INSA INSTITUT NATIONAL DES SCIENCES APPLIQUÉES LYON

LIRIS

Lyon 1

Accio: A Location Privacy Framework and Tool

Vincent Primault

DRIM seminar, January 27th 2017

What is Accio?

Accio is a scientific workflow management tool…
… designed towards spatio-temporal data…
… aimed a studying privacy.

Scientific workflow management

Location privacy algorithms

Scientific workflow management

A Survey of Data-Intensive Scientific Workflow Management. Liu, Ji and Pacitti, Esther and Valduriez, Patrick and Mattoso, Marta. *Journal of Grid Computing, Volume 13, Issue 14, December 2015.*

Location privacy study

- Only one work I am aware of!

- Implemented as an open source tool (C++).
- Quantification of privacy only (not utility).
- Strict probabilistic framework.

Quantifying Location Privacy. Shokri, Reza and Theodorakopoulos, George and Le Boudec , Jean-Yves and Hubaux , Jean-Pierre. *Security & Privacy, May 2011.*

Accio key features

Simple yet powerful DSL

Easy results analysis

Multi-tenant & scalable

Battery-packed for privacy study

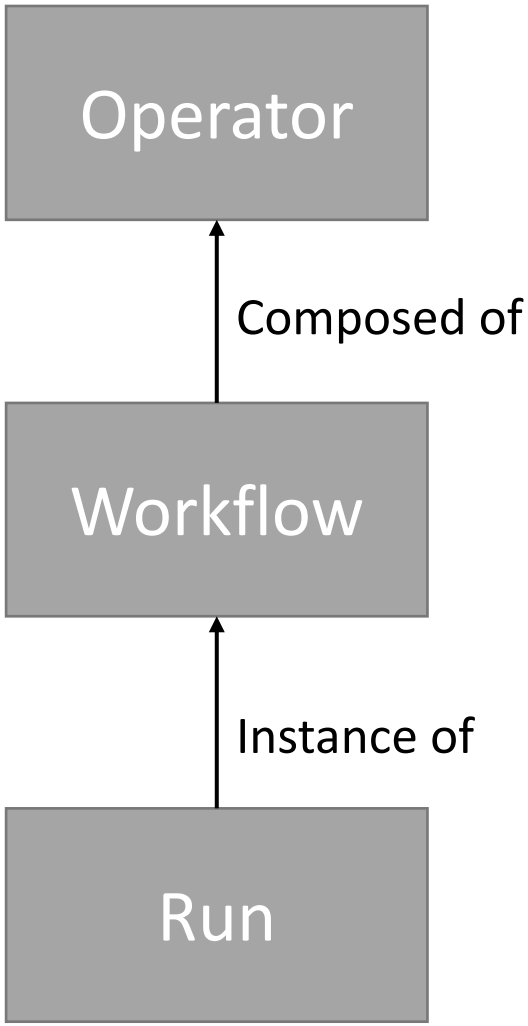
Outline

1. Introduction
2. **Accio, a workflow management tool**
3. Accio, a location privacy tool
4. Hands-on

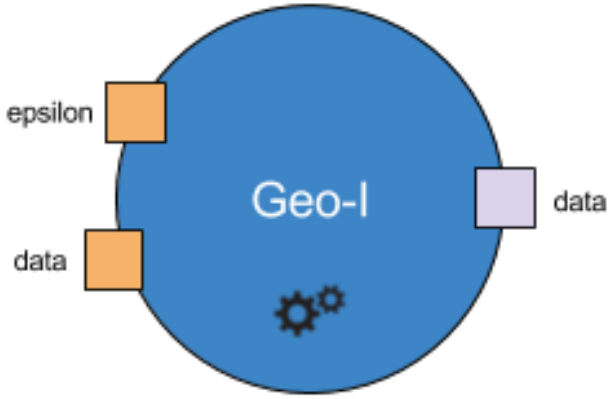
History

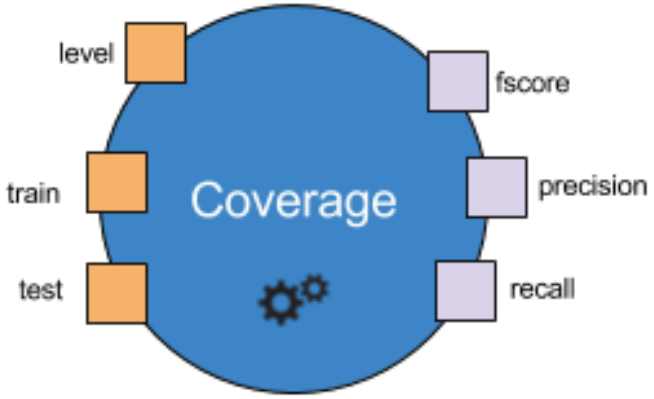
- **April 2016:** ALP, first configurable tool for a paper. Still very specific concepts.
- **September 2016:** Accio v1, direct successor of ALP. Sophie begins using it.
- **November 2016:** Accio v1.5, more generic concepts. Mohamed begins using it.
- **January 2017:** Accio v2, distributed version.

Core concepts



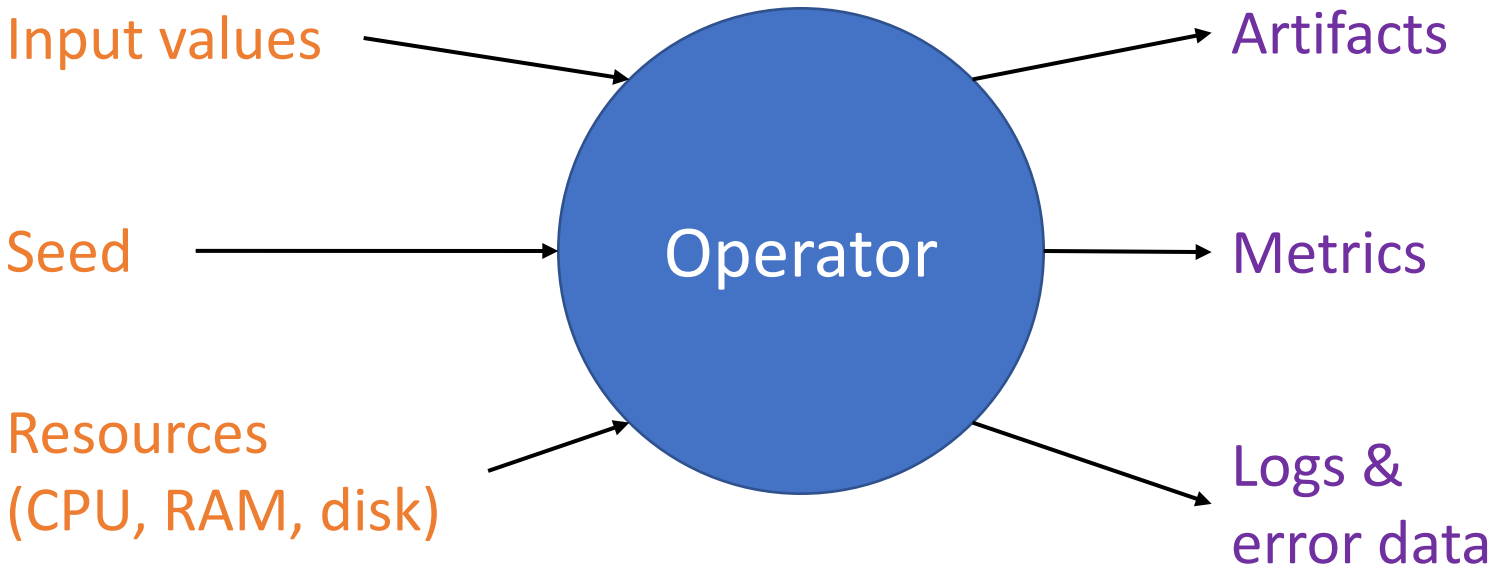
Operators





Input ports and Output ports

Operators


Input values
Seed
Resources
(CPU, RAM, disk)
Operator
Artifacts
Metrics
Logs &
error data

Type system

- Atomic types:
  - Numeric: byte, integer, long, double
  - Pseudo-numeric: **distance**, **duration**
  - Other: string, boolean, **location**, **timestamp**, **dataset**

- Collection types :
  - List (of atomic types)
  - Set (of atomic types)
  - Map (of atomic types)

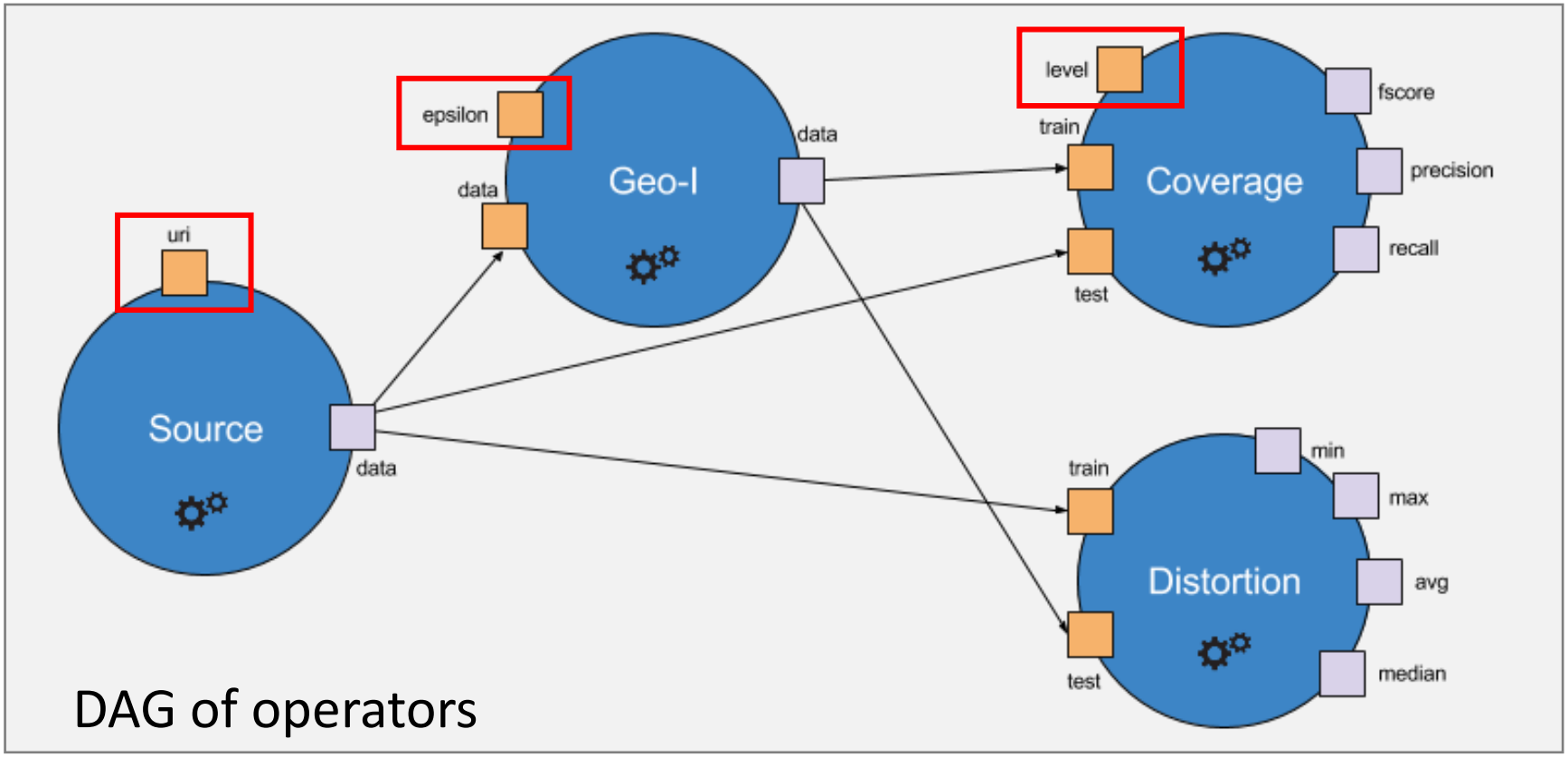
Operator authoring

```
@Op(
  category = "lppm",
  help = "Enforce geo-indistinguishability guarantees on traces.",
  unstable = true,
  cpu = 4,
  ram = "2G")
class GeoIndistinguishabilityOp
extends Operator[GeoIndistinguishabilityIn, GeoIndistinguishabilityOut] {

  override def execute(in: GeoIndistinguishabilityIn, ctx: OpContext): GeoIndistinguishabilityOut = {
    val input = read[Trace](in.data)
    val rnd = new Random(ctx.seed)
    val seeds = input.keys.map(key => key -> rnd.nextLong()).toMap
    val output = input.map(trace => new Laplace(in.epsilon, seeds(trace.id)).transform(trace))
    GeoIndistinguishabilityOut(write(output, ctx.workDir))
  }
}

case class GeoIndistinguishabilityIn(@Arg epsilon: Double = 0.001, @Arg data: Dataset)
case class GeoIndistinguishabilityOut(@Arg data: Dataset)
```

Workflows



DAG of operators

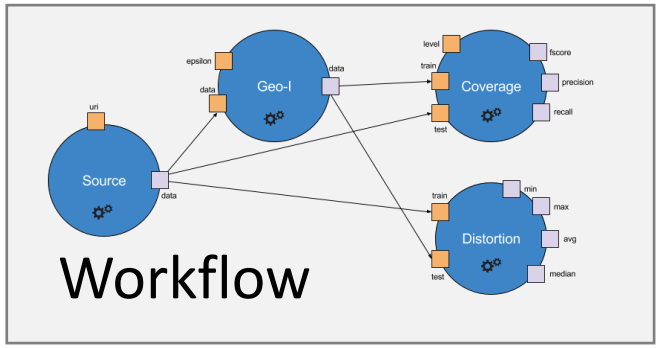
+ Parameters

+ Metadata: name, owner

Workflow DSL (JSON)

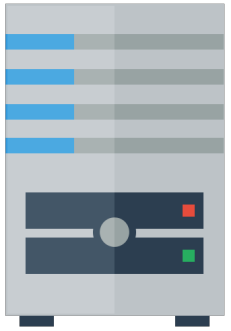
```
{
  "id": "geoind_workflow",
  "name": "Geo-indistinguishability workflow",
  "owner": "John Doe <john.doe@gmail.com>",
  "params": [
    { "name": "epsilon", "kind": "double"}
  ],
  "graph": [
    {
      "op": "GeoIndistinguishability",
      "inputs": {
        "epsilon": {"param": "epsilon"},
        "data":  "/path/to/my/dataset"
      }
    },
    {
      "op": "PoisRetrieval",
      "inputs": {
          "diameter": "200.meters",
          "duration": "15.minutes",
          "threshold": "100.meters",
          "train": "/path/to/my/dataset",
          "test": {"reference": "GeoIndistinguishability/data"}
       }
      }
  ]
}
```

Runs



+ Parameters (possibly a sweep)
+ Repeat times
+ Seed
+ Metadata: name, owner, tags…

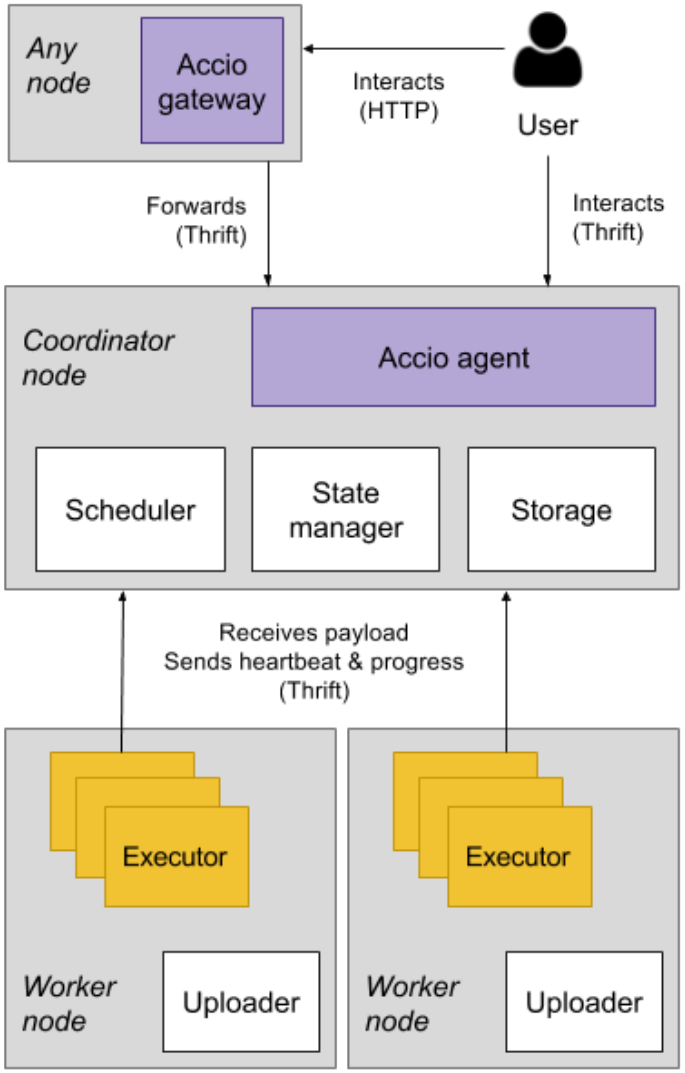
One or many runs, scheduled on a cluster



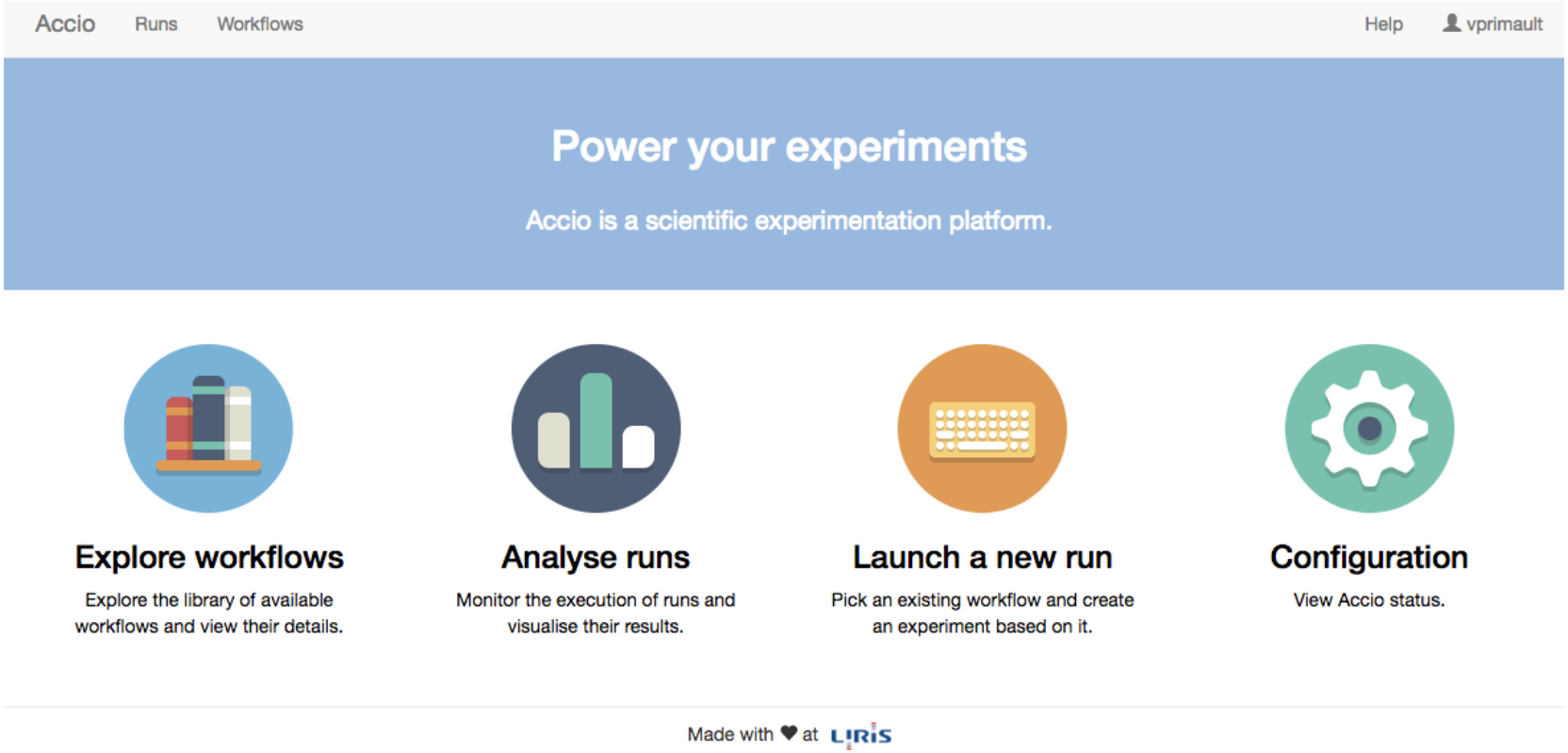
Run DSL (JSON)

```
{
  "workflow": "my_awesome_workflow",
  "repeat": 3,
  "name": "My brand new experiment",
  "tags": ["brand", "new"],
  "params": {
    "epsilon": {
      "from": 0.00001,
      "to": 1,
      "log": true
    },
    "duration": "10.seconds"
  }
}
```

System architecture



Web UI


Accio
Runs
Workflows
Help
vprimault
Power your experiments
Accio is a scientific experimentation platform.
Explore workflows
Explore the library of available workflows and view their details.
Analyse runs
Monitor the execution of runs and visualise their results.
Launch a new run
Pick an existing workflow and create an experiment based on it.
Configuration
View Accio status.
Made with ♥ at LIRIS

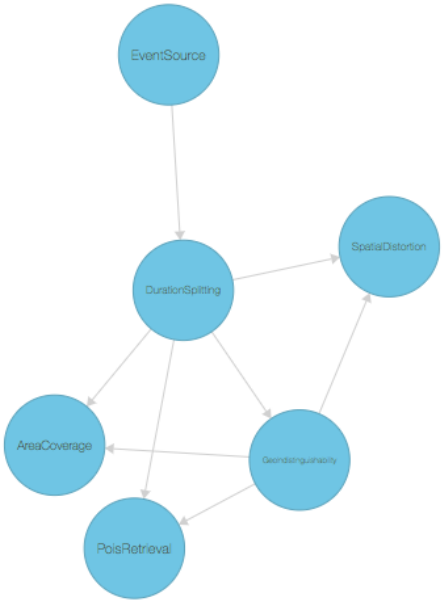
Web UI

Workflow details

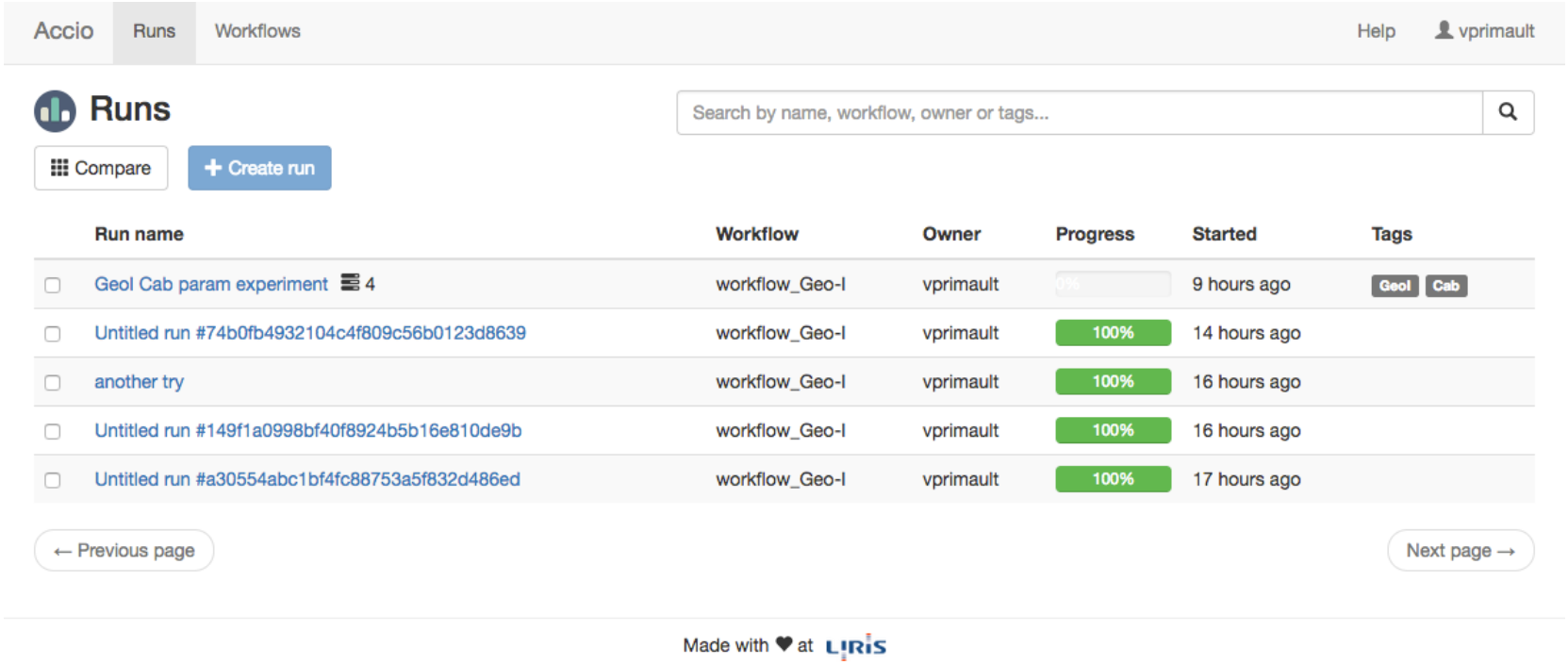
Name	Geo-I Cab nominal workflow
Created	Jan 26th 2017, 01:03pm
Owner	Sophie Cerf <sophie.cerf@gipsa-lab.fr>

Workflow parameters

Operators graph

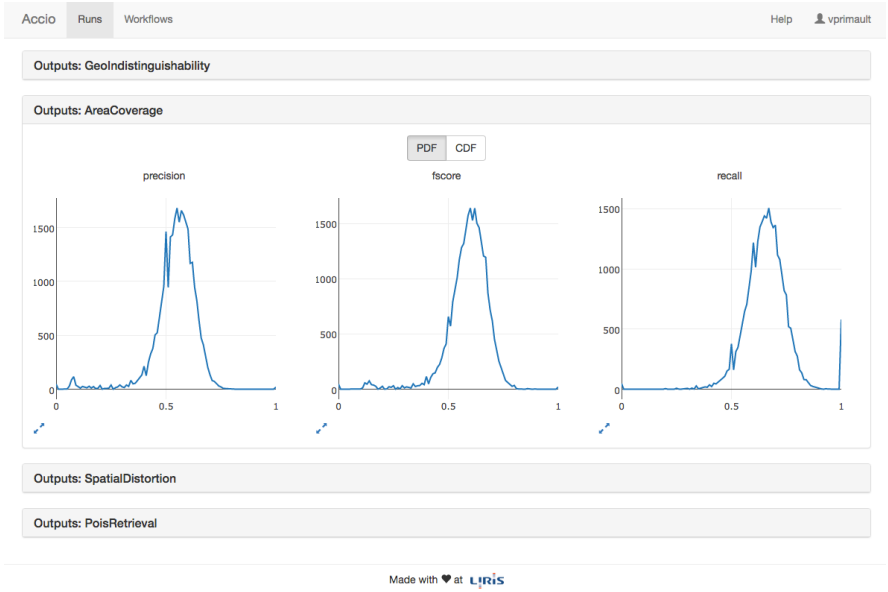


Web UI


Accio
Runs
Workflows
Help
vprimault
Runs
Search by name, workflow, owner or tags...
Compare
+ Create run
Run name
Workflow
Owner
Progress
Started
Tags
Geol Cab param experiment 4
workflow_Geo-I
vprimault
9 hours ago
Geol
Cab
Untitled run #74b0fb4932104c4f809c56b0123d8639
workflow_Geo-I
vprimault
100%
14 hours ago
another try
workflow_Geo-I
vprimault
100%
16 hours ago
Untitled run #149f1a0998bf40f8924b5b16e810de9b
workflow_Geo-I
vprimault
100%
16 hours ago
Untitled run #a30554abc1bf4fc88753a5f832d486ed
workflow_Geo-I
vprimault
100%
17 hours ago
← Previous page
Next page →
Made with ♥ at LIRIS

Results analysis

- Monitoring and results preview via CLI.
- Export of artifacts and metrics via CLI.
- Monitoring and interactive analysis via web UI.
- REST API for more advanced usages.



Implementation notes

- Scala (backend & client): 10,840 LOC.
- NodeJS (web UI): 1,700 LOC.

- Storage: Local, Elasticsearch.
- State manager: Local, Zookeeper.
- Scheduler: Local, GridEngine (TBD), Nomad (TBD).
- Uploader: Local, SCP (TBD), S3 (TBD).

Outline

1. Introduction
2. Accio, a workflow management tool
3. **Accio, a location privacy tool**
4. Hands-on

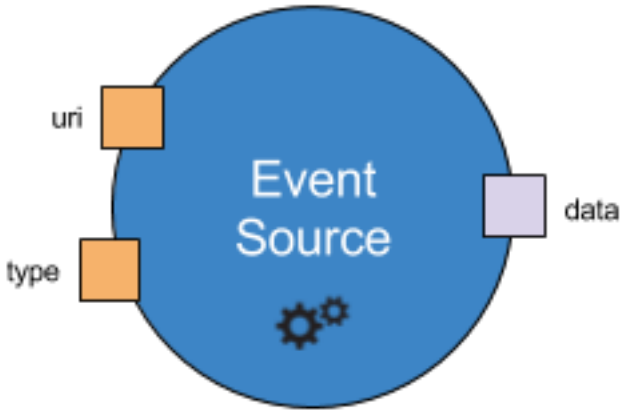
Data model

- Event: A triplet (user, location, timestamp).
- Trace: An ordered sequence of events belonging to the same user.

- Traces are written as CSV.
- Sparkle library for parallel processing of traces.

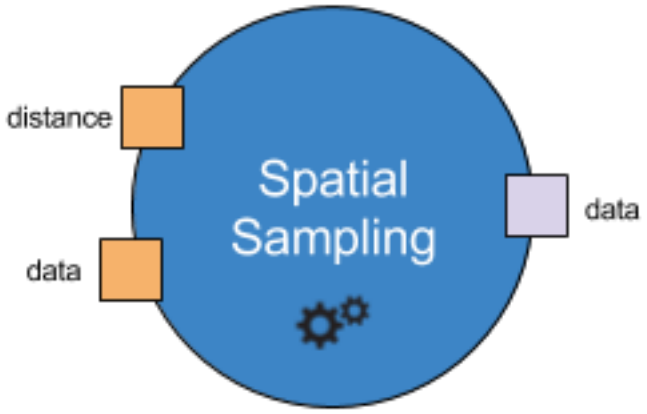
Source operators

- EventSource



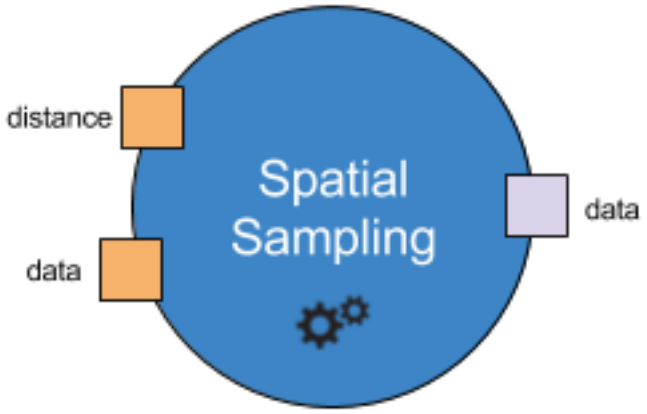
Transform operators

- CollapseTemporalGaps
- DurationSplitting
- EnforceDuration
- EnforceSize
- GaussianKernelSmoothing
- ModuloSampling
- SequentialSplitting



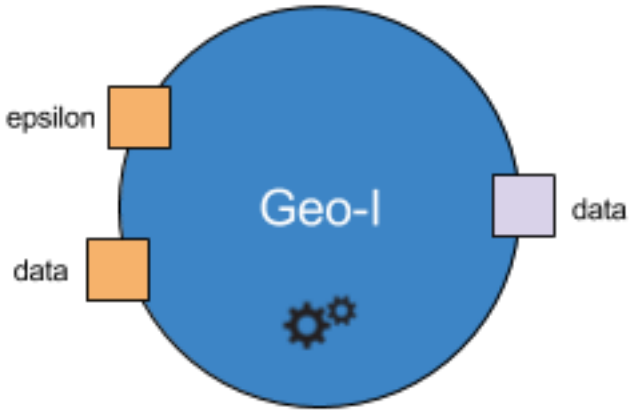
Transform operators (cont'd)

- SizeSplitting
- SpatialGapSplitting
- SpatialSampling
- TemporalGapSplitting
- TemporalSampling
- UniformSampling
- **PoisExtraction**



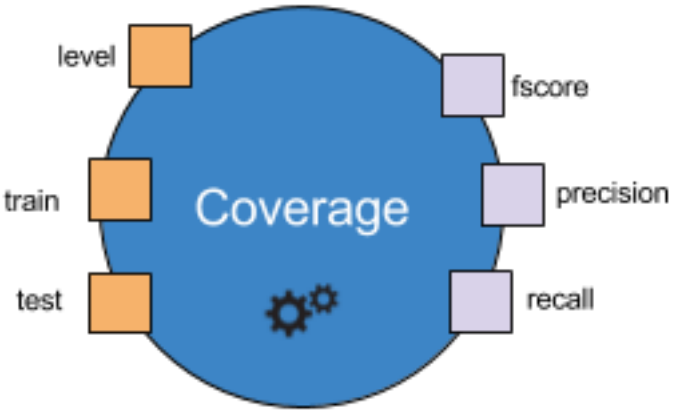
LPPM operators

- GeoIndistinguishability
- Promesse
- Wait4Me



Metric operators

- AreaCoverage
- CountQueriesDistortion
- DataCompleteness
- PoisReident
- PoisRetrieval
- SpatialDistortion
- TransmissionDelay



Outline

1. Introduction
2. Accio, a workflow management tool
3. Accio, a location privacy tool
4. **Hands-on**