Discoverable Metadata for System Monitoring Data

S. Leak, A. Greiner, A. Gentile, J. Brant

U.S. DEPARTMENT OF ENERGY | Office of Science
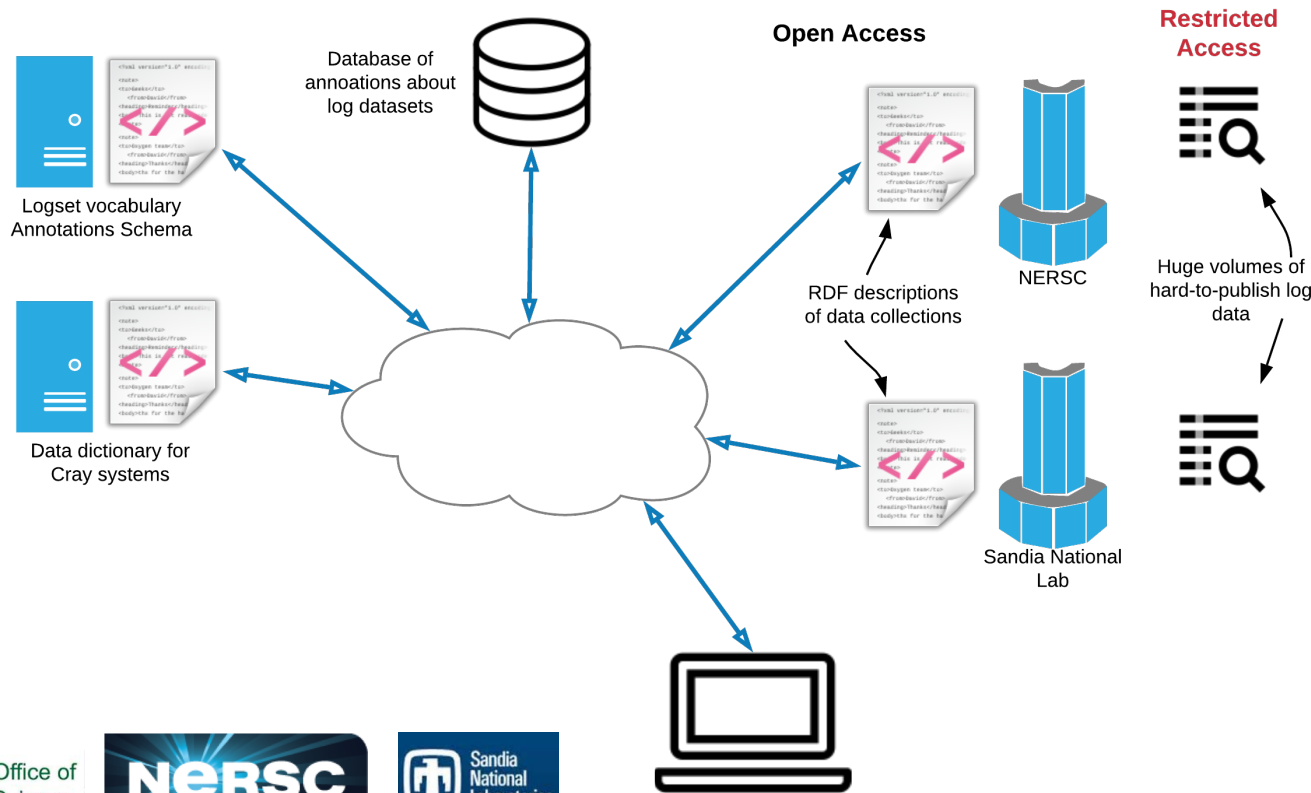
NERSC

Sandia National Laboratories

Context and Premise

- Much log data being collected, hard to coordinate who is collecting what, where, how

- DOE Resilience Project: how to make log data navigable for investigating fault propagation

- Idea: Machine-readable metadata to bypass many of the problems

Linked data connecting disparate independent metadata into a single global query-able graph

Database of annoations about log datasets

Logset vocabulary Annotations Schema

Data dictionary for Cray systems

Open Access

Restricted Access

RDF descriptions of data collections

NERSC

Huge volumes of hard-to-publish log data

Sandia National Lab

(Contrived) Example

What might have caused this unexpected job failure around 8am on 2015-03-27?

```
./logs.py -q '
SELECT ?title ?contact ?page WHERE {
  ?logset a logset:LogSet .
  ?logset dct:title ?title .
  ?logset dcat:contactPoint ?contact .
  ?logset dcat:landingPage ?page .
  ?logset dct:temporal/logset:startDate ?start .
  ?logset dct:temporal/logset:endDate ?end .
  FILTER (?start < "2015-03-27T07:00:00"^^xsd:dateTime) .
  FILTER (?end > "2015-03-27T09:00:00"^^xsd:dateTime) .
  snl:mutrino (logset:hasPart|logset:affects)* ?thing .
  ?logset logset:isViewOf*/dcat:distribution*/logset:subject
?thing .
}
```

Query returns a few things

Someone has log data covering that!

Better still: annotations (smaller/easier to download)

```
['"annotations for Mutrino dataset release ...
['"annotations for Mutrino dataset release ...
['"annotations for Mutrino dataset release ...
['"Mutrino dataset release SAND2016-2449 ...
['"Mutrino dataset release SAND2016-2449 ...
['"Mutrino dataset release SAND2016-2449 ...
```

Download and query the Annotations

```
./anno.py -s 2015-03-27T07:00:00 -e 2015-03-27T09:00:00
----------
[772743] by abc on system None
Time: 2015-03-27 07:52:33 to 2015-03-27 07:52:33
Start state: None ; End state: None
Description: Machine check events (MCE) logged. This occurs when the system
processor detected a hardware error.
Manually invoked? False ; System down?:  False
Components: ["c0-0c1s8n1"]
Tags:
LogDiver category group: NO
Baler pattern ID: 3801
Relevant log files: console
----------
[772744] by abc on system None
Time: 2015-03-27 08:15:33 to 2015-03-27 08:15:33
[snip]
```

Progress and Challenges

- RDF vocabulary and sqlite3 schema for annotations

- Some prototype tools for working with them

- Annotated a sample of logs

Progress and Challenges

- Graph queries are really powerful
  - .. But can be really slow too!
- Impedance mismatch between formats
  - Can be mitigated, but still laborious
- Some of the tools for RDF/Linked Data can be a little clunky