SOSC 2024 Sixth International School on Open Science Cloud

2–6 Dec 2024
University of Bologna. Department of Physics and Astronomy
Europe/Rome timezone

Enter your search term

SOSC24: a bit of history and some Logistics

Daniele Spiga, INFN-Perugia
spiga@pg.infn.it

2 December 2024 to 6 December 2024
INFN Perugia

Welcome to Bologna and to SOSC 2024

SOSC (School on Open Science Cloud) is jointly organized by INFN, University of Perugia and University of Bologna.

- This is the Sixth edition and is held in Bologna
- 28 students, few auditors from INFN and University staff, **a now traditional blending among theory and practice**

spiga@pg.infn.it

What is the SOSC?

yet another school? (Hopefully no!)

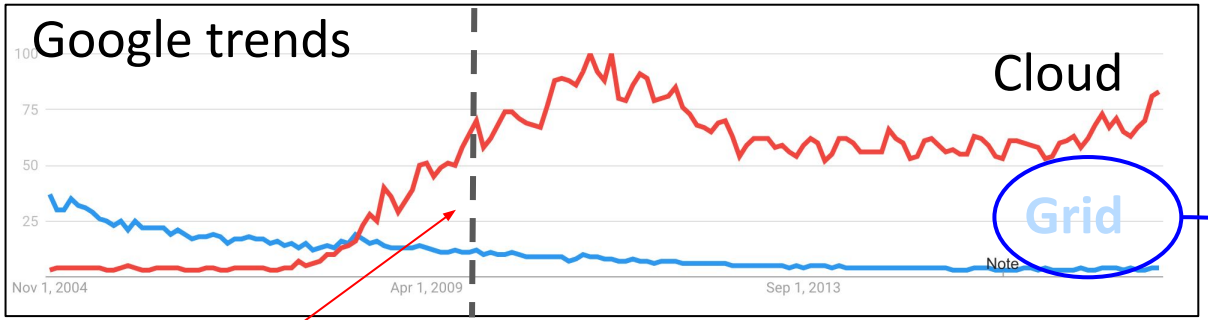
SOSC is designed to facilitate the adoption of

- new technologies
- new paradigms to support scientific data analysis workflows.

Born from the observation of an objective lack of schools that dealt with both components!

spiga@pg.infn.it

Context & Background



Since 2010 INFN is active in cloud-related projects and R&D activities.

R&D activities: the vision

To **develop software components and solutions** to facilitate (or simply make possible) the **exploitation of distributed cloud and storage resources through public or private infrastructures.**

- Tailored to science and targeting **multi-disciplinary scientific communities**

To **develop software components and solutions** to facilitate (or simply make possible) the **exploitation of distributed cloud and heterogeneous providers (i.e. HTC/HPC/Cloud)**

- Tailored to science and targeting **multi-disciplinary scientific communities**

Since then we raised the bar expanding the phase space

2015-2017

INDIGO - DataCloud

A successful EU project coordinated by INFN (PI D.Salomoni)

Istituto Nazionale di Fisica Nucleare

SOSC mission

Contributes to democratize the computing resource access.

- The value is the science so researchers need to effectively use distributed and heterogeneous computing infrastructures

Aims to build a community (of young researchers) that share experiences and knowledge in modern/promising technologies and data analysis techniques

But also **an opportunity for us** (organizers) **to get feedbacks** from users

- To learn how to keep improving our middleware solutions to enable a better science

spiga@pg.infn.it

A few SOSC keywords

Infrastructure as a Code; Declarative approach; Automation

- Don't waste effort doing and redoing intricate configurations/setup

Managing **user tailored computing runtime**

- Remember our target: distributed & heterogeneous environment

Adoption of **open source industry de jure/de facto standards**

- Let's not reinvent the wheel, better contributing upstream

SOSC perspective is on **end-user data processing** (aka analysis step)

NOT a school on ML/AI. It uses ML

- Not just because it's sexy!
- Rapidly **gaining traction in many scientific areas**.
- Potential for **cross-fertilization** of knowledge with **"externals/industry"**
- Represents a **computational challenge**

Showcase real science experiment

- CYGNO (see more later)

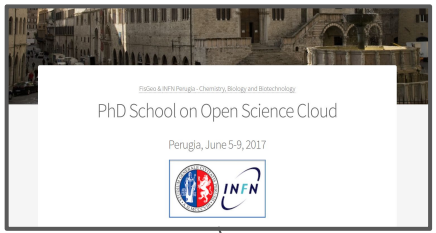
spiga@pg.infn.it

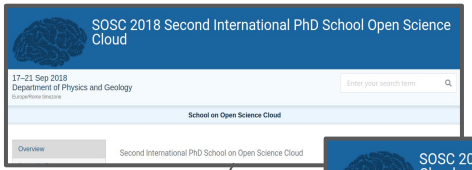
Story line

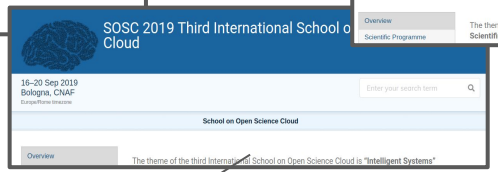
Data Science oriented

Consolidating the essence of SOSC

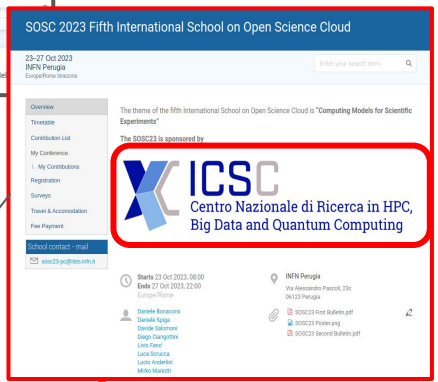
Exploratory edition



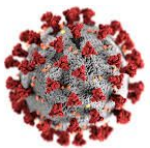




Exploring the new format



2017 2018 2019 2020 2021 2022 2023 **2024**



Covid19 Pandemic

We're now running the sixth edition

spiga@pg.infn.it

Theme and structure of the School

The SOSC 2024 theme is **“Computing Models for Scientific Experiments”**; the school programme is organized in three main tracks:

1. **Fundamentals ->** **Monday and Tuesday**
2. **Machine Learning ->** **Tuesday and Wednesday**
3. **Automation, Workflow Management and pipelines->** **Thursday**

The school is structured with **lectures in the morning** and **hands-on sessions (mainly but not exclusively) in the afternoon**.
A student project will be developed from Monday afternoon up until Thursday afternoon

Please refer to the SOSC 2024 agenda for details and for the official timetable. All presentations will be uploaded there.

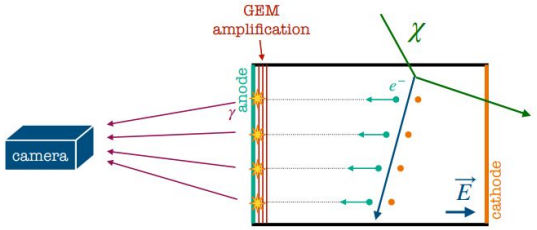
spiga@pg.infn.it

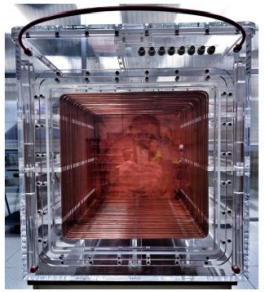
SOSC24 project is inspired by a real experiment

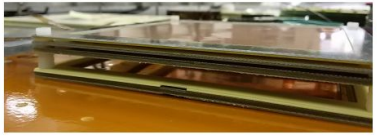
The CYGNO project

- **Aiming for** a large detector for high precision **3D tracking of rare low energy nuclear recoils** (keV) possibly induced by **dark matter** (DM) particles and solar neutrinos
- **Experimental challenges:** rate O(evt/kg/year), background rejection, and energy threshold (keV)
- **Strategy: photograph nuclear recoils** in a (1 atm) He:CF_4 TPC with a GEM amplification stage

➡ low energy events in 1 atm gas ⇒ visible tracks







Courtesy Stefano Piacentini

We/you will use Cygno images

Dott. Giorgio Dho will tell us all the interesting details about the Cygno Experiment

About The individual project

The SOSC24 foresees the development of a personal project.

- Upon successful completion of the project the School will issue an official **“SOSC School of Computing Diploma"**

The project is structured with Jupyter Notebooks that will include also exercises. While the first part of the notebooks will be guided by tutors, **the exercises will be developed just by you we will use a cloud environment (Cloud@INFN) to:**

- Simulate a Data Acquisition System (DAQ), or better the interaction with a DAQ
- pre-process acquired data to the analyze it
- analyze data
- automate the pipeline

You will need to provide us with your Jupyter Notebooks (each day you will have one to deliver)

NOTE: An **attendance certificate will be provided to the students attending the whole School programme**

spiga@pg.infn.it

Hands-on: tools

You've been asked to **register** here:

- if you have not done so yet, please tell us and do it asap.

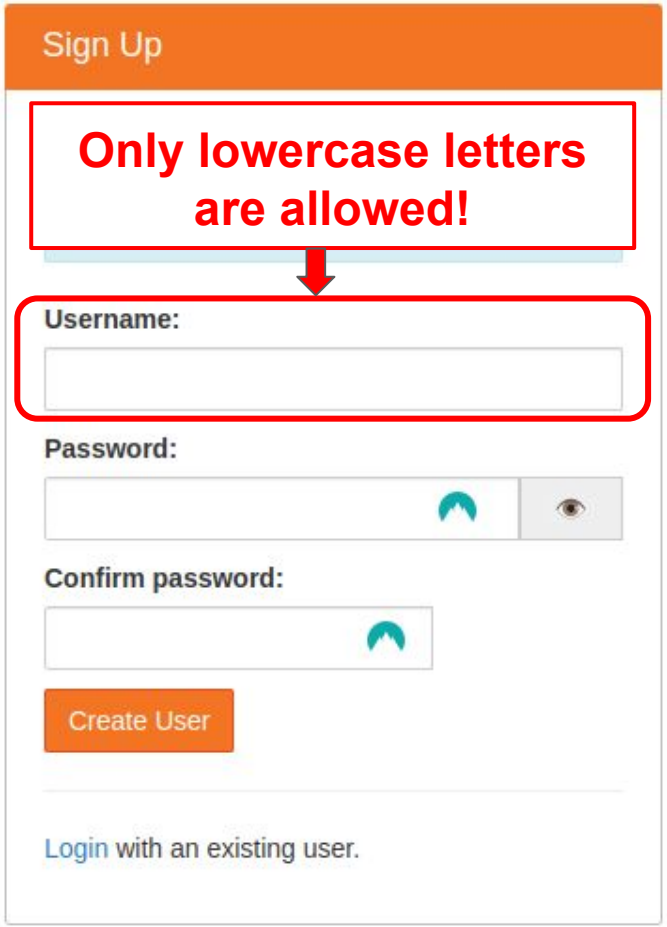
The provided cloud platform has been deployed via INFN-Cloud The platform will provide you with

- **A Jupyter environment**

A Cloud storage service (MinIO)
A Cloud native messaging system (Nats)

All will be in form of "as a Service" systems.

You will learn how to interact and operate with those services. We'll teach you.. That's part of the Fundamentals



Hands-on: Guided Exercise

The hands-on are meant as guided exercises that will treat most of the topics that will be required to develop your individual project. There will be hands-on several topics:

- Software packaging and software management
  - Container and docker, a pillars in the cloud-native applications
- Image handling in python and data pre-processing
  - SOSC23 data are images (Cygno images). We will learn how to manipulate images in Python. This is needed to prepare our data to be treated in a meaningful way.
- Manage data in a Cloud Environment Storage and message queue
  - We're in the cloud :) and thus we focus on S3 Cloud Storage
  - We also learn how to exchange data, segmented in the form of messages. Focus on publish and subscribe as a way to process data as they come from our camera/sensor
- Machine Learning technique to treat our data.
  - We will have a classification problem… need to distinguish electron recoil events from Nuclear Recoil events… in the end we're looking for Dark Matter
- Automation in a cloud environment
  - We will build a pipeline making our analysis reproducible

All this should provide you with material (code) needed to develop the individual project

spiga@pg.infn.it

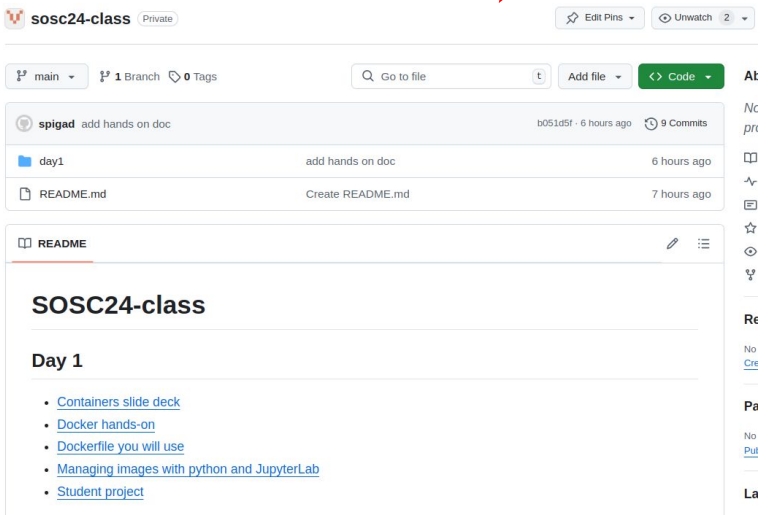
Hands-on and Individual Project: Material

All the material for the guided hands-on as well as the tracks of the individual project will be in form of Jupyter Notebooks. Everything will be made available via the [github repo of the SOSC24]

- You'll be guided to interact with git software repositories

If not already present, we will push material the day itself

- Before the school session will start



During the Closing session on Friday

On Friday we shall have:

- An overall SOSC 2024 **evaluation questionnaire**.
- The delivery of a **SOSC 2024 certificate to the students attending** the whole school program.
- The delivery of the **Official SOSC School of Computing Diploma.**
  - For the students successfully completing the project

NOTE: we will keep the cloud platform accessible for a week (20.12.2024). After that, the systems will be destroyed. At the end of the school you should be able to gather all your material, however ask us for any support

spiga@pg.infn.it

Logistic and Events

Wireless access is available either via **eduroam, or via a special SSID**.

- If you do not have eduroam, contact us and we will provide your credentials.

All the sessions of the school will be held in the “**Aula bp-2b**”

- feel free to choose your working station and keep it until friday

We assume all of you have a laptop to connect to the cloud provided services

- Services will be available even from outside

Given we also have international students, **the lectures will be in English**

spiga@pg.infn.it

Logistic and Events (cont.)

Each morning you should **sign the presence**

Coffee breaks will be served here in this room at mid-morning and mid-afternoon (**included in the Fee**)

Events:

1. A welcome event with buffet **this night** in conjunction with a focus event on FPGA (**included in the Fee**)

2. Social Dinner on Thursday (**included in the Fee**)

Lunch Breaks are not included in the fee → plenty of Bar/Restaurants

- Ask us for suggestions if you need them

spiga@pg.infn.it

Tonight the first event: Wine & FPGA

Where: It will be at Lab 16

When: Tonight at 19.00 sharp

- We will meet directly there

How to reach the Lab 16? **Google Maps is your friend. (less than 20 min by walk)**

A welcome event to relax and socialize. We will talk about FPGA for Machine Learning.

- **(food for nerds :))**

spiga@pg.infn.it

Social Dinner

Where: It will be at Osteria La Traviata

When: Thursday 5th at 20.00

- We will meet directly there

How to reach the restaurant? **Google Maps is still your friend**

spiga@pg.infn.it

Recap

Useful info and links

1. SOSC24 **github** here: https://github.com/SOSC-School/SOSC24-class
2. SOSC24 hands-on **Platform** here: https://sosc.131.154.98.182.myip.cloud.infn.it/
3. SOSC24 **agenda** of the school https://web.infn.it/SOSC24

And now let's start the SOSC24

Welcome again to SOSC 2024

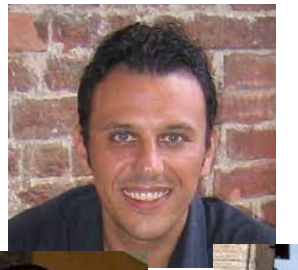
Hope you will have a fruitful week. Enjoy the school and the social events, enjoy Bologna and, of course, do not forget to have fun!

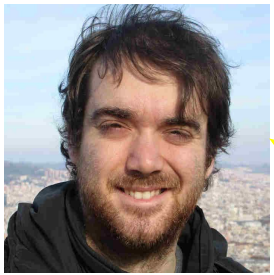
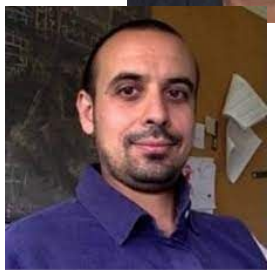
For any further information, see us here, or email the SOSC 2024 Program Committee at sosc24-pc@lists.infn.it

The SOSC 2024 Program Committee:

- **Lucio Anderlini** INFN Firenze
- **Daniele Bonacorsi**, University of Bologna
- **Diego Ciangottini** INFN Perugia
- **Mirko Mariotti**, University of Perugia
- **Davide Salomoni**, ICSC
- **Luca Scrucca**, University of Bologna
- **Daniele Spiga**, INFN Perugia

Our faces!





To handle with care

Our FPGA Expert

CYGNO
Expert: invited

spiga@pg.infn.it