Introducing the Zeuthen Data Science Seminar

Jakob van Santen
2020-02-25

HELMHOLTZ RESEARCH FOR GRAND CHALLENGES

DESY.

Rationale

- Much of our work involves writing and using software to analyze data. As data volumes grow, this is becoming more and more complex.

- This is almost never what we talk about in existing seminars and colloquia.

- The Data Science Seminar is an opportunity to exchange knowledge about problems we have (or have solved!) in data analysis, computing, and software development.

Format

- Mix of informal, club-style talks, hands-on tutorials, and invited talks from external speakers

- We want you to learn:

  - What are my colleagues doing **on a technical level**?

  - What can I appropriate for **my own work**?

  - What tools are out there that could **make my work easier?**

- Only one rule: **no science results!**

Organizers

Thorsten Kuhl (ATLAS)

Gernot Maier (CTA, VERITAS)

Stefan Ohm (CTA, HESS)

Jakob van Santen (IceCube)

Prospective topics

Evolving in response to your input

Computing
- Grid computing: DESY-internal and European consortia
- Biggish Data: parallel analysis of larger-than-RAM datasets
- Parallel computing with GPUs: typical applications, pitfalls, useful libraries
- Performance profiling best practices

Software development
- Setting up a development environment for beginners
- Choosing the right text editor
- Shell plumbing

Collaborative software development
- Development workflows with Git
- Automated testing and why you need it
- Continuous Integration services
- Sustainable development and software quality

Numerical methods
- Approaches and pitfalls in numerical minimization

Prospective topics

Evolving in response to your input

Computing

- Grid computing: DESY-internal and European consortia
- Biggish Data: parallel analysis of larger-than-RAM datasets
- Parallel computing with GPUs: typical applications, pitfalls, useful libraries
- Performance profiling best practices

Software development

- Setting up a development environment for beginners
- Choosing the right text editor
- Shell plumbing

Collaborative software development

- Development workflows with Git
- Automated testing and why you need it
- Continuous Integration services
- Sustainable development and software quality

Numerical methods

- Approaches and pitfalls in numerical minimization

Is there a topic you'd like to hear more about? One you'd like to present? Contact us!

Recurring segments
(Our initial ideas. There may be more.)

- **Tip of the Week**: share a useful package or language feature you've discovered. 5 minutes, 2 slides max. **Every meeting.**

- **Machine learning applications in <my experiment>**: share what you do! What works? What is difficult? **As often as is practical.**

- **DESY Alum Talk**: ex-DESYans return to talk about their work outside academia. **Every ~2 months.**

Recurring segments
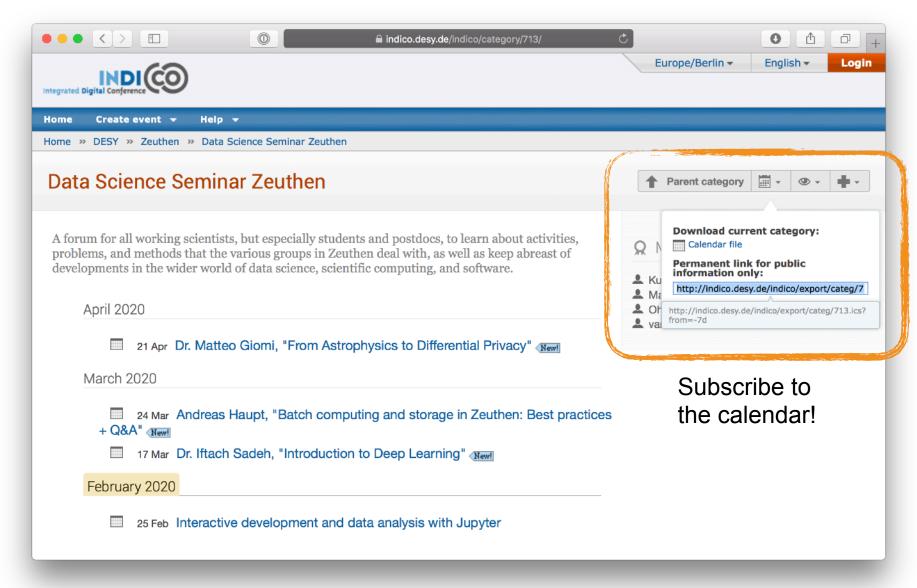**(Our initial ideas. There may be more.)**

Help wanted!

- **Tip of the Week**: share a useful package or language feature you've discovered. 5 minutes, 2 slides max. **Every meeting.**

- **Machine learning applications in <my experiment>**: share what you do! What works? What is difficult? **As often as is practical.**

- **DESY Alum Talk**: ex-DESYans return to talk about their work outside academia. **Every ~2 months.**

Upcoming events

- **March 17**: "Introduction to Deep Learning" with Iftach Sadeh

- **March 24**: "Batch computing and storage in Zeuthen: Best Practices + Q&A" with Andreas Haupt

- **April 21**: "From Astrophysics to Differential Privacy" with Matteo Giomi (DESY Alum Talk)

DESY.

Schedule on Indico

Subscribe to the calendar!

https://indico.desy.de/indico/category/713/

Slack workspace

https://zeuthendatasc-fxm8846.slack.com

(follow up on examples, discuss future topics)

DESY.

Questions?

Literal string interpolation in Python (≥ 3.6)

My first Tip of the Week

```
In [1]: blerh, blah = 3.14159, 6e-12
```

Traditional printf() style: hard to read

```
In [2]: "pi is %.2f, %.1e is small" % (blerh, blah)
Out[2]: 'pi is 3.14, 6.0e-12 is small'
```

Python 2.6: better, but quite verbose

```
In [3]: "pi is {blerh:.2f}, {blah:.1e} is small".format(blerh=blerh, blah=blah)
Out[3]: 'pi is 3.14, 6.0e-12 is small'
```

Python 3.6: nice and compact!

```
In [4]: f"pi is {blerh:.2f}, {blah:.1e} is small"
Out[4]: 'pi is 3.14, 6.0e-12 is small'
```

Bonus: any one-liner can be expanded

```
In [5]: f"pi is {blerh*2:.2f}, {blah+1:.1e} is small"
Out[5]: 'pi is 6.28, 1.0e+00 is small'
```

https://www.python.org/dev/peps/pep-0498/