NASA STRATEGY FOR CRITICAL DATA THROUGH KNOWLEDGE MANAGEMENT

Office of the
CHIEF KNOWLEDGE OFFICER

David Meza – Chief Knowledge Architect

Knowledge Management Technologies Event
June 9, 2015

CHALLENGES IN BUILDING KM SYSTEMS

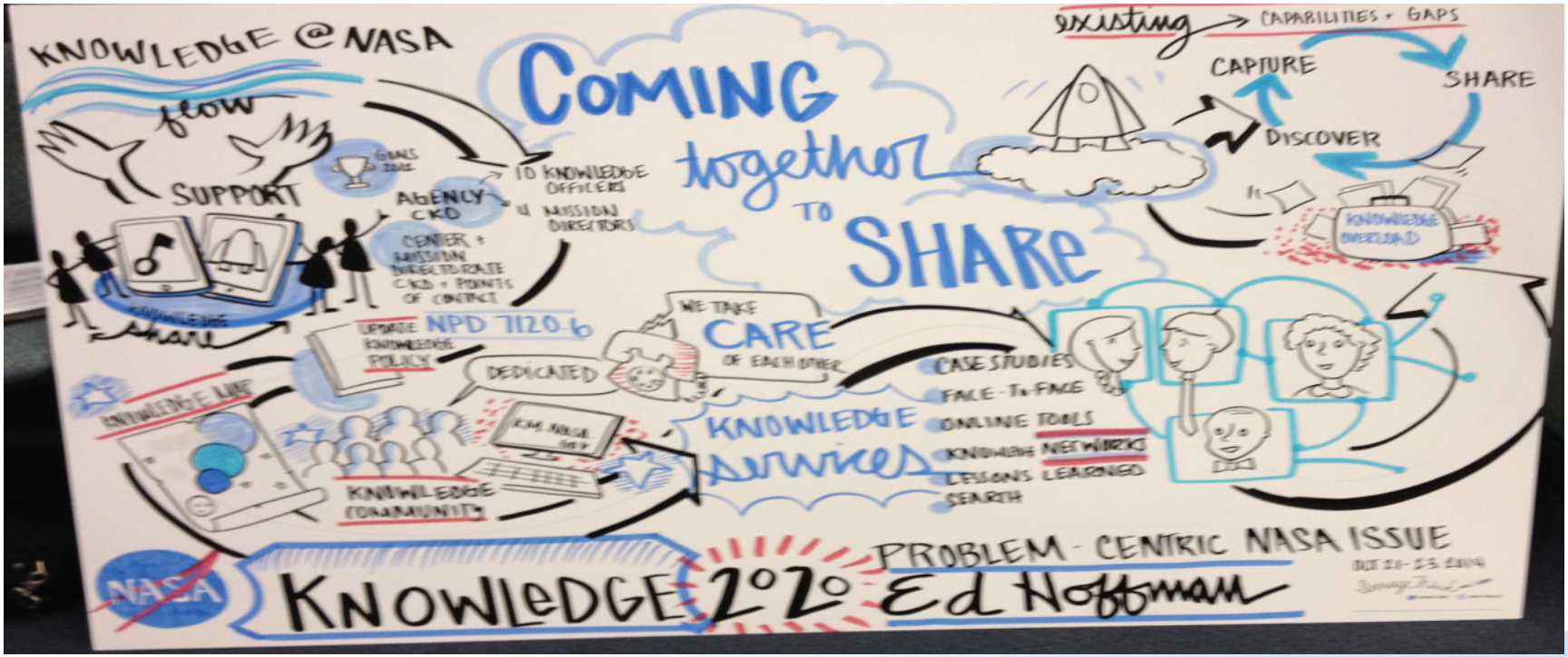
- **Culture**
  - Getting people to share knowledge
- **Knowledge evaluation**
  - Assessing the worth of knowledge across the organization
- **Knowledge processing**
  - Turning tacit knowledge into explicit knowledge
- **Knowledge Implementation**
  - Organizing knowledge and integrating into your processes

ORGANIZATIONAL LEARNING

Organizational learning incorporates the sharing of information, knowledge and good work practices that foster an organizational culture committed to continuous improvement.

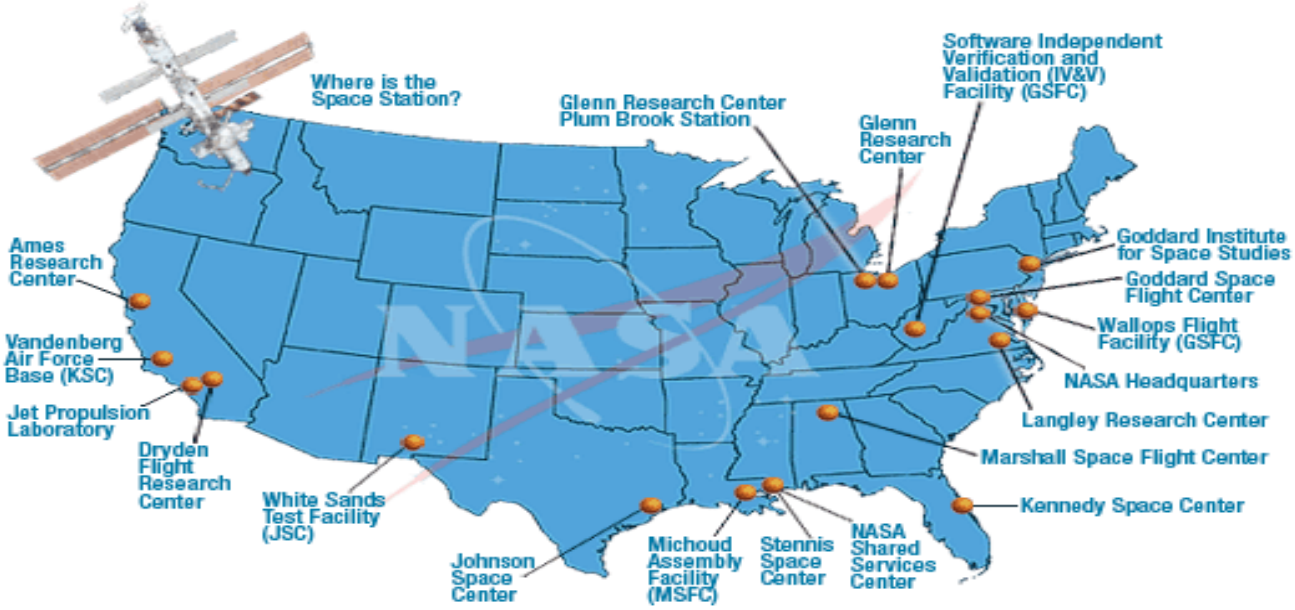
LEARNING TECHNIQUES AND TOOLS

- Pause and Learn (PAL)
- CoP A community of practice
- Wiki
- Best Practice
- Peer Assist (learning before)
- After-action Review (AR) (learning during)
- Retrospect (learning after)
- Knowledge Asset



Current Situation

- Hundreds of millions of documents, reports, project data, lessons learned, scientific research, medical analysis, geo spatial data, IT logs, etc., are stored nation wide
- The data is growing in terms of variety, velocity, volume, value and veracity
- Accessibility to Engineering data sources
- Visibility is limited
- Access to end users for use case development
- How do we turn the data into Knowledge?



ACCESSIBILITY AND VISIBILITY

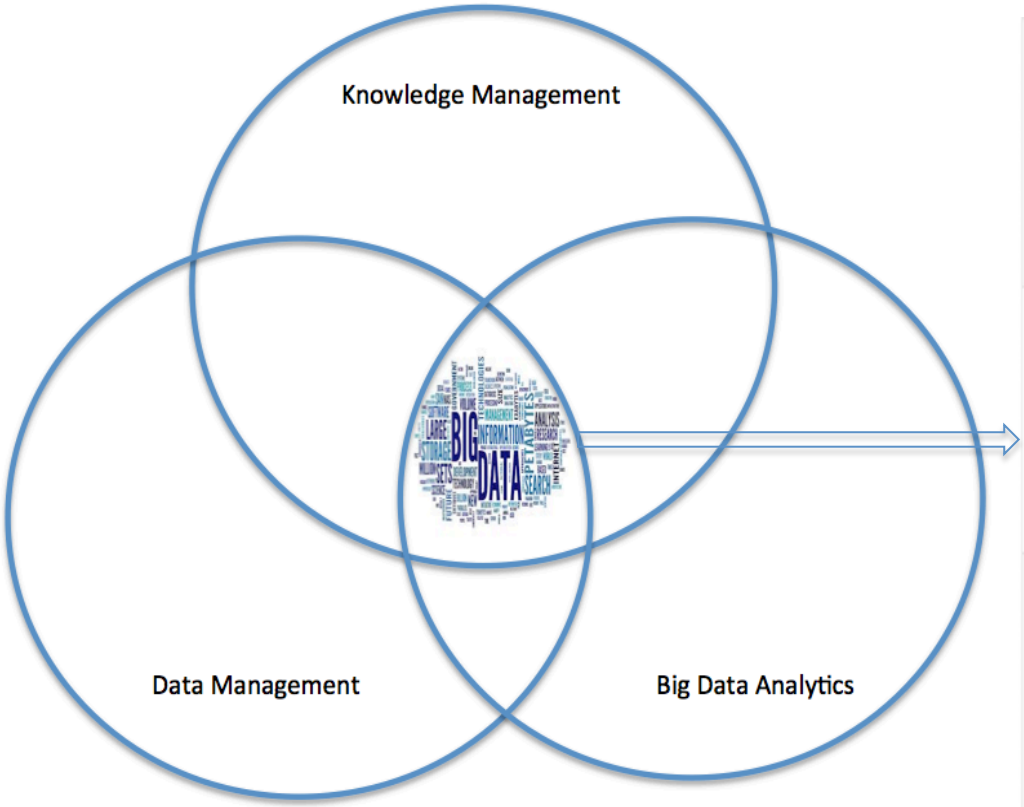
- Identify and Prioritize Agency Data types, repositories, uses and interaction
- Identify User search preference and habits
- Develop Search use cases
- **Review and evaluate search methods**
  - **Semantic**
  - **Faceted**
  - **Computational**
  - **Cognitive**
  - **Natural Language query**
- Explore Presentation Mechanisms

KEY TAKEAWAYS

- There is No One Solution
  - Data Driven Visualization
- Master Data Management Plan is essential
  - Structure
  - Format
  - Metadata
- Identify Critical Data
- Develop Standards for Government and Contractor created data
  - Content Management
  - Storage
  - Accessibility
- We are losing vast amounts of critical data

Approach

To convert data to knowledge a convergence of Knowledge Management, Data Management and Big Data Analytics is necessary



Strategy

- Data Management
  - Framework for setting common procedures and formats for creating and storing data
- Big Data Analytics
  - Provides tools and methodology to convert data into knowledge
- Knowledge Management
  - Tacit to explicit, Search, Business Intelligence

End User is the Priority

- Their primary concern is having access and visibility to critical data as near real time as possible

- Visualization is the KEY to meeting their needs

Emphasis on Visual Analytics

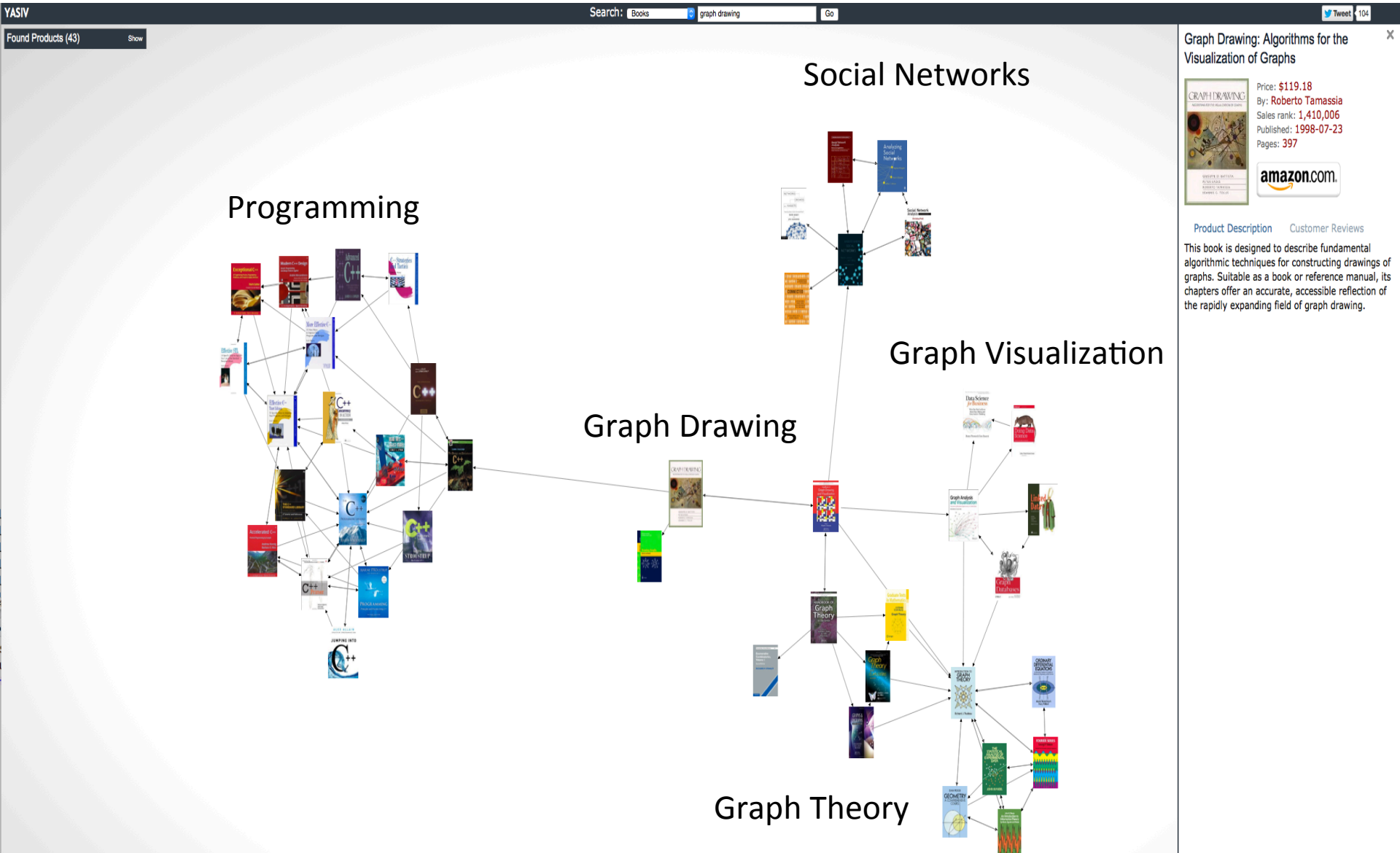
- Visual Analytics shortens the time frame for end users to find information
- Helps users make connections between data
- Anticipating their needs through predictive modeling
- Improves Risk based decisions

Available Tools

- MapReduce
- NoSQL databases
  - Neo4j (graph database), mongoDB (document database)
- Visibility
  - Semantic search, cognitive search, computational search, faceted search
- Analytics
  - Text Analytics, Natural language processing, R Statistical Programming, SAS, Mathematica, Probabilistic Graph Models
- Visualization
  - Visual Analytics, Shiny, D3, Tom Sawyer

Examples

Find a Document…



Find Experts…

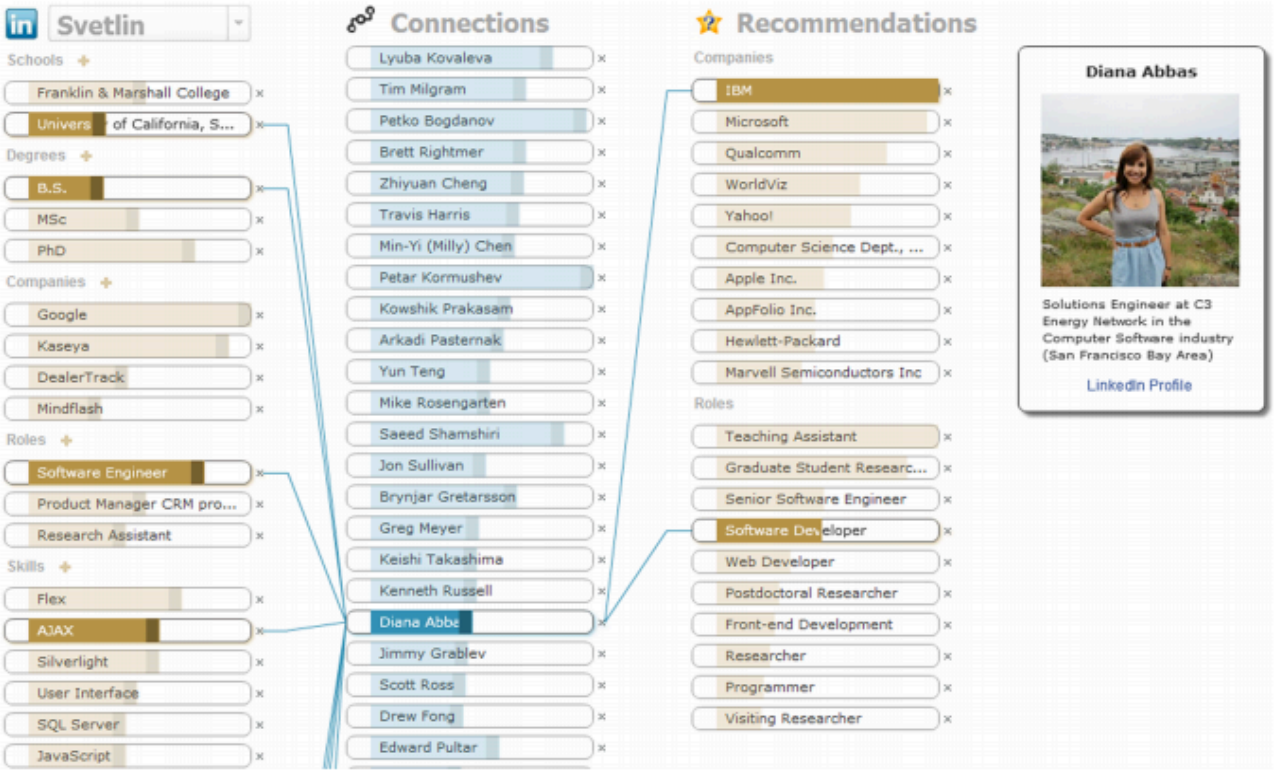
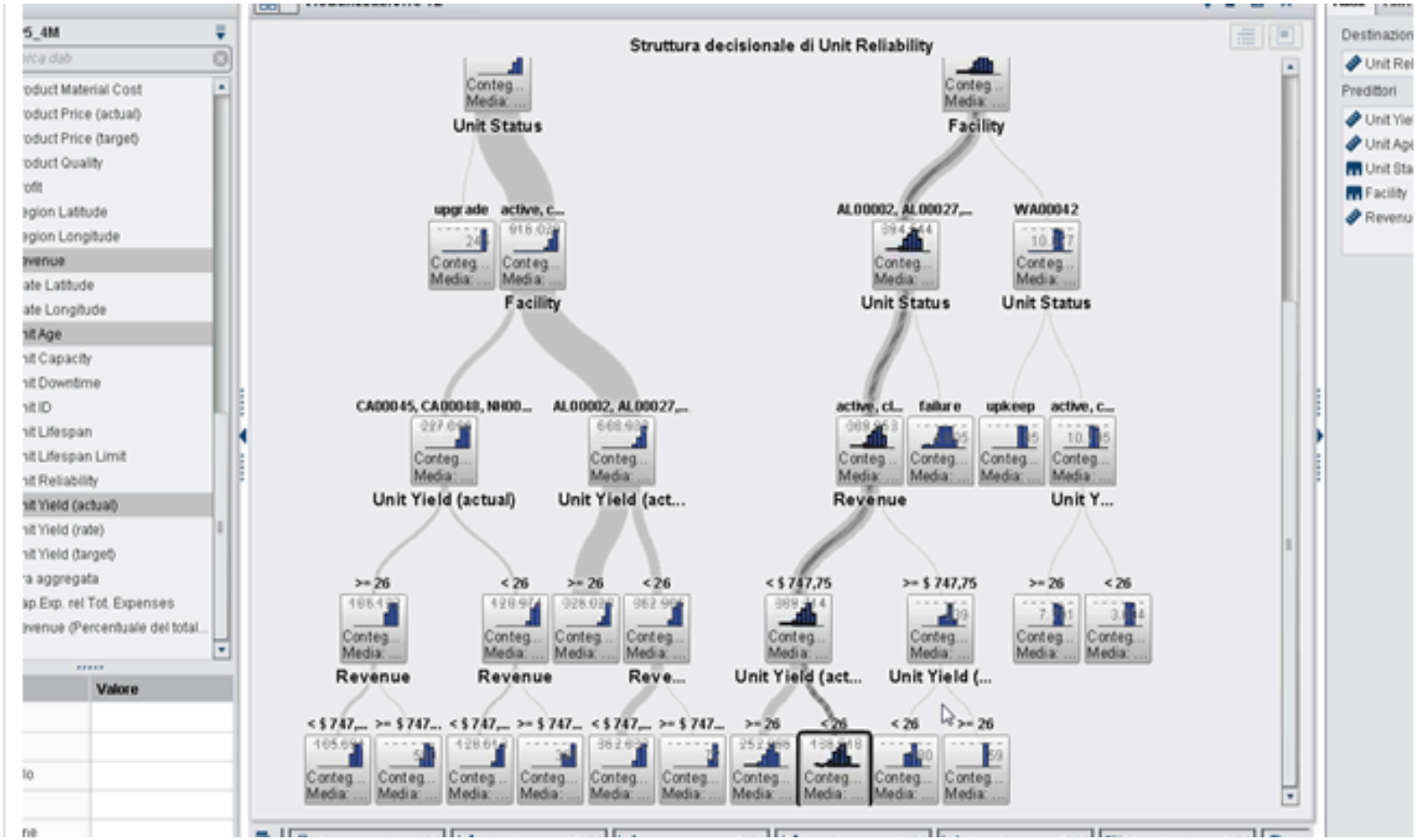


Figure 1. Screenshot of the LinkedVis Recommender System.

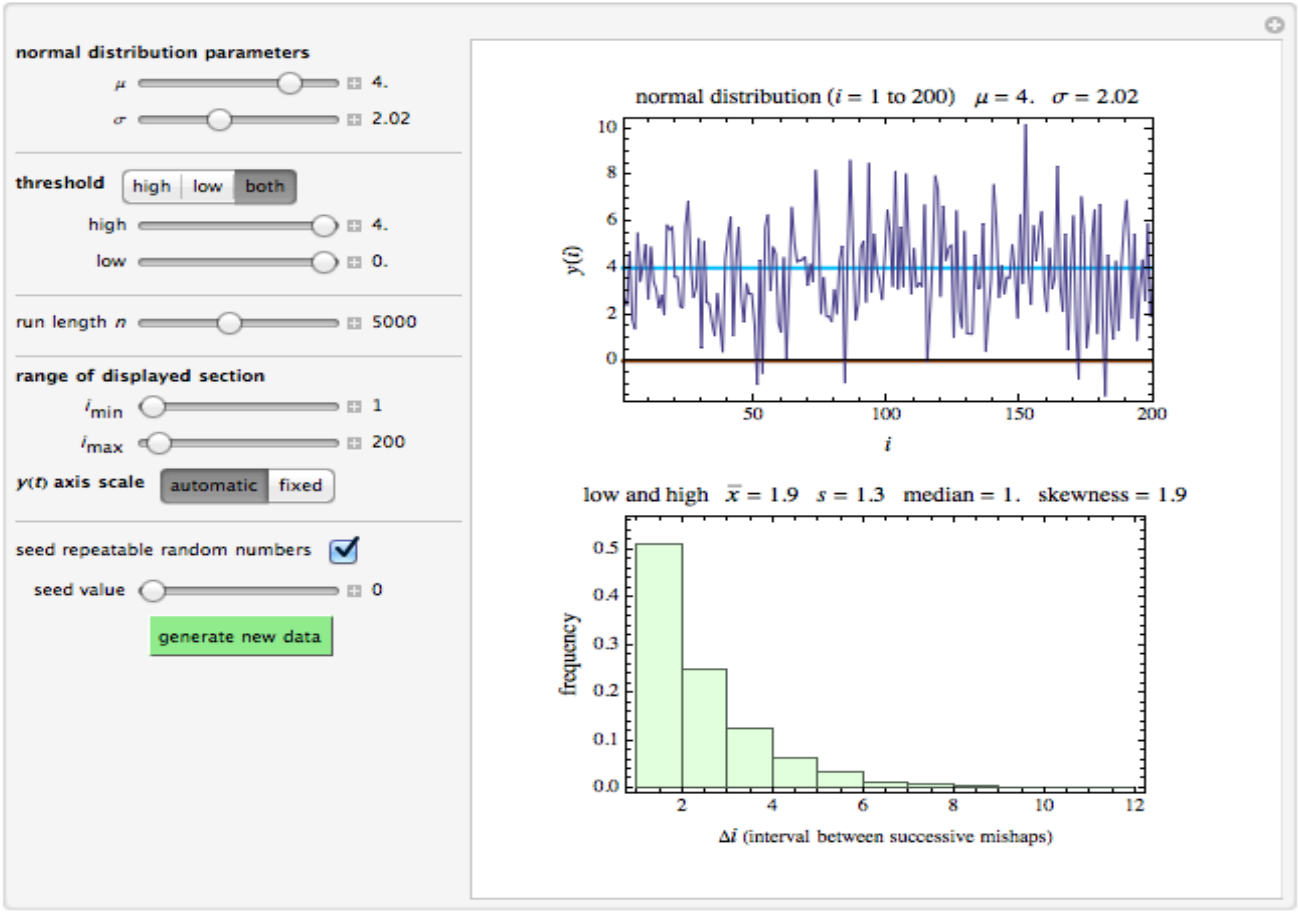
Source: http://cs.ucsb.edu/~jod/papers/C-6-LinkedVis-IUI2013.pdf

Make decisions…



INTERACTIVE DATA REPORTS

Estimating the Time between Mishaps from Quality Control Data

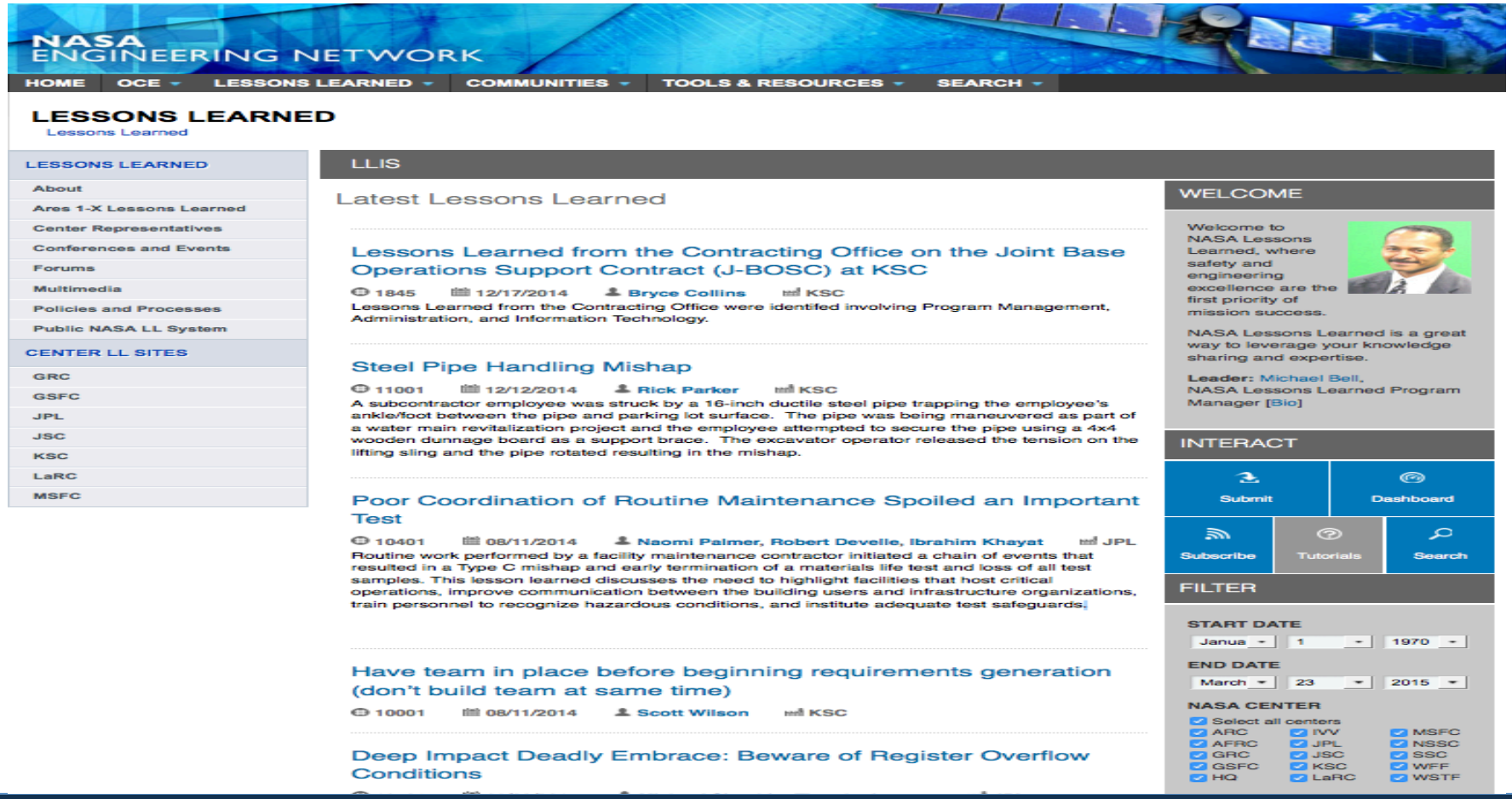


Knowledge Discovery Through Topic Modeling

Office of the
CHIEF KNOWLEDGE OFFICER

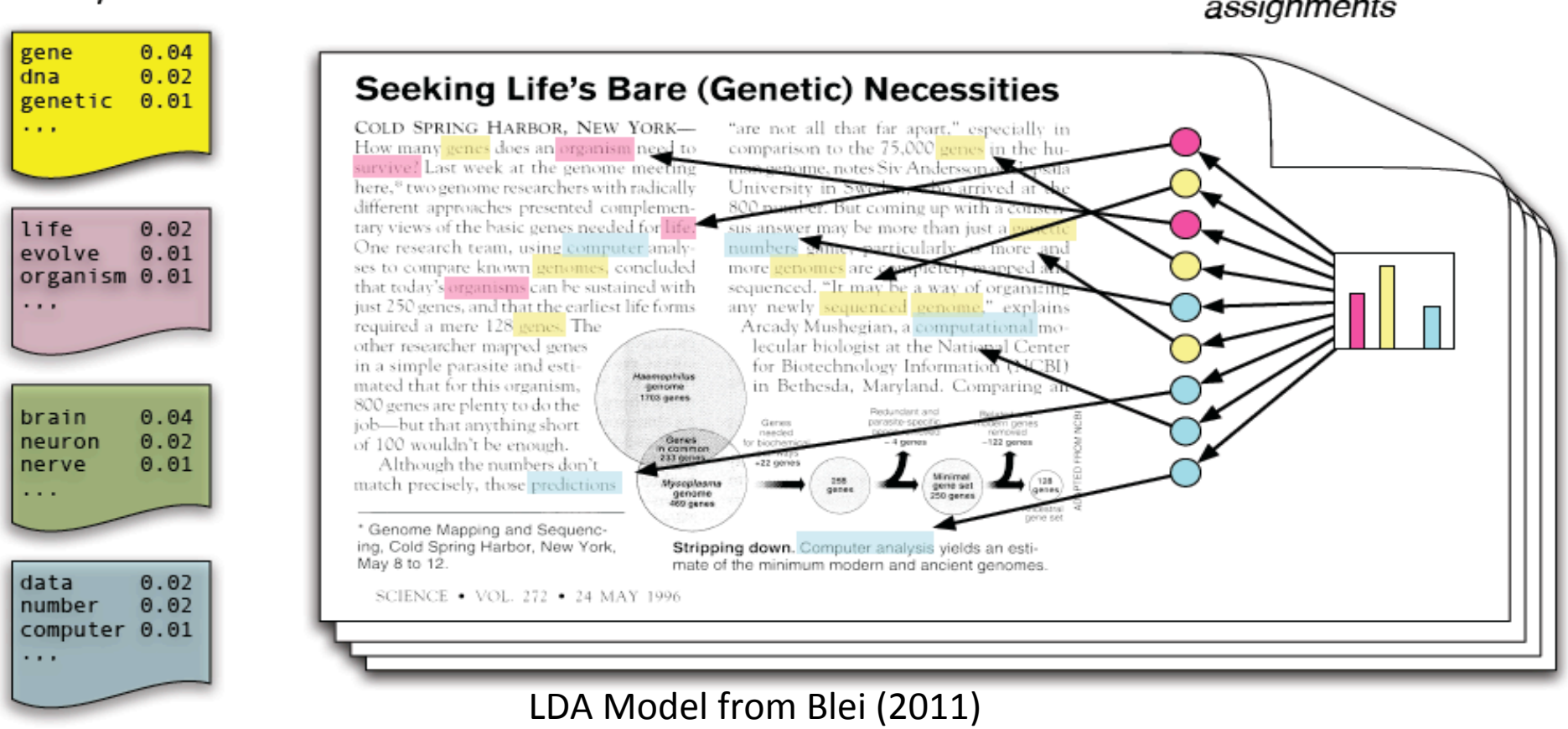
David Meza - Chief Knowledge Architect

LESSON LEARNED DATABASE



2031 lessons submitted across NASA. Filter by date and Center only.
Useful information stored in database.

TOPIC MODELING



LDA Model from Blei (2011)

Topic models are based upon the idea that documents are mixtures of topics, where a topic is a probability distribution over words.

Blei, David M. 2011. "Introduction to Probabilistic Topic Models." *Communications of the ACM*. David Blei homepage - http://www.cs.columbia.edu/~blei/topicmodeling.html

TOPIC MODELING PROCESS STEPS

- Create the Document Term Matrix
  - Store the word frequency per document
- Create the Model
  - Uses LDA to assign each document to a topic
- Explore your model
  - Uncover hidden associations

	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7	Topic 8	Topic 9	Topic 10
1	human	fuel	check	damag	shuttl	technolog	tool	crew	loss	chamber
2	task	water	leak	pump	abort	vendor	heat	iss	command	optic
3	unit	valv	electron	air	main	higher	radiat	comput	fault	delay
4	weight	technician	fluid	vacuum	extens	recommend	exposur	pilot	status	ignit
5	core	cell	leakag	improp	ssme	procur	align	upgrad	sequenc	fiber

Using the tm package, create the corpus from the source
Create the document term matrix, reducing to words occurring in at least five documents

CATEGORIES

```
FirstCategoryByYear <- doBy::orderBy(~-Year, data = FirstCategoryByYear)
head(table(FirstCategoryByYear))
```

```
##                                                 Year
## Category                                        2006 2007 2008 2009 2010
##   Categories                                      12    5    6    5    5
##   Engineering Design                               2    5    6    2    2
##   Integration and Testing                          2    2    1    2    4
##   Manufacturing and Assembly                       2    0    3    1    3
##   Missions and Systems Requirements Definition     1    1    0    0    2
##   Program Management                               1    3    0    0    7
##                                                 Year
## Category                                        2011 2012 2013 2014
##   Categories                                       9    0    0    0
##   Engineering Design                               2    0    0    0
##   Integration and Testing                          9    8    4    2
##   Manufacturing and Assembly                       0    4    0    0
##   Missions and Systems Requirements Definition     3    0    1    0
##   Program Management                               8    0    0    0
```

```
tally(group_by(FirstCBY_tbl, Year))
```

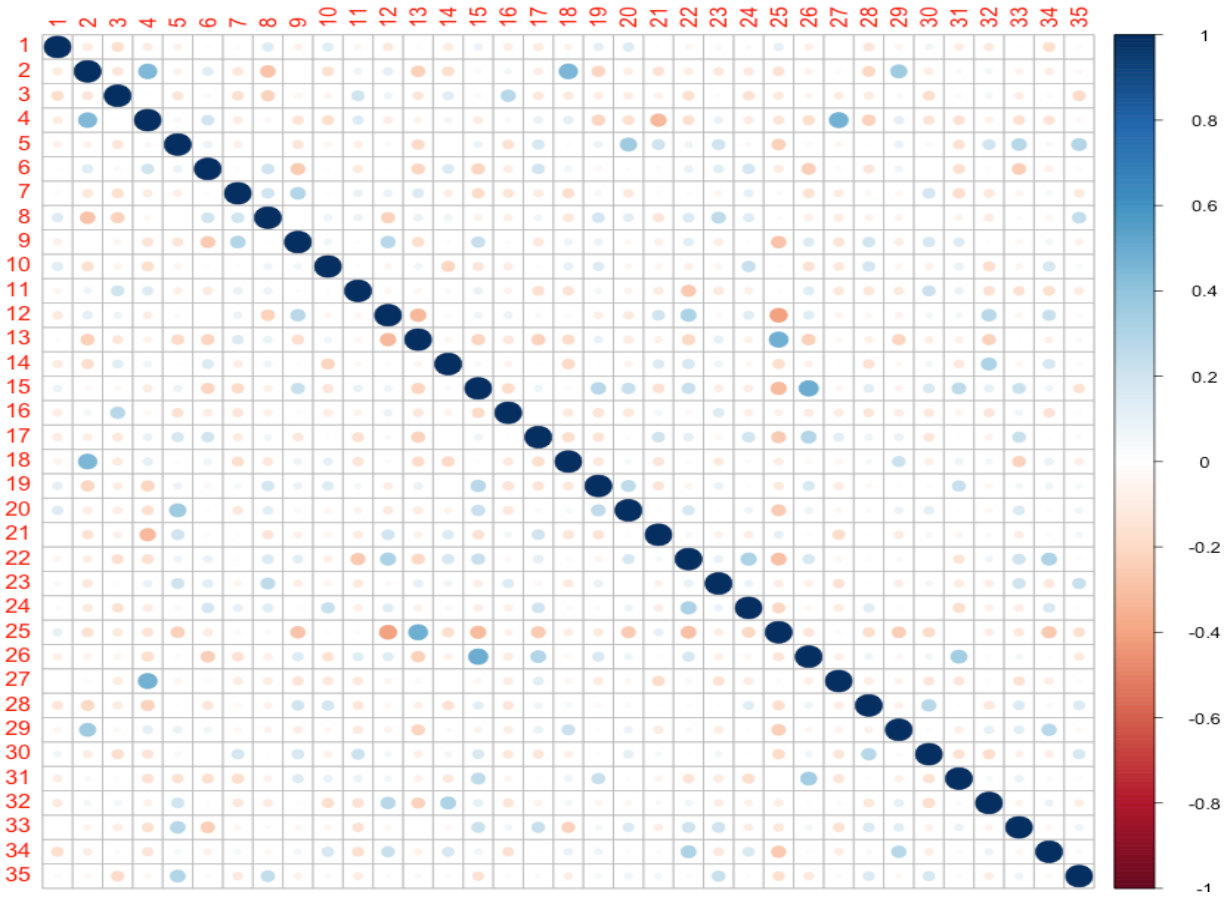
```
## Source: local data frame [9 x 2]
##
##    Year   n
## 1  2006  22
## 2  2007  17
## 3  2008  30
## 4  2009  28
## 5  2010 158
## 6  2011  82
## 7  2012  24
## 8  2013  18
## 9  2014   3
```

```
FirstCBY_tbl %>% count(Category) # Count of Category
```

```
## Source: local data frame [74 x 2]
##
##                                        Category  n
## 1                                    Categories 42
## 2                             Engineering Design 19
## 3                         Integration and Testing 34
## 4                      Manufacturing and Assembly 13
## 5  Missions and Systems Requirements Definition  8
## 6                              Program Management 19
## 7                    Safety and Mission Assurance  3
## 8                          Emergency Preparedness  3
## 9                  Administration / Organization  5
## 10                                      Computers  3
```

The original data frame had additional metadata, Category, Project Phase, Safety Issue, and others.

CORRELATION BY CATEGORY



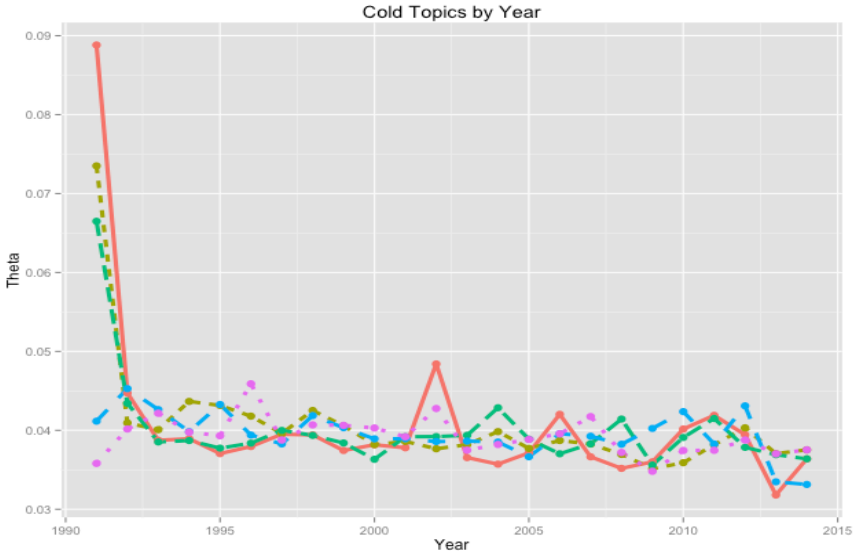
To find the per-document probabilities we extract theta from the fitted model's topic posteriors

TOPICS FOR CATEGORIES

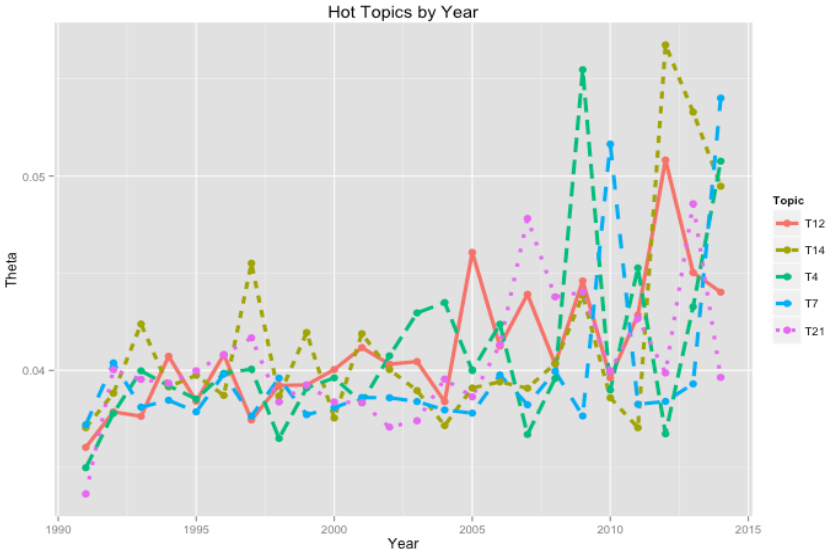
Energy	2
Missions and Systems Requirements Definition	26
Flight Equipment	28
Program and Project Management	24
Review boards	32
Administration / Organization	31
Program level review processes	7
Requirements critical to costing and cost credibility	10
NASA Standards	17
Communication Systems	17
Integration and Testing	3
Mission Operations and Ground Support Systems	13
Flight Operations	24
Facilities	17
Pressure Vessels	10
Security	17
Aerospace Safety Advisory Panel	25
Disposal	21
Manufacturing and Assembly	21
Independent Verification and Validation	28

With mean theta available, it is possible to select the most representative topics of each category.

TOPIC TRENDS



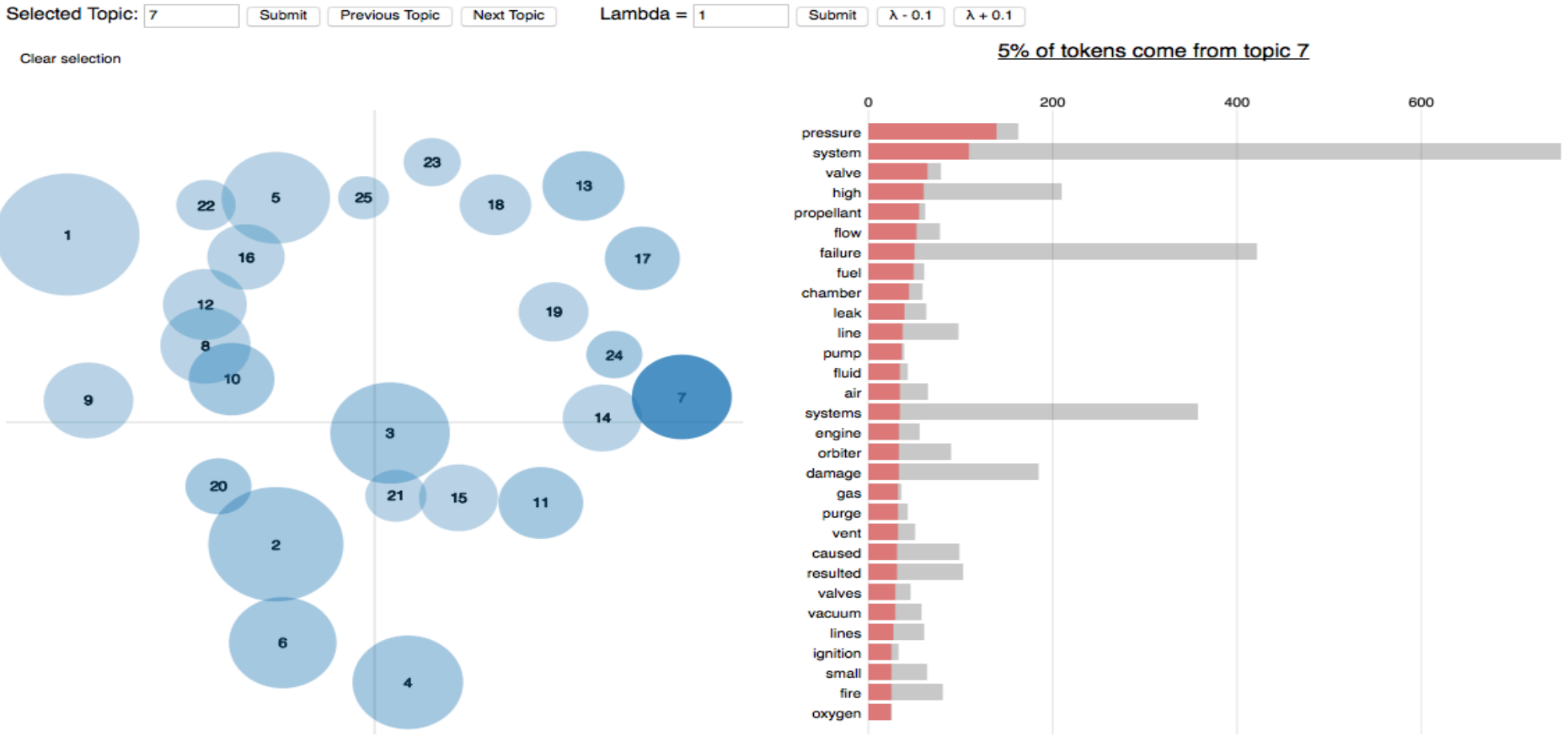
Topic 22	Topic 20	Topic 10	Topic 2	Topic 11
fuel	circuit	access	damag	pump
valv	payload	posit	electr	chamber
gse	rocket	stop	connector	air
water	electr	platform	wire	ignit
cell	connect	switch	improp	surface
technician	termin	secur	inadvert	delay
bay	fire	extens	pin	vacum
sting	solid	transport	mate	filter
angl	short	deploy	catastroph	oxygen
opf	measur	door	shown	failures
delug	power	supervisor	esd	source
emerg	electrostat	bridg	har	combust



Topic 12	Topic 14	Topic 4	Topic 7	Topic 21
lesson	materi	human	install	fire
hazard	thermal	task	vent	jpl
prime	excess	research	panel	batteri
oversight	contamin	protect	blanket	drive
analysis	temperatur	hose	brimg	serious
satellit	clean	weight	back	display
care	properti	visual	post	emerg
bolt	hot	flex	brspan	mar
performed	margin	higher	fit	damage
plate	coat	spaceflight	sound	backup
crew	materials	carri	caviti	late
handl	spot	short	see	msl

Using mean of theta by years to trend topics

TOPIC VISUALIZATION



Next Phase in the process, to provide users an easy method to search through the lessons. Visualization created using LDAvis and Shiny package

QUESTIONS?

Menu | Find | Edit data ON OFF

Nodes | Edges | Paths | Patterns

Type to search nodes...

1 SELECTED NODES

- Fuel and Oxidizer Storage Tank Relief Valves. #6202

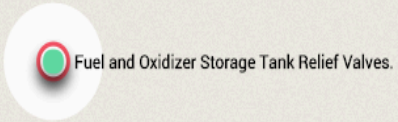
Lesson

Select... | Deselect | Hide | Expand...

PROPERTIES

Find a property...

abstract	
id	146
lesson	Lack of relief valve isolation has resulted in excessive toxic vapor releases to the environment.
org	
safety	false
title	Fuel and Oxidizer Storage Tank Relief Valves.



Menu | Find | Edit data ON OFF

Nodes | Edges | Paths | Patterns

Type to search nodes...

1 SELECTED NODES

• Fuel and Oxidizer Storage Tank Relief Valves. #6202

Lesson

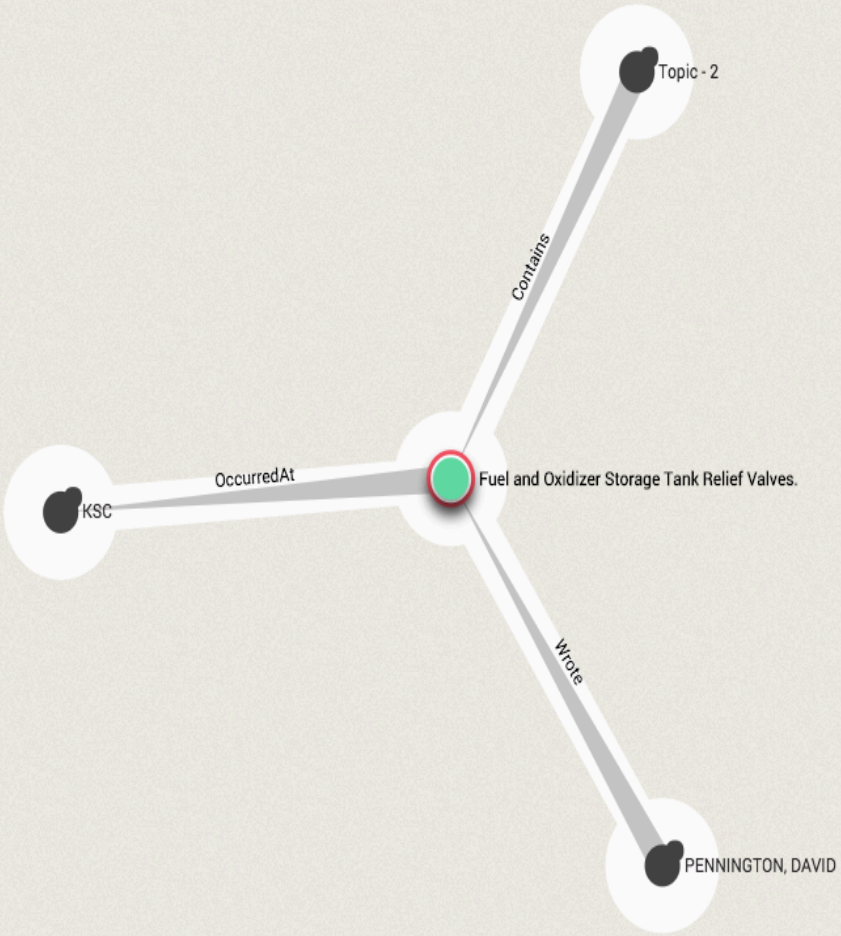
Select... | Deselect | Hide | Collapse

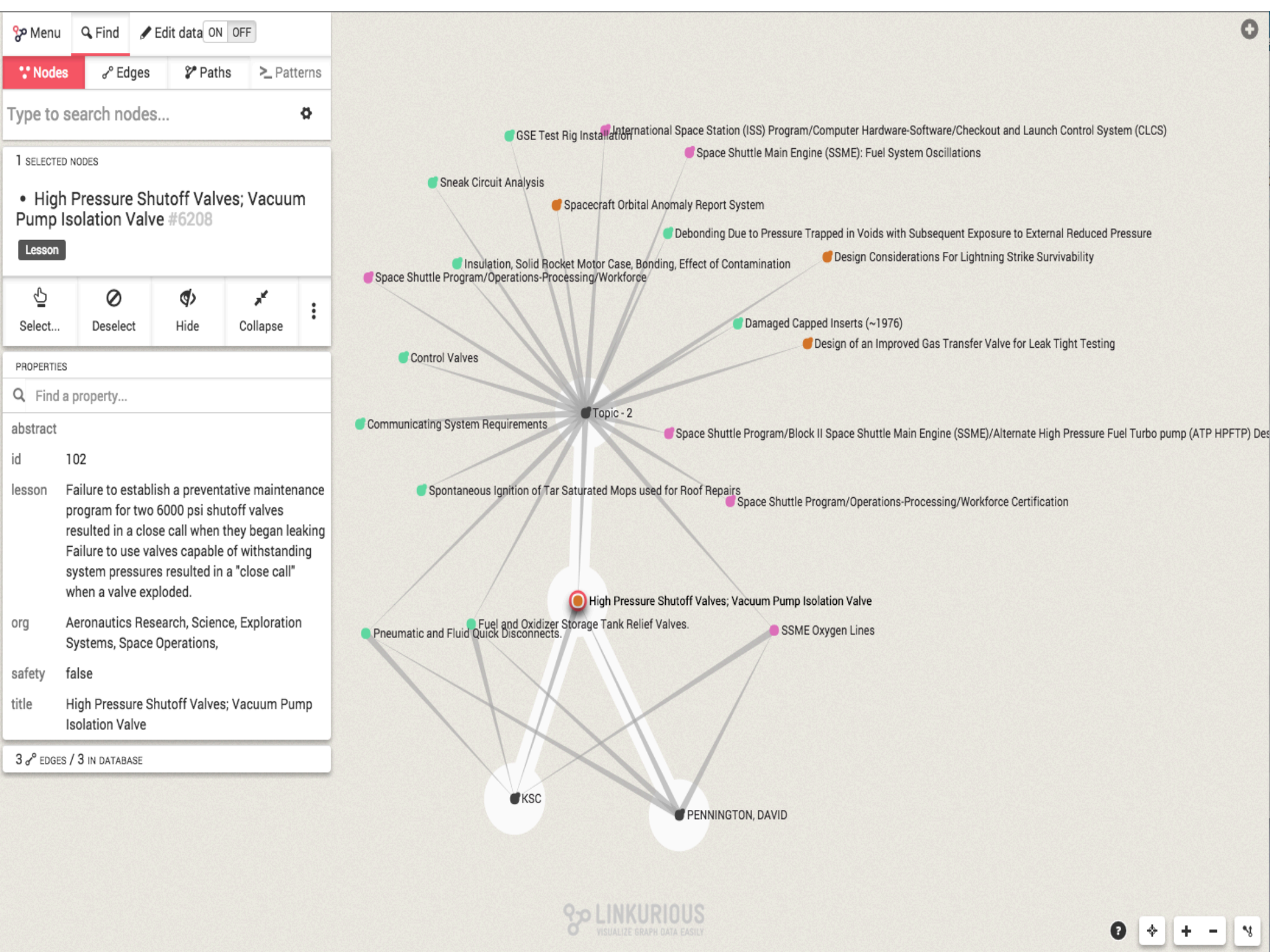
PROPERTIES

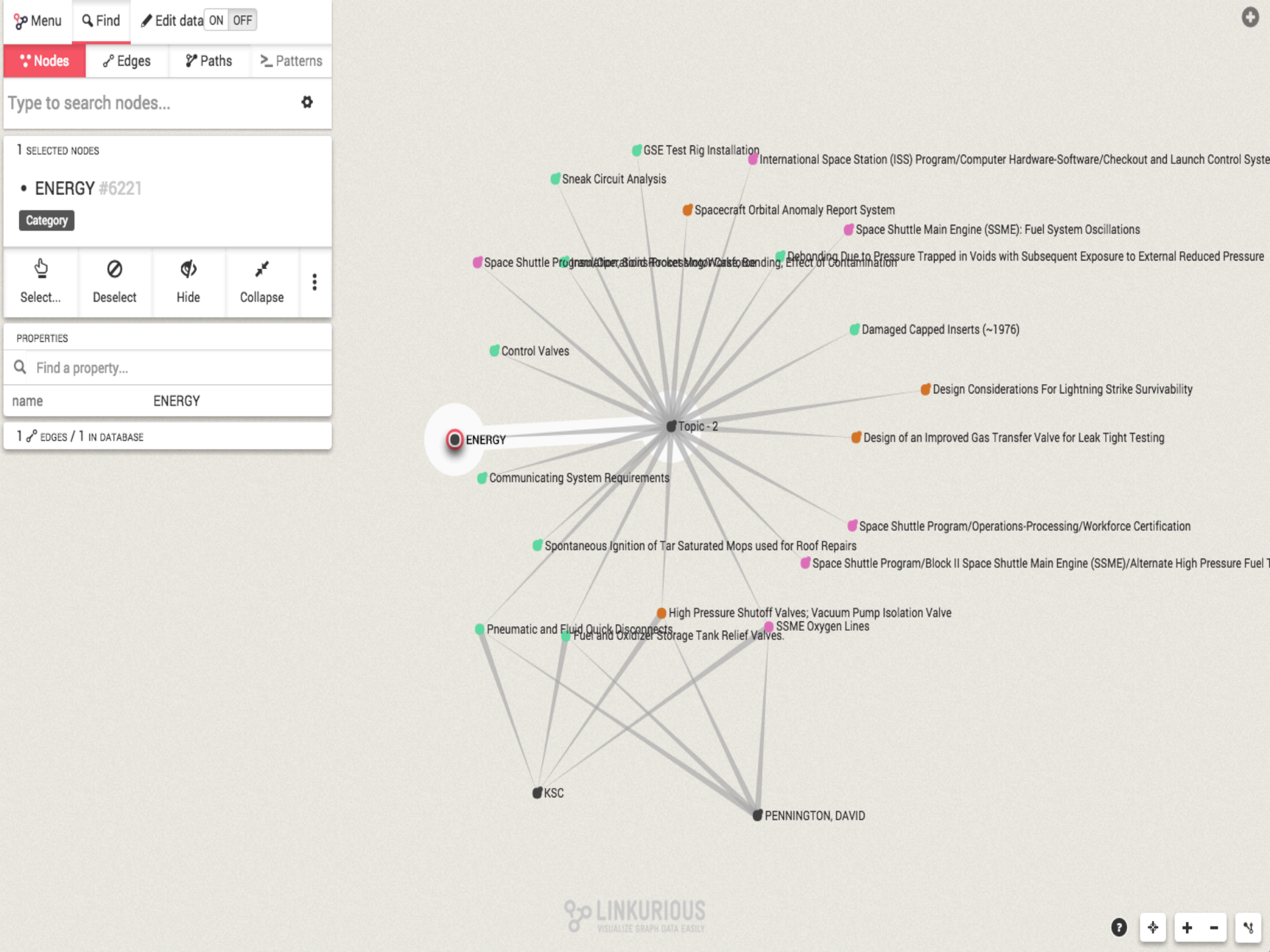
Find a property...

abstract	
id	146
lesson	Lack of relief valve isolation has resulted in excessive toxic vapor releases to the environment.
org	
safety	false
title	Fuel and Oxidizer Storage Tank Relief Valves.

3 EDGES / 3 IN DATABASE




Menu
Find
Edit data
ON
OFF
Nodes
Edges
Paths
Patterns
Type to search nodes...
1 SELECTED NODES
• High Pressure Shutoff Valves; Vacuum Pump Isolation Valve #6208
Lesson
Select...
Deselect
Hide
Collapse
PROPERTIES
Find a property...
abstract
id 102
lesson Failure to establish a preventative maintenance program for two 6000 psi shutoff valves resulted in a close call when they began leaking Failure to use valves capable of withstanding system pressures resulted in a "close call" when a valve exploded.
org Aeronautics Research, Science, Exploration Systems, Space Operations,
safety false
title High Pressure Shutoff Valves; Vacuum Pump Isolation Valve
3 EDGES / 3 IN DATABASE
GSE Test Rig Installation
International Space Station (ISS) Program/Computer Hardware-Software/Checkout and Launch Control System (CLCS)
Space Shuttle Main Engine (SSME): Fuel System Oscillations
Sneak Circuit Analysis
Spacecraft Orbital Anomaly Report System
Debonding Due to Pressure Trapped in Voids with Subsequent Exposure to External Reduced Pressure
Insulation, Solid Rocket Motor Case, Bonding, Effect of Contamination
Design Considerations For Lightning Strike Survivability
Space Shuttle Program/Operations-Processing/Workforce
Damaged Capped Inserts (~1976)
Design of an Improved Gas Transfer Valve for Leak Tight Testing
Control Valves
Topic - 2
Communicating System Requirements
Space Shuttle Program/Block II Space Shuttle Main Engine (SSME)/Alternate High Pressure Fuel Turbo pump (ATP HPFTP) Des
Spontaneous Ignition of Tar Saturated Mops used for Roof Repairs
Space Shuttle Program/Operations-Processing/Workforce Certification
High Pressure Shutoff Valves; Vacuum Pump Isolation Valve
Fuel and Oxidizer Storage Tank Relief Valves.
Pneumatic and Fluid Quick Disconnects.
SSME Oxygen Lines
KSC
PENNINGTON, DAVID
LINKURIOUS
VISUALIZE GRAPH DATA EASILY


Menu
Find
Edit data
ON
OFF
Nodes
Edges
Paths
Patterns
Type to search nodes...
1 SELECTED NODES
• ENERGY #6221
Category
Select...
Deselect
Hide
Collapse
PROPERTIES
Find a property...
name
ENERGY
1 EDGES / 1 IN DATABASE
GSE Test Rig Installation
International Space Station (ISS) Program/Computer Hardware-Software/Checkout and Launch Control Syste
Sneak Circuit Analysis
Spacecraft Orbital Anomaly Report System
Space Shuttle Main Engine (SSME): Fuel System Oscillations
Debonding Due to Pressure Trapped in Voids with Subsequent Exposure to External Reduced Pressure
Damaged Capped Inserts (~1976)
Control Valves
Design Considerations For Lightning Strike Survivability
Topic - 2
ENERGY
Design of an Improved Gas Transfer Valve for Leak Tight Testing
Communicating System Requirements
Space Shuttle Program/Operations-Processing/Workforce Certification
Spontaneous Ignition of Tar Saturated Mops used for Roof Repairs
Space Shuttle Program/Block II Space Shuttle Main Engine (SSME)/Alternate High Pressure Fuel T
High Pressure Shutoff Valves; Vacuum Pump Isolation Valve
Pneumatic and Fluid Quick Disconnects
SSME Oxygen Lines
KSC
PENNINGTON, DAVID
LINKURIOUS
VISUALIZE GRAPH DATA EASILY