A semantic framework for modelling and analysing software supply chain through SBOMs

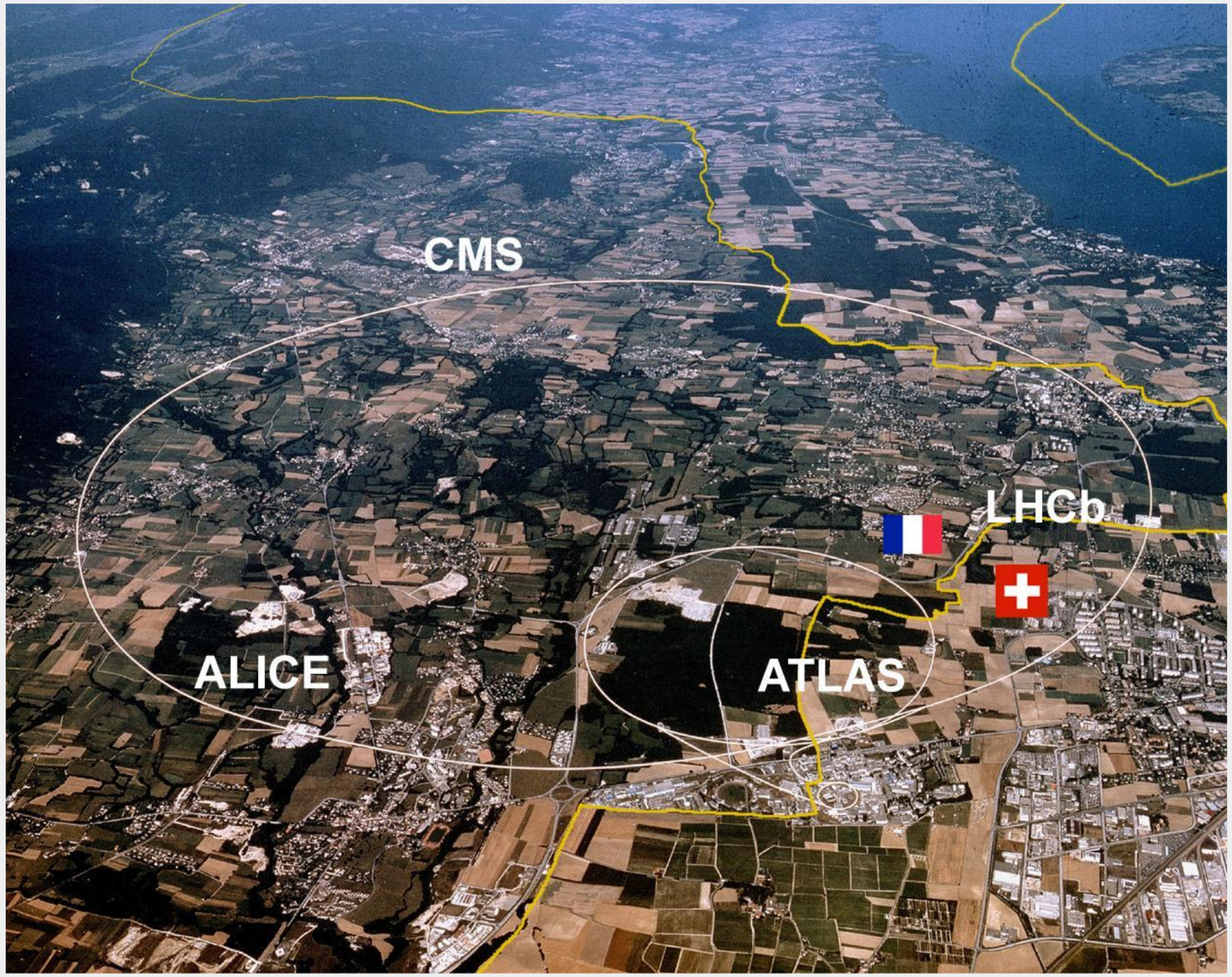
Giacomo Tenaglia (CERN)
Gianluca De Bonis (Bologna University)

FOSDEM - 1st February 2026

CERN Open Science

CERN | OSPO

CERN Mission



- Perform world-class research in fundamental physics;
- Provide particle accelerator facilities in an environmentally responsible and sustainable way;
- Unite people from all over the world to push the frontiers of science and technology;
- Train new generations of physicists, engineers and technicians;
- Engage all citizens in research and in the values of science.

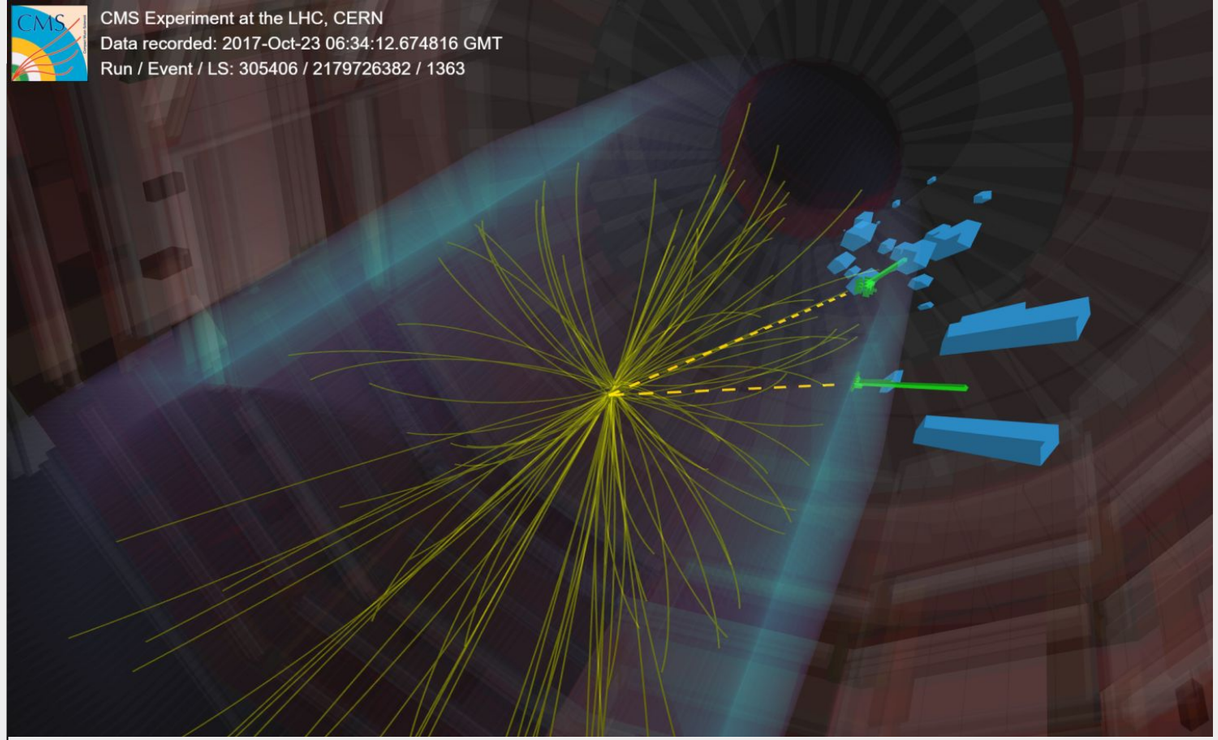
CERN
Organisation and Statistics

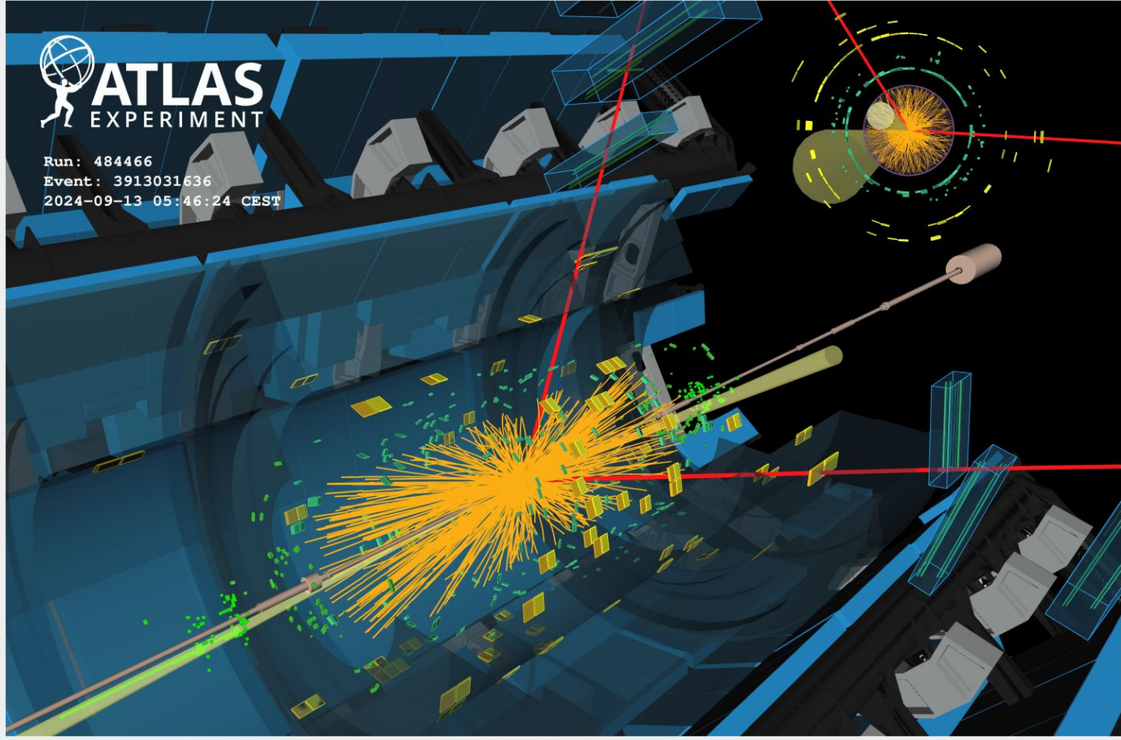
- 25 Member States
- 1 Pre-stage Membership States
- 8 Associate Member States
- 4 Observers

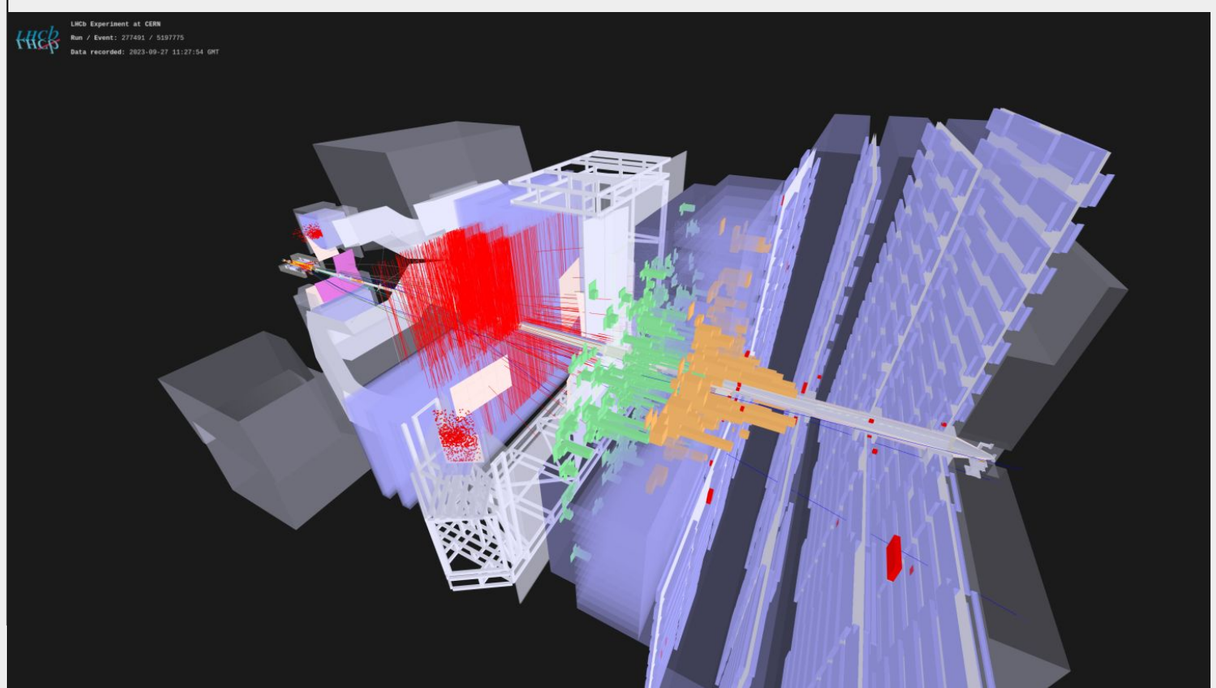
- 3,800+ Personnel
- 2,000+ Contractors
- 15,000+ Engineers, Scientists and Researchers
- 110+ Nationalities

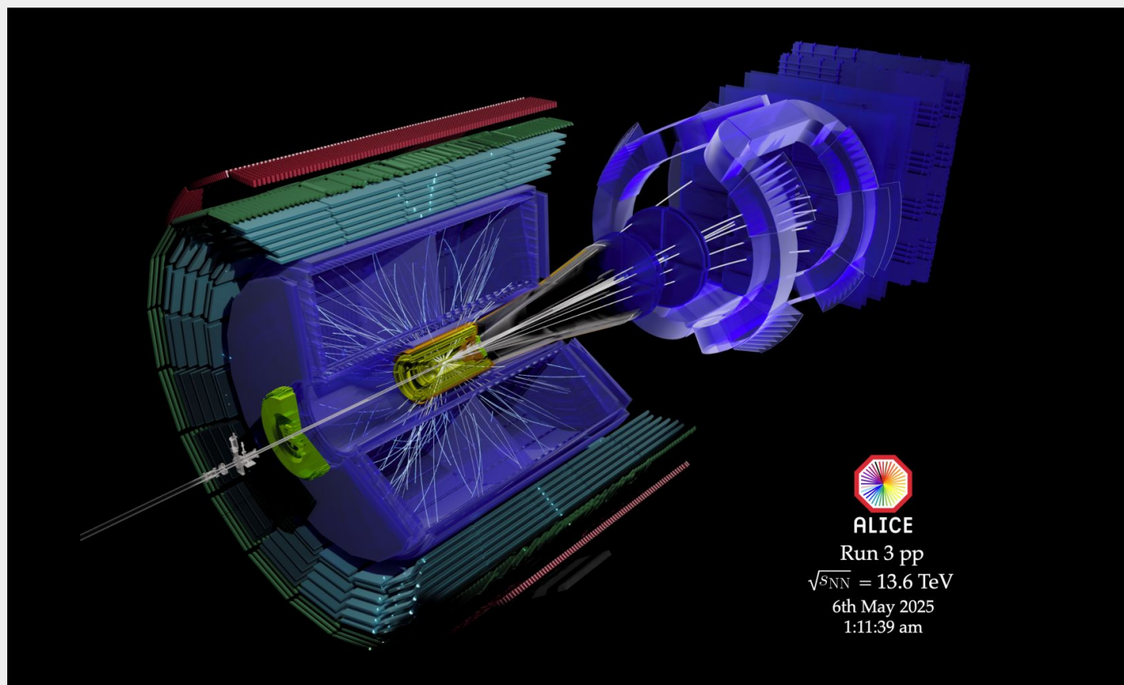
CERN | OSPO

CERN Open Science





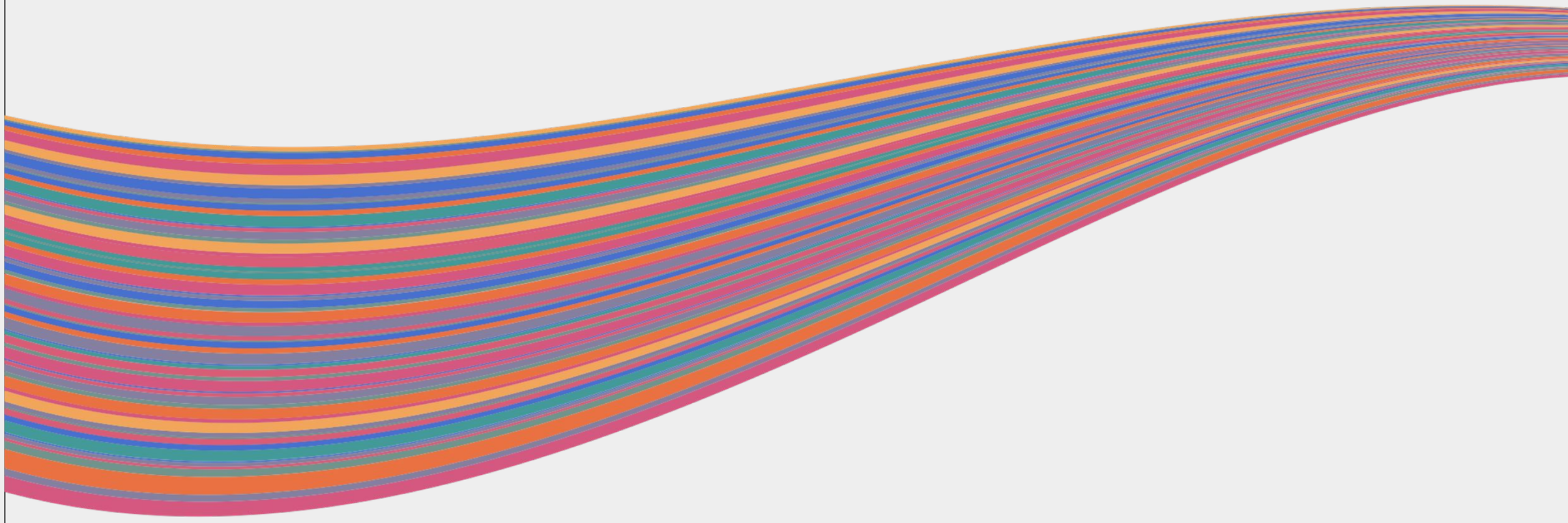




G. Tenaglia, G. De Bonis

“… and the results of its experimental and theoretical work shall be published or otherwise made generally available.”

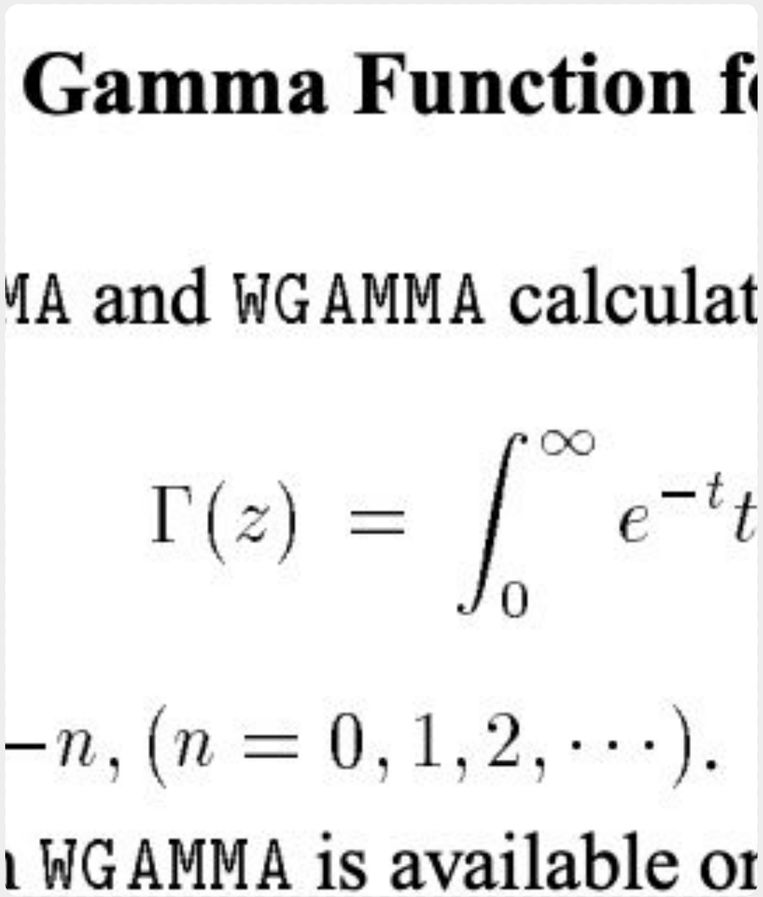
CERN CONVENTION, 1953



Free/Open Source SW and HW at CERN
Milestones

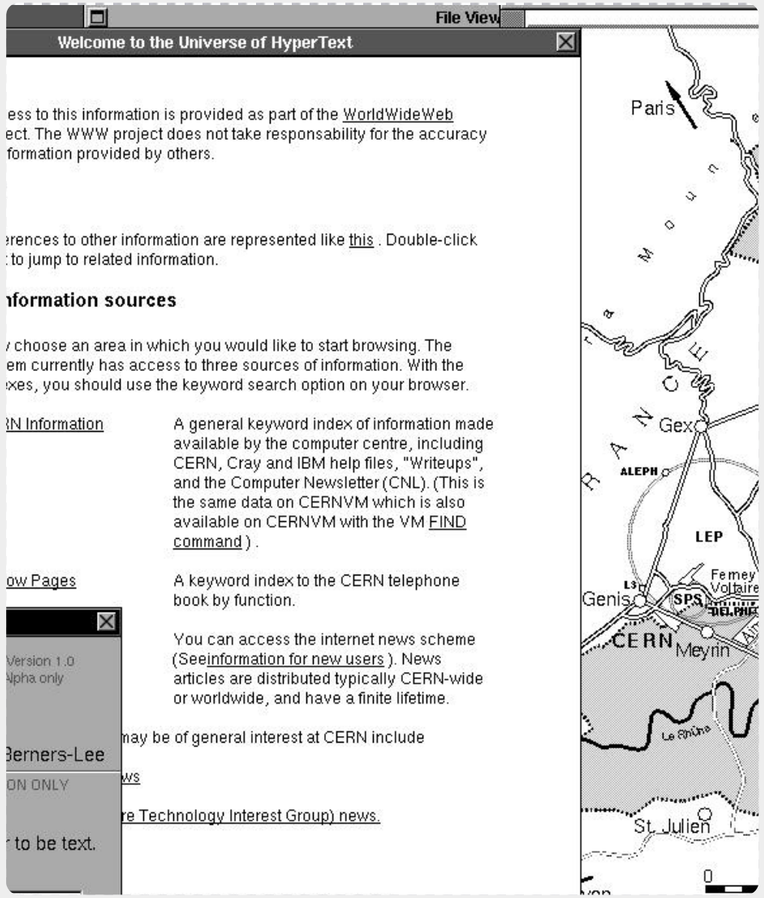
1970

CERN School of Computing



1983-4

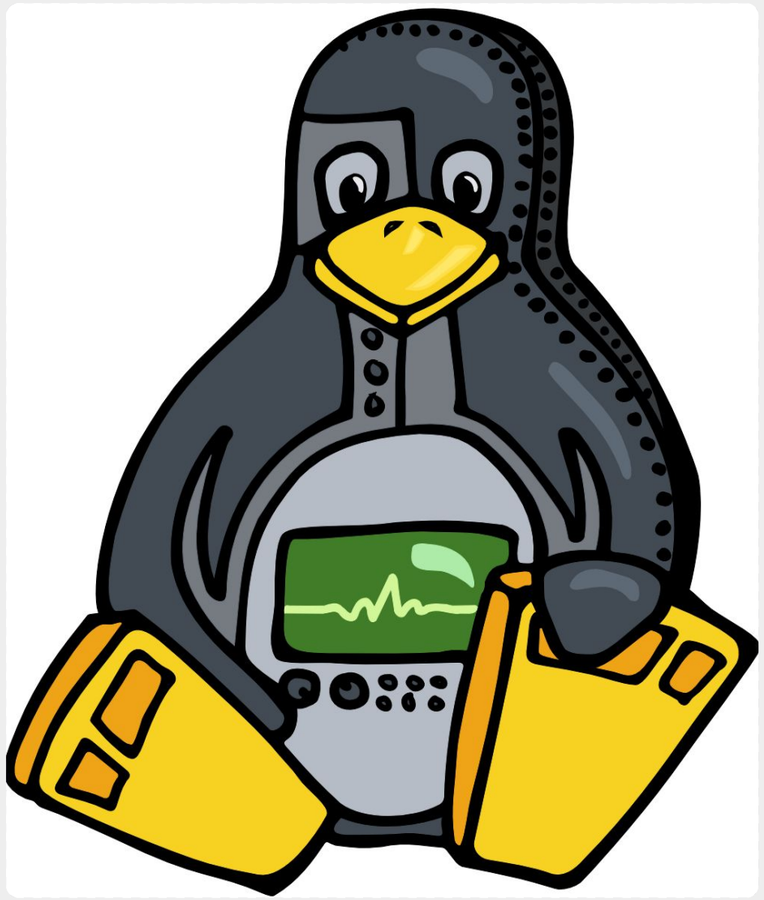
HEPVM & cernlib



1994

Licensing the web as FOSS

Free/Open Source SW and HW at CERN
Milestones

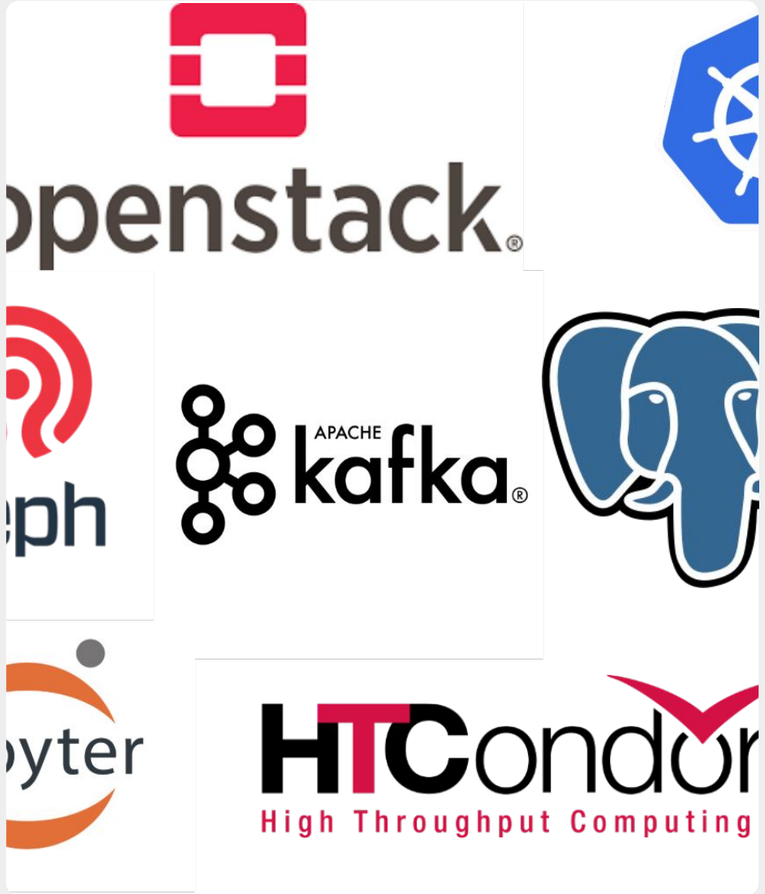


2011

CERN Open Hardware Licence

2012

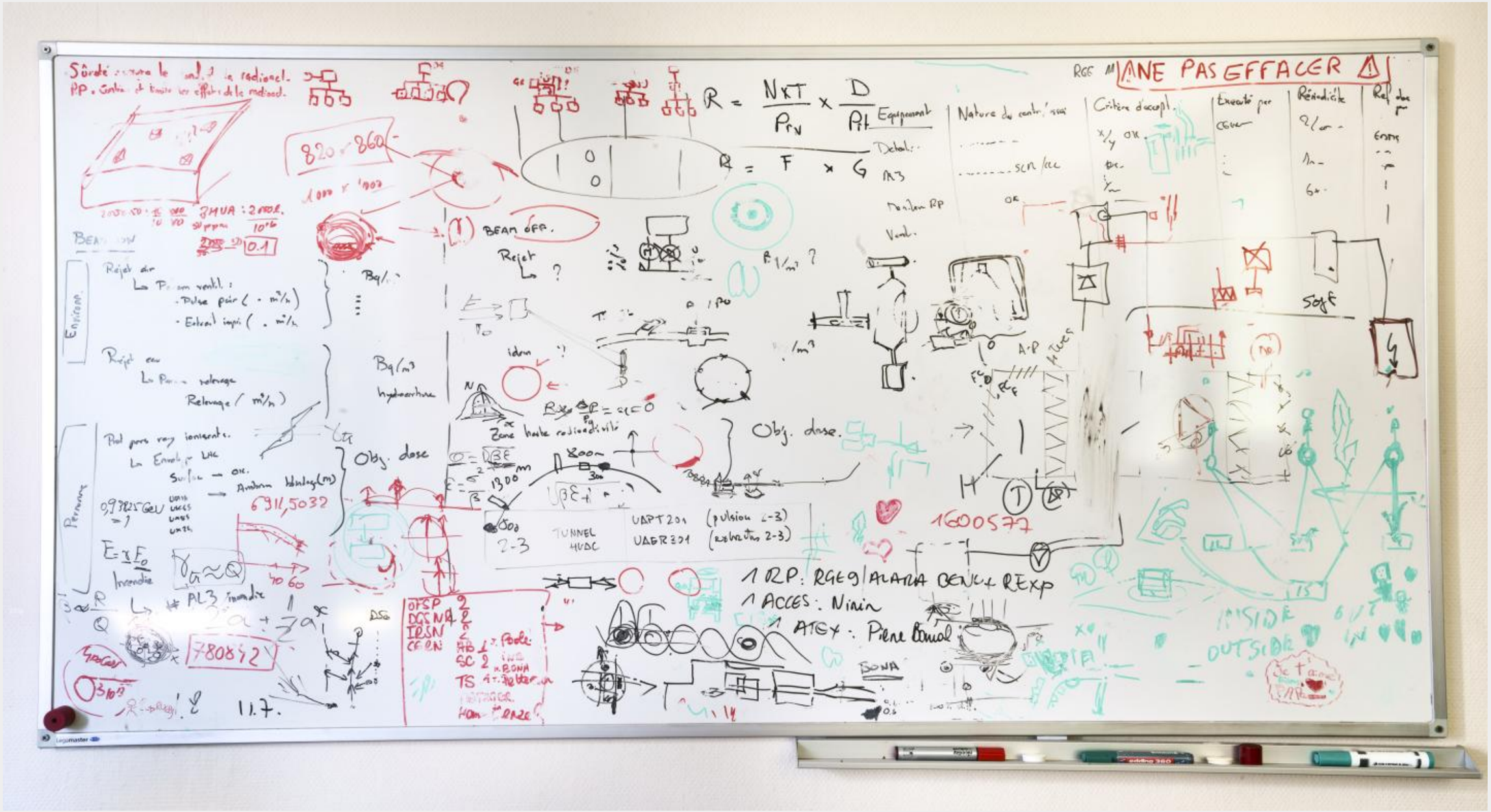
Open Source License TF



2013+

Massive FOSS tools adoption

A systemic approach to Free/Open Source SW/HW



CERN | OSPO

CERN Open Science

Open Source Program Office Mandate (2023)

INTERNAL MANDATE

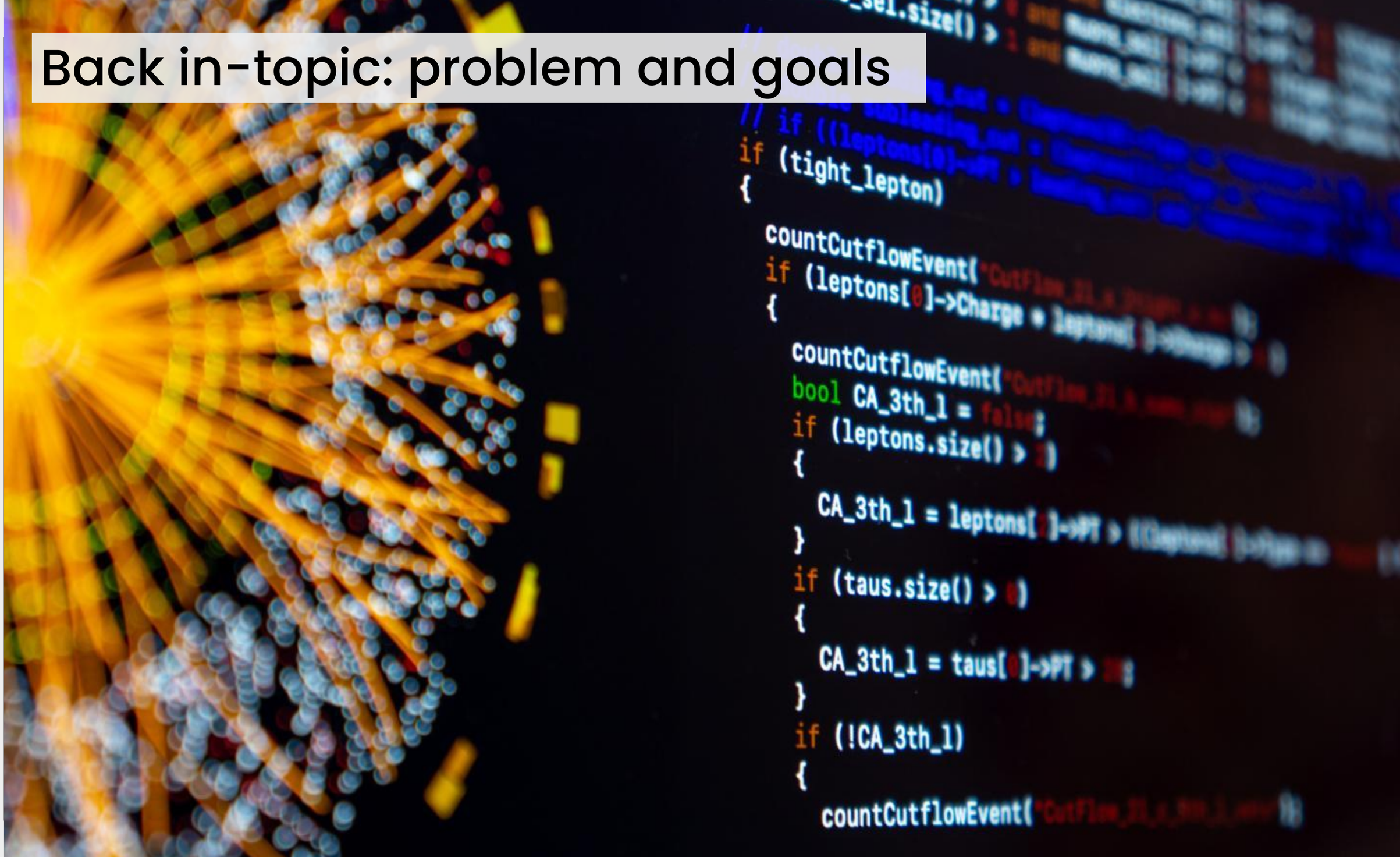
- Consult, advise, train on Open Source best practices, tools licenses, etc.
- Advises on open-sourcing CERN software and hardware.
- **Identify dependencies and compatibility for critical services.**
- Advises CERN on Open Source matters.

EXTERNAL MANDATE

- Showcase CERN's Open Source contributions
- Facilitate partnerships with external entities, e.g. companies.
- Promote CERN as an Open Source lab.

Mandate: cds.cern.ch/record/2879995

Back in-topic: problem and goals



Back in-topic: problem and goals

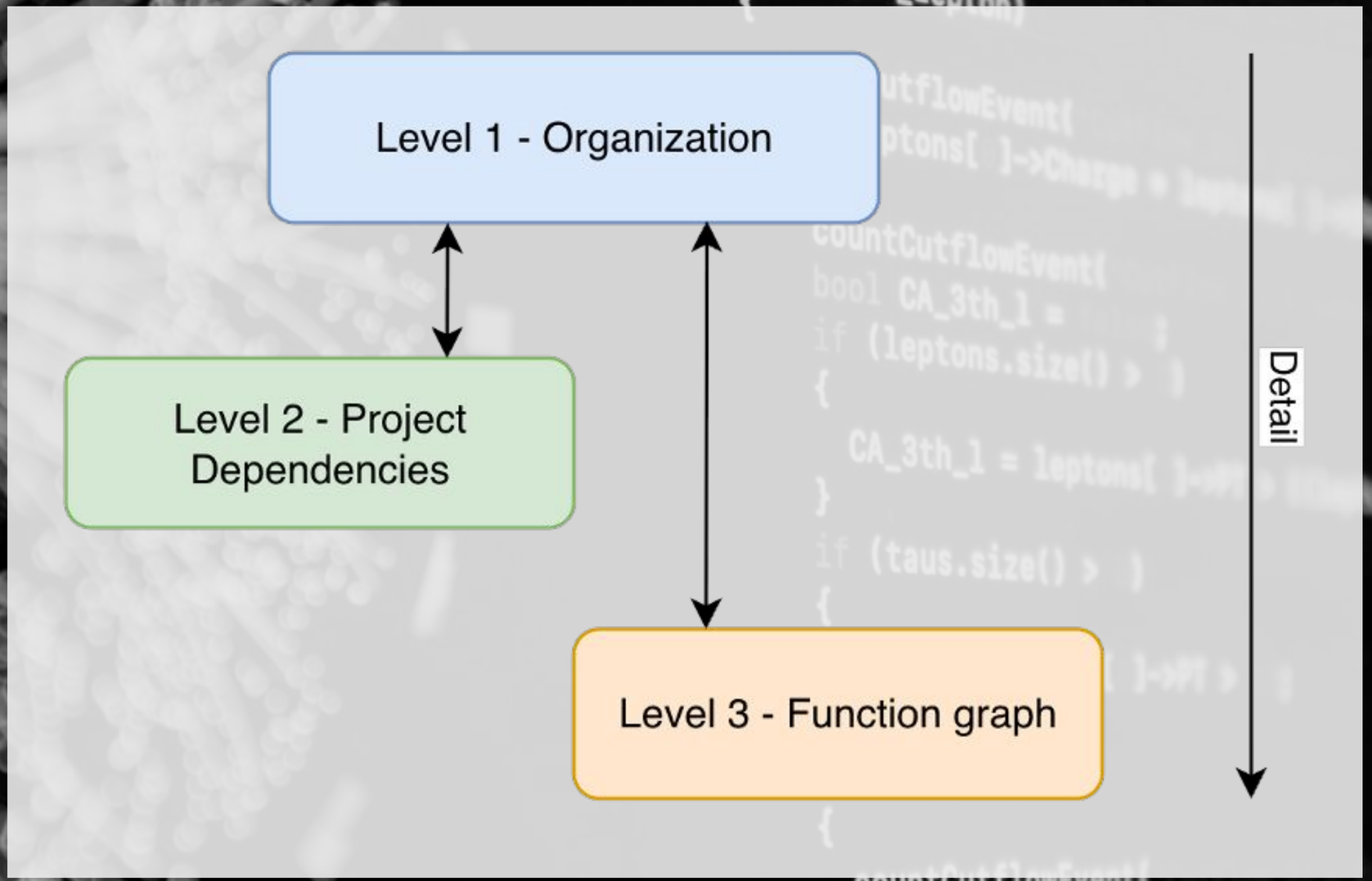
Problem

- **331** software services.
- **140.000+** internal repositories.
- Diverse programming language ecosystem.
- Distributed development teams.
- “Academic freedom” tradition.

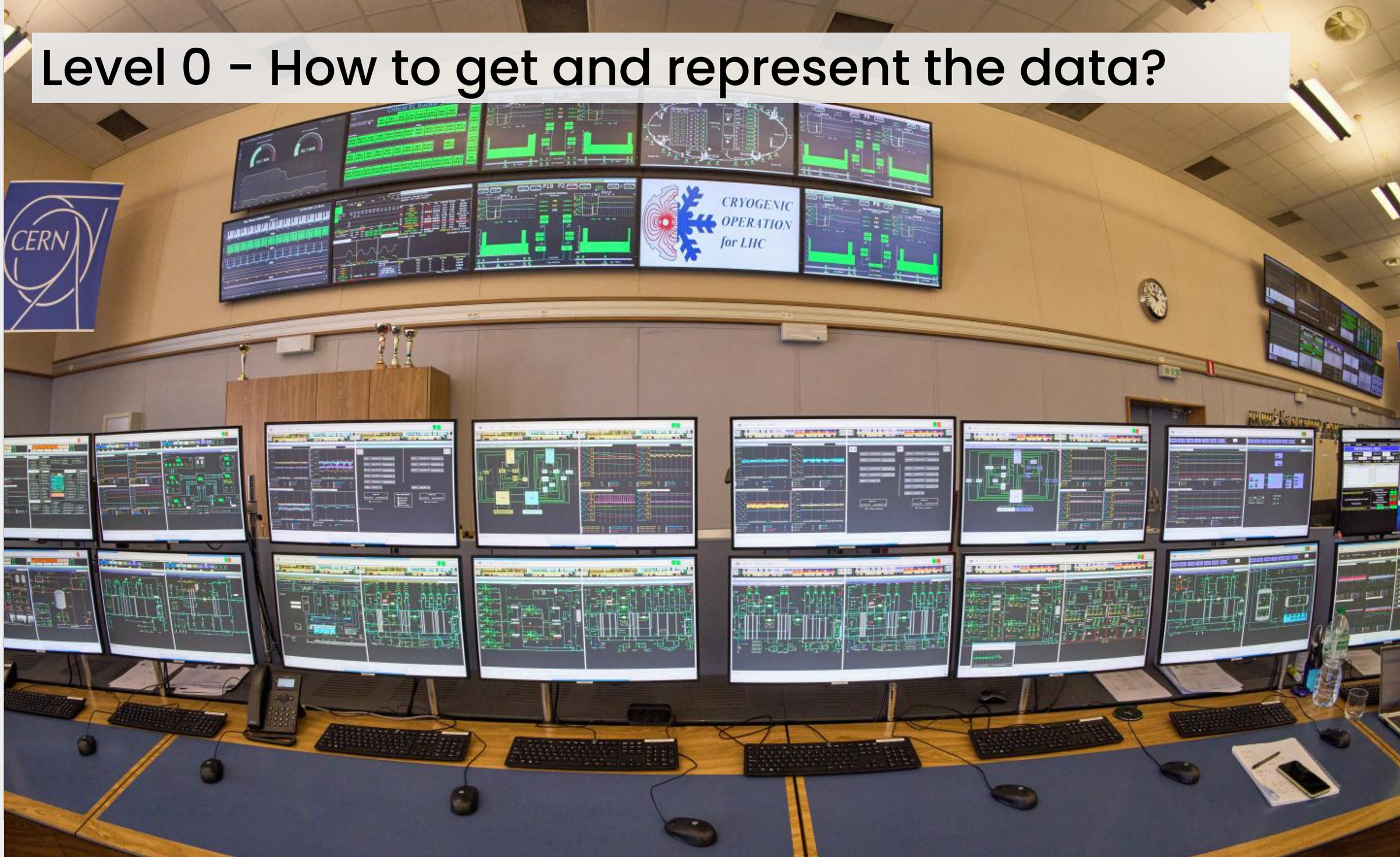
Goals

- **Correlate** with services and org structure.
- **Security angle:** SBOM investigation/PoC.
- **OSPO angle:** “big picture” view of the FOSS landscape and dependencies at CERN.

Back in-topic: a knowledge representation approach



Level 0 - How to get and represent the data?



Level 0 - How to get the data?

CERN Open-Source Software Dependency Survey 2025

Submission modes:
- Upload SBOM
- Upload survey form
  - CSV
  - Free-text

Email	**Organizational unit** (Department-Group-Section)	**Service element**	**Application name**	**Main open-source package used**	**Any other relevant open-source packages used**

G. Tenaglia, G. De Bonis

Level 0 - How to represent the data?

Ontology: *formal model of knowledge defining concepts, relationships and rules describing a domain.*

- **Concepts (Classes):** main entities in the domain
  - Package, Version, Vulnerability, License, …
- **Relationships (Properties):** how those entities are connected
  - depends_on, affected_by, has_license, …
- **Semantics:** explicit and machine-readable meaning (!= JSON/DB schema)
- Formalisation: OWL/RDF, enables reasoning/interoperability

Knowledge Graphs: *network of entities/relationships representing knowledge in a structure and interconnected way, can use ontologies as schema layer.*

```
[Component A] --depends_on--> [Component B] --affected_by--> [CVE-2024-XXXX]
```

Level 0 - How to represent the data?

Competency questions: *natural language questions outlining and constraining the scope of knowledge represented in an ontology.*

Some examples:

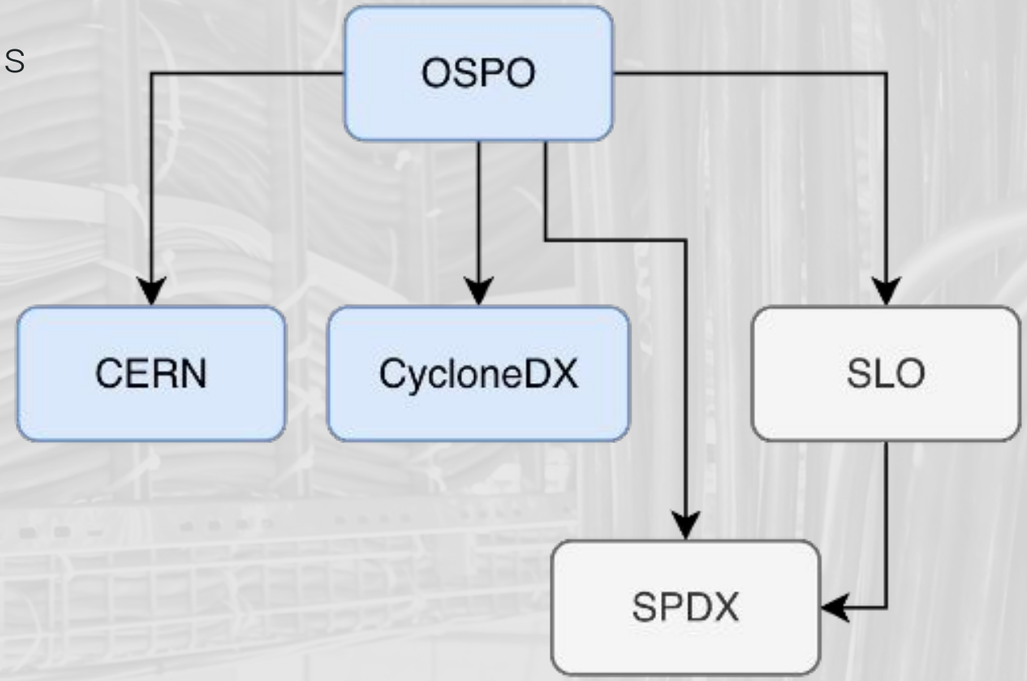
1. What are the **most used libraries/languages/frameworks**? Which ones are the most used **in a certain department/group**?
2. Which projects have a **certain known vulnerability**?
3. What are the **known vulnerabilities** present in a certain **IT Service**?
4. What **licenses** are declared in a project?
5. Which **tools** have been used to **generate the SBOMs**?

Level 1 - Organizational Level

Level 1 - Organizational Level

Custom ontology definitions:

- **CycloneDX**
  - Components, dependencies, vulnerabilities
- **CERN**
  - Org structure, service taxonomy
- **OSPO**
  - Glues everything together to answer competency questions



And import of:

- SPDX ontology
- Software License Ontology (SLO)
- SKOS, DCTERMS, FOAF, …

Raw data → Knowledge Graph

SPARQL queries to answer Competency Questions.

ospo-ont (http://example.org/ospo-ont#)

Active ontology × | Entities × | Individuals by class × | DL Query ×

Classes | Object properties | Data properties | Annotation properties | Datatypes | Individuals

Individuals: slo:GPL-3.0

- :SBOMType-Design
- :SBOMType-Operational
- :SBOMType-Runtime
- :SBOMType-Source
- :SBOMTypeScheme
- slo:0BSD
- slo:AFL-3.0
- slo:AGPL-3.0
- slo:APACHE-2.0
- slo:ARTISTIC-2.0
- slo:BSD-2-CLAUSE
- slo:BSD-3-CLAUSE
- slo:BSD-3-CLAUSE-CLEAR
- slo:BSD-4-CLAUSE
- slo:BSL-1.0
- slo:CC-BY-4.0
- slo:CC-BY-SA-4.0
- slo:CC0-1.0
- slo:CECILL-2.1
- slo:CommercialUsePermission
- slo:Copyleft
- slo:CopyrightNoticeCondition
- slo:CreativeWorkRecommendation
- slo:DatabaseRecommendation
- slo:DataRecommendation
- slo:DiscloseSourceCondition
- slo:DistributionPermission
- slo:ECL-2.0
- slo:EPL-1.0
- slo:EPL-2.0
- slo:EUPL-1.1
- slo:EUPL-1.2
- slo:GPL-2.0
- **slo:GPL-3.0**
- slo:ISC
- slo:LGPL-2.1
- slo:LGPL-3.0
- slo:LiabilityLimitation
- slo:LicenseNoticeCondition
- slo:LPPL-1.3c
- slo:MIT
- slo:ModificationPermission
- slo:MPL-2.0
- slo:MS-PL
- slo:MS-RL
- slo:NCSA
- slo:NetworkUseIsDistributionCondition

slo:GPL-3.0 — http://www.softwarelicenseontology.org#GPL-3.0

Annotations | Usage

Annotations: slo:GPL-3.0

Annotations

Description: slo:GPL-3.0

Types
- slo:SoftwareLicense

Same Individual As

Different Individuals

Property assertions: slo:GPL-3.0

- slo:hasCondition slo:CopyrightNoticeCondition
- slo:hasCondition slo:DiscloseSourceCondition
- slo:hasCondition slo:LicenseNoticeCondition
- slo:hasCondition slo:SameLicenseCondition
- slo:hasCondition slo:StateChangesCondition
- slo:hasLicenseType slo:Copyleft
- slo:hasLimitation slo:LiabilityLimitation
- slo:hasLimitation slo:WarrantyLimitation
- slo:hasPermission slo:CommercialUsePermission
- slo:hasPermission slo:DistributionPermission
- slo:hasPermission slo:ModificationPermission
- slo:hasPermission slo:PatentUsePermission
- slo:hasPermission slo:PrivateUsePermission
- slo:isRecommendedFor slo:SoftwareRecommendation

Data property assertions
- slo:hasChoosealicenseDescription "Permissions of this strong copyleft license are conditioned on making available complete source code of licensed works and modifications, which include larger works using a licensed work, under the same license. Copyright and license notices must be preserved. Contributors provide an express grant of patent rights."
- slo:isFsfLibre true
- slo:isOsiApproved true
- slo:isOsiPopular true

Git: **master**

To use the reasoner click Reasoner > Start reasoner | Show Inferences

Level 1 - Data analysis

SBOM -> RDF

- Formalise org data in RDF form
- Get missing SBOMs
- Convert survey data to CSV
- Normalise the CSV
- Convert normalised CSV + SBOMs to RDF
- Run LIMES
- Run ontology query engine

Level 1 - Data analysis

```sql
# Most used libraries used by a given organisational unit
PREFIX ospo: <http://example.org/ospo-ont#>
PREFIX cern: <http://example.org/cern#>
PREFIX dcterms: <http://purl.org/dc/terms/>
SELECT ?orgUnit ?library ?libraryTitle (COUNT(?library) AS ?usageCount)
WHERE {
    ?orgUnit a cern:OrgUnit ;
        ospo:usesSoftware ?app .
    ?app ospo:usesLibrary ?library .
    ?library dcterms:title ?libraryTitle .
}
GROUP BY ?orgUnit ?libraryTitle
ORDER BY DESC(?usageCount)
LIMIT 50

# Get all versions of a certain library and the projects where they are used
PREFIX ospo: <http://example.org/ospo-ont#>
PREFIX dcterms: <http://purl.org/dc/terms/>
SELECT ?library ?libraryVersion (GROUP_CONCAT(?applicationTitle; separator=", ") AS ?applications)
WHERE {
    ?library a ospo:SoftwareVersion ;
        dcterms:title $LIB_NAME ;
        ospo:version ?libraryVersion .
    ?application ospo:dependsOn ?library ; # TODO: link to cern:Application instead of other libraries.
        dcterms:title ?applicationTitle .
}
GROUP BY ?library ?libraryVersion
```

Level 2 - Project Dependencies Level

CERN Open Science

CERN | OSPO

Level 2 - Project Dependencies Level

Based on **Vulnerability-Dependency Graph (VDGraph)**

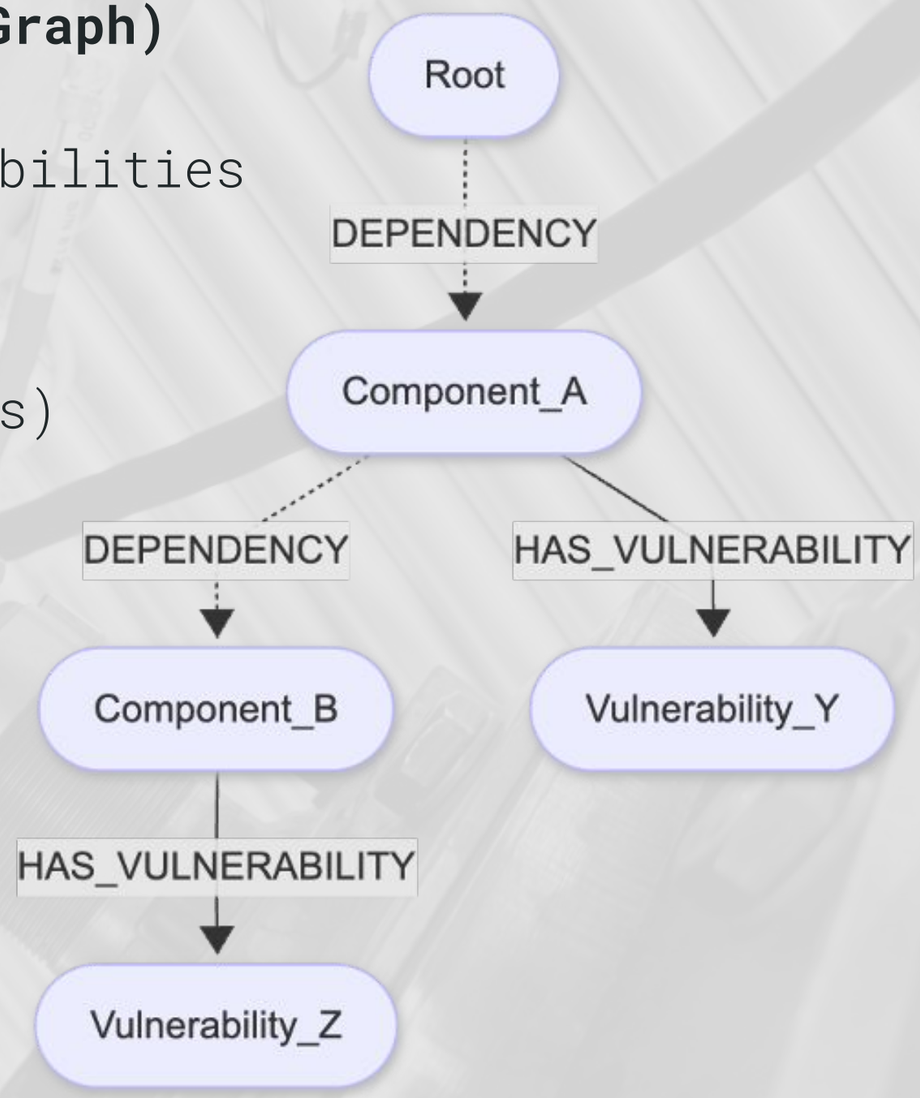
- https://arxiv.org/abs/2507.20502
- Holistic view on dependencies and vulnerabilities

Queried databases:

- **https://deps.dev/** (transitive dependencies)
- **https://osv.dev/** (known vulnerabilities)

Graph database for storage and analysis:

- **Neo4j** to store the data
- **Cypher** to query the data



neo4j Desktop

Feedback | Fully-managed cloud database Neo4j AuraDB

Instance: VDGraph | Database: vdgraph-db CYPHER 5 | User: neo4j

Data services

- Local instances
- Remote connections
- Import

Tools

- Query
- Explore

About

- Settings

Database information

Nodes (1,786)

* component Component Resource root Root vulnerability Vulnerability

Relationships (3,698)

* DEPENDENCY HAS_VULNERABILITY

Property keys

_applyNeo4jNaming _classLabel _classNamePropName _dataTypePropertyLabel _domainRel _handleMultival _handleRDFTypes _handleVocabUris _keepCustomDataTypes _keepLangTag _objectPropertyLabel _rangeRel _relNamePropName _subClassOfRel _subPropertyOfRel bom_ref data group id modified n4sch n4sch__comment n4sch__name n4sch__propCharacteristics name nodes onPropertyName onPropertyURI published

Last update: 12:51:42

vdgraph-db$

```
1 MATCH (n)
2 WHERE (n:Root OR n:Component)
3   AND EXISTS {
4     MATCH (n)-[:DEPENDENCY]->(:Component)
5   }
6 WITH COLLECT(n) AS nodes
7 UNWIND nodes AS n
8 MATCH (n)-[r]-(m)
9 WHERE m IN nodes
10 RETURN DISTINCT n, r, m;
```

Graph | Table | RAW

Node details

Component

Key	Value
<id>	4:fc44f2ca-9ebe-4ece-a058-e587fbd02235:343
bom_ref	"pkg:npm/array-buffer-byte-length@1.0.1"
group	""
name	"call-bind"
purl	"pkg:npm/call-bind@1.0.8"
source	"sbom"
system	"npm"
version	"1.0.8"

Started streaming 3,990 records after 381 ms and completed after 808 ms.

```
1 MATCH (c:Component)-[h:HAS_VULNERABILITY]->(v:Vulnerability)
2 RETURN c, v, h;
```

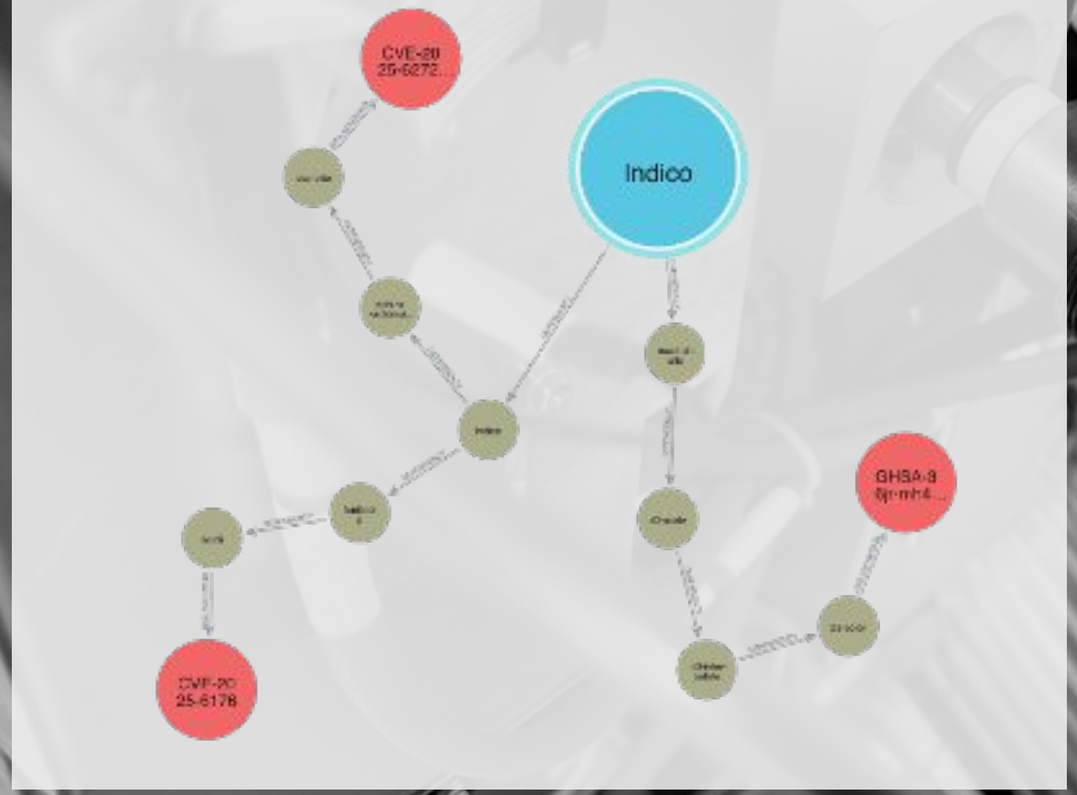
Graph | Table | RAW

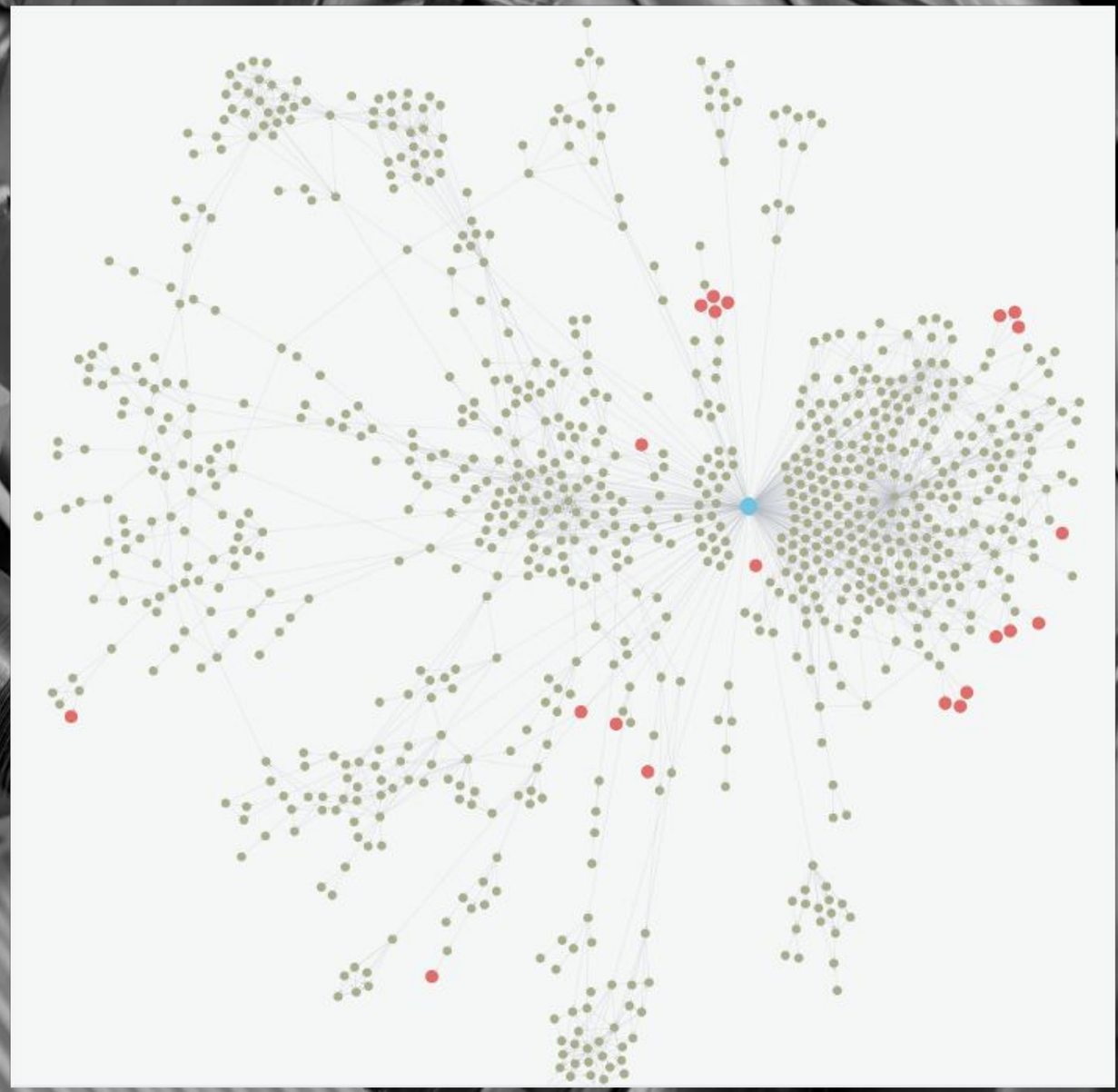
Node details

Component

Level 2 - VDGraph

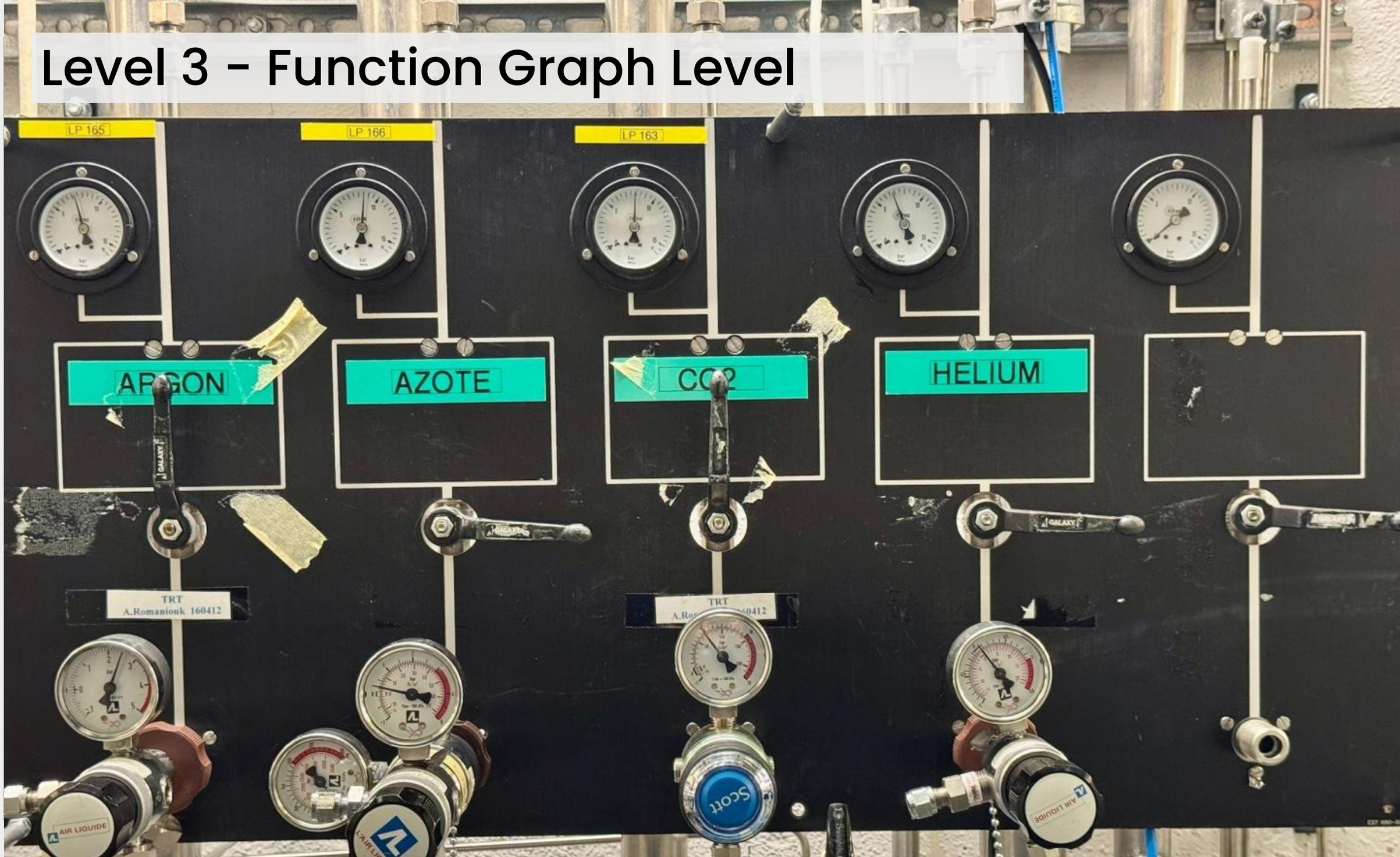
Example query: vulnerabilities with high severity score having a distance greater than 3 to the root node.





Example: **indico** / event management
https://getindico.io/

Level 3 - Function Graph Level



Level 3 - Function Graph Level

Graph Attention Network (GAT)

Based on *Profile of Vulnerability Remediations in Dependencies Using Graph Analysis* (*https://arxiv.org/abs/2403.04989*)

- **Structure:** functions as nodes, calls as edge
- **Purpose:** highlights critical/highly used functions depending on vulnerable libraries
- **Metrics:** centrality degree, betweenness, GAT score -> Risk score

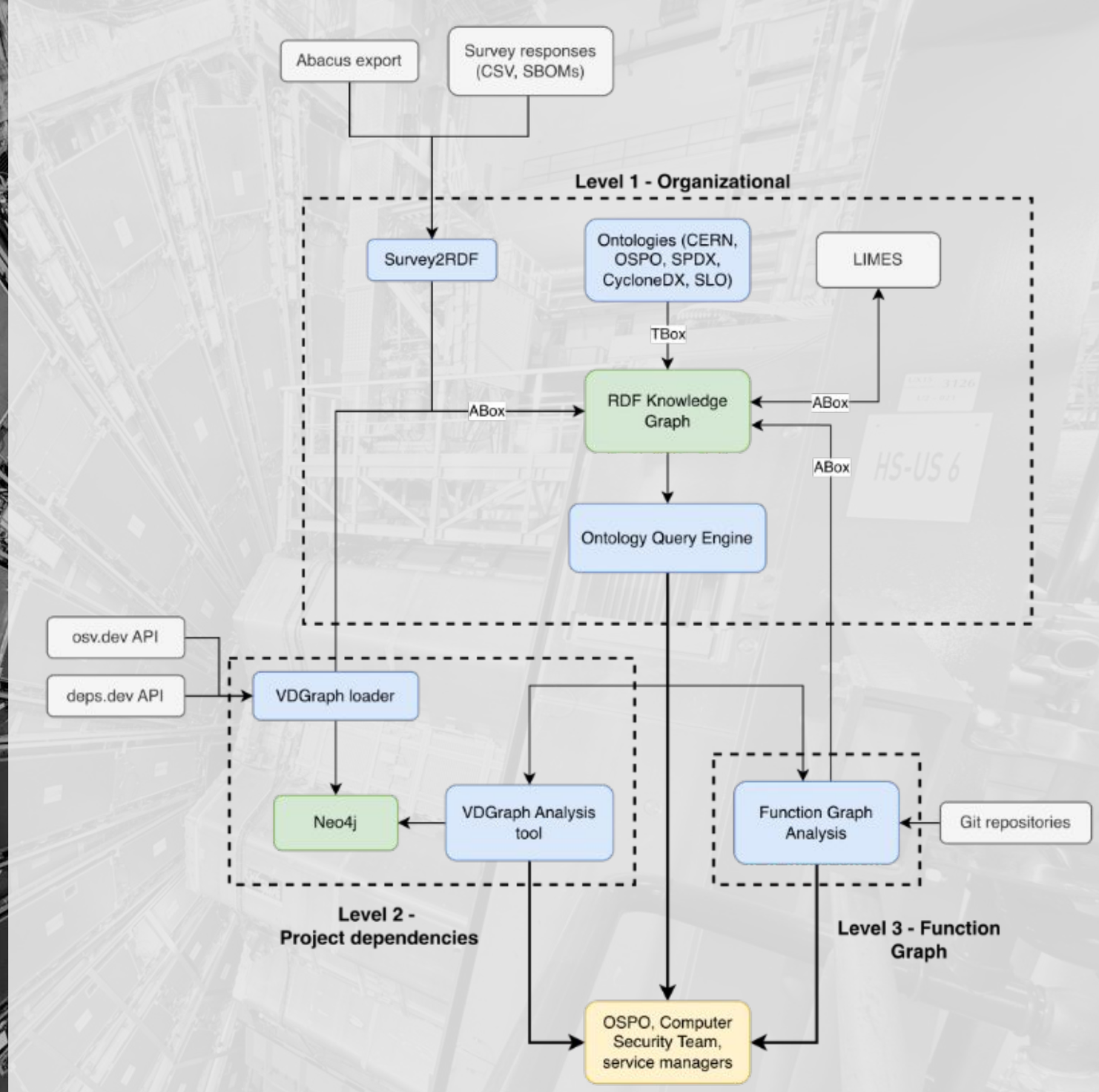
Function	Risk	GAT	Deg	Betw.	Vuln IDs
`report/ virus_report.py: prepare_context`	0.912	0.134	4	2.75e-04	GHSA-cpwx-vrp4-4pq7 GHSA-q2x7-8rv6-6q7h GHSA-gmj6-6f8f-6699

Final framework architecture

CERN | OSPO
CERN Open Science

Final framework architecture

- Blue: tools developed
- Green: data storage
- Grey: data sources
- Yellow: users



Conclusions and future work

Conclusions and future work

Lessons learned:
- Semantic framework: unify heterogeneous SBOM + survey data
- Ontologies: foundation for org-wide Knowledge Graph
- VDGraph / Function Graph: project/code level insight
- The supply chain FOSS tooling ecosystem is amazing (and vast!)

Next steps:
- Reproducibility:
  - Automation
  - Storage centralisation
  - Dashboard/easy way to query
- Technology evolution:
  - https://ecosyste.ms/ for dependencies
  - Neo4j / Cypher alternative

Thank You!

https://opensource.cern

Open.Source@cern.ch

Giacomo.Tenaglia@cern.ch
Gianluca.De.Bonis@cern.ch

Student? Apply now!
https://careers.cern/jobs/tsc-co/

CERN Open Science

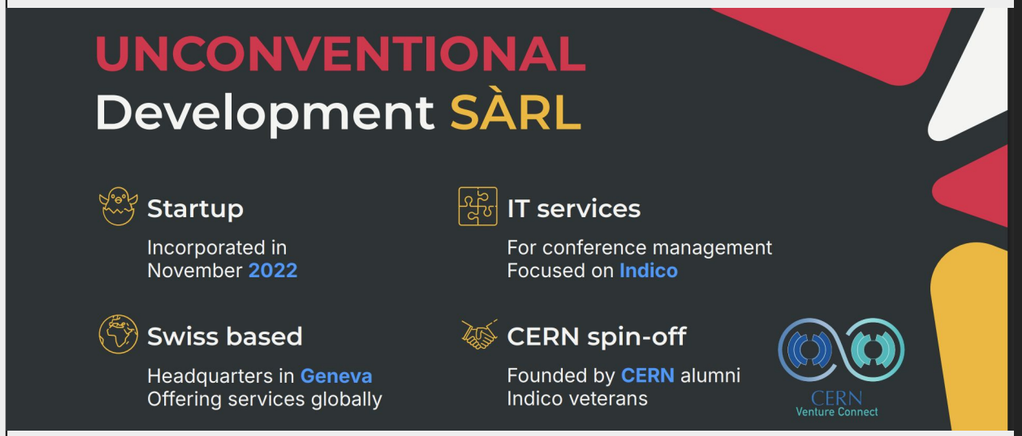
CERN | OSPO

Backup slides

CERN-made OSS

Indico https://getindico.io/

- From ”meeting organisation website” to general event management tool
- Hundreds of instances worldwide
- CERN leads the development
- Partnerships with SMEs / spinoffs for consultancy / development



CERN | OSPO

CERN Open Science

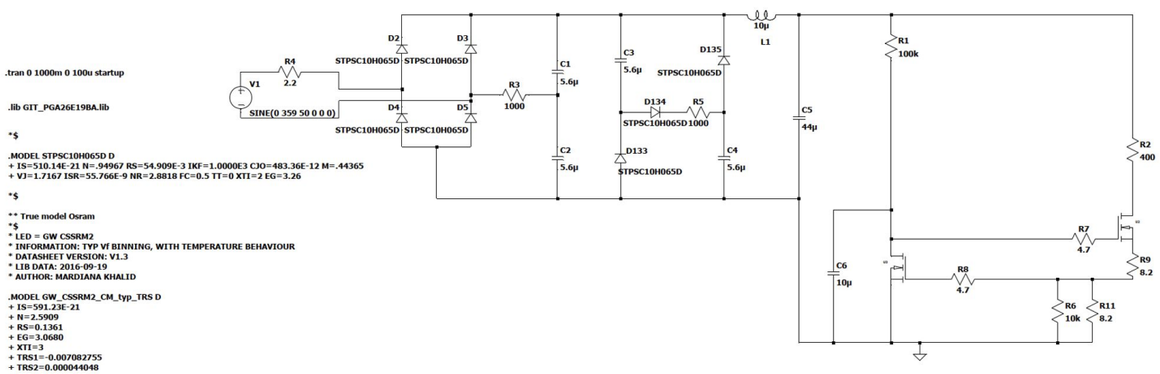
CERN-made OSHW

Radiation-tolerant lightning

- Need to replace 1k+ lights on CERN's Proton-Synchrotron accelerator
- BAU vs "specialist route" (4k / each) vs DIY
- Design -> CE certification -> procurement
- 415,- each
- Reused in Australia and the US facilities

Linear regulator in GaN with SiC diodes.

```
.tran 0 1000m 0 100u startup

.lib GIT_PGA26E19BA.lib

*$

.MODEL STPSC10H065D D
+ IS=510.14E-21 N=.94967 RS=54.909E-3 IKF=1.0000E3 CJO=483.36E-12 M=.44365
+ VJ=1.7167 ISR=55.766E-9 NR=2.8818 FC=0.5 TT=0 XTI=2 EG=3.26

*$

** True model Osram
*$
* LED = GW CSSRM2
* INFORMATION: TYP VF BINNING, WITH TEMPERATURE BEHAVIOUR
* DATASHEET VERSION: V1.3
* LIB DATA: 2016-09-19
* AUTHOR: MARDIANA KHALID

.MODEL GW_CSSRM2_CM_typ_TRS D
+ IS=591.23E-21
+ N=2.5909
+ RS=0.1361
+ EG=3.0680
+ XTI=3
+ TRS1=-0.007082755
+ TRS2=0.000044048
+ CJO=1.0000E-12

*$
```

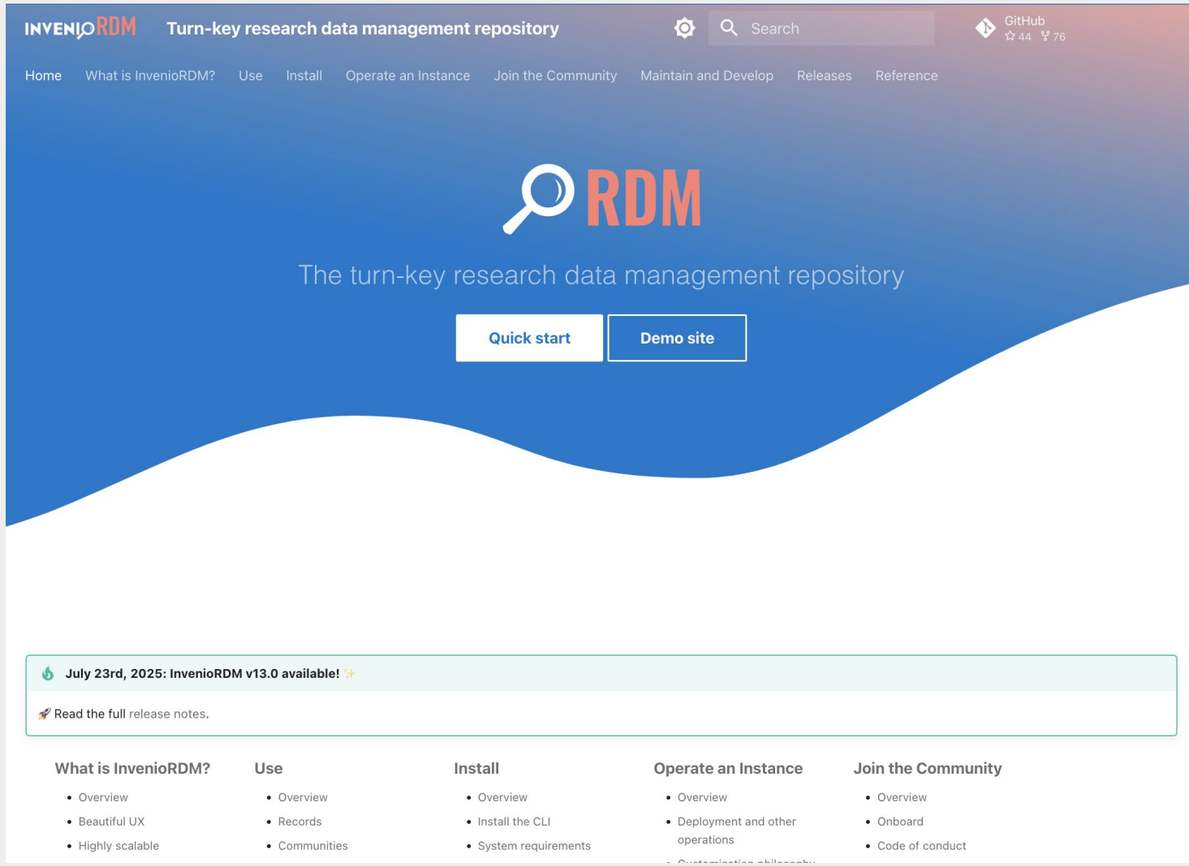


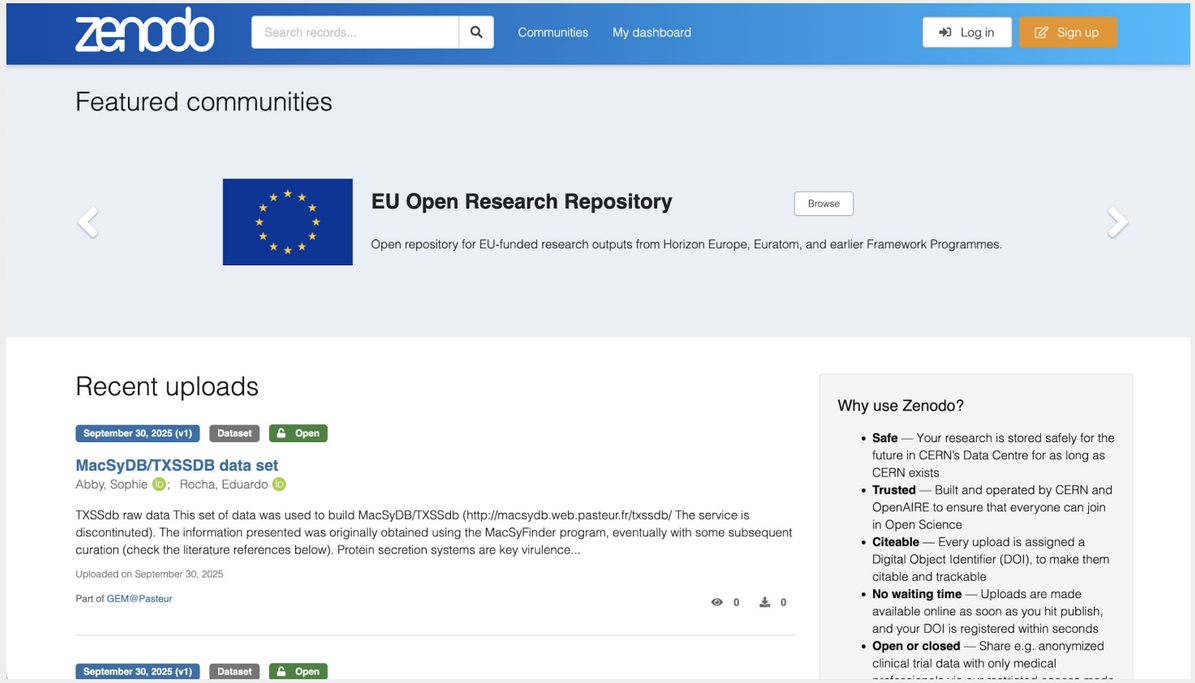
CERN-made OSS

InvenioRDM

https://inveniordm.docs.cern.ch/

- Research Data Management repository
- The technology underlying Zenodo.org
- From popular service to turn-key solution





Establish and foster collaborations

The White Rabbit collaboration

- Sub-nanosecond time synchronisation over Ethernet
- Part of IEEE 1588 PTP standard
- Reference implementation fully open-source hardware, gateware, firmware and software
- Application: finance sector, telecom, telescopes, quantum networks, …
- The Collaboration offers: technical support, training/workshops, like-minded community, possibility to use WR trademark for qualified products, …

White Rabbit
COLLABORATION

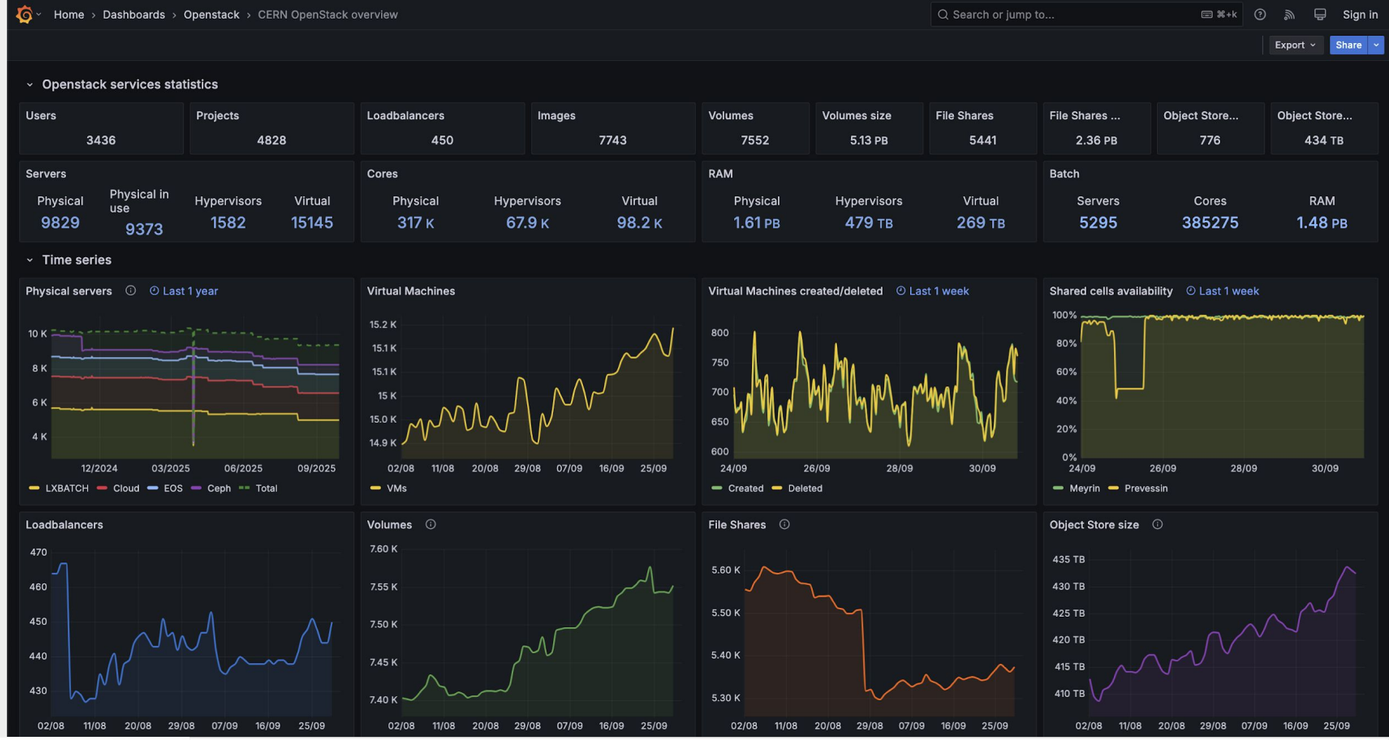
Members of the Collaboration

Nikhef · CERN · iol InterOperability Laboratory · SAFRAN · Sync-Tech · gmv INNOVATING SOLUTIONS · LMS LIQUID-MARKETS-SOLUTIONS The Power of the Possible™ · jumptrading · DEUTSCHE BÖRSE GROUP · MYRRHA · nu QUANTUM · IQD · SURF · Quincy Data · UNIVERSITÉ DE FRANCHE-COMTÉ · GSI FAIR · INFN · Fraunhofer HHI · UNIVERSITY OF MARYLAND · Fermilab · creotech

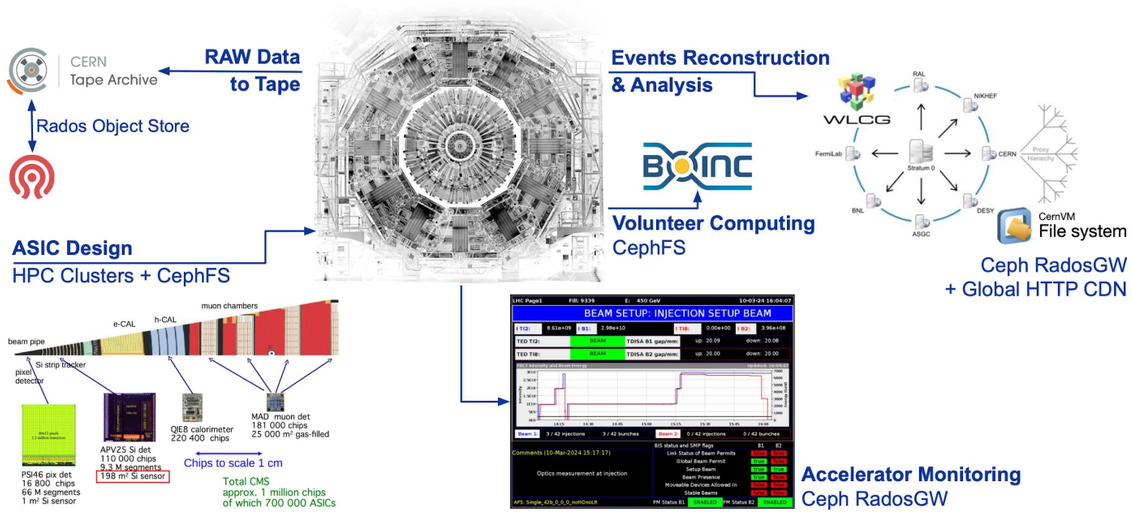
Scaling Open Source to CERN needs

Openstack and Ceph

- Powering the CERN Cloud since 2012
- From early adopter to power user
- Contributions based on extensive operational experience
- Significant community actor



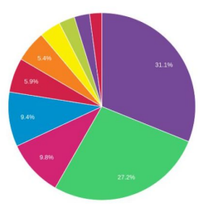
How Ceph + OpenStack support CERN's mission



How do we interact with the OpenStack community?

➔ We contribute code / docs!

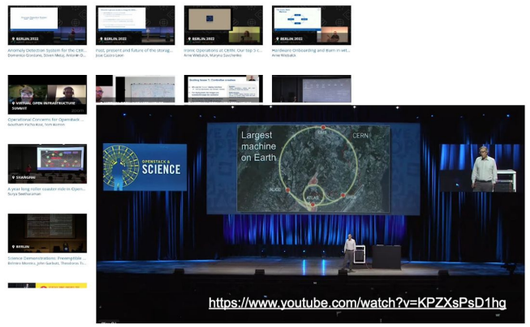
- We proposed and added features we needed
- From March 14, 2012 … to Jan 23, 2025
- 140'000 LOC, >1000 commits, 38 contributors



https://www.stackalytics.io/?release=all&metric=filed-bugs&company=cern

➔ We share *what we do* with it!

- Presentations at summits, user meetings, …
  "By contributing to open source you help us find the lost 96% of the universe!"
- https://www.openstack.org/videos/search?search=cern



Scaling Open Source to CERN needs

KiCAD

- 2011: CERN joins KiCAD development
- 2013: start of the donation programme
- 2015: EDA devroom @ FOSDEM
- 2023: stop of the donation programme, procuring standard support contract via CERN IT