

# Introducing the Zeuthen Data Science Seminar

Jakob van Santen  
2020-02-25

# 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

11 AM in SR3

- **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)

# Schedule on Indico

The screenshot shows the Indico website interface. The browser address bar displays `indico.desy.de/indico/category/713/`. The page title is "Data Science Seminar Zeuthen". Below the title, a description states: "A forum for all working scientists, but especially students and postdocs, to learn about activities, problems, and methods that the various groups in Zeuthen deal with, as well as keep abreast of developments in the wider world of data science, scientific computing, and software." The event schedule is organized by month: April 2020, March 2020, and February 2020. An orange box highlights a dropdown menu for "Download current category:", which includes a "Calendar file" option and a "Permanent link for public information only:" option with a URL: `http://indico.desy.de/indico/export/categ/713.ics?from=-7d`.

**Data Science Seminar Zeuthen**

A forum for all working scientists, but especially students and postdocs, to learn about activities, problems, and methods that the various groups in Zeuthen deal with, as well as keep abreast of developments in the wider world of data science, scientific computing, and software.

**April 2020**

- 21 Apr Dr. Matteo Gioni, "From Astrophysics to Differential Privacy" **New!**

**March 2020**

- 24 Mar Andreas Haupt, "Batch computing and storage in Zeuthen: Best practices + Q&A" **New!**
- 17 Mar Dr. Iftach Sadeh, "Introduction to Deep Learning" **New!**

**February 2020**

- 25 Feb Interactive development and data analysis with Jupyter

**Download current category:**

- Calendar file
- Permanent link for public information only:  
`http://indico.desy.de/indico/export/categ/713.ics?from=-7d`

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)

# Questions?

# Literal string interpolation in Python ( $\geq 3.6$ )

## My first Tip of the Week

	<pre>In [1]: blerh, blah = 3.14159, 6e-12</pre>
<b>Traditional printf() style: hard to read</b>	<pre>In [2]: "pi is %.2f, %.1e is small" % (blerh, blah) Out[2]: 'pi is 3.14, 6.0e-12 is small'</pre>
<b>Python 2.6: better, but quite verbose</b>	<pre>In [3]: "pi is {blerh:.2f}, {blah:.1e} is small".format(blerh=blerh, Out[3]: 'pi is 3.14, 6.0e-12 is small'</pre>
<b>Python 3.6: nice and compact!</b>	<pre>In [4]: f"pi is {blerh:.2f}, {blah:.1e} is small" Out[4]: 'pi is 3.14, 6.0e-12 is small'</pre>
<b>Bonus: any one-liner can be expanded</b>	<pre>In [5]: f"pi is {blerh*2:.2f}, {blah+1:.1e} is small" Out[5]: 'pi is 6.28, 1.0e+00 is small'</pre>

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