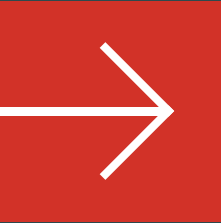


What are your Data Dreams?



bit.ly/datadreams



Help us build OARS!

Tell us your data questions and desires for the future of Analytics on Open edX.

Demo:

oars-demo.openedx.org

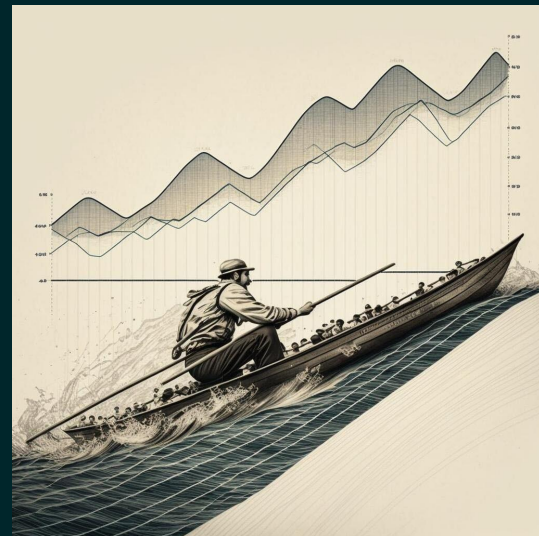
Slides: bit.ly/oars-2023-slides

Open edX Conference 2023

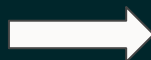
Navigate Your Data Lake with OARS

Axim

Jillian Vogel, OpenCraft | Brian Mesick, ~~CRIL~~



“Oarville, the data object lesson”



Survey: bit.ly/datadreams

Test Instance: oars-demo.openedx.org



Agenda

1. **Introductions**
2. Who are you?
3. Data Landscape
4. The Open Analytics Reference System (OARS)
5. Next Steps

Survey: bit.ly/datadreams

Test Instance: oars-demo.openedx.org

Jill

Open Source Developer @ OpenCraft

Online Learning DevOps @ U of Adelaide

Scuba diver



Brian

Platform Architect, Data @ Axim

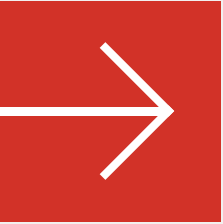
Platform & Data Engineering @ edX

Data Engineering @ Covera Health

Amateur shepherd



But first, the elephant in the room...



The Open Analytics Reference System

A successor to Insights to enable reporting for small-to-medium installs.

- **Open** - Open source tools and configuration, deploy anywhere
- **Analytics** - Data transformed for analysis
- **Reference System** - A stepping off point, not a one-size-fits-all solution

xAPI + Ralph + ClickHouse + Superset + Tutor Config = OARS v1(?)

The Open Analytics Reference System

Today...

Complete

- ✓ Discovery
- ✓ Tech Selection
- ✓ Prototyping
- ✓ 22 Event Types

The Open Analytics Reference System

Tomorrow(ish)...

Complete

- ✓ Discovery
- ✓ Tech Selection
- ✓ Prototyping
- ✓ 22 Event Types

v1 - Next 90 Days

- ↻ Course Structure*
- ↻ Data Modeling
- ↻ More Events & Reports
- ↻ Feedback & Iteration

The Open Analytics Reference System

The near future...

Complete

- ✓ Discovery
- ✓ Tech Selection
- ✓ Prototyping
- ✓ 22 Event Types

v1 - Next 90 Days

- ↻ Course Structure*
- ↻ Data Modeling
- ↻ More Events & Reports
- ↻ Feedback & Iteration

v2 - This year

- 🌐 Round-trip to CMS
- 🌐 Enrichment
- 🌐 More Reports
- 🌐 Message Bus

Agenda

1. Introductions
2. **Who are you?**
3. Data Landscape
4. The Open Analytics Reference System (OARS)
5. Next Steps

Survey: bit.ly/datadreams

Test Instance: oars-demo.openedx.org

Show of Hands 🖐️

Are you a...

- Learner
- Educator
- Learning designer
- Data analyst
- Program administrator
- Site operator
- Site developer
- Other?



235+

Distinct
Tracking Log
event types

**Tracking logs store
events from all over the
platform, not just
learner data!**

* edx.org 30 days, 2022

Show of Hands 🖐️

What do you use for OeX analytics?

- Insights
- Google Analytics
- Tools like Cairn or Figures
- Homegrown tools
- Raw database queries or spreadsheets
- Other?



68%

Of all tracking
log events are
page view
events

**We want to add page
view events for v2, but
need to prepare for the
big increase in volume!**

* edx.org 30 days, 2022

Show of Hands 🖐️

How many learners do you support?

- Hundreds
- Thousands
- Millions
- Tens of millions+



13.7%

Of events would currently be captured in OARS

With the addition of grades and page views in v2 we should be over 90%!

* edx.org 30 days, 2022

Survey

What are your data questions?

bit.ly/datadreams



9.9%

Of all events
are video
related

**Video events are in
OARS. Tell us your
ideas for how to use
them!**

* edX.org 30 days, 2022

Agenda

1. Introductions
2. Who are you?
- 3. Data Landscape**
4. The Open Analytics Reference System (OARS)
5. Next Steps

Survey: bit.ly/datadreams

Test Instance: oars-demo.openedx.org

Data Use Cases

"Protecting learner privacy and personal data is critical."

"What content needs work?"

"Aggregated data across all courses and programs."

"Summary statistics like weekly enrollment counts, attendance, completion."

"Which learners need help?"

Event Data OARS Currently Captures

Enrollment

- Enrollment activated
- Enrollment deactivated

Navigation

- Outline Selected
- “Next” Selected
- “Previous” selected
- Tab selected
- Link clicked*

Video

- Loaded
- Played
- Stopped
- Paused
- Searched
- CC Shown
- CC Hidden
- Transcript Shown
- Transcript Hidden
- Speed changed

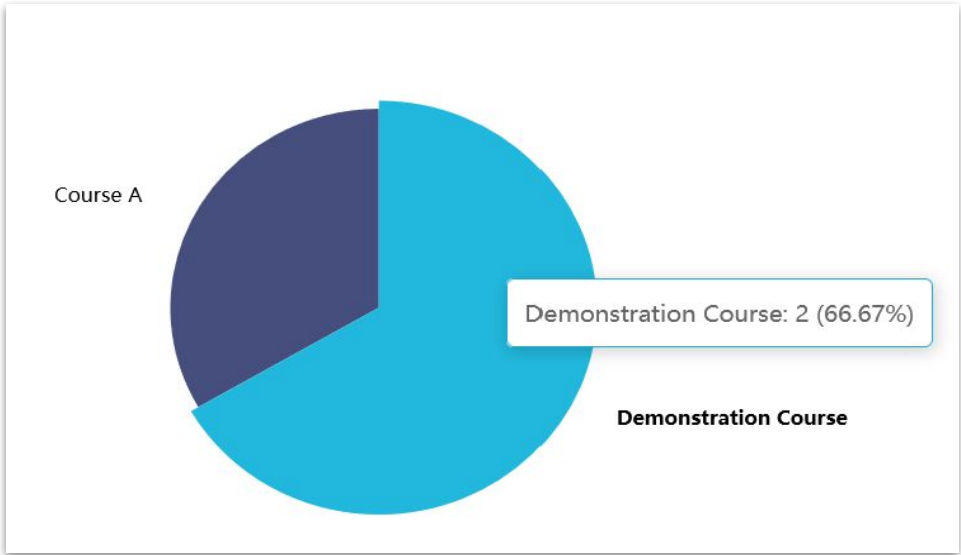
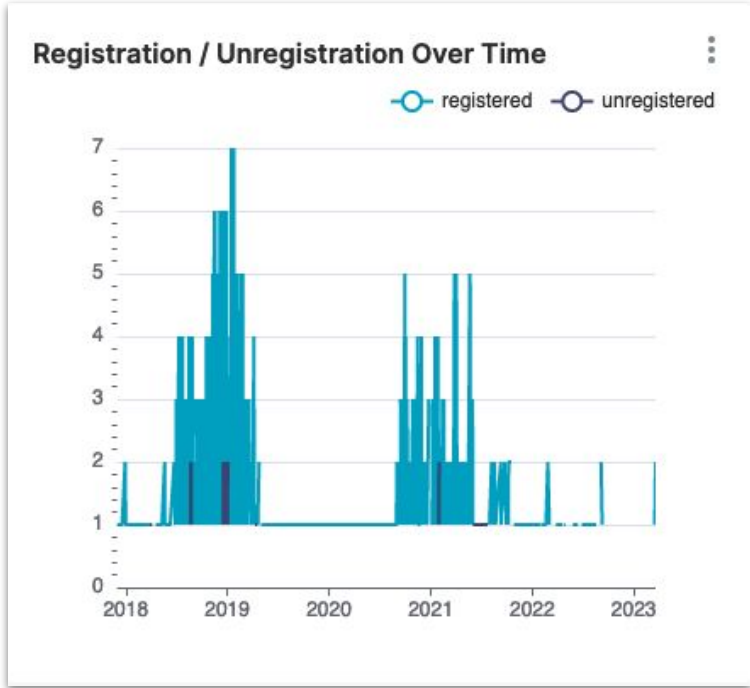
Problem

- Problem check (server)
- Problem check (browser)
- Show answer
- Hint displayed

Grading

- Passed first time

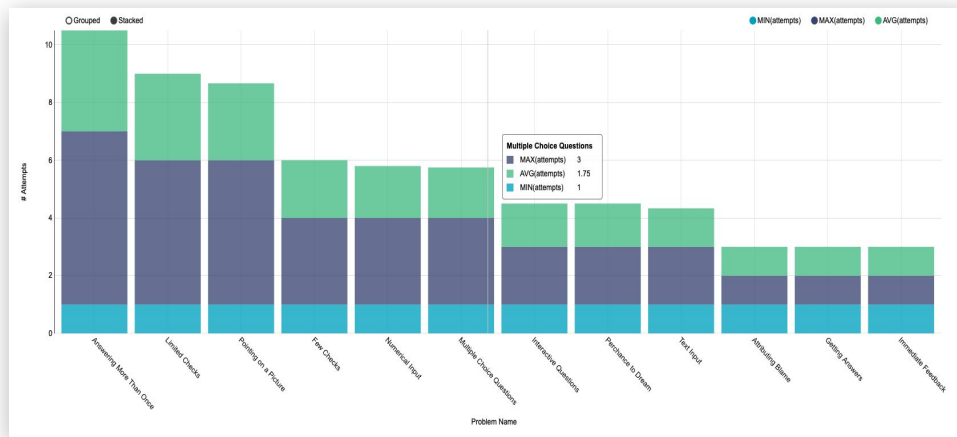
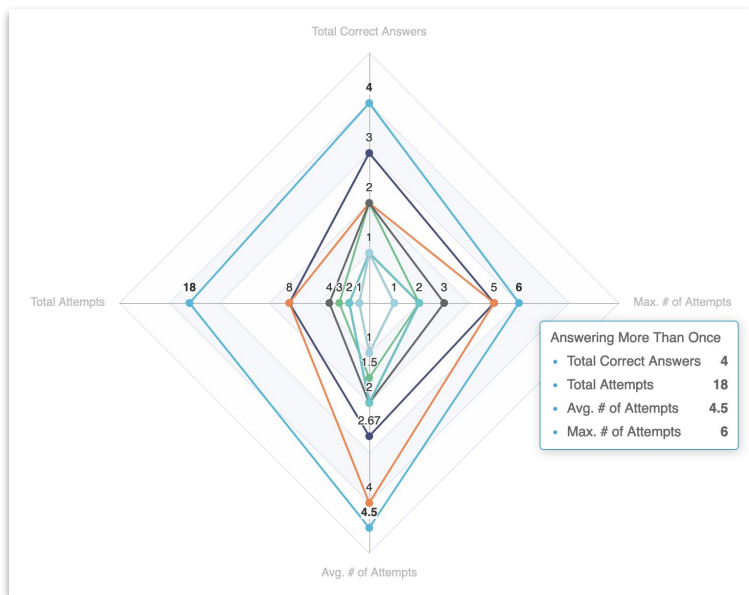
Superset: Enrollments



Superset: Problem Difficulty



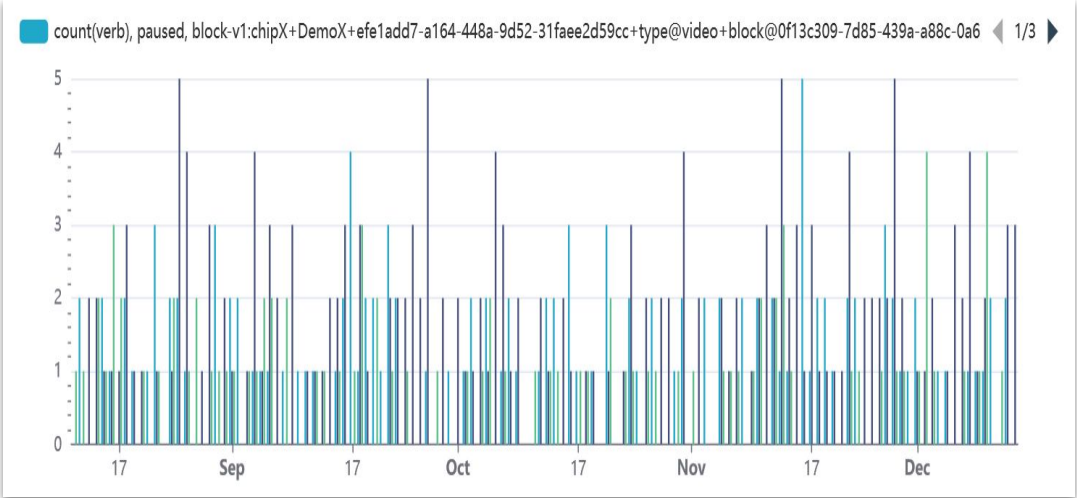
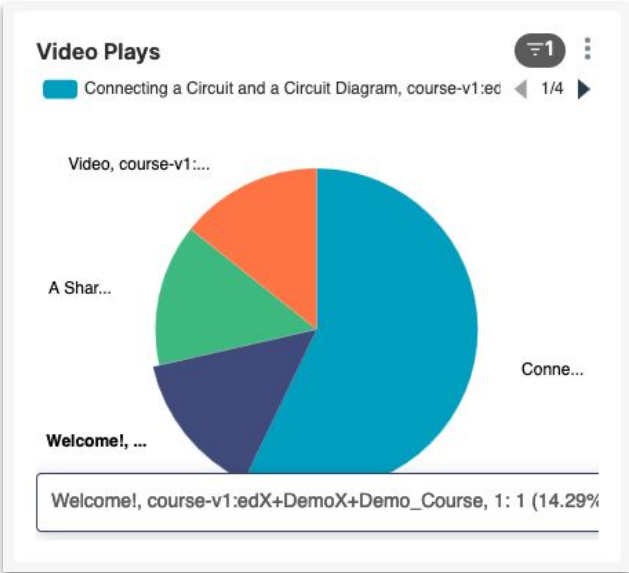
Number of attempts made for each problem: min, max, average.



Superset: Video events



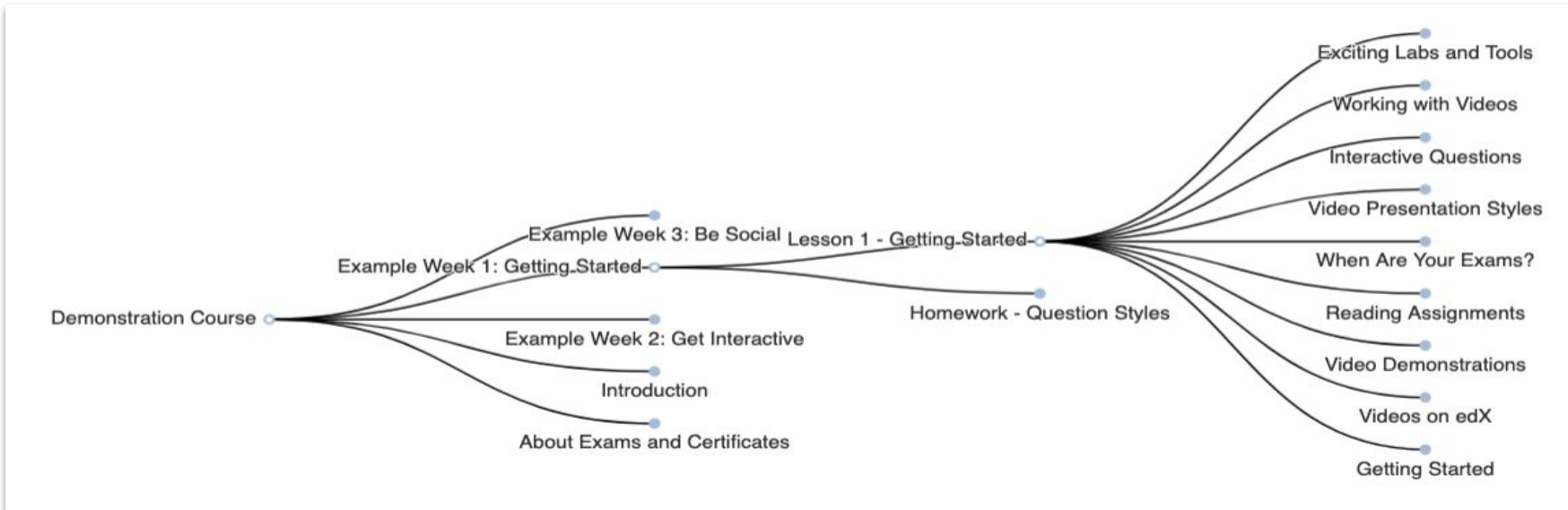
Shows count of video events by type + block



Superset: Course Data



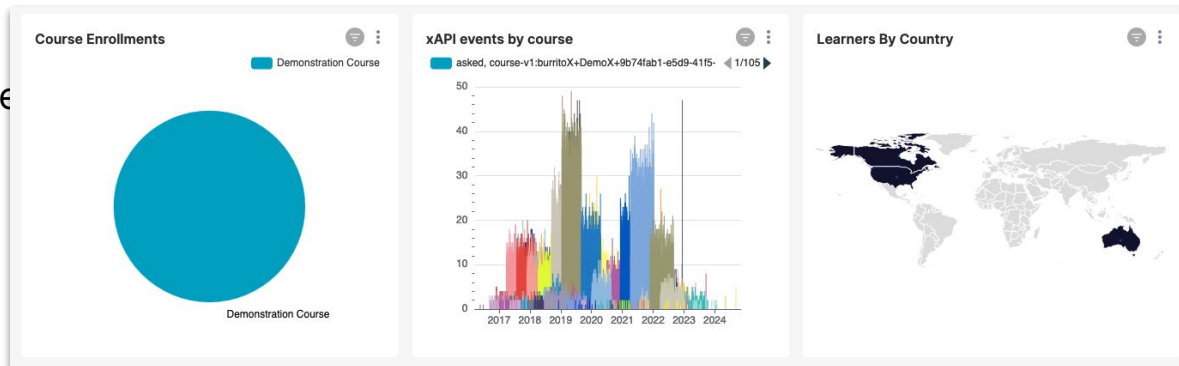
Near-real-time updates of course structure



Superset: Admin Dashboards



Aggregated information across the whole instance...



Agenda

1. Introductions
2. Who are you?
3. Data Landscape
- 4. The Open Analytics Reference System (OARS)**
5. Next Steps

Survey: bit.ly/datadreams

Test Instance: oars-demo.openedx.org

Ok but what actually *is* OARS?

OARS is...

1. Opinionated configuration of...
2. Curated tools for...
3. A near-real-time data pipeline to...
4. Enable data exploration and reporting...
5. And share data and learnings.

OARS Technologies

xAPI (Experience API)

- Open standard for sharing learning data
- Created by the US government's Advanced Distributed Learning Initiative
- Allows flexible and extensible expressions of learning data to be used in a consistent way
- (It's JSON, using the Actor / Verb / Object paradigm)

We transform Open edX tracking log events to xAPI learning statements via...

www.adlnet.gov/projects/xapi/

OARS Technologies

Event-routing-backends (LMS plugin)

- Open source plugin for the edx-platform
- Developed by **Arbisoft / edX** to implement [OEP-26: Real-time Events](#) (created in 2018!)
- Allows a high degree of flexibility in how learning events are transformed and where they are sent

Specifically OARS sends them to...

event-routing-backends.readthedocs.io

OARS Technologies

Ralph (Learning Record Store++)

- Open source collection of tools for storing, retrieving, and transforming learning log data
- Developed by Open edX partners **France Université Numérique**
- Parses, transforms, stores and serves event logs and learning traces

OARS uses it primarily as a way to store and deduplicate xAPI statements, which Ralph can then store in...

openfun.github.io/ralph

OARS Technologies

ClickHouse (Analytics Database)

- Open source column-oriented database
- Developed by ClickHouse, Inc.
- Designed to “generate analytical reports in real-time from non-aggregated data which also grows in real-time”

OARS uses ClickHouse’s unique architecture to provide near real-time data for use in...

clickhouse.com

Apache Superset (Data Exploration & Visualization)

- Open source
- Developed by the Apache Software Foundation
- Dashboards, reports, alerts, import/export, and API

OARS will come with a growing set of dashboards and reports, as well as security plugins and configuration to allow SSO against the LMS and row-level permissions to unlock a greater ability to self-explore without compromising privacy.

superset.apache.org

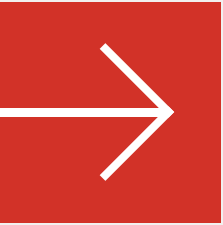
Privacy by default

- xAPI Actor ID is anonymized by default
- Limited access to MySQL / Mongo data
- Course authors can only see their courses

BUT

There is no “perfect privacy”.

Users and site administrators will still need to be mindful of the privacy and legal implications of which data gets stored and how they make it available!



Time for a live demo! 🤯

Join in on the test instance!
oars-demo.openedx.org

What *isn't* OARS?

- Complete :)
- A grading solution
- A web analytics package
- An enterprise analytics package
 - Superset is still rough around the edges
- Ready for huge sites
 - Got a pile of tracking logs and want to help?
- Set in stone

Agenda

1. Introductions
2. Who are you?
3. Data Landscape
4. The Open Analytics Reference System (OARS)
- 5. Next Steps**

Survey: bit.ly/datadreams

Test Instance: oars-demo.openedx.org

Next Steps (for us)

1. Production configurations
2. More data types! More charts!
3. Community sharing
4. Event bus – for scale & resiliency
5. Event replay for disaster recovery
6. Near real time course data
7. Tracking events cleanup for reliability

Next Steps (for you!)

Test Instance:

oars-demo.openedx.org

1. Try it out! Send us feedback on the [Forums](#), [Slack](#), GitHub, or email!
2. Monitor the [OARS project board](#)
3. Check out [State of Data WG](#)
4. Join the Data WG meet-up this **Friday 31 March 2:30pm**

Acknowledgements

Special thanks to:

- Régis @ Overhang.IO for Tutor and pointing us to ClickHouse
- Nimisha for OEP-26 (2018!) and everything else
- Zia et al @ Arbisoft for event-routing-backends
- Julien, Quitterie, and Wilfried @ FUN for Ralph & xAPI help
- Bob @ 2U for tracking log event counts
- Jenna @ tCRIL for turning this into a real project
- Kyle and Dave @ tCRIL for last minute help with Coursegraph
- The Data WG and everyone who took time to offer use cases, suggestions, and feedback!



References

- [OARS project board](#)
- [Data Working Group](#)
- [Data Use Cases](#)

Tutor plugins for...

- [OARS](#)
- [Ralph](#)
- [Clickhouse](#)
- [Superset](#)



Thank you!

More questions?

Data Working Group Forums:
discuss.openedx.org/c/working-groups/data/

Jill Vogel: jill@opencraft.com
Brian Mesick: bmesick@tcrl.org