.63 ±0.06 (a)	4.6 (a)	(2.56 ±0.05) $\times 10^{-8}$ (a)	0.39 $\times 10^{8}$
	π^0	0	135.04 ±0.16 (a)		< 4 $\times 10^{-16}$ (d)	> 2.5 $\times 10^{15}$
	K^+	0	494.0 ±0.2 (g)	0.4±1.8	(1.224±0.013) $\times 10^{-8}$ (h)	0.815 $\times 10^{8}$
	K^0	0	494.4 ±1.8 (i)		K_1: (0.95 ±0.08) $\times 10^{-10}$ (e)	1.05 $\times 10^{10}$
					K_2: (4< τ <13) $\times 10^{-8}$ (c)	(0.07< τ <0.25) $\times 10^{8}$
Baryons	p	½	938.213 ±0.01 (a)		stable	0.0
	n	½	939.506 ±0.01 (a)		(1.04 ±0.13) $\times 10^{+3}$ (a)	0.96 $\times 10^{-3}$
	Λ	½	1115.2 ±0.14 (j)		(2.77 ±0.15) $\times 10^{-10}$ (k)	0.36 $\times 10^{10}$
	Σ^+	½	1189.4 ±0.25 (l)	7.1 ± 0.4	(0.83 $^{+.06}_{-.05}$) $\times 10^{-10}$ (m)	1.21 $\times 10^{10}$
	Σ^-	½	1196.5 ±0.5 (n)	$6.0^{+1.4}_{-0.9}$	(1.67 ±0.17) $\times 10^{-10}$ (o)	0.60 $\times 10^{10}$
	Σ^0	½	$1190.5^{+0.9}_{-1.4}$ (p)		(<0.1) $\times 10^{-10}$ (b) theoretically ~10^{-19}	>10 $\times 10^{10}$ theoretically ~10^{19}
	Ξ^-	?	1320.4 ± 2.2 (q)		(4.6< τ <200) $\times 10^{-10}$ (f)	(>0.005, <0.2) $\times 10^{10}$
	Ξ^0	?	?		?	

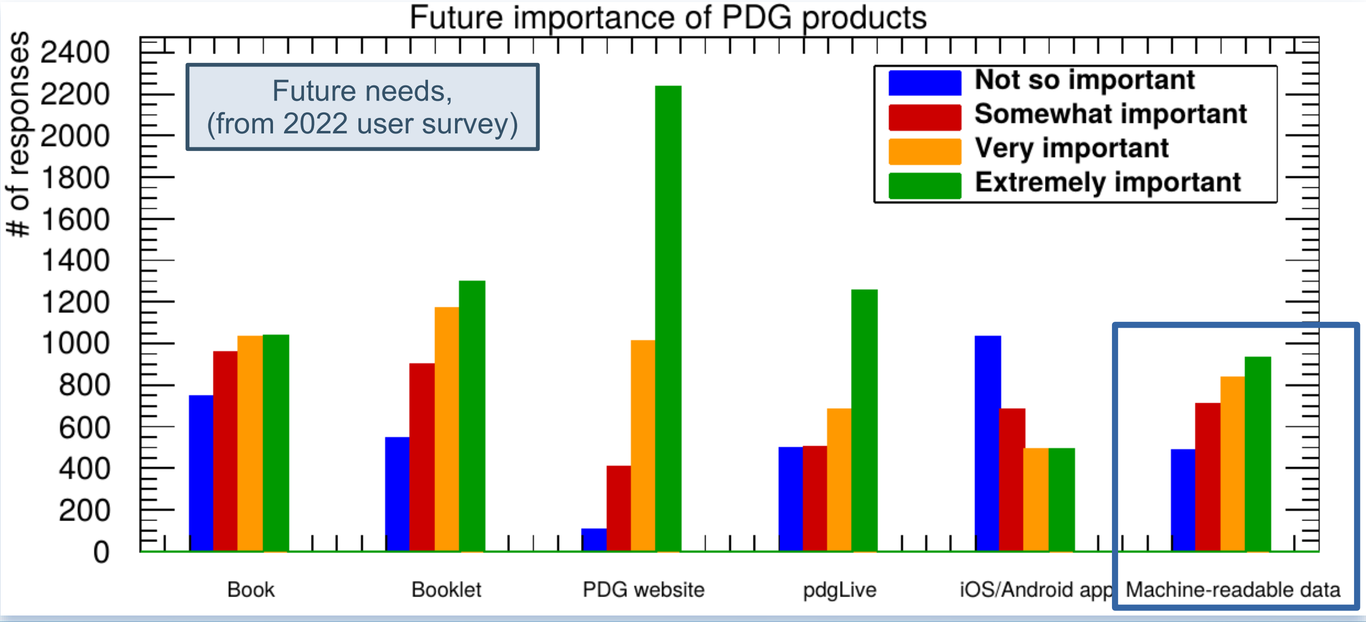
Table IV



PDG Books as the primary format for many decades

Available in many formats today





New PDG API with Three Tools

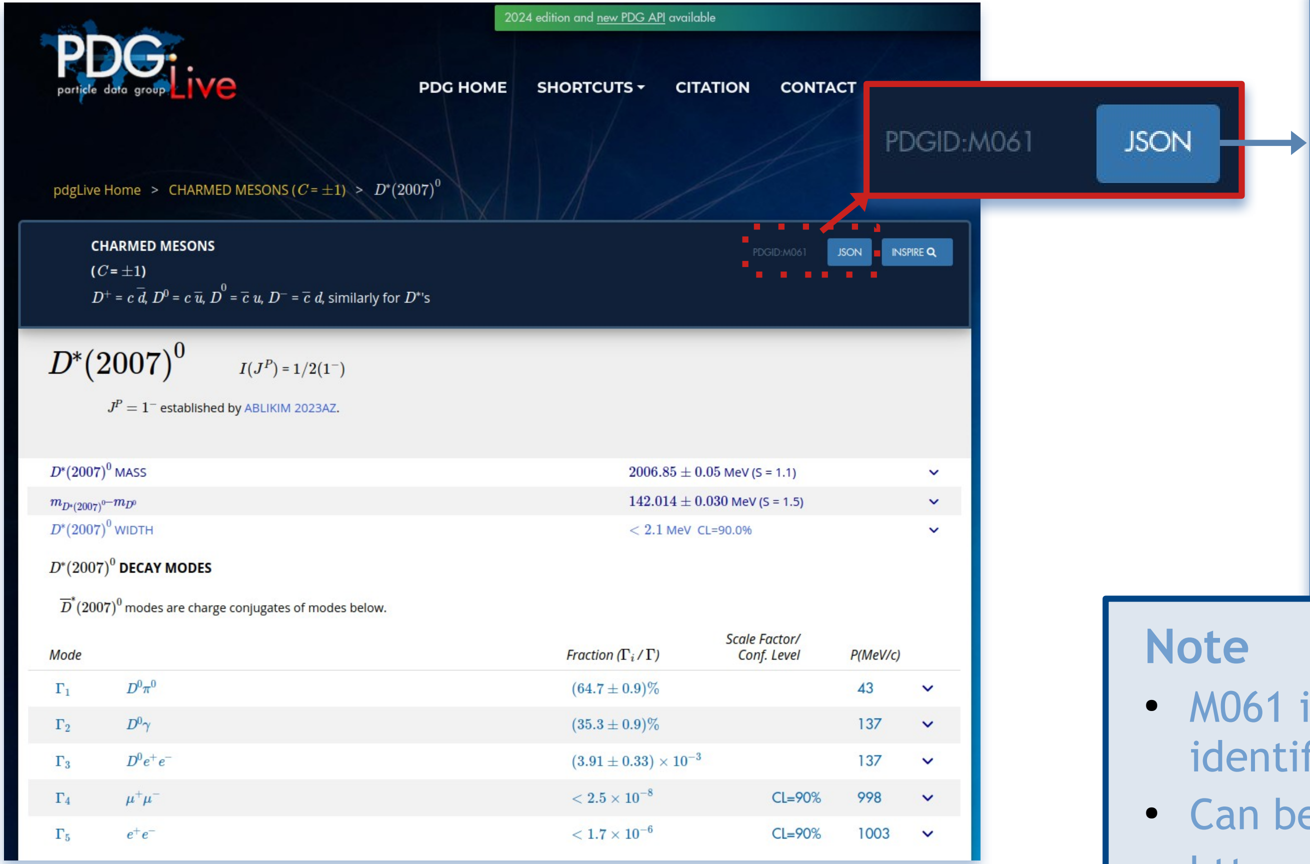
To make PDG data available in machine-readable format, PDG has been developing a new API with three related tools, aimed at different use cases

- **REST API**
  - Download data shown by pdgLive in JSON format
  - Can also be used in scripts/programs w/o installing any PDG-specific software
  - Intended for incidental, rate-limited use
- **Python API**
  - High-level API for accessing PDG data
  - Includes local data store (SQLite database file)
- **Database files**
  - SQLite files with part of or whole PDG dataset
  - Low-level access for software developers

Status

- First production release with 2024 online publication (beta version since June 2023)
- Access to PDG world averages and limits (including branching fractions)
- Access to measurements from Particle Listings coming soon
- All tools will be updated with each new edition of the *Review of Particle Physics*

REST API – Get Data in JSON Format



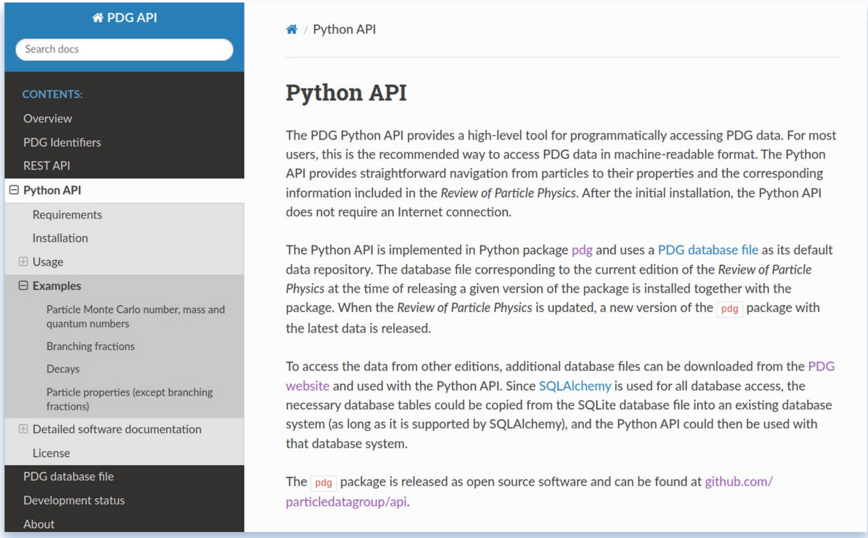

```
{
  "status_code": 200,
  "status_message": "OK",
  "request_timestamp": "2024-07-10 15:48:09 PST",
  "request_url": "https://pdgapi.lbl.gov/summaries/M061",
  "edition": "2024",
  "about": "For further information see https://pdg.lbl.gov/api",
  "pdgid": "M061",
  "description": "D^*(2007)0",
  "summaries": {
    "properties": [
      {
        "pdgid": "M061M",
        "description": "D*(2007)0 MASS",
        "pdg_values": [
          {
            "value": 2006.852502141979,
            "error_positive": 0.05458406870966528,
            "error_negative": 0.05458406870966528,
            "value_text": "2006.85+-0.05",
            "unit": "MeV",
            "scale_factor": 1.127246,
            "type": "OUR FIT"
          }
        ]
      },
      {
        "pdgid": "M061DM",
        "description": "mass(D*(2007)0)-mass(D0)",
```

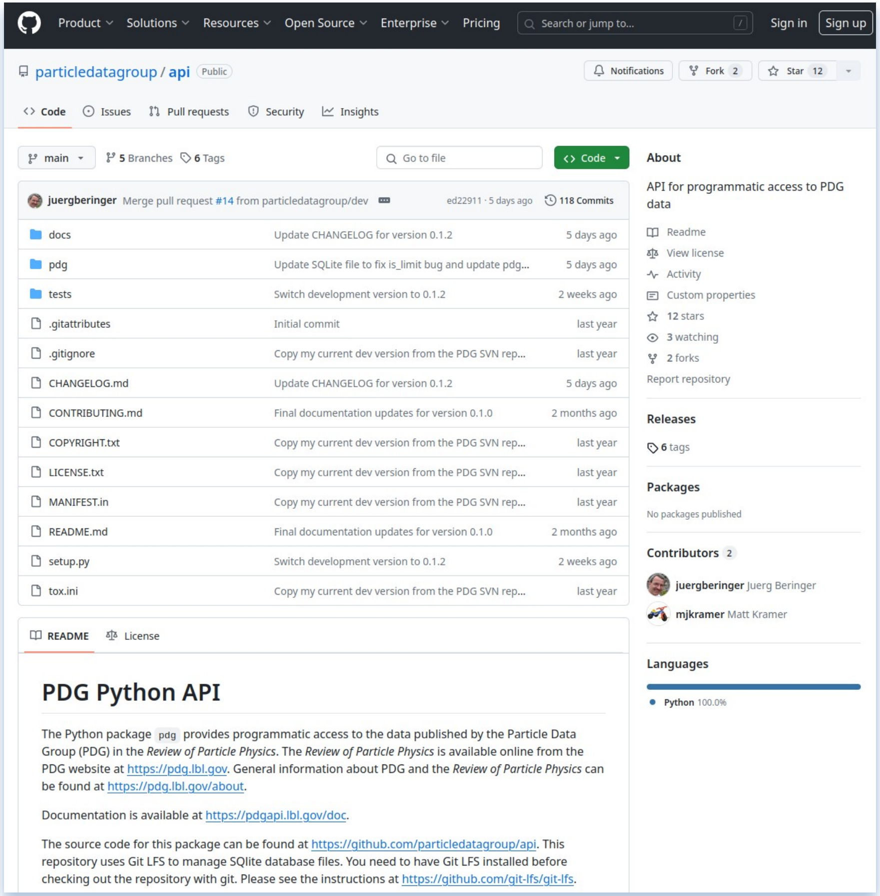

Note

- M061 is the PDG ID (PDG digital object identifier) for $D^*(2007)^0$
- Can be used to retrieve data via REST API: https://pdgapi.lbl.gov/summaries/M061

Python API

- **Implemented in Python package [pdg](pdg)**
  - Installation: `pip install pdg`
  - Supports Python 3 (and for now also 2.7)
  - github.com/particledatagroup/api
- **Documentation at pdgapi.lbl.gov**





Using the Python API

```
[1]: import pdg
     api = pdg.connect()
     print(api)
```

Import package, connect to the default database, and show what we're using

```
2024 Review of Particle Physics, data release 2024-05-31 02:00:13 PDT, API version 0.1.2
S. Navas et al. (Particle Data Group), Phys. Rev. D 110, 030001 (2024)
(C) Particle Data Group (PDG), data released under CC BY 4.0
For further information see https://pdg.lbl.gov/api
```

```
[2]: dstar0 = api.get_particle_by_mcid(423)
     # dstar0 = api.get_particle_by_name('D^*(2007)0')
     # api.get('M061')[0]
```

Get particle data in one of several ways

```
[3]: dstar0.description, dstar0.mcid, dstar0.mass, dstar0.is_meson, dstar0.quantum_I
```

Get some particle properties

```
[3]: ('D^*(2007)0', 423, 2.006852502141979, True, '1/2')
```

```
[4]: for bf in dstar0.branching_fractions():
         print(f'{bf.description:30}', bf.display_value_text)
```

Iterate over and print all branching fractions

```
D^*(2007)0 --> D0 pi0          (64.7+-0.9)%
D^*(2007)0 --> D0 gamma        (35.3+-0.9)%
D^*(2007)0 --> D0 e+ e-        (3.91+-0.33)E-3
D^*(2007)0 --> mu+ mu-         <2.5E-8
D^*(2007)0 --> e+ e-           <1.7E-6
```

See documentation for more features and examples

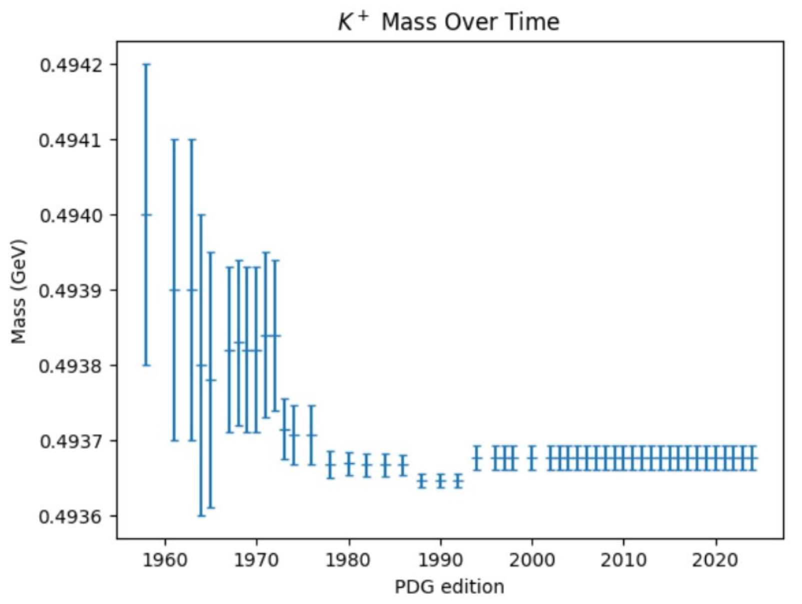
Making a PDG History Plot in Python

```
[1]: import pdg
     import matplotlib.pyplot as plt
     import numpy as np

[2]: # Use a PDG database file with historical data
     api = pdg.connect("sqlite:///pdgall-2024-v0.1.2.sqlite")

[3]: xs, ys, yerrs = [], [], []
     for edition in api.editions:
         p = api.get_particle_by_name('K+', edition=edition)
         if p.has_mass_entry and p.mass is not None:
             xs.append(int(edition))
             ys.append(p.mass)
             yerrs.append(p.mass_error)

     plt.errorbar(xs, ys, yerrs, marker='_', capsize=2, linestyle='')
     plt.xlabel('PDG edition')
     plt.ylabel('Mass (GeV)')
     plt.title('$K^+$ Mass Over Time');
```



Current search (using Google):

SMEFT operators

Results provided by Google

About 4 results (0.08 seconds) Sort by: Relevance

61. Top Quark - Particle Data Group
pdg.lbl.gov › reviews › rpp2024-rev-top-quark
File Format: PDF/Adobe Acrobat
May 31, 2024 ... Model implementations are available for tree-level and even one-loop Monte Carlo simulations. The definitions of the **SMEFT operators** can be ...

55 page PDF file

Tests of Conservation Laws - Particle Data Group
pdg.lbl.gov › reviews › rpp2024-rev-conservation-laws
File Format: PDF/Adobe Acrobat
there exists a tower of **operators** in the **SMEFT** Lagrangian (12), containing only SM fields, that break one or both of these symmetries. We briefly review

19 page PDF file

Can we do better with recent AI tools?

- Paragraph-level results and preview
- Fine-grained references
  - Link to specific page in review's PDF file, rather than to whole file
- Get values from Summary Tables
  - "Value of B0 mass?"
- Summarize information
  - "How is the top quark mass measured?"
- ...

Possible Implementation Blueprint

- **Use Large Language Model (LLM) to determine what user asks for**
  - Specific value of a quantity? Figure or table from a review?
  - General question about a topic covered in PDG review?
- **For PDG reviews ("semantic search")**
  - Split review into small chunks (paragraph or smaller) while keeping association to pages in the PDF file
  - Using embedding model and vector database, represent each chunk and user's query in a high-dimensional vector space
  - Retrieve relevant chunks based on a similarity score
- **For data from Summary Tables or Listings**
  - Use a LLM to determine what quantity (PDG ID) user asks for
  - Then retrieve that data using new PDG API
- **If user wants, use LLM to summarize search results**
  - Only information from PDG is used in the summary
  - Retrieval Augmented Generation (RAG)

Must avoid hallucinations

- Text found via search comes verbatim from PDG review articles
- Values are from Summary Tables and Particle Listings, retrieved via PDG API
- Summaries based only on text and values from PDG (RAG)

Challenges and Tools

- **PDG reviews authored in LaTeX with PDG-specific LaTeX style file, distributed as PDF files**
- **How to convert to more suitable, structured format such as Markdown (or HTML)?**
  - From LaTeX sources (preferred), but so far not successful with standard tools (pandoc, LaTeXML)
  - From PDF files – several powerful tools have recently become available
    - Nougat: Neural Optical Understanding for Academic Documents (arXiv:2308.13418) mostly succeeds with excellent accuracy, but correctness is not guaranteed and there are occasional mistakes
- **Math support in Markdown dependent on Markdown flavor**
  - mathpix-markdown-it adds advanced equation and table support via LaTeX syntax, used by Nougat
- **Some other tools**
  - Custom code for semantic chunking
  - PDFFigures 2.0 to extract figures
  - Embedding model (e.g. bge-base-en-v1.5 with 768 embedding dimensions)
  - Chroma vector database
  - Langchain
  - A LLM (e.g. GPT-4o)

A First Prototype

Version: ngbase-bgebase (RPP 2024)

AI Search Assistant

SMEFT operators | Search | Summarize

61. Top Quark, page 3 (TABLE, chunk 12) — **Score = 0.801**

From section Couplings:

Operator	Field content	Operator	Field content
Four-quark		Two-quark-two-lepton	
$O_{qq}^{1(ijkl)}$	$(\bar{q}_i\gamma^\mu q_j)(\bar{q}_k\gamma_\mu q_l)$	$O_{lq}^{1(ijkl)}$	$(\bar{l}_i\gamma^\mu l_j)(\bar{q}_k\gamma^\mu q_l)$
$O_{qq}^{3(ijkl)}$	$(\bar{q}_i\gamma^\mu\tau^I q_j)(\bar{q}_k\gamma_\mu\tau^I q_l)$	$O_{lq}^{3(ijkl)}$	$(\bar{l}_i\gamma^\mu\tau^I l_j)(\bar{q}_k\gamma^\mu\tau^I q_l)$
$O_{qu}^{1(ijkl)}$	$(\bar{q}_i\gamma^\mu q_j)(\bar{u}_k\gamma_\mu u_l)$	$O_{lq}^{(ijkl)}$	$(\bar{l}_i\gamma^\mu l_j)(\bar{u}_k\gamma^\mu u_l)$
$O_{qu}^{8(ijkl)}$	$(\bar{q}_i\gamma^\mu T^A q_j)(\bar{u}_k\gamma_\mu T^A u_l)$	$O_{eq}^{(ijkl)}$	$(\bar{e}_i\gamma^\mu e_j)(\bar{q}_k\gamma^\mu q_l)$
$O_{qd}^{1(ijkl)}$	$(\bar{q}_i\gamma^\mu q_j)(\bar{d}_k\gamma_\mu d_l)$	$O_{eu}^{(ijkl)}$	$(\bar{e}_i\gamma^\mu e_j)(\bar{u}_k\gamma^\mu u_l)$
$O_{qu}^{8(ijkl)}$	$(\bar{q}_i\gamma^\mu T^A q_j)(\bar{d}_k\gamma_\mu T^A d_l)$	$^\ddagger O_{lequ}^{1(ijkl)}$	$(\bar{l}$ …
$O_{uu}^{(ijkl)}$	$(\bar{u}_i\gamma^\mu u_j)(\bar{u}_k\gamma_\mu u_l)$	$^\ddagger O_{lequ}^{3(ijkl)}$	$(\bar{l}$ …
$O_{ud}^{1(ijkl)}$	$(\bar{u}_i\gamma^\mu u_j)(\bar{d}_k\gamma_\mu d_l)$	$^\ddagger O_{lequ}^{(ijkl)}$	$(\bar{l}$ …
$O_{ud}^{8(ijkl)}$	$(\bar{u}_i\gamma^\mu T^A u_j)(\bar{d}_k\gamma_\mu T^A d_l)$		
$^\ddagger O_{quqd}^{(ijkl)}$	$(\bar{q}_i u_j)\,\varepsilon\,(\bar{q}_k d_l)$		
$^\ddagger O_{qud}^{8(ijkl)}$	$(\bar{q}_i T^A u_j)\,\varepsilon\,(\bar{q}_k T^A d_l)$		
Two-quark operators		Baryon- and lepton-number-violating	
$^\ddagger O_{u\nu}^{(ij)}$	$\bar{q}_i u_j \bar{\nu}\,(\varphi^\dagger\varphi)$	$^\ddagger O_{qqu}^{(ijkl)}$	$(\overline{q^c}_{i\alpha}\varepsilon q_{j\beta})(\overline{u^c}_{k\nu}c_l)\ \epsilon^{\alpha\beta\gamma}$
$O_{\nu q}^{1(ij)}$	$(\varphi^\dagger \overleftrightarrow{D}_\mu\varphi)(\bar{q}_i\gamma^\mu q_j)$	$^\ddagger O_{qqq}^{(ijkl)}$	$(\overline{q^c}_{i\alpha}\varepsilon q_{j\beta})(\overline{q^c}_{k\gamma}\varepsilon l_l)\ \epsilon^{\alpha\beta\gamma}$
$O_{\nu q}^{3(ij)}$	$(\varphi^\dagger \overleftrightarrow{D}_\mu\varphi)(\bar{q}_i\gamma^\mu r^I q_j)$	$^\ddagger O_{qqq}^{3(ijkl)}$	$(\overline{q^c}_{i\alpha}\tau^I\varepsilon q_{j\beta})(\overline{q^c}_{k\gamma}\tau^I\varepsilon l_l)\ \epsilon^{\alpha\beta\gamma}$
$O_{\nu u}^{(ij)}$	$(\varphi^\dagger \overleftrightarrow{D}_\mu\varphi)(\bar{u}_i\gamma^\mu u_j)$	$^\ddagger O_{duu}^{(ijkl)}$	$(\overline{d^c}_{i\alpha}u_{j\beta})(\overline{u^c}_{k\gamma}\varepsilon l_l)\ \epsilon^{\alpha\beta\gamma}$
$^\ddagger O_{\varphi ud}^{(ij)}$	$(\tilde{\varphi}^\dagger iD_\mu\varphi)(\bar{u}_i\gamma^\mu d_j)$		
$^\ddagger O_{uW}^{(ij)}$	$(\bar{q}_i\sigma^{\mu\nu}T^I u_j)\ \tilde{\varphi}W_{\mu\nu}^I$		
$^\ddagger O_{W}^{(ij)}$	$(\bar{q}_i\sigma^{\mu\nu}T^I d_j)\ \tilde{\varphi}W_{\mu\nu}^I$		
$^\ddagger O_{uB}^{(ij)}$	$(\bar{q}_i\sigma^{\mu\nu}u_j)\ \tilde{\varphi}B_{\mu\nu}$		
$^\ddagger O_{uG}^{(ij)}$	$(\bar{q}_i\sigma^{\mu\nu}T^A u_j)\ \tilde{\varphi}G_{\mu\nu}^A$		

Table 6.1.1: List of SMEFT operators at dimension six, assuming $U(2)_q \times U(2)_u \times U(2)_d$ flavor symmetry. For more details on notation and conventions, see reference in the text.

Preview of content, even for advanced math

91. Searches for Quark and Lepton Compositeness, page 5 (TEXT, chunk 9) — **Score = 0.765**

From section 91.1 Limits on contact interactions:

Eq. (91.1) can also be regarded as a part of more general dimension-six operators in the context of low-energy standard-model effective field theory (SMEFT). For a complete list of SM gauge-invariant dimension-six operators, see [30, 31]. A computation of the one-loop
…s. [35, 36, 37] for recent reviews. Ref. [17]
…ork, using the CMS data of the inclusive jet
…CMS $t\bar{t}$ cross section data. The results are translated into a 95% C.L. exclusion limit on the quark compositeness $\Lambda_{LL} > 24$ TeV, which is shown as the second solid line in brown in Figure 91.2. In models where the SM fermions get their masses though the mixing with the composite states, the top-quark is expected to show compositeness properties [38] resulting in the $t\bar{t}t\bar{t}$ contact interaction operators in the SMEFT context. An enhancement of the four-top-quark production cross section is expected at hadron colliders in these models [39]. The ATLAS Collaboration has extracted 95% C.L. observed (expected) limits on the $t\bar{t}t\bar{t}$ contact interaction operator $|C_{4t}|/\Lambda^2 < 1.9$ TeV $^{-2}$ (1.6 TeV $^{-2}$) using their 36.1 fb $^{-1}$ data at $\sqrt{s} = 13$ TeV [40]. Consistent SMEFT reinterpretations of top-quark data, and high energy non-resonant dijet and dilepton data are provided, e.g., in Refs. [41, 42].

Link that opens PDF file on the relevant page

61. Top Quark, page 4 (TEXT, chunk 15) — **Score = 0.749**

From section Couplings:

The definitions of the SMEFT operators can be organised in four categories: Four-quark, two-quark-two-lepton, and baryon-lepton-number violating operators. The overwhelming number of four-fermion operators is tamed by adopting simplifying assumptions about beyond-the-standard-model flavor structures. A baseline flavor scenario in the quark sector and motivated by the minimal flav… [15, 16, 17] corresponds to imposing a $U(2)_q \times U(2)_u \times U(2)_d$ symmetry among the first two generations. In this c… of degrees of freedom are produced for the operators of each category of field content:

four heavy quarks	$11 + 2$ CPV
two light and two heavy quarks	14
two heavy quarks and bosons	$9 + 6$ CPV
two heavy quarks and two leptons	$(8 + 3$ CPV$) \times 3$ lepton flavors

where we counted separately CP-conserving and CP-violating (CPV) parameters. They are collected in Table 61.1. Other less restricted scenarios, such as that obtained by imposing $U(2)_{q+u+d}$ symmetry featuring additional $10 + 10$ CPV degrees of freedom, or more restricted ones, such as the *top-philic* scenario where it is assumed that new physics couples dominantly to the left-handed doublet and right-handed up-type quark singlet of the third generation as well as to bosons, featuring only 19+6 (CPV) degrees of freedom, are often considered. It is also customary to analyse top-quark flavor-changing neutral currents (FCNCs) separately as, at the tree level, they enter only quadratically. More details can be found in Ref. [12].

Similarity score

61. Top Quark, page 4 (TEXT, chunk 14) — **Score = 0.732**

From section Couplings:

The currently adopted parametrization for SMEFT interpretations of top quark measurements relies on the Warsaw basis of gauge-invariant dimension-six operators [11] and it is presented in Tab. 61.1 as detailed in Ref. [12] (see [13, 14] for early discussions of top-quark related operators). For convenience, often specific degrees of freedom are identified from combinations of Warsaw-basis operator coefficients aligned with the directions of the EFT parameter space which appear in given processes, in interferences with SM amplitudes, and in top-quark interactions with some of the gauge boson mass eigenstates. Model implementations are available for tree-level and even one-loop Monte Carlo simulations.

Conclusions

- **The new PDG API provides programmatic access to PDG data via three tools, aimed at different use cases**
  - Data in JSON format from pdgLive (or via REST API)
  - **Python API**
  - SQLite database files

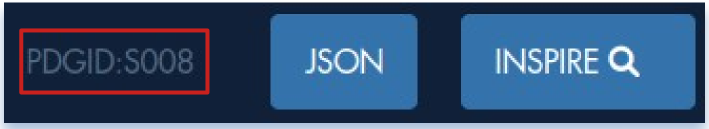
- **Exploring recent AI (LLM) methods with the goal to provide enhanced searching of PDG data and reviews**
  - Results from a first prototype are encouraging and are used to learn what matters and where improvement is needed
  - One of the hurdles is the 100% accurate conversion of PDG reviews into a suitable structured format such as Markdown

Extra Slides

PDG Identifiers

How can one refer to desired particle physics quantities in a program?

- **Easy for some things, e.g. for the charged pion π^+**
  - ASCII name: pi+
  - MC particle number: 211
- **How about $B^0 \to J/\Psi(1S)K^*(892)^0\pi^+\pi^-$?**
- **PDG defines digital object identifiers ("PDG Identifiers")**
  - Case-insensitive, alphanumeric strings
    - First 4 alphanumeric characters typically denote particle, additional characters for properties
    - STRING.NUMBER for branching fractions
  - Examples
    - S008 for π^+
    - S042.214 for the above decay
  - In most cases can e.g. iterate over quantities, but if
  - PDG Identifier is needed, can look it up e.g. in pdgLive



Python API: Ambiguities and Pedantic Mode

- **For properties such as mass, by default the Python API returns the “best” value**
  - Sometimes this is ambiguous
- **In “pedantic mode”, assumptions are avoided and one gets an exception in case of ambiguity**

t

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

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

Mass (direct measurements) $m = 172.57 \pm 0.29$ GeV [a,b] $(S = 1.5)$
Mass (from cross-section measurements) $m = 162.5^{+2.1}_{-1.5}$ GeV [a]
Mass (Pole from cross-section measurements) $m = 172.4 \pm 0.7$ GeV

```
[1]: import pdg

[2]: api = pdg.connect()
     top = api.get_particle_by_name('t')
     [m.display_value_text for m in top.masses()]

[2]: ['172.57+-0.29', '162.5+2.1-1.5', '172.4+-0.7']

[3]: top.mass

[3]: 172.5746930968864

[4]: api = pdg.connect(pedantic=True)
     top = api.get_particle_by_name('t')
     top.mass

---------------------------------------------------------------------------
PdgAmbiguousValueError                    Traceback (most recent call last)
Cell In[4], line 3
      1 api = pdg.connect(pedantic=True)
      2 top = api.get_particle_by_name('t')
----> 3 top.mass
```

PDG Database Files

- **PDG data in a single file in SQLite format**
  - Different versions can be downloaded from [pdg.lbl.gov/api](pdg.lbl.gov/api)
  - "pdgall" files contain historical summary data

Database files

The following database file are currently available (see documentation):

Database file	Edition	Description
pdg-2024-v0.1.2.sqlite	2024	Minimal data file as provided by Python package v0.1.2
pdgall-2024-v0.1.2.sqlite	2024	Ditto, but with historical Summary Table data

- **Python API can use a downloaded file instead of the default data source, e.g.**

```
api = pdg.connect("sqlite:///pdgall-2024-v0.1.2.sqlite")
len(api.editions), api.editions[:5]
```

```
(52, ['2024', '2023', '2022', '2021', '2020'])
```