



ess

EUROPEAN
SPALLATION
SOURCE



The Instrument Data Scientist / Computational Instrument Scientist

Indispensable Expertise for User Operation

PRESENTED BY AARON FINKE

Agenda



1. The IDS/CIS Role

- What we do
- What we don't do
- What to expect of us for User Ops

2. AI from the IDS/CIS perspective

- How it can help
- How to make the best use of it
- How ESS can make the best use of it for scientists

DMSC @ ESS: Borderless Science



For successful user operation, we need to work together like this border does not exist.

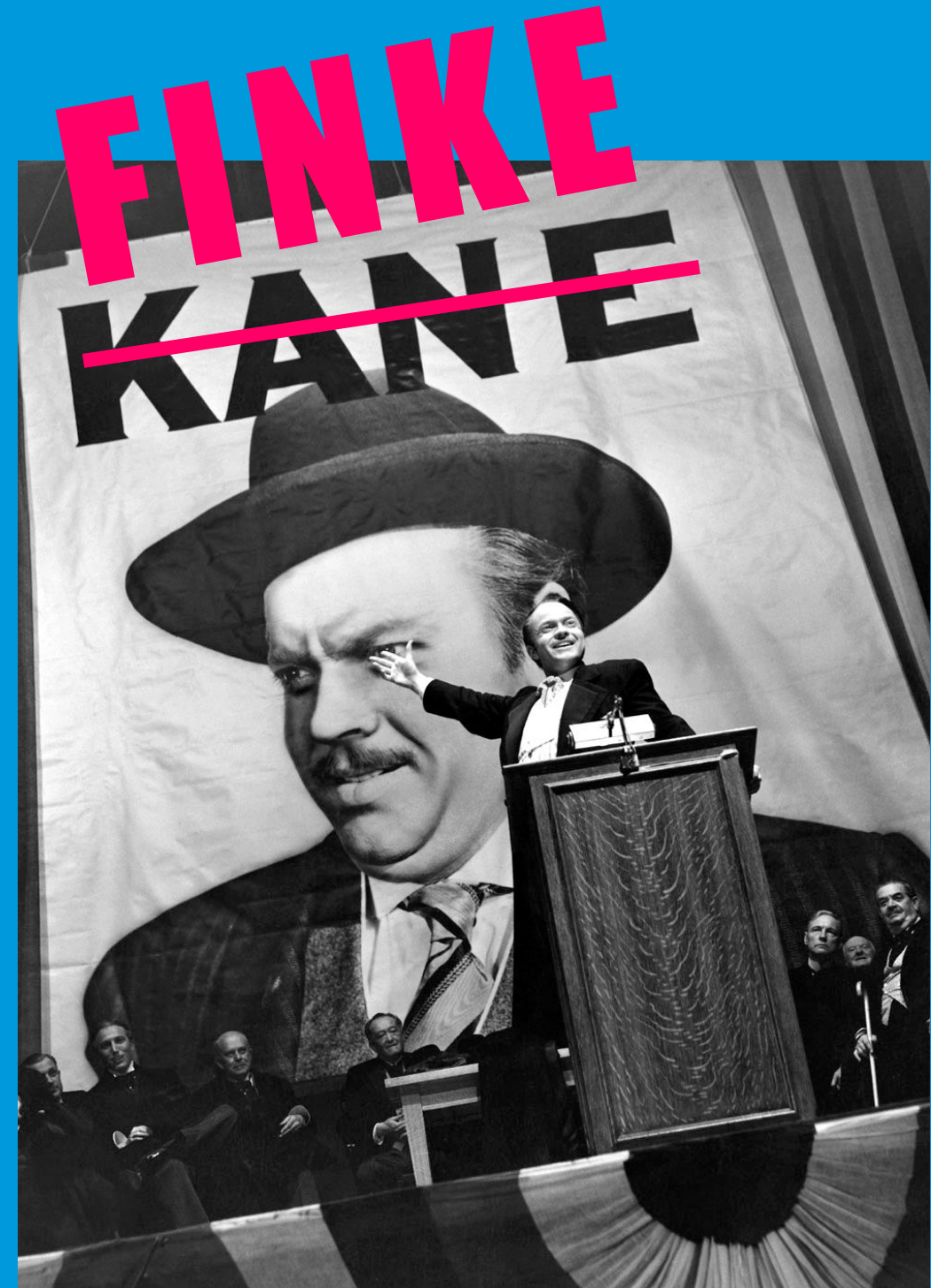
The decision to put DMSC in Denmark was largely *political*.

However, our presence in Denmark gives us access to Danish resources (£€)
and researchers at top Danish universities.

Many at DMSC are already taking advantage of our Danish affiliation
(I have an adjunct affiliation at Aarhus Uni and am on the NEUDATA4LIFE Lighthouse grant).

Who am I?

1. IDS/CIS for NMX
2. Spent ten years as a beamline scientist in macromolecular crystallography (MX)
 - Swiss Light Source (PX I-III)
 - Cornell High Energy Synchrotron Source (ID7B2/FlexX)
 - MAX IV (BioMAX)
3. MX beamlines are *high-volume* user facilities
 - Users have lots of samples and limited time
 - Real-time data analysis went from a luxury to an expectation in the last 15 years
 - I did lots of software development

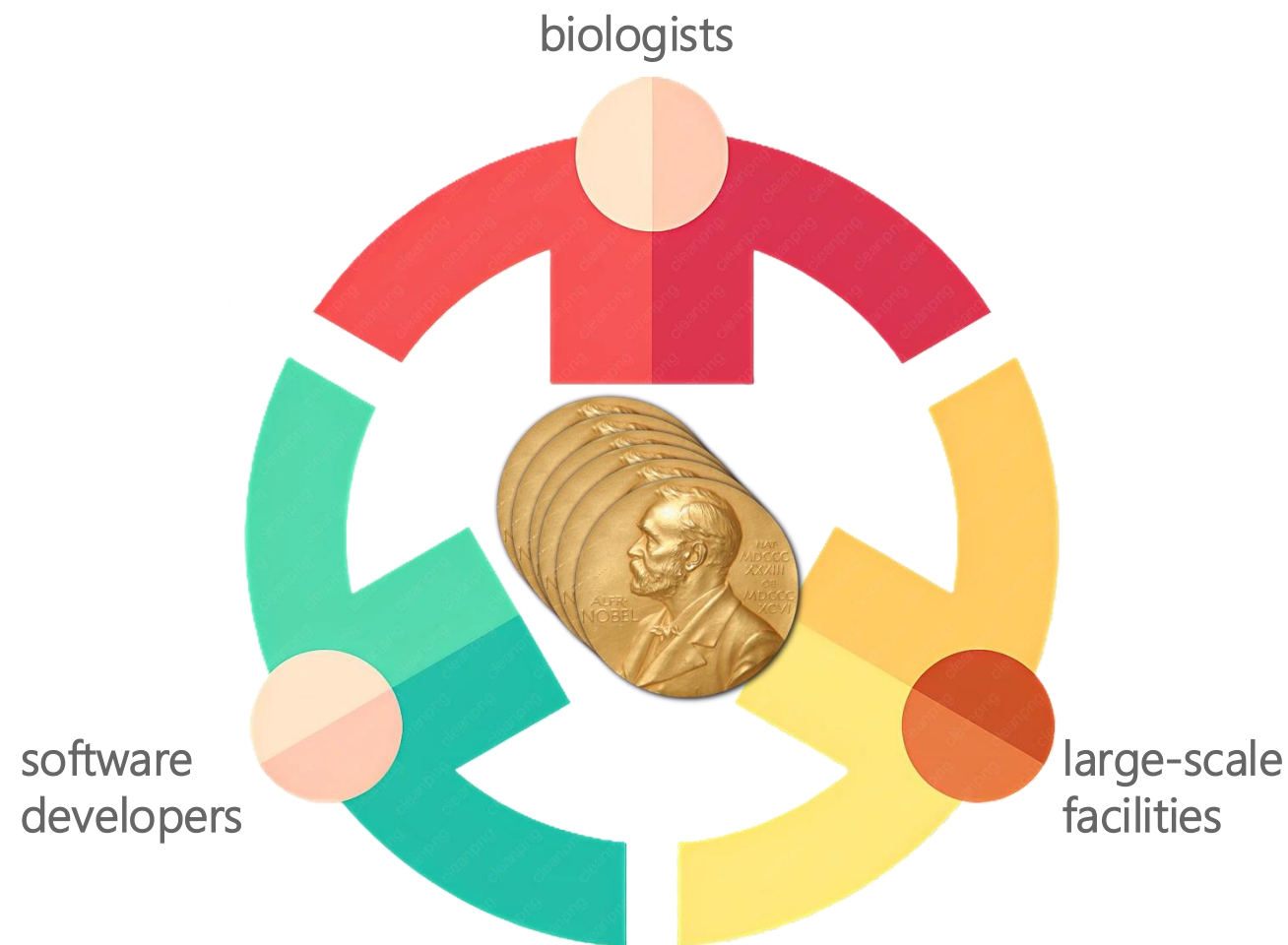


The “Unusual Success” of MX

Tempered by an open spirit of collaboration from seemingly disparate fields



- At least seven Nobel Prizes in Chemistry have been awarded to MX researchers.
- That’s a lot for a relatively niche field. Why?
 - Important problems/discoveries in biological chemistry
 - Advances in facilities/technology (especially synchrotron light sources)
 - Algorithmic development in MX data (molecular replacement, anomalous phasing, solvent flattening, FFT indexing, etc. etc.)



And all of these people working closely together!

1.

What is an IDS/CIS?



The IDS/CIS

The R2A2



- Liaise with the relevant instrument teams and technology groups to establish and maintain an effective interface with DMSC activities
- Act as expert in software, data processing, data science, simulations and scientific computing for relevant instrument teams
- Work with the user community and instrument teams to find suitable solutions for their scientific computing needs
- Represent DMSC at relevant instrument STAP meetings
- Coordinate and share knowledge with other instrument data scientists and DMSC technology groups to facilitate sustainable solutions, avoid duplication of work, and foster a collaborative spirit
- Contribute to the development of open-source software where needed and appropriate



The IDS/CIS

The R2A2



- Oversee, review, evaluate and take ownership of DMSC deliverables to relevant instruments, specifically the full data processing pipeline
- Act as the user community's advocate and assume an active approach
- Provide constructive feedback to DMSC development teams where relevant
- Maintain and update roadmap and / or project schedule for DMSC deliverables to relevant instruments
- Conduct scientific or technical research and author scientific or technical publications in the field of scientific computing and neutron scattering
- Take part in peer review of scientific publications, grant proposals, research, and facilities
- Represent ESS at appropriate conferences and workshops



The IDS/CIS

The R2A2



- Establish and maintain constructive contacts with the research and innovation community
- Present scientific and technical research at relevant conferences and workshops, and in seminars and lectures
- Supervise post docs and students and participate in teaching, as appropriate
- Keep relevant instrument models up to date at an appropriate level of fidelity
- Spend up till (*sic*) 20% of time on local contacting for user experiments in alignment with instrument team
- Support users with treating, analyzing, modelling and simulating data from experiments at ESS in alignment with instrument team
- Provide training on the DMSC software suite to instrument teams and users



The IDS/CIS

The tl;dr



- We are **instrument scientists**
 - Experts in the relevant technique(s)
 - Assist with user operation
 - Maintain and optimize the data pipeline
 - Involved in many instrument-related projects
- We are **data scientists**
 - Develop new approaches to data analysis
 - Perform research in our respective fields
- We are **team players**
 - We work with you!



The IDS/CIS

What we are not



- **We are not tech support.**
 - We are not affiliated with IT.
 - We are not sysadmins.
 - We have staff access to the data systems and compute architecture. With a few exceptions, we have the same permissions on these systems as instrument scientists.
- IDS/CIS liaise with DST (DMSC IT) to ensure compute/data are working properly. But fixes that require admin access generally must be done by DST.
- Other IT issues are resolved *via* JIRA.



The IDS/CIS

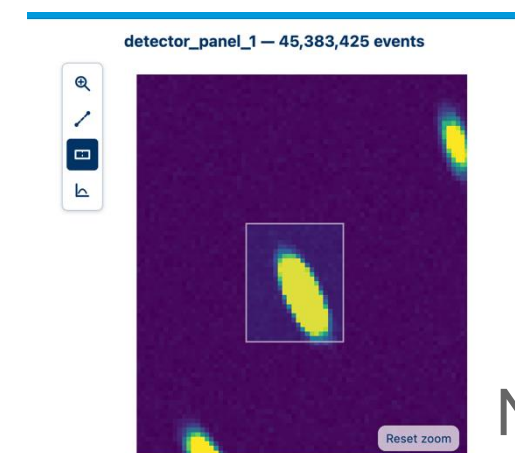
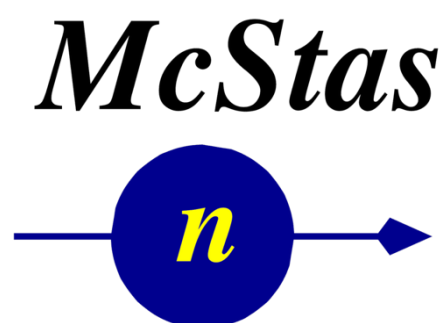
A necessary role

- At most facilities, instrument-specific, data-related work is delegated to instrument scientists.
- I personally spent most of my time at MAX IV writing software, but I was still a “Beamline Scientist” on paper.
- Delegating the data pipeline tasks to a specific position ensures that the data/software pipelines are not neglected.
- By ensuring a connection between software developers and instruments, messy bloatware like Mantid can be avoided.

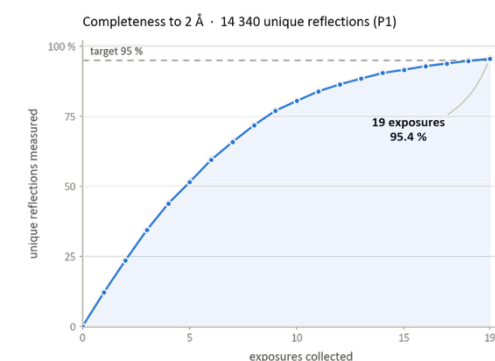


The IDS/CIS

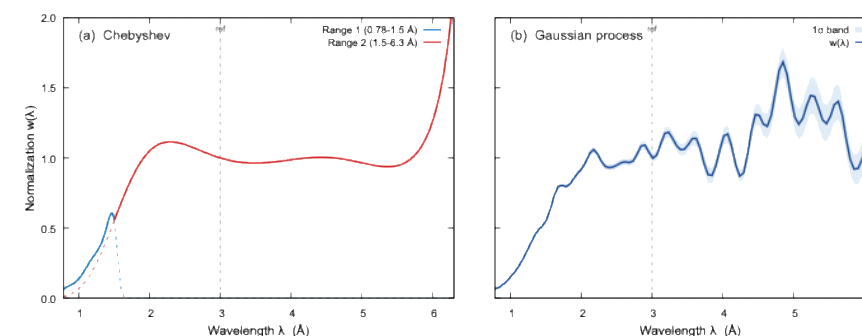
What does the NMX IDS do?



NMXViewer



NMXStrategy



Lawless
(post-integration
scaling)

Duct Tape Scripting

You know what it is. You might not call it that, but we're here to avoid it



- If an instrument required a piece of duct tape adjusted in just the right position to work properly, it becomes a *single point of failure* to fix a temporary issue.
- Temporary solutions become permanent ones when expedience outweighs reliability.
- The same is true with compute processes. "Duct tape scripts" are scripts that are hastily developed to solve a problem in the data/compute pipeline. Without proper testing, they can cause major issues later.
- IDS/CIS exist so temporary solutions are replaced with reliable, persistent ones.



VISA

The User-Facing Platform of DMSC

- VISA is a virtual computing system that provides user access to compute resources with access to data and data processing software, curated by DMSC.
- Linux desktop and Jupyter notebook instances are provided.
 - Jupyter comes with instrument-tailored *conda* instances.
- All instrument staff should already be using VISA now for data processing.
- **I am the product owner of VISA.** Please contact me or your IDS/CIS if you have suggestions or concerns.



<https://visa.ess.eu>



Data Analysis, in the cloud

VISA (Virtual Infrastructure for Scientific Analysis) makes it simple to create compute instances on the data analysis infrastructure to analyse your experimental data using just your web browser

[Sign in with your user account](#)

[instance](#) and use your web browser to access a Remote Desktop or using your experimental data

ir team

ance with other members of your team to [collaborate together](#) in real time

ftware

come with pre-installed [data analysis software](#) so you can start analysing your diately



Questions or feedback?

VISA

The User-Facing Platform of DMSC

- In many ways, **VISA is the most important product provided by DMSC**, as we expect all users to interact with it, and it hosts most other DMSC products.
- Because of this, it is important that we know about VISA issues immediately.
- IDS/CIS should be the first point of contact for VISA issues during user operation.
- DST has a separate Jira ticket point for VISA-related issues.
- Hardware updates are in the pipeline but currently without a timeline due to the volatile computing hardware market.



<https://visa.ess.eu>



Data Analysis, in the cloud

VISA (Virtual Infrastructure for Scientific Analysis) makes it simple to create compute instances on the data analysis infrastructure to analyse your experimental data using just your web browser

[Sign in with your user account](#)

[instance](#) and use your web browser to access a Remote Desktop or using your experimental data

ir team

ance with other members of your team to [collaborate together](#) in real time

ftware

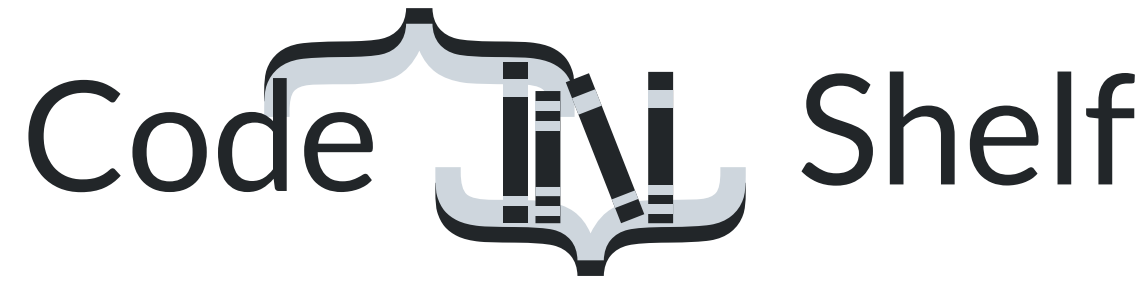
come with pre-installed [data analysis software](#) so you can start analysing your diately



Questions or feedback?

The Code Shelf

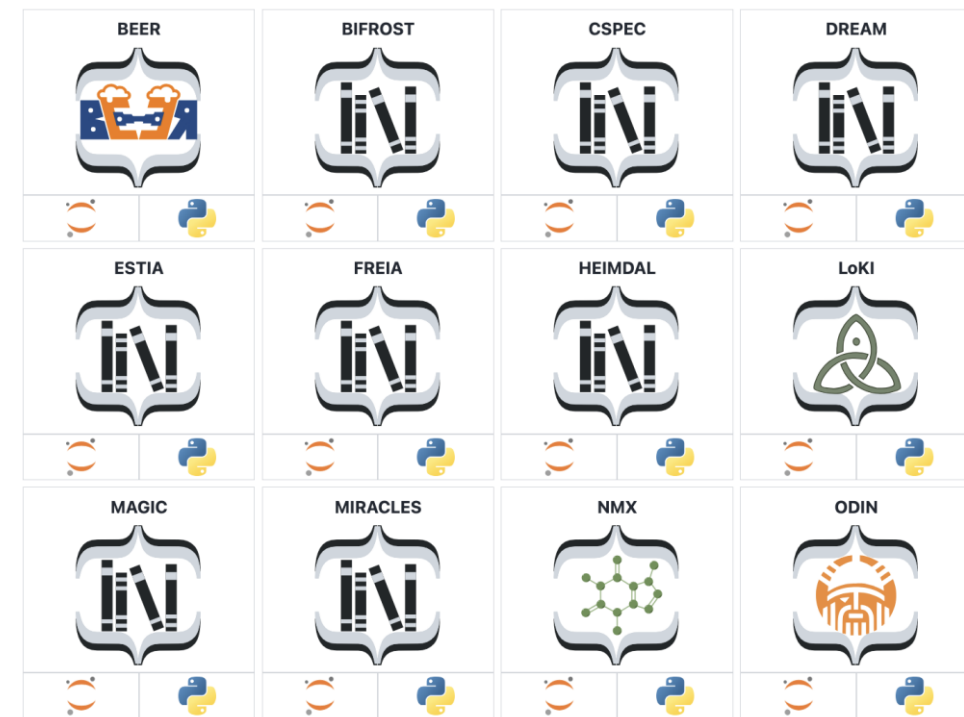
The repository for instrument-specific analysis scripts and notebooks



<https://dram.docs.esss.dk/code-shelf/code-shelf/>

- Code Shelf is a space where instrument-specific notebooks and scripts can be accessed, copied, and kept up to date.
- Code Shelf notebooks contain metadata for tracking when files were created and modified.
- The IDS/CIS is the primary responsible person for instrument-specific Code Shelf content.
- Discuss use of Code Shelf with your IDS/CIS.

Instruments



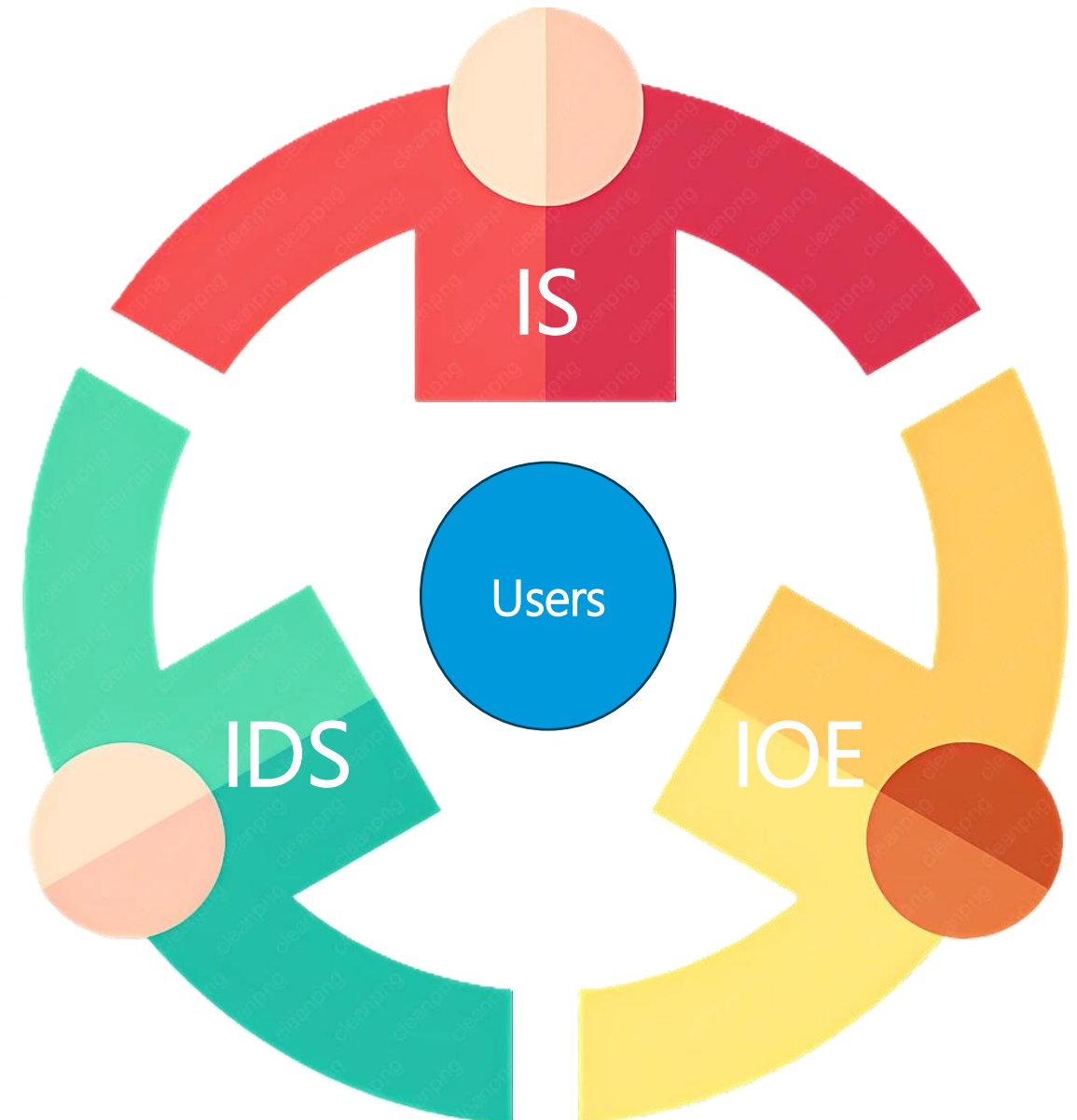
(+ the other instruments...)

IDS/CIS during User Operation

All Instrument Staff are User Resources



- All instrument staff should be able to assist users in any user-oriented task, whether before, during, or after the experiment.
 - This includes ISes, IDS/CISes and IOEs.
- Instrument scientists should be familiar with all data processing procedures for users. IDS/CIS should be able to fully operate the instrument with users.
- No single points of failure during routine user operation.



IDS/CIS after User Operation

All Instrument Staff are User Resources



- IDS/CIS are the primary contacts for data analysis pipelines.
- User analysis should be done on VISA. We have HPC resources if needed for compute-intensive jobs.
- IDS/CIS will *assist* users with data analysis as needed. If IDS/CIS are analyzing the data for users, then collaborative credit should be made in publications.
- While IDS/CIS should be heading the effort for data analysis with users, IS/IOEs should also be able to assist users in this task (much like IDS/CIS should be able to run the instrument with users!)



IDS or CIS?

That is the question...



- About a year ago a decision was made to change the Instrument Data Scientist (IDS) name to Computational Instrument Scientist (CIS).
- Reasons:
 - De-emphasize “data science” role (data science is involved, but not the way most self-described “data scientists” think it does– attracting less-than-ideal candidates...
 - Emphasize “instrument science” role
- You can call us either IDS or CIS for now. We’ll probably migrate to all-CIS later.





EUROPEAN
SPALLATION
SOURCE

Thank you!

aaron.finke@ess.eu



ess.eu



@ess.neutron



@ess-neutron



@essneutron