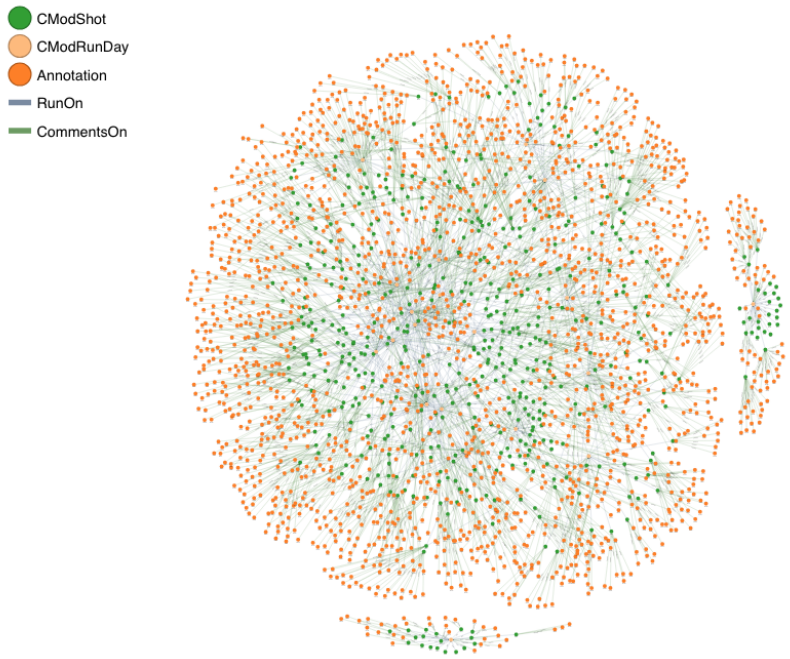
Navigational Data Management

A general approach to representation and exploitation of relationships in scientific data sets.

Joshua Stillerman
Martin Greenwald
Jason Stillerman
John Wright

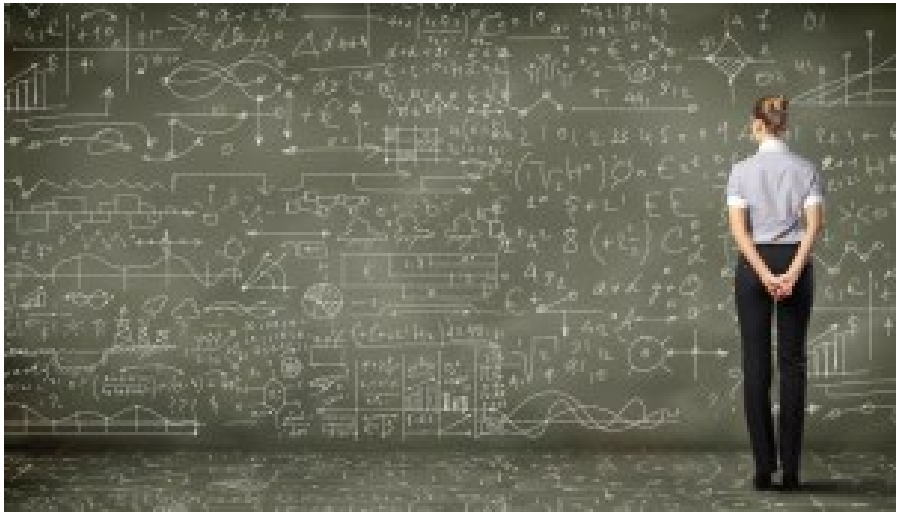
MIT Plasma Science and Fusion Center



Alcator C-Mod Runs, Shots and Annotations (**one month**)

“I have a system for storing my data and getting it back, aren’t I done?”

- Collecting data has never been easier, but…
- We’re struggling to keep up with the rapidly growing volume and complexity of scientific data.



Our Thesis

- The challenge is all about giving this mountain of data meaning and putting it into context
- Context is about metadata and relationships among data objects – “navigational metadata”
- This is not specific to one science domain

- In general, our approach to capturing and exploiting this class of metadata has been ad hoc and inadequate
- This hampers data discovery and the ability to assemble coherent, complete, useful data sets.

Discovering and Understanding Data Is Largely About Context

- Context is metadata about relationships between data
- Data discovery relies on “adjacency” to find other interesting data
- In the more distant past when things were smaller and simpler, we could keep that context in our heads
  - or in our colleague's heads
- Historically we’ve each build a set of ad-hoc, domain specific tools to store, explore, and retrieve this relationship metadata.
- Similar issues confront all data intensive areas of research.

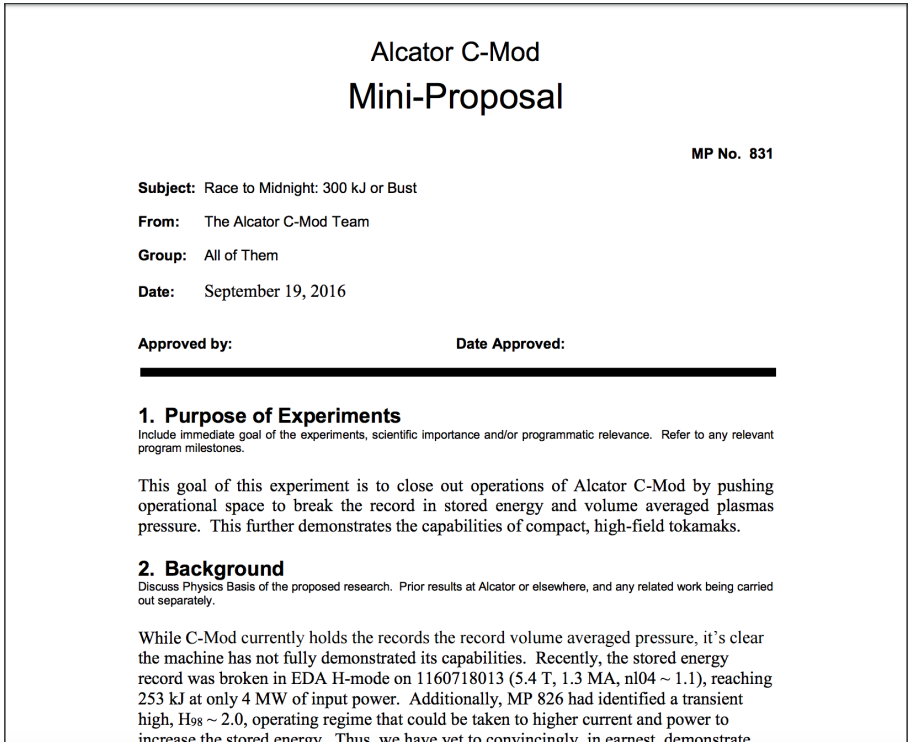
Organizing knowledge is an old problem

- Can we solve these problems in our own domain?
- Can we generalize these to provide solutions across a broader set of domains?

Complexity: What Sorts Of Data Might Exist From A Typical Experiment?

- Hierarchical data stores with raw and processed data (~10^5 named data objects per shot)
- Relational databases with “high level” results
- Electronic logbooks & annotation
- Experimental proposals
- Run Plans & Summaries
- Data provenance systems
- Data catalogs
- Data dictionaries
- Information about experimental campaigns & plans
- Publications & presentations
- Information about researchers, authors
- Simulation inputs & outputs
- Source code management systems
- Facility information, with details of experiment, measurement systems
- Document, drawing management systems
- QA, QC information
- WBS for projects

All Of Those Data Are Linked In Multiple and Complex Ways



Alcator C-Mod

Mini-Proposal

MP No. 831

Subject: Race to Midnight: 300 kJ or Bust

From: The Alcator C-Mod Team

Group: All of Them

Date: September 19, 2016

Approved by: **Date Approved:**

1. Purpose of Experiments

Include immediate goal of the experiments, scientific importance and/or programmatic relevance. Refer to any relevant program milestones.

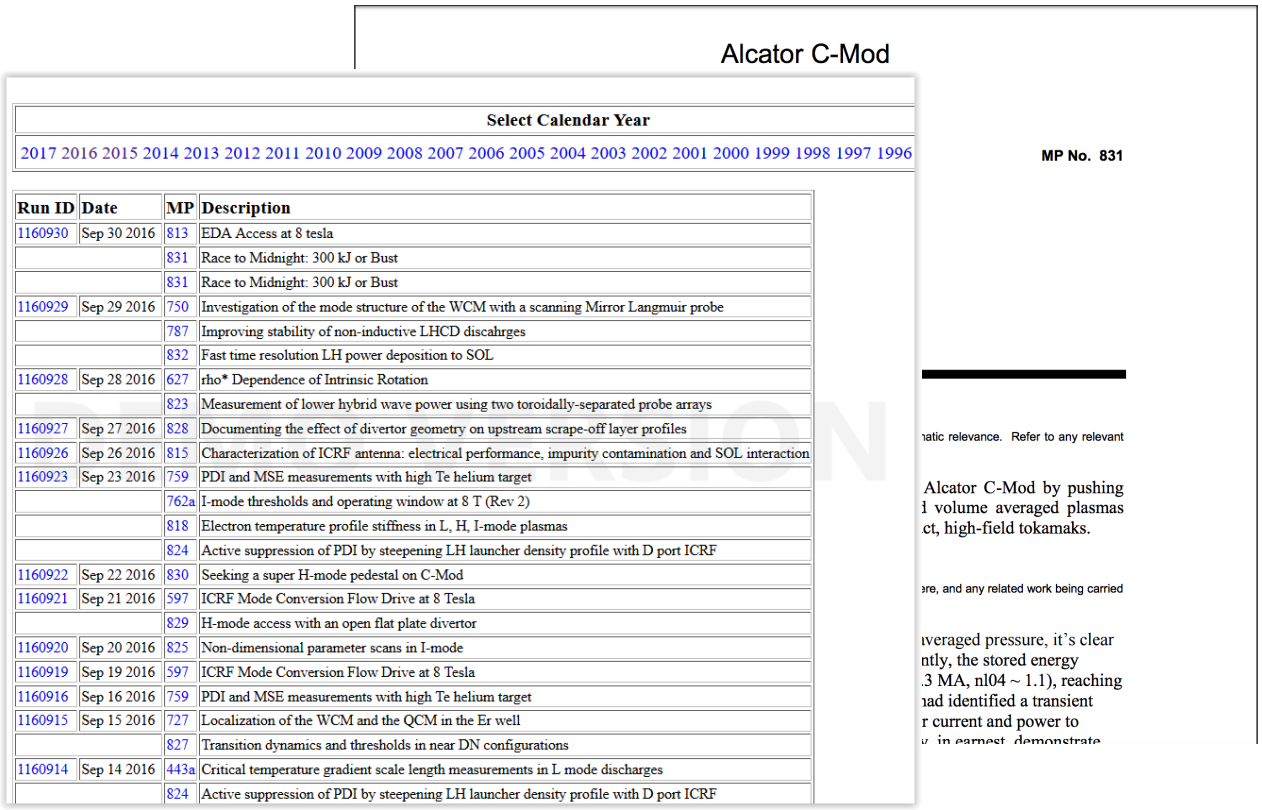
This goal of this experiment is to close out operations of Alcator C-Mod by pushing operational space to break the record in stored energy and volume averaged plasmas pressure. This further demonstrates the capabilities of compact, high-field tokamaks.

2. Background

Discuss Physics Basis of the proposed research. Prior results at Alcator or elsewhere, and any related work being carried out separately.

While C-Mod currently holds the records the record volume averaged pressure, it's clear the machine has not fully demonstrated its capabilities. Recently, the stored energy record was broken in EDA H-mode on 1160718013 (5.4 T, 1.3 MA, nl04 ~ 1.1), reaching 253 kJ at only 4 MW of input power. Additionally, MP 826 had identified a transient high, $H_{98} \sim 2.0$, operating regime that could be taken to higher current and power to increase the stored energy. Thus, we have yet to convincingly, in earnest, demonstrate

All Of Those Data Are Linked In Multiple and Complex Ways



Alcator C-Mod

Select Calendar Year

2017 2016 2015 2014 2013 2012 2011 2010 2009 2008 2007 2006 2005 2004 2003 2002 2001 2000 1999 1998 1997 1996

Run ID	Date	MP	Description
1160930	Sep 30 2016	813	EDA Access at 8 tesla
		831	Race to Midnight: 300 kJ or Bust
		831	Race to Midnight: 300 kJ or Bust
1160929	Sep 29 2016	750	Investigation of the mode structure of the WCM with a scanning Mirror Langmuir probe
		787	Improving stability of non-inductive LHCD discahrges
		832	Fast time resolution LH power deposition to SOL
1160928	Sep 28 2016	627	rho* Dependence of Intrinsic Rotation
		823	Measurement of lower hybrid wave power using two toroidally-separated probe arrays
1160927	Sep 27 2016	828	Documenting the effect of divertor geometry on upstream scrape-off layer profiles
1160926	Sep 26 2016	815	Characterization of ICRF antenna: electrical performance, impurity contamination and SOL interaction
1160923	Sep 23 2016	759	PDI and MSE measurements with high Te helium target
		762a	I-mode thresholds and operating window at 8 T (Rev 2)
		818	Electron temperature profile stiffness in L, H, I-mode plasmas
		824	Active suppression of PDI by steepening LH launcher density profile with D port ICRF
1160922	Sep 22 2016	830	Seeking a super H-mode pedestal on C-Mod
1160921	Sep 21 2016	597	ICRF Mode Conversion Flow Drive at 8 Tesla
		829	H-mode access with an open flat plate divertor
1160920	Sep 20 2016	825	Non-dimensional parameter scans in I-mode
1160919	Sep 19 2016	597	ICRF Mode Conversion Flow Drive at 8 Tesla
1160916	Sep 16 2016	759	PDI and MSE measurements with high Te helium target
1160915	Sep 15 2016	727	Localization of the WCM and the QCM in the Er well
		827	Transition dynamics and thresholds in near DN configurations
1160914	Sep 14 2016	443a	Critical temperature gradient scale length measurements in L mode discharges
		824	Active suppression of PDI by steepening LH launcher density profile with D port ICRF

MP No. 831

natic relevance. Refer to any relevant

Alcator C-Mod by pushing
l volume averaged plasmas
ct, high-field tokamaks.

ere, and any related work being carried

iveraged pressure, it's clear
ntly, the stored energy
.3 MA, nl04 ~ 1.1), reaching
had identified a transient
r current and power to

All Of Those Data Are Linked In Multiple and Complex Ways

Select Calendar Year

2017 2016 2015 2014 2013 2012 2011 2010 2009 2008 2007 2006 2005 2004 2003 2002 2001 2000

Run ID	Date	MP	Description
1160930	Sep 30 2016	813	EDA Access at 8 tesla
		831	Race to Midnight: 300 kJ or Bust
		831	Race to Midnight: 300 kJ or Bust
1160929	Sep 29 2016	750	Investigation of the mode structure of the WCM with a scanning Mirror Langmuir probe
		787	Improving stability of non-inductive LHCD discahrges
		832	Fast time resolution LH power deposition to SOL
1160928	Sep 28 2016	627	rho* Dependence of Intrinsic Rotation
		823	Measurement of lower hybrid wave power using two toroidally-separated probe arrays
1160927	Sep 27 2016	828	Documenting the effect of divertor geometry on upstream scrape-off layer profiles
1160926	Sep 26 2016	815	Characterization of ICRF antenna: electrical performance, impurity contamination and SOL
1160923	Sep 23 2016	759	PDI and MSE measurements with high Te helium target
		762a	I-mode thresholds and operating window at 8 T (Rev 2)
		818	Electron temperature profile stiffness in L, H, I-mode plasmas
		824	Active suppression of PDI by steepening LH launcher density profile with D port ICRF
1160922	Sep 22 2016	830	Seeking a super H-mode pedestal on C-Mod
1160921	Sep 21 2016	597	ICRF Mode Conversion Flow Drive at 8 Tesla
		829	H-mode access with an open flat plate divertor
1160920	Sep 20 2016	825	Non-dimensional parameter scans in I-mode
1160919	Sep 19 2016	597	ICRF Mode Conversion Flow Drive at 8 Tesla
1160916	Sep 16 2016	759	PDI and MSE measurements with high Te helium target
1160915	Sep 15 2016	727	Localization of the WCM and the QCM in the Er well
		827	Transition dynamics and thresholds in near DN configurations
1160914	Sep 14 2016	443a	Critical temperature gradient scale length measurements in L mode discharges
		824	Active suppression of PDI by steepening LH launcher density profile with D port ICRF

1160930041	jwhughes	SESSION_LEADER	11:29:06:267PM

Shot 41

We are attempting to shorten up the pre-H period of RF
Also increased the N2 puff in the beginning

Data system hang, followed by no power shot

1160930042	marmar	PHYSICS_OPERATOR	Sep 30 2016 11:37:13:997PM

Ip to 1.4 MA

1160930042	jwhughes	SESSION_LEADER	Sep 30 2016 11:57:45:183PM

Shot 42

After a long delay, got a 1.4MA shot! 235kJ

We're going to try for one more!

1160930043	marmar	PHYSICS_OPERATOR	Oct 1 2016 12:05:23:313AM

nl_04 to 9e19
nitrogen to 90 msec (from 70)

plasma, disrupt in rampdown at 1.3 seconds

that's it folks!

1160930043	jwhughes	SESSION_LEADER	Oct 1 2016 12:06:46:987AM

Shot 43

That's all folks

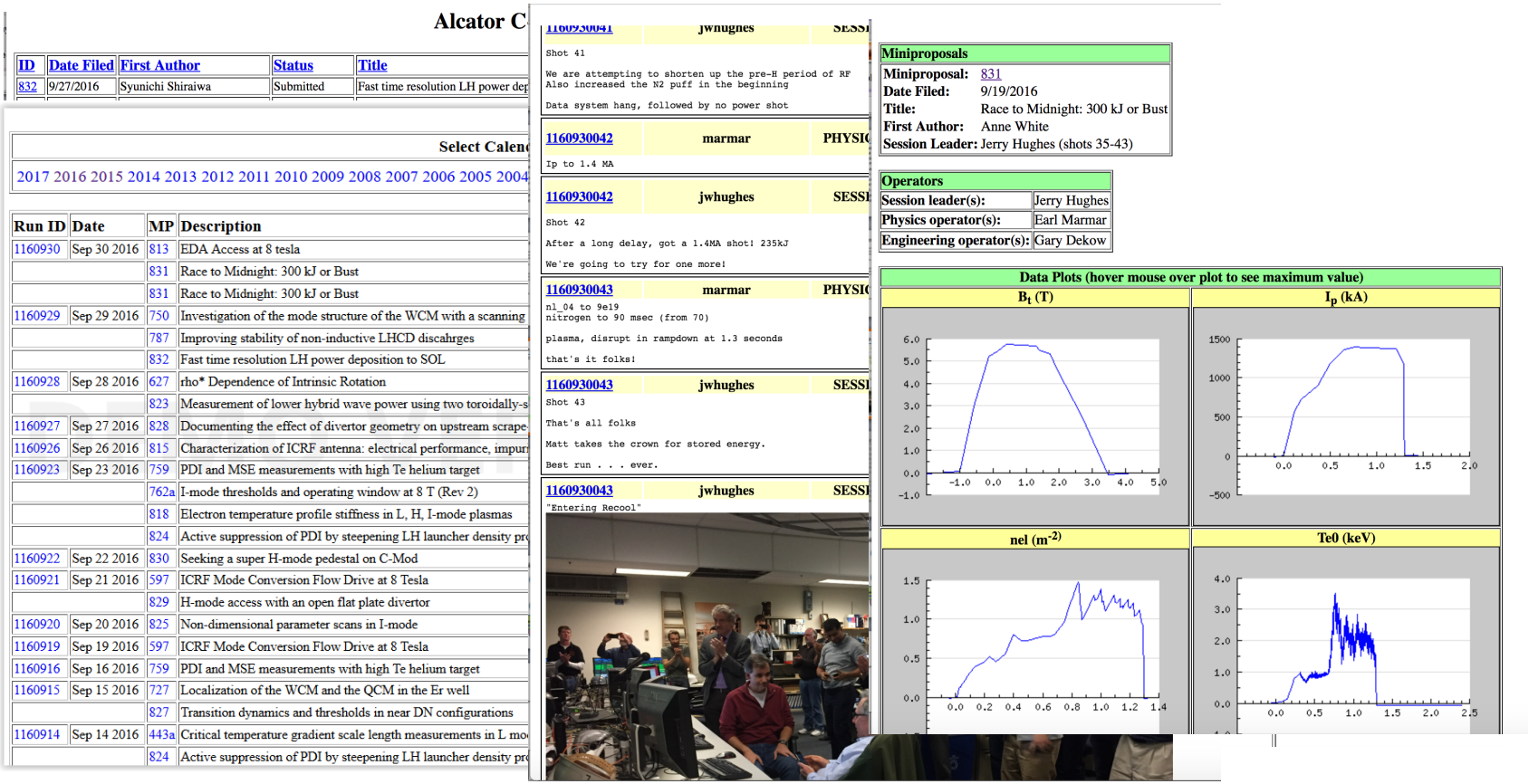
Matt takes the crown for stored energy.

Best run . . . ever.

1160930043	jwhughes	SESSION_LEADER	Oct 5 2016 12:19:48:210PM

"Entering Recool"

All Of Those Data Are Linked In Multiple and Complex Ways



Alcator C

ID	Date Filed	First Author	Status	Title
832	9/27/2016	Syunichi Shiraiwa	Submitted	Fast time resolution LH power dep

Select Calen

2017 2016 2015 2014 2013 2012 2011 2010 2009 2008 2007 2006 2005 2004

Run ID	Date	MP	Description
1160930	Sep 30 2016	813	EDA Access at 8 tesla
		831	Race to Midnight: 300 kJ or Bust
		831	Race to Midnight: 300 kJ or Bust
1160929	Sep 29 2016	750	Investigation of the mode structure of the WCM with a scanning
		787	Improving stability of non-inductive LHCD discahrges
		832	Fast time resolution LH power deposition to SOL
1160928	Sep 28 2016	627	rho* Dependence of Intrinsic Rotation
		823	Measurement of lower hybrid wave power using two toroidally-s
1160927	Sep 27 2016	828	Documenting the effect of divertor geometry on upstream scrape
1160926	Sep 26 2016	815	Characterization of ICRF antenna: electrical performance, impur
1160923	Sep 23 2016	759	PDI and MSE measurements with high Te helium target
		762a	I-mode thresholds and operating window at 8 T (Rev 2)
		818	Electron temperature profile stiffness in L, H, I-mode plasmas
		824	Active suppression of PDI by steepening LH launcher density pr
1160922	Sep 22 2016	830	Seeking a super H-mode pedestal on C-Mod
1160921	Sep 21 2016	597	ICRF Mode Conversion Flow Drive at 8 Tesla
		829	H-mode access with an open flat plate divertor
1160920	Sep 20 2016	825	Non-dimensional parameter scans in I-mode
1160919	Sep 19 2016	597	ICRF Mode Conversion Flow Drive at 8 Tesla
1160916	Sep 16 2016	759	PDI and MSE measurements with high Te helium target
1160915	Sep 15 2016	727	Localization of the WCM and the QCM in the Er well
		827	Transition dynamics and thresholds in near DN configurations
1160914	Sep 14 2016	443a	Critical temperature gradient scale length measurements in L mo
		824	Active suppression of PDI by steepening LH launcher density pr

1160930041	jwhughes	SESSI
Shot 41		
We are attempting to shorten up the pre-H period of RF Also increased the N2 puff in the beginning		
Data system hang, followed by no power shot		
1160930042	marmar	PHYSIC
Ip to 1.4 MA		
1160930042	jwhughes	SESSI
Shot 42		
After a long delay, got a 1.4MA shot! 235kJ		
We're going to try for one more!		
1160930043	marmar	PHYSIC
nl_04 to 9e19 nitrogen to 90 msec (from 70)		
plasma, disrupt in rampdown at 1.3 seconds		
that's it folks!		
1160930043	jwhughes	SESSI
Shot 43		
That's all folks		
Matt takes the crown for stored energy.		
Best run . . . ever.		
1160930043	jwhughes	SESSI
"Entering Recool"		

Miniproposals

Miniproposal:	831
Date Filed:	9/19/2016
Title:	Race to Midnight: 300 kJ or Bust
First Author:	Anne White
Session Leader:	Jerry Hughes (shots 35-43)

Operators

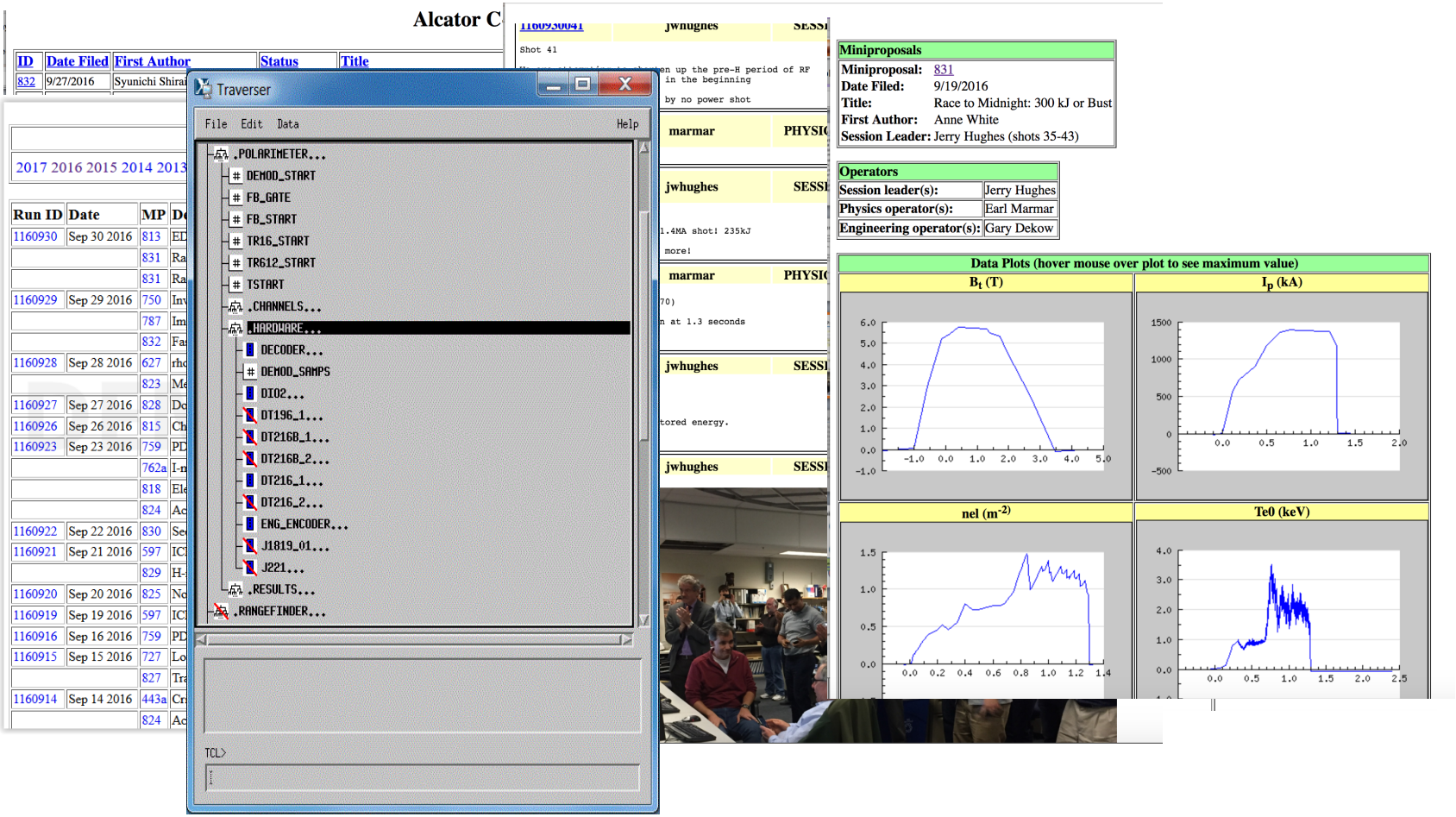
Session leader(s):	Jerry Hughes
Physics operator(s):	Earl Marmar
Engineering operator(s):	Gary Dekow

Data Plots (hover mouse over plot to see maximum value)

B_t (T) | I_p (kA)

nel (m^{-2}) | Te0 (keV)

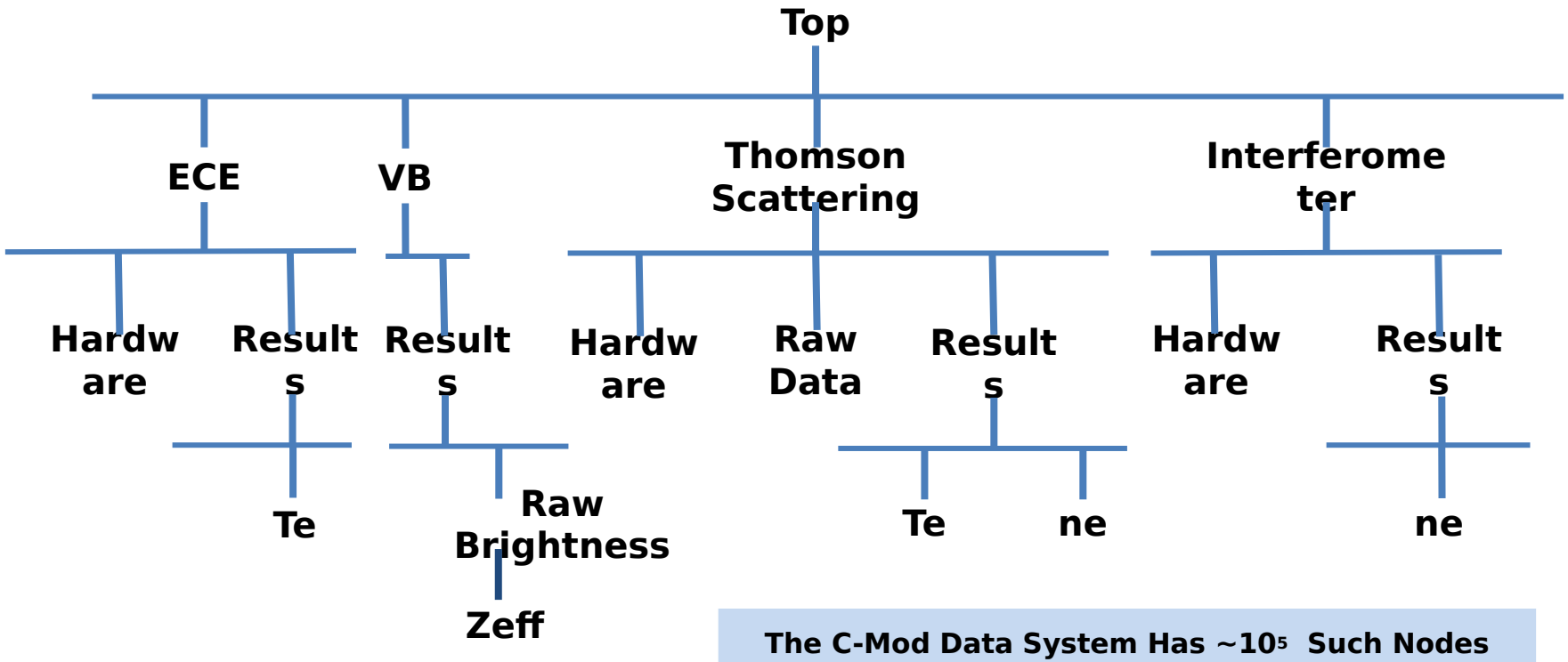
All Of Those Data Are Linked In Multiple and Complex Ways



Relationship Web Is Incomplete, Ad Hoc, Asymmetric, Singularly Organized

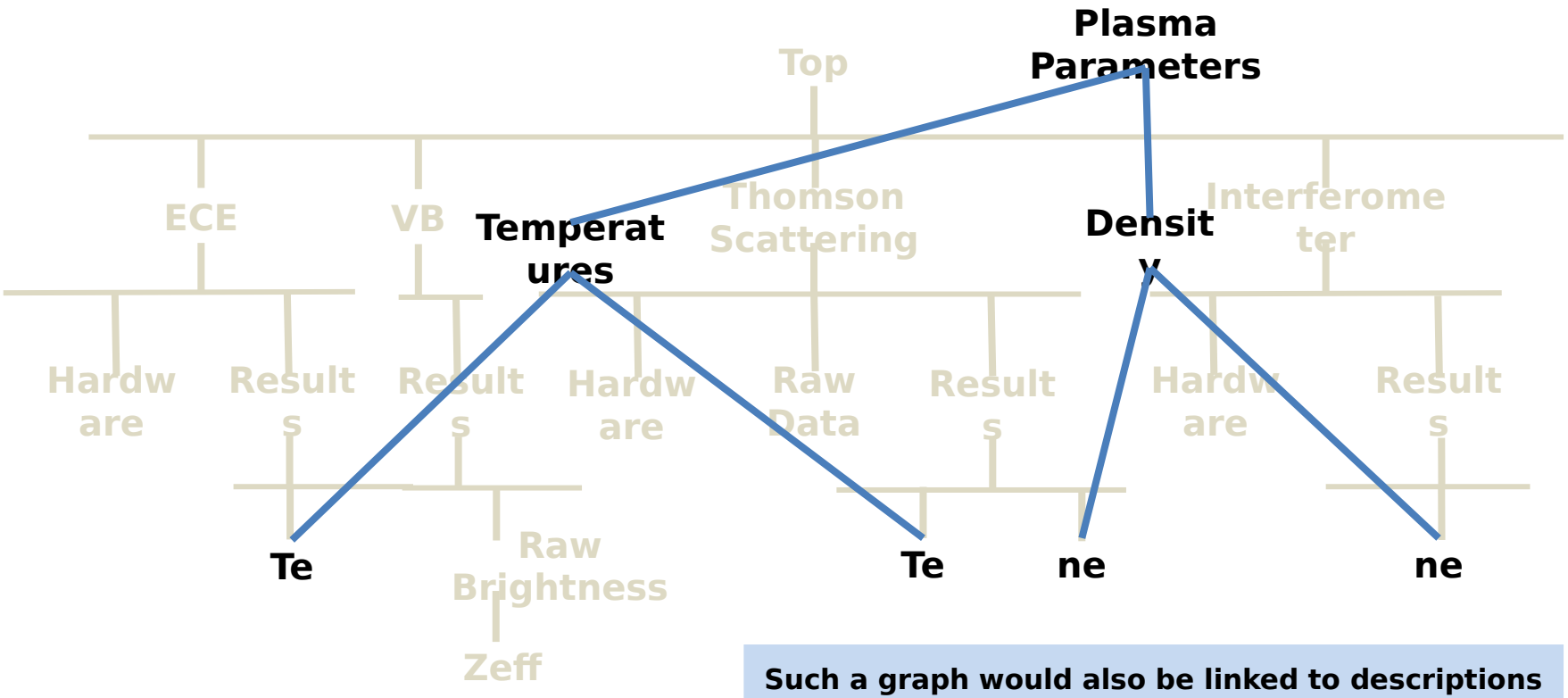
- Incomplete
  - Some relationships are explicitly represented in databases
  - Some are implicit in data or text
  - Some are only known by particular users
  - Some are not recorded and are lost forever
- Ad Hoc
  - We've added this information as needs arise
  - Schemas, vocabulary are not always consistent
  - Level of detail is uneven
- Asymmetric
  - Example: We point to interesting data from the logbook (annotation); but do not point to annotation from data (many, many other examples)
- Singularly organized
  - Trees, Tables (columns, indices), Directory/File Names

Organization of Data – By Diagnostic System



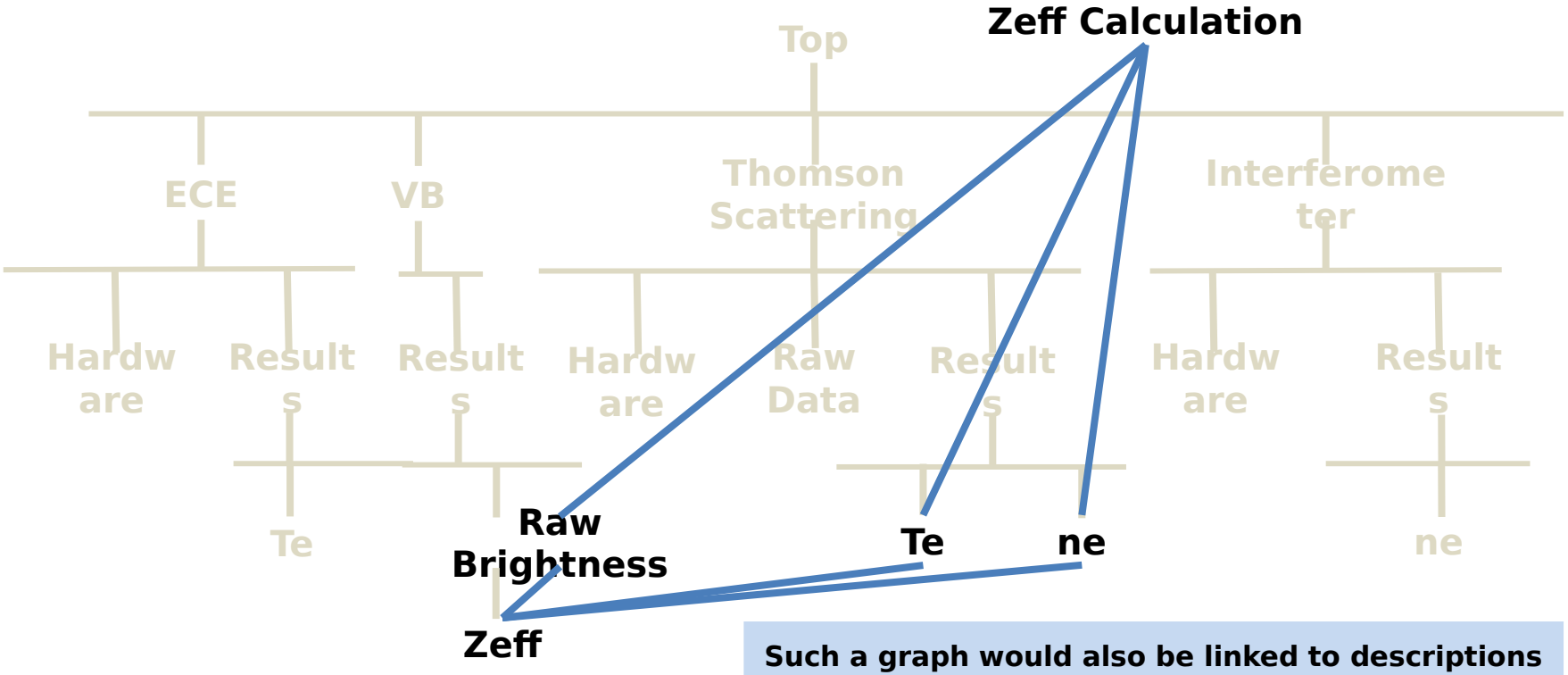
The C-Mod Data System Has ~10^5 Such Nodes With Significant Metadata For Each Node

Organization of Data – By Physics Parameter



Such a graph would also be linked to descriptions of experiments, annotation, etc.

Organization of Data – By Data Provenance



Such a graph would also be linked to descriptions of analysis codes, annotation, etc.

Approach

- Use graphs to describe relationships between data
- Schema defined using schema.JSON
  - Nodes
    - Who, what, when, history
    - List of properties appropriate to their type
    - [URI to have objects stored in other systems]
      - Protocol://location/specifiers
    - GUID
  - Edges
    - Who, what, when, history
    - Allowable SRCs, Destinations
    - Properties (if needed)

Name	Color	SuperClasses
Annotation		_NDMobject, V
CModRunDay		_RunDay
CModShot		_Shot
DropboxImage		_ExternalReference
MDSplus		_ExternalReference

Name	Color	SuperClasses
MiniProposal		_ExternalReference
PDFReference		_ExternalReference
Person		V
Topic		_NDMobject, V
V		

Name	Color	SuperClasses
CommentsOn		_NDMobject, E
E		
HasImage		_NDMobject, E
HasTopic		_NDMobject, E
RunOn		_NDMobject, E

Approach

- Use graphs to describe relationships between data
- Schema defined using schema.JSON
  - Nodes
    - Who, what, when, history
    - List of properties appropriate to their type
    - [URI to have objects stored in other systems]
      - Protocol://location/specifiers
    - GUID
  - Edges
    - Type
    - Allowable SRCs, Destinations
    - Properties (if needed)

```
$ cat PDFReference.json
{
    "id": "http://ndm.mit.edu/ndm/schemas/PDFReference",
    "$schema": "http://json-schema.org/draft-06/schema#",
    "title": "PDFReference",
    "description": "Link to a web accessible PDF",
    "definitions": {},
    "type": "object",
    "allOf": [{ "$ref": "_ExternalReference"} ],
    "properties": {
        "name": { "type": "STRING” },
        "type": {
            "type": "STRING",
            "defaultvalue":"PDF"
        }
    },
    "required": ["name"],
    "metadata": (
        "schemaMetadata": {
            "title" : "{{@class}} {{$name}}",
            "brief" : [],
            "body"  : ["URI"],
            "links" : []
    }
  }
}
```

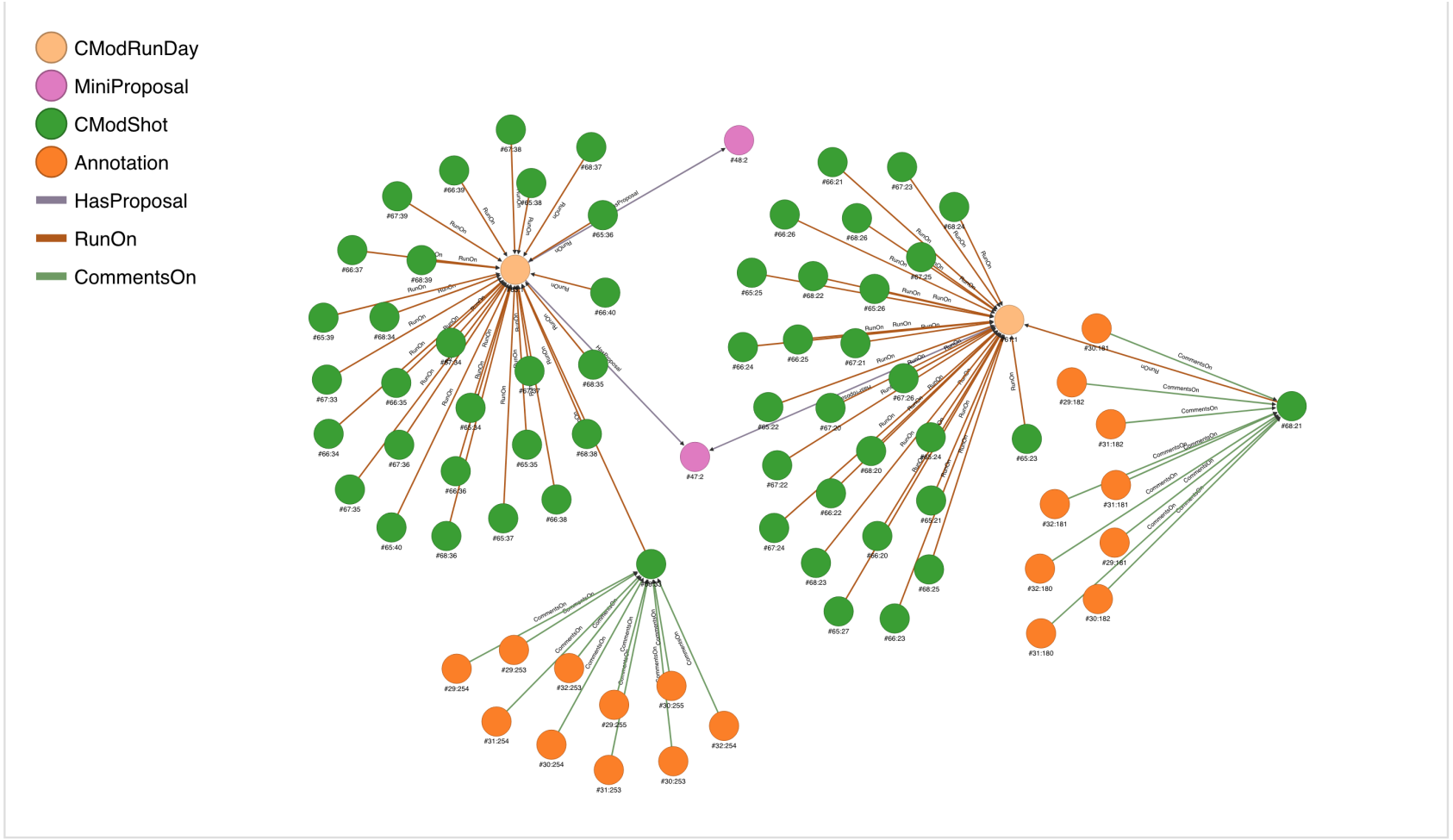
Approach

- Graph database - OrientDB
- Javascript SPA
- VUE.js – web frontend xxx
- Auth0
- Docker-compose (microservices)
  - DB server
  - Authenticator Proxy
  - Web Server
  - Notification server
- PM2 – lightweight process orchestration

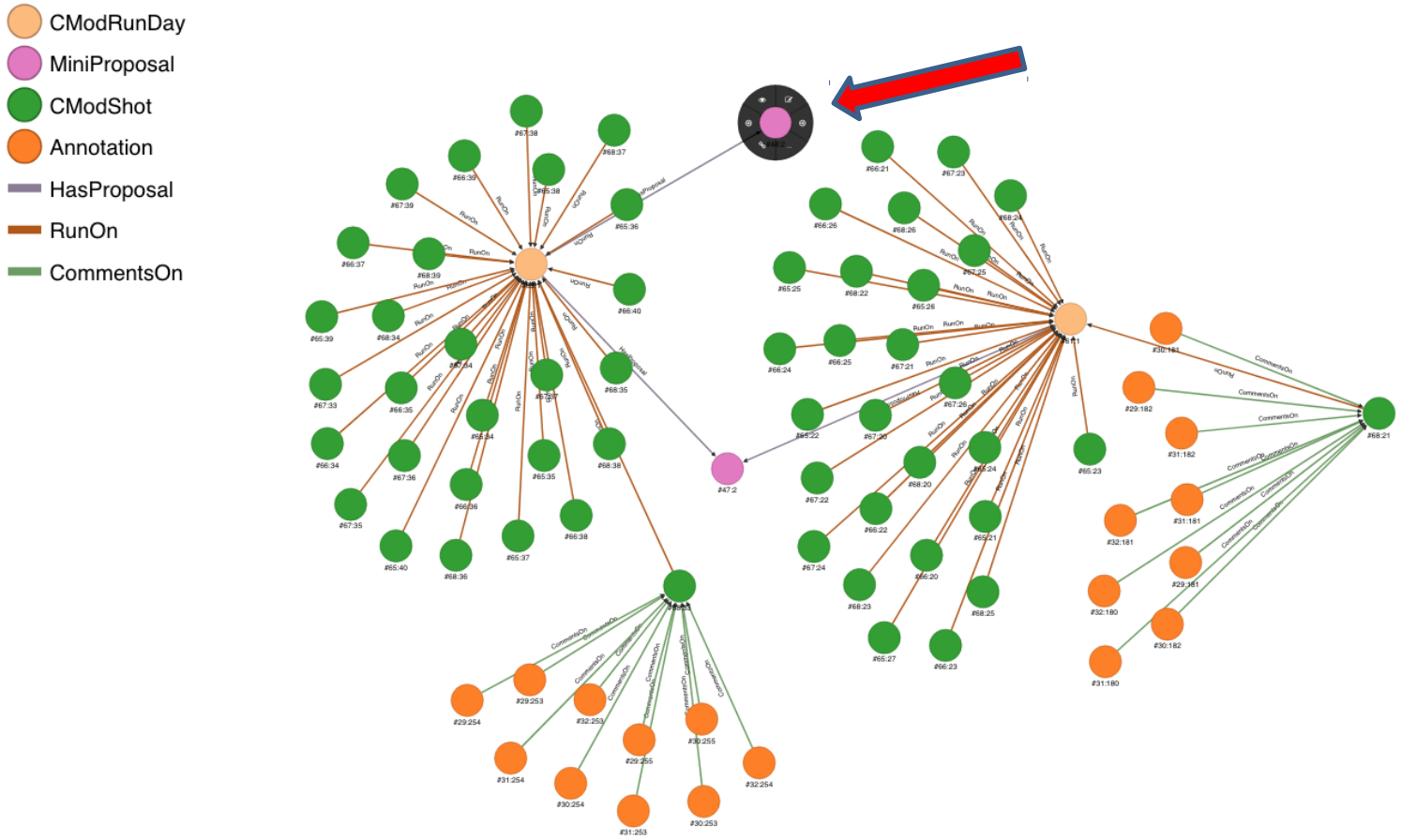
```
$ cat docker-compose.yml
version: '3'

services:
    web:
      build:
        context: './Client'
        args:
          - dbname=${ORIENT_DBNAME}
          - dropboxkey=${DROPBOX_KEY}
      ports:
        - 80:80
      depends_on:
        - proxy
    proxy:
      build: './Proxy'
      volumes:
        - ./Proxy/db:/usr/src/app/db
...
```

Proposals, Run Days, Shots and Annotations



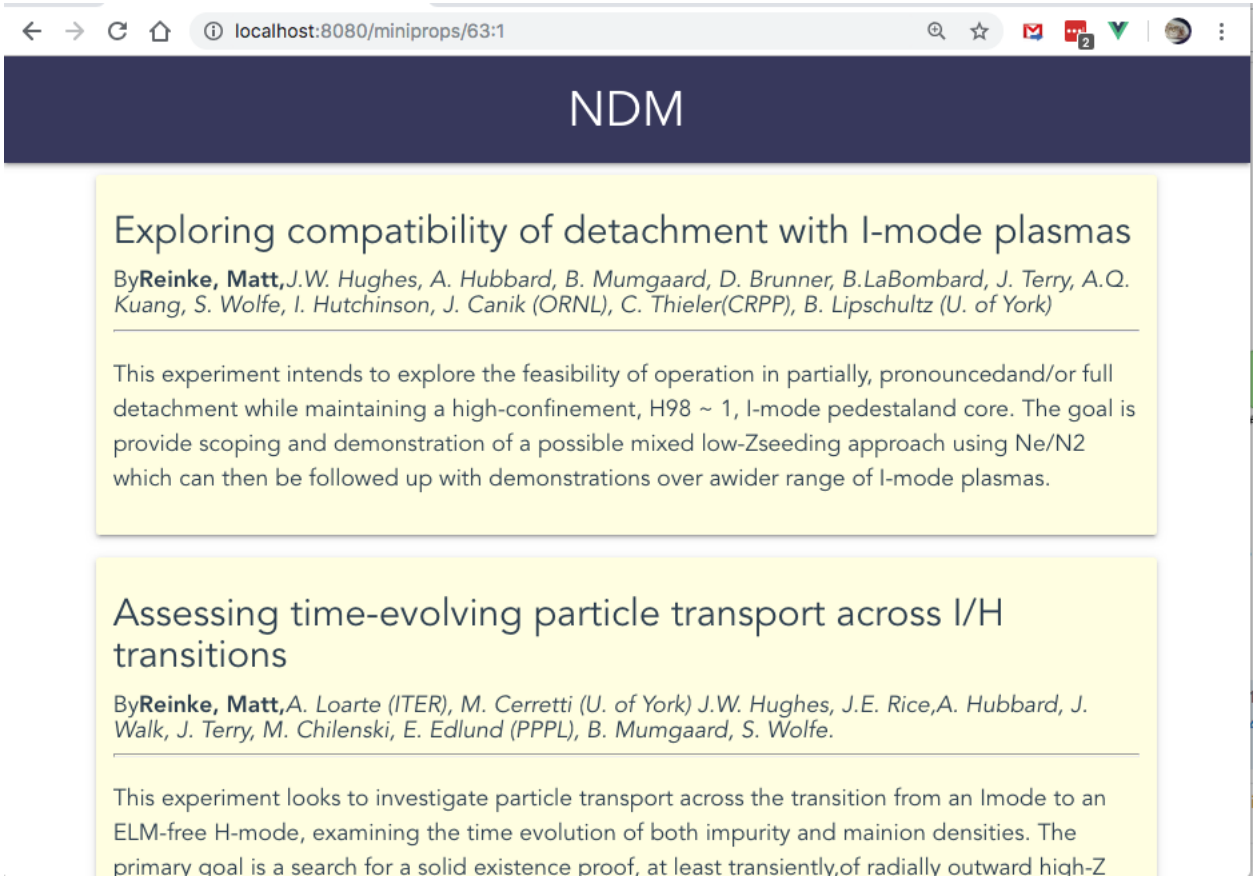
Proposals, Run Days, Shots and Annotations



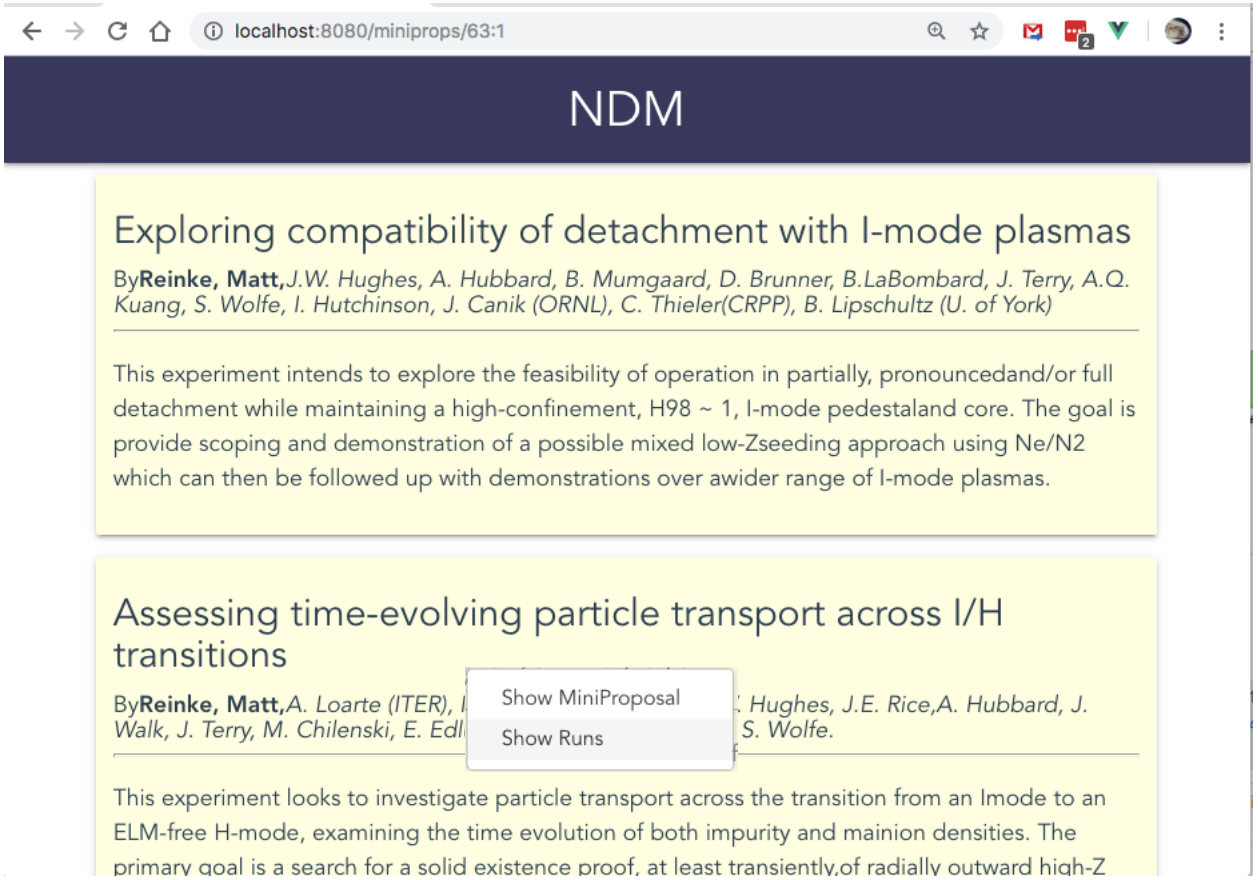
Proposals, Run Days, Shots and Annotations

Abstract	This experiment intends to explore the feasibility of operation in partially, pronouncedand/or full detachment while maintaining a high-confinement, H98 ~ 1, I-mode pedestaland core. The goal is provide scoping and demonstration of a possible mixed low-Zseeding approach using Ne/N2 which can then be followed up with demonstrations over awider range of I-mode plasmas.
CoAuthors	J.W. Hughes, A. Hubbard, B. Mumgaard, D. Brunner, B.LaBombard, J. Terry, A.Q. Kuang, S. Wolfe, I. Hutchinson, J. Canik (ORNL), C. Thieler(CRPP), B. Lipschultz (U. of York)
submitDate	2019-04-29 16:41:53

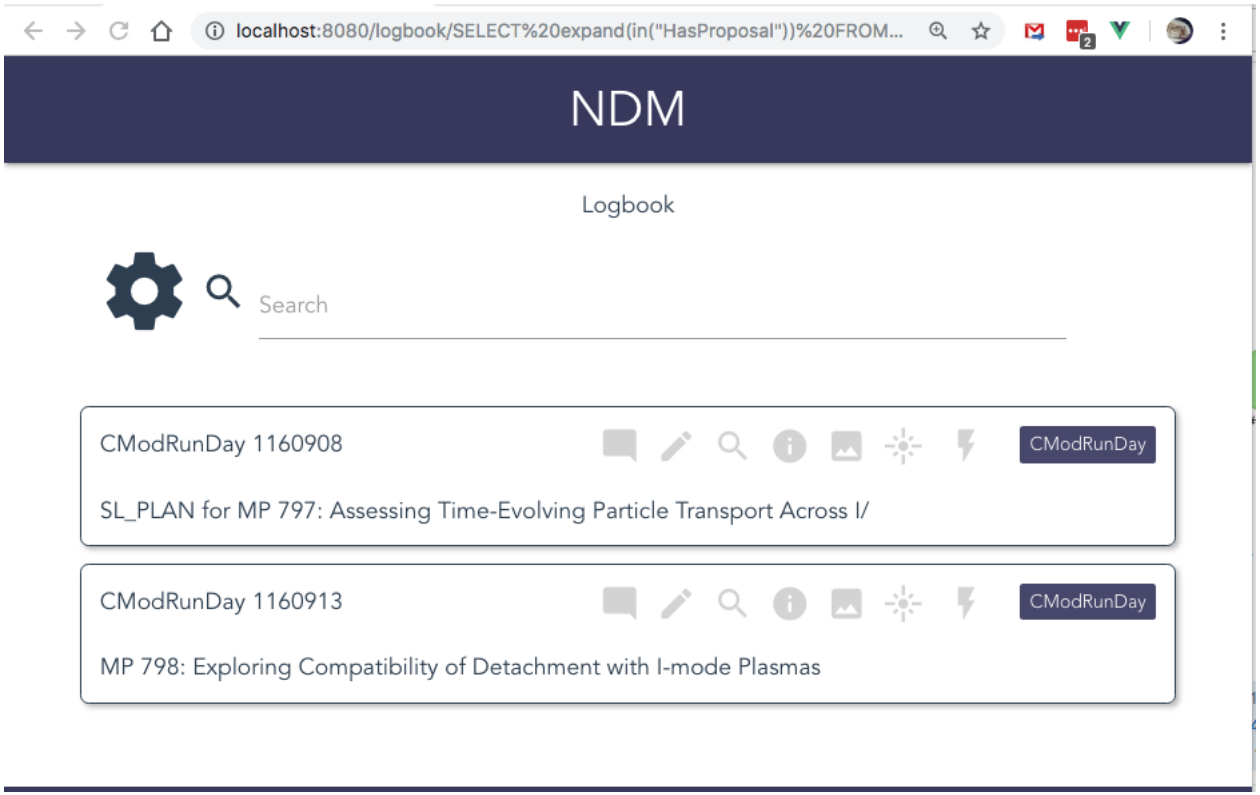
Proposals, Run Days, Shots and Annotations



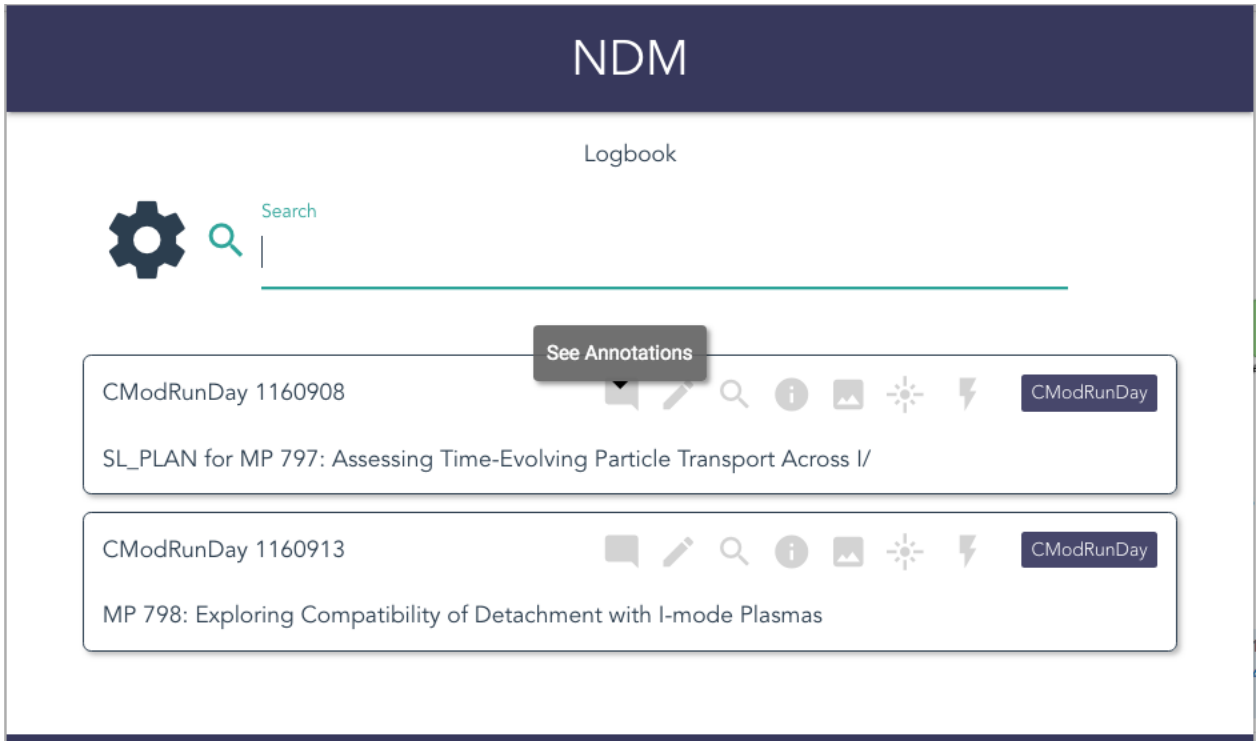
Proposals, Run Days, Shots and Annotations



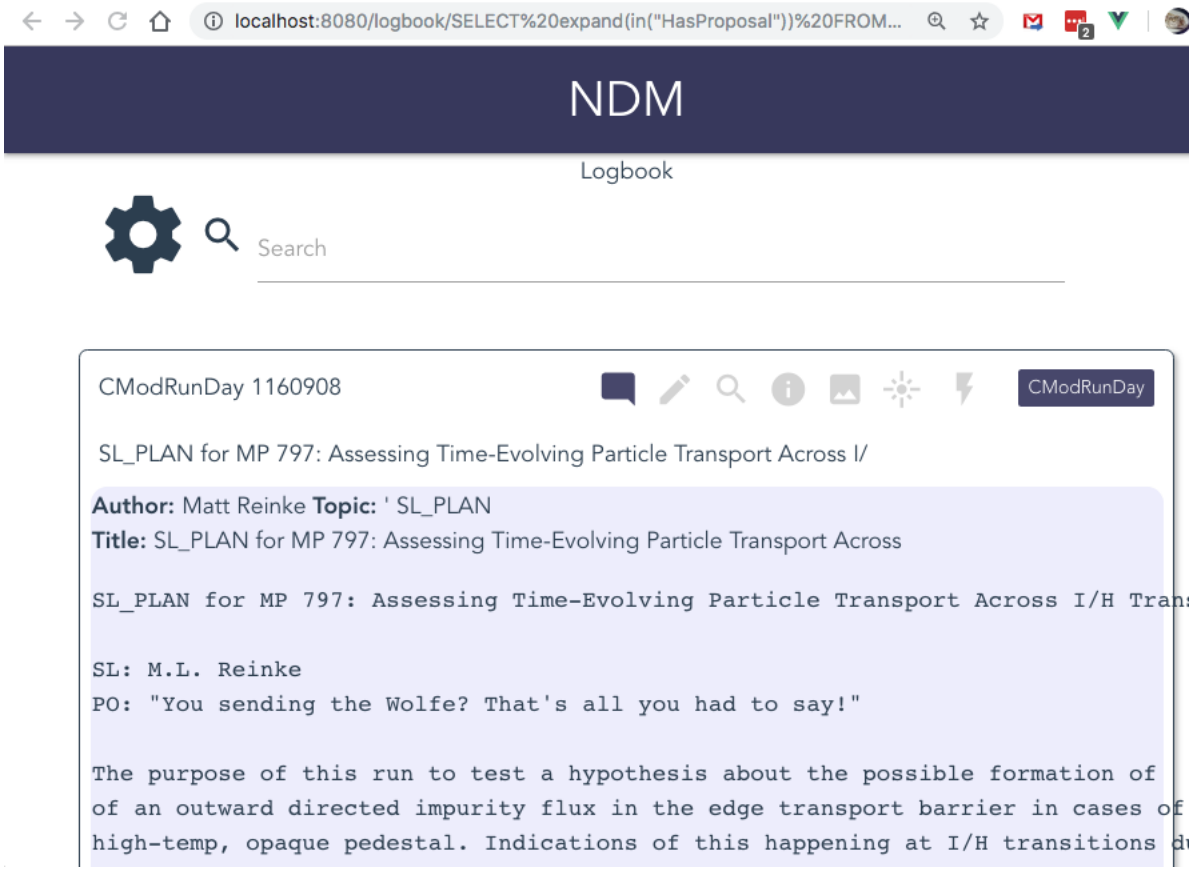
Proposals, Run Days, Shots and Annotations



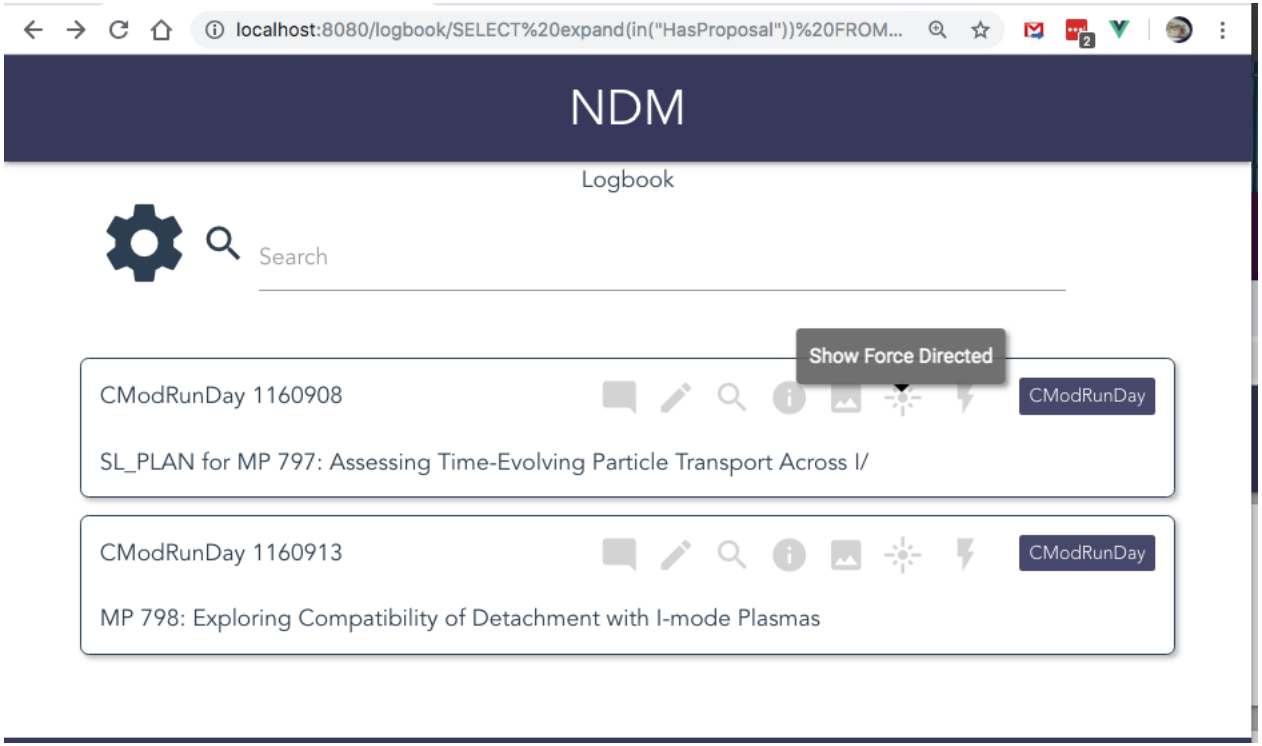
Proposals, Run Days, Shots and Annotations



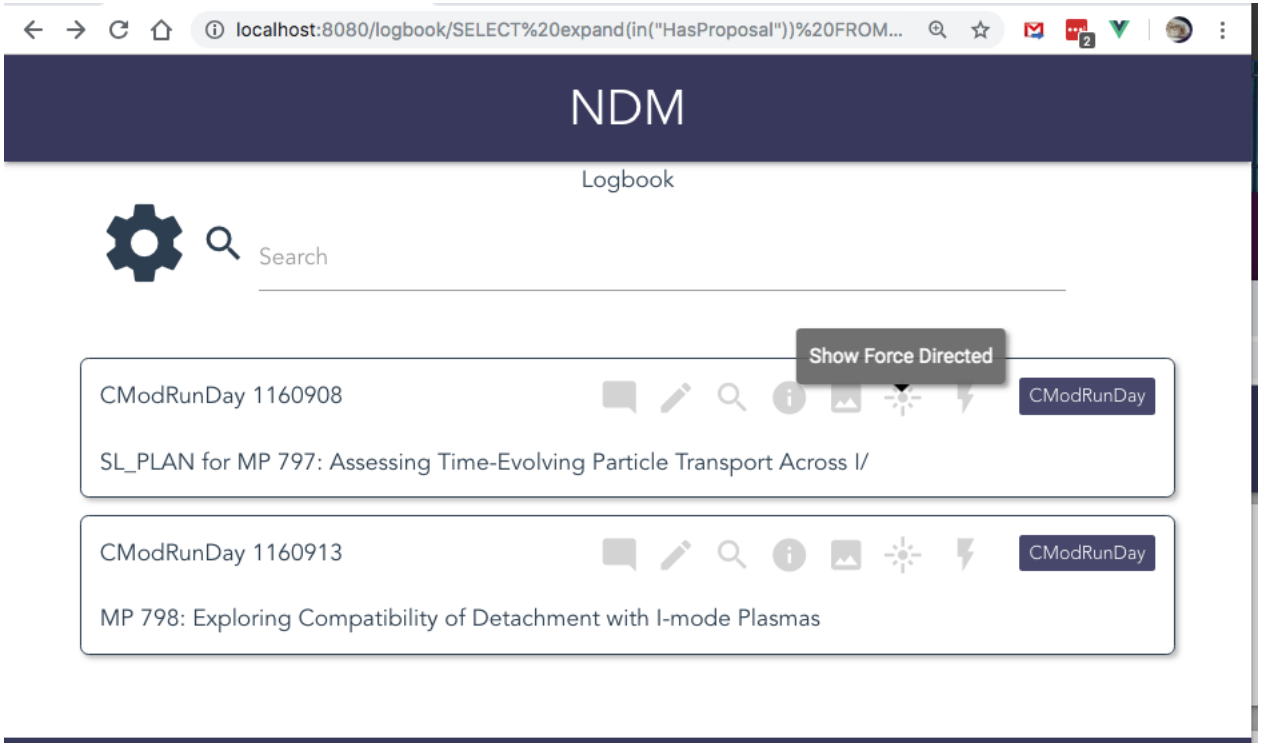
Proposals, Run Days, Shots and Annotations



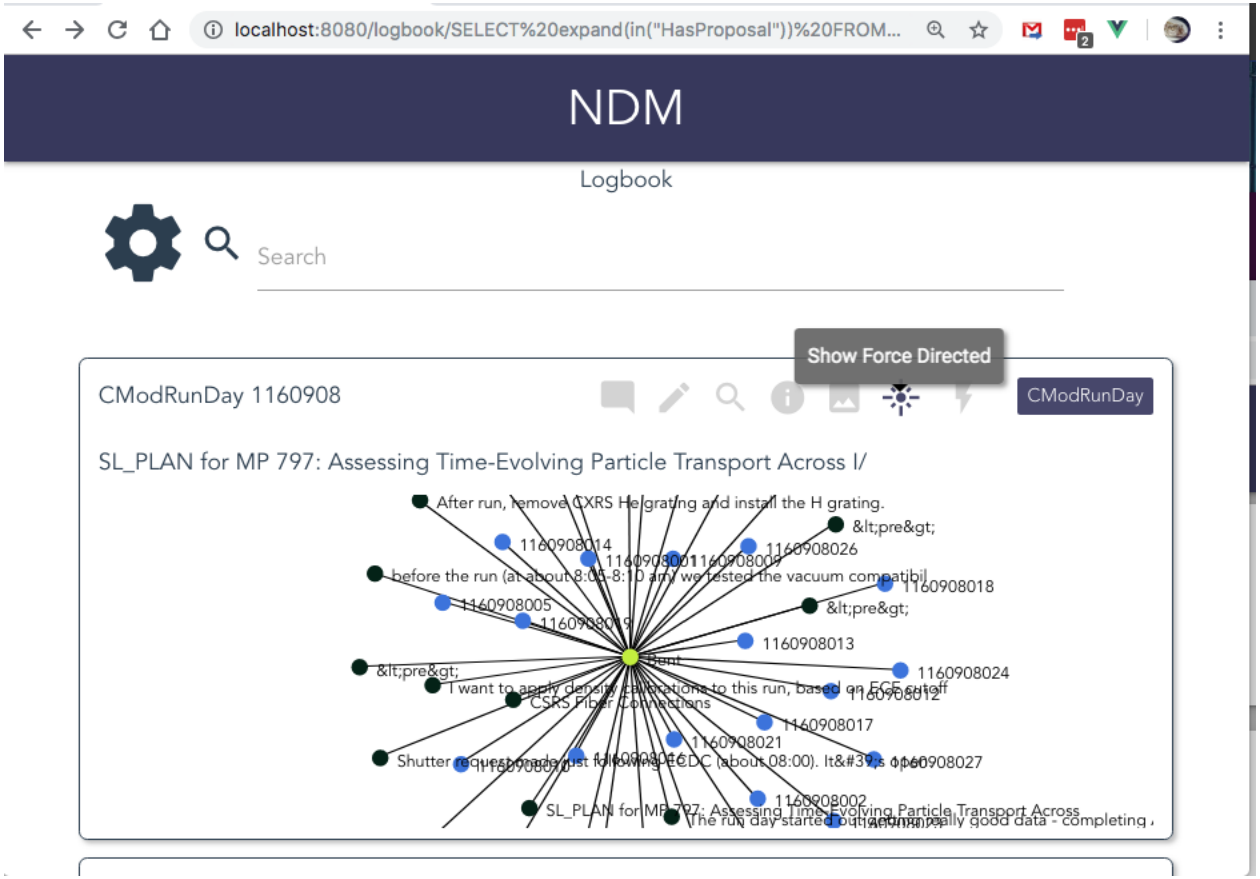
Proposals, Run Days, Shots and Annotations



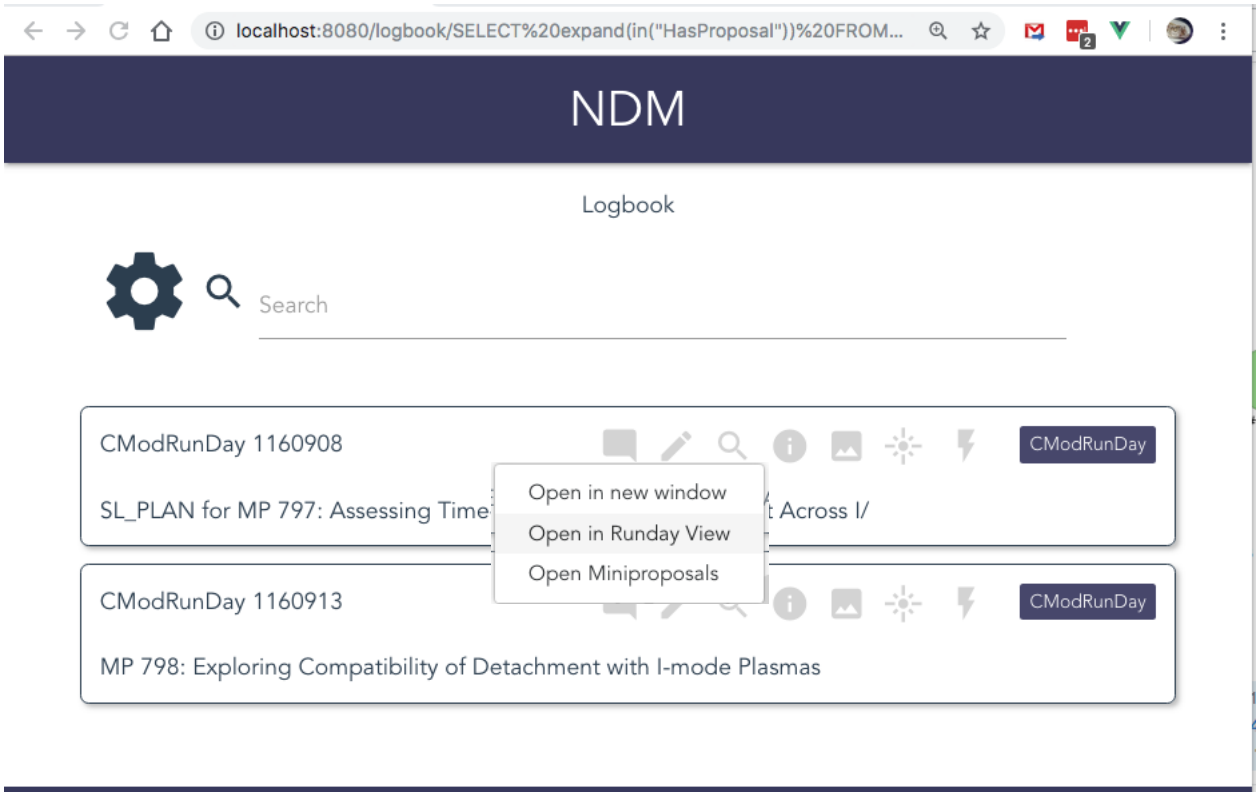
Proposals, Run Days, Shots and Annotations



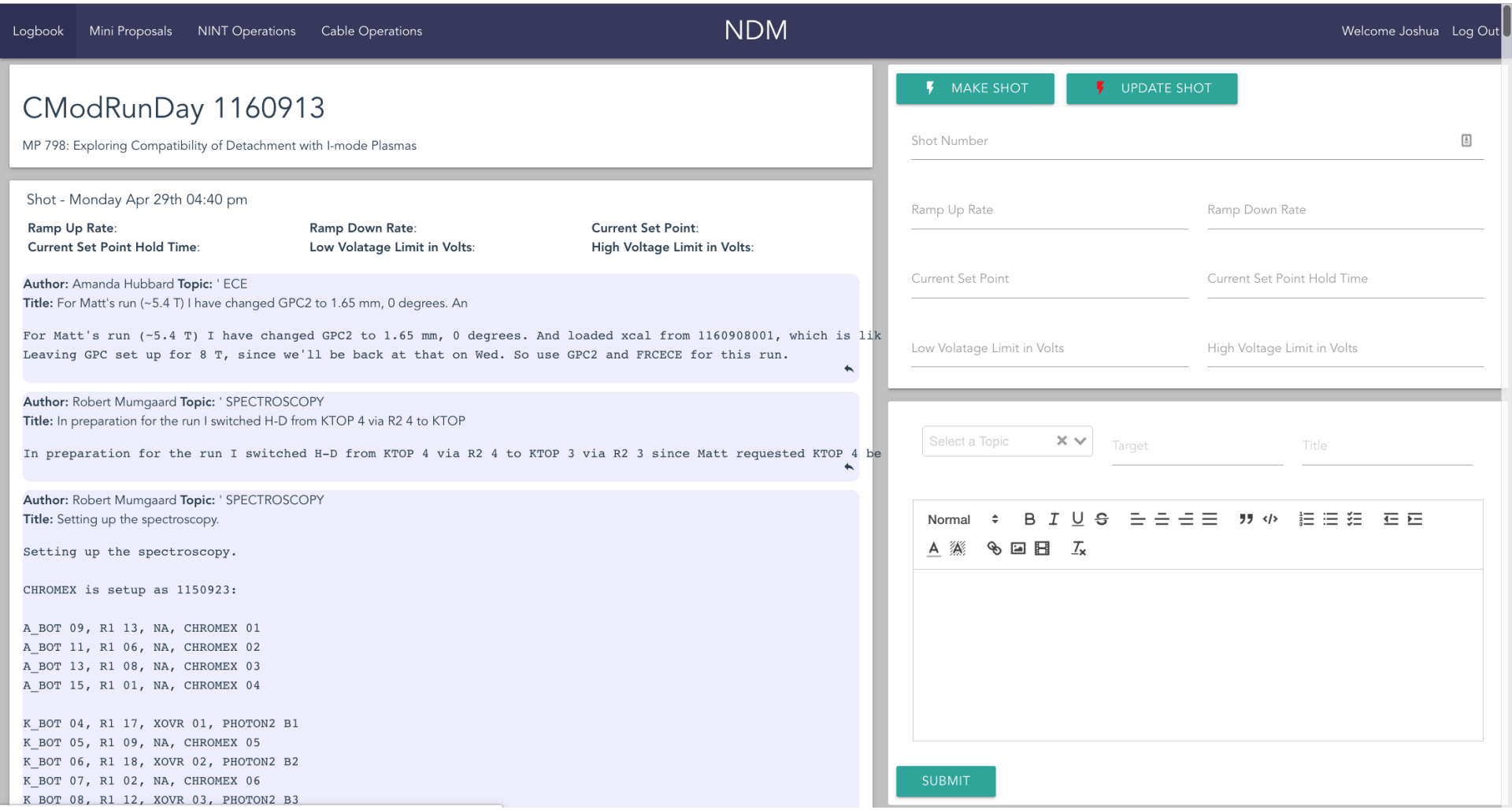
Proposals, Run Days, Shots and Annotations



Proposals, Run Days, Shots and Annotations



Proposals, Run Days, Shots and Annotations



Conclusions

- Start with simple user interfaces
  - Too much complexity and fanciness inhibits iteration
  - Complexity must be carried along
  - Each user facing function has to implement the shared UI
- Applications need to be customized to their tasks
  - General backend with application specific front-end / user interfaces
  - Sharing underlying data structures and APIs
    - facilitates the development of applications
    - Allows for mixing and traversing of information domains
  - Application specific code is needed to achieve needed functionality and usability
- Early users are critical
  - Fancy initial user Interfaces complicates this
- Authentication is difficult
  - Authorization is even harder
- Initial project (and funding) winding down, will continue development.
- We need this to integrate the disparate information about our research.

END