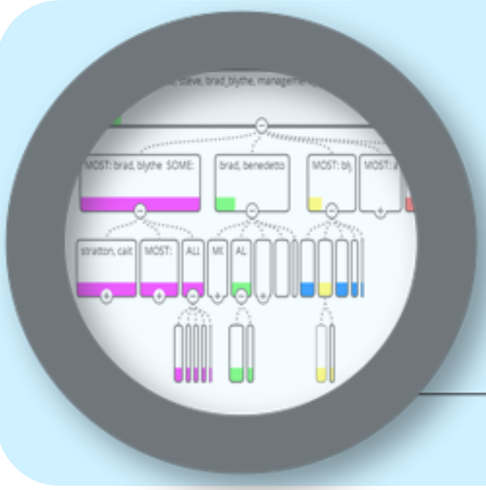
Matthew Brehmer

Stephen Ingram

Jonathan Stray

Tamara Munzner



Overview

the design, adoption, and analysis of a visual document mining tool for investigative journalists

IEEE InfoVis
Nov 14, 2014

Matthew Brehmer

Stephen Ingram

Jonathan Stray

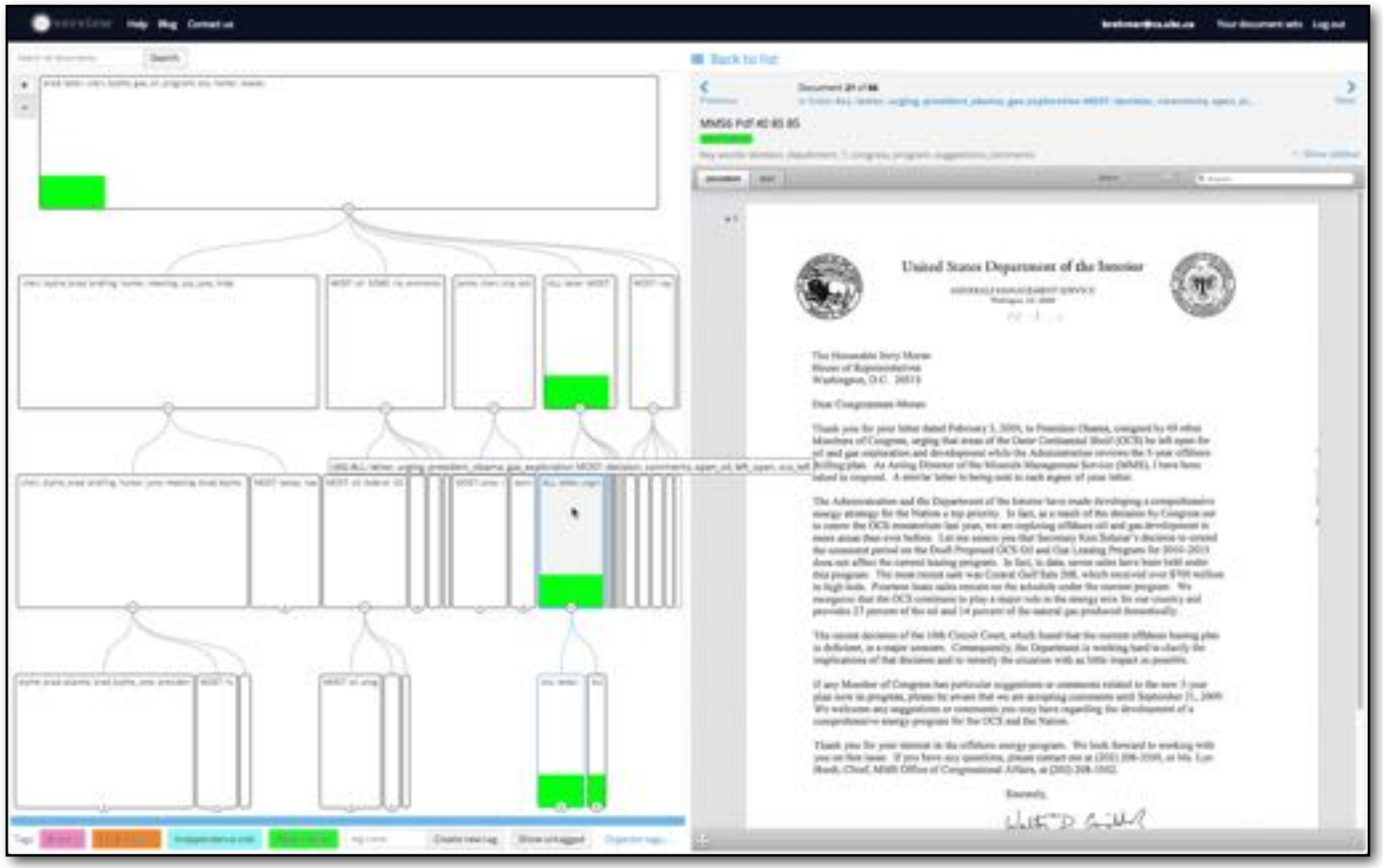
Tamara Munzner

Overview

the design, adoption, and analysis of a visual document mining tool for investigative journalists

IEEE InfoVis
Nov 14, 2014

Overview



overviewproject.org

Police misconduct hidden from public by secrecy law, weak oversight
By Sandra Peddie and Adam Playford
Newsday

What did private security contractors do in Iraq?
by Jonathan Stray on 02/21/2012 | 3 | Edit
AP

TPD working through flawed mobile system
By JARREL WADE World Staff Writer on Jun 3, 2012, at 2:19 AM
TULSA WORLD

RYAN ASKED FOR FEDERAL HELP AS HE CHAMPIONED CUTS
By JACK GILLUM — Oct. 12 7:20 PM EDT
AP

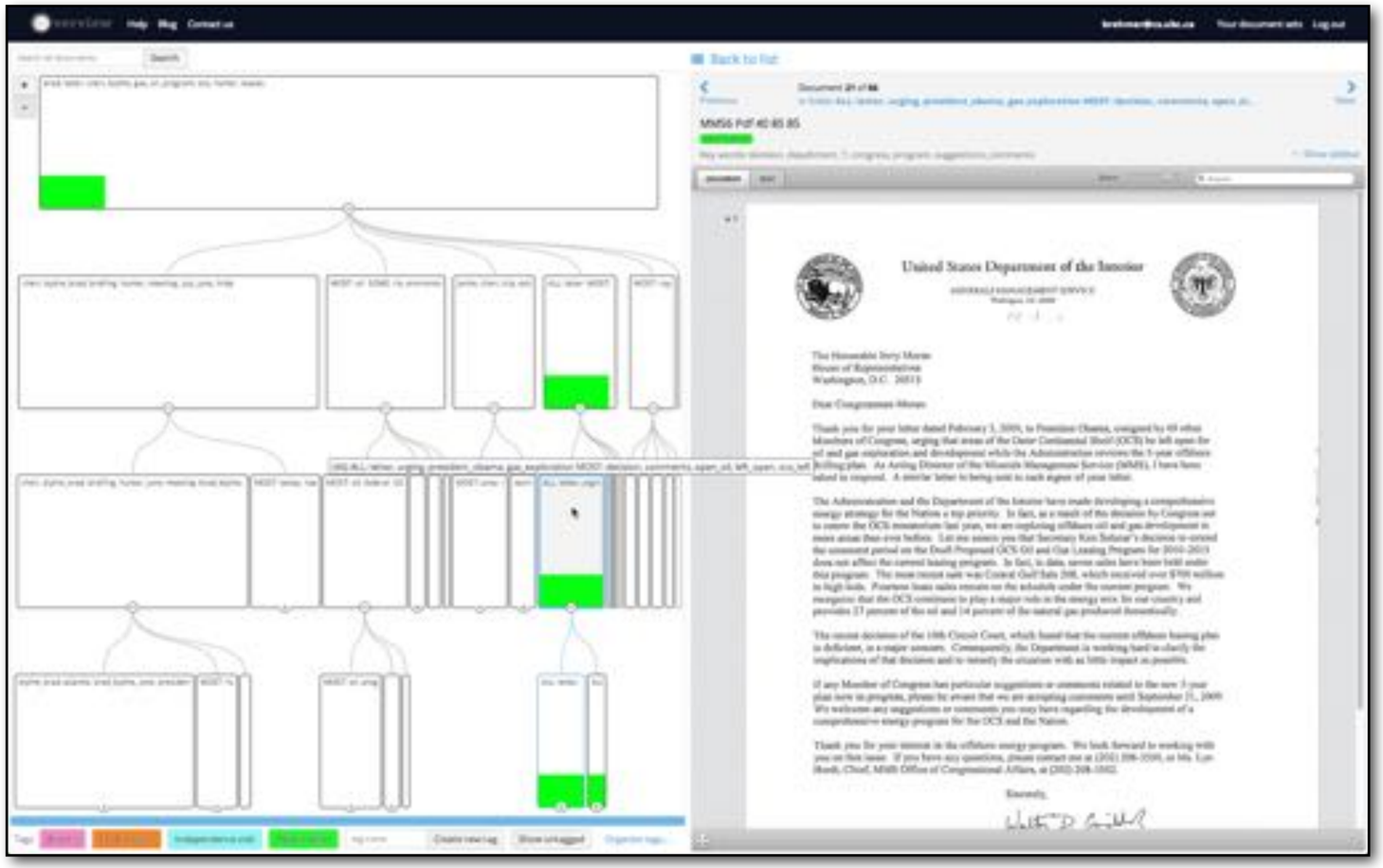
INTERACTIVE 12.22.12
Own a Gun? Tell Us Why
THE DAILY BEAST
Michael Keller

The Brilliance of Louis C.K.'s Emails: He Writes Like a Politician
Where campaign strategy and comedy marketing collide
ADRIENNE LAFRANCE | JUL 16 2014, 6:10 AM ET
the Atlantic

Surprise! Many credit card agreements allow repossession
Analysis: 'Security interest' clause present on 200 cards
By Fred O. Williams
CreditCards.com

Private memo reveals winding tale involving John McCain, the NRA and ... condors
by Nancy Watzman | SEPT. 18, 2014, 1:59 P.M.
SUNLIGHT FOUNDATION

Overview



overviewproject.org

Police misconduct hidden from public by secrecy law, weak oversight
By Sandra Peddie and Adam Playford
Newsday

What did private security contractors do in Iraq?
by Jonathan Stray on 02/21/2012 | 3 | Edit
AP

TPD working through flawed mobile system
By JARREL WADE World Staff Writer on Jun 3, 2012, at 2:19 AM
TULSA WORLD

RYAN ASKED FOR FEDERAL HELP AS HE CHAMPIONED CUTS
By JACK GILLUM — Oct. 12 7:20 PM EDT
AP

INTERACTIVE 12.22.12
Own a Gun? Tell Us Why
THE DAILY BEAST
Michael Keller

The Brilliance of Louis C.K.'s Emails: He Writes Like a Politician
Where campaign strategy and comedy marketing collide
ADRIENNE LAFRANCE | JUL 16 2014, 6:10 AM ET
the Atlantic

Surprise! Many credit card agreements allow repossession
Analysis: 'Security interest' clause present on 200 cards
By Fred O. Williams
CreditCards.com

Private memo reveals winding tale involving John McCain, the NRA and ... condors
by Nancy Watzman | SEPT. 18, 2014, 1:59 P.M.
SUNLIGHT FOUNDATION

image: commons.wikimedia.org

A VISUALIZATION DESIGN STUDY

image: commons.wikimedia.org

real journalists, real documents, real investigations...

A VISUALIZATION DESIGN STUDY

image: commons.wikimedia.org

real journalists, real documents, real investigations...

...**rationale** for visual encoding and interaction design choices...

A VISUALIZATION DESIGN STUDY

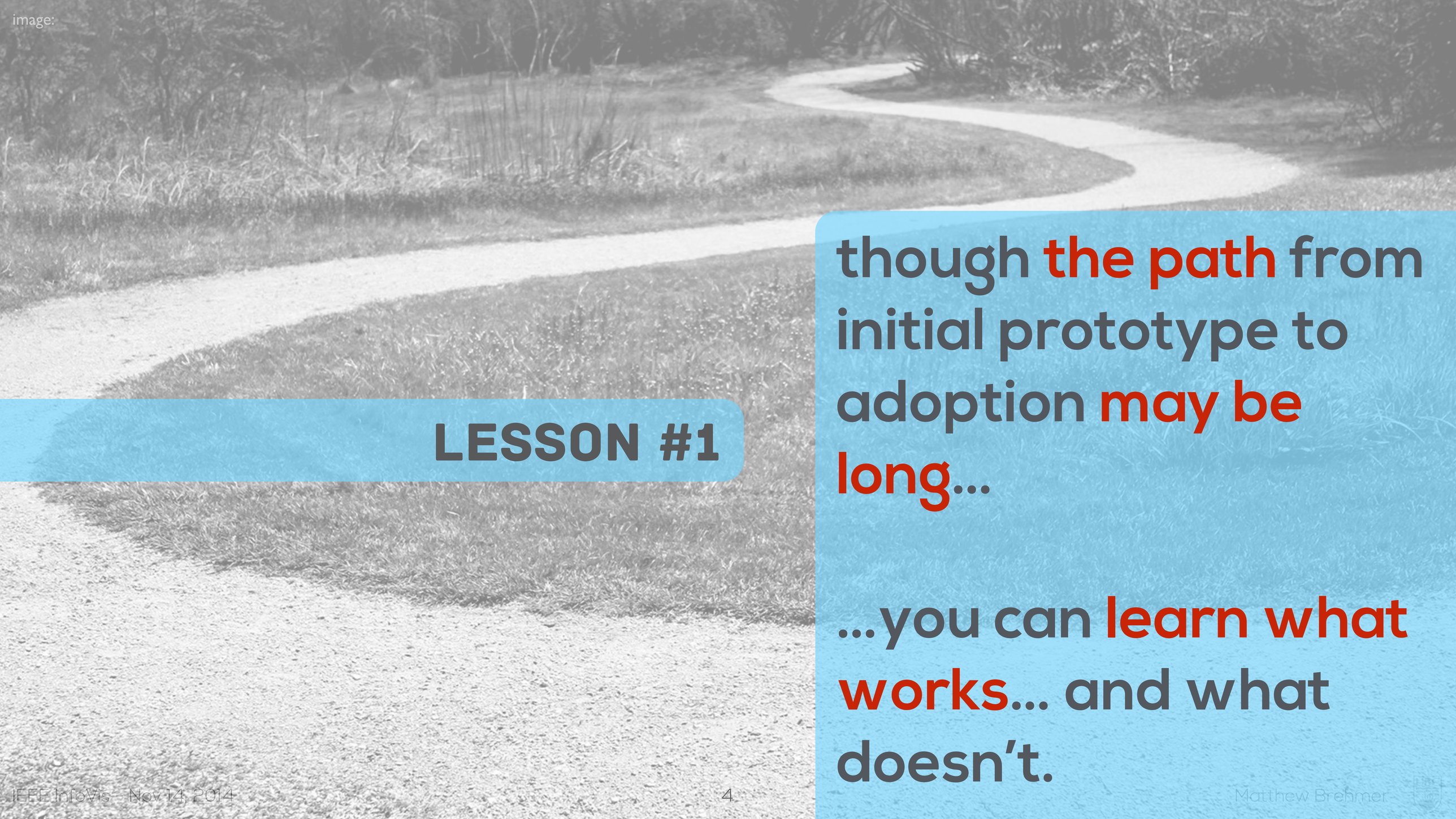
image: commons.wikimedia.org

real journalists, real documents, real investigations...

...**rationale** for visual encoding and interaction design choices...

...generalizable **lessons** for visualization design methodology

A VISUALIZATION DESIGN STUDY

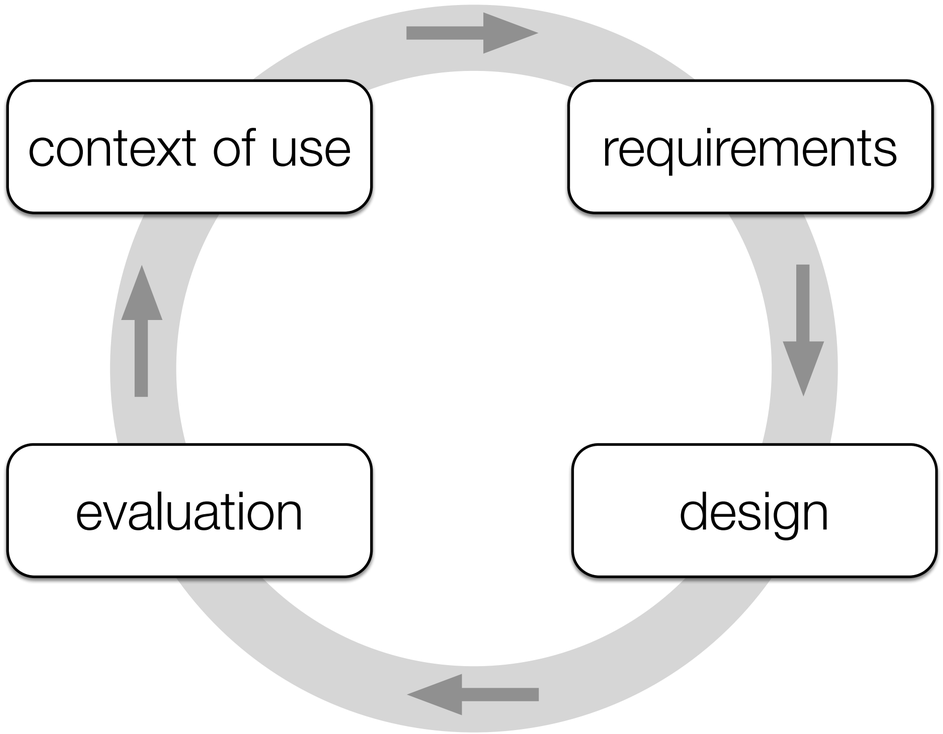


LESSON #1

though the path from initial prototype to adoption may be long...

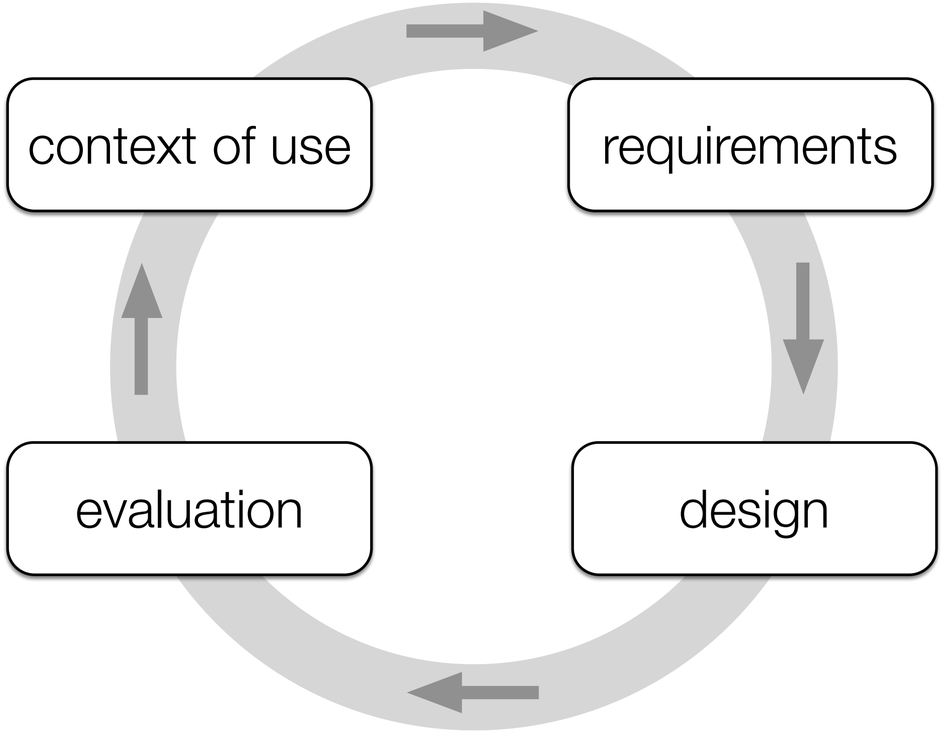
...you can learn what works... and what doesn't.

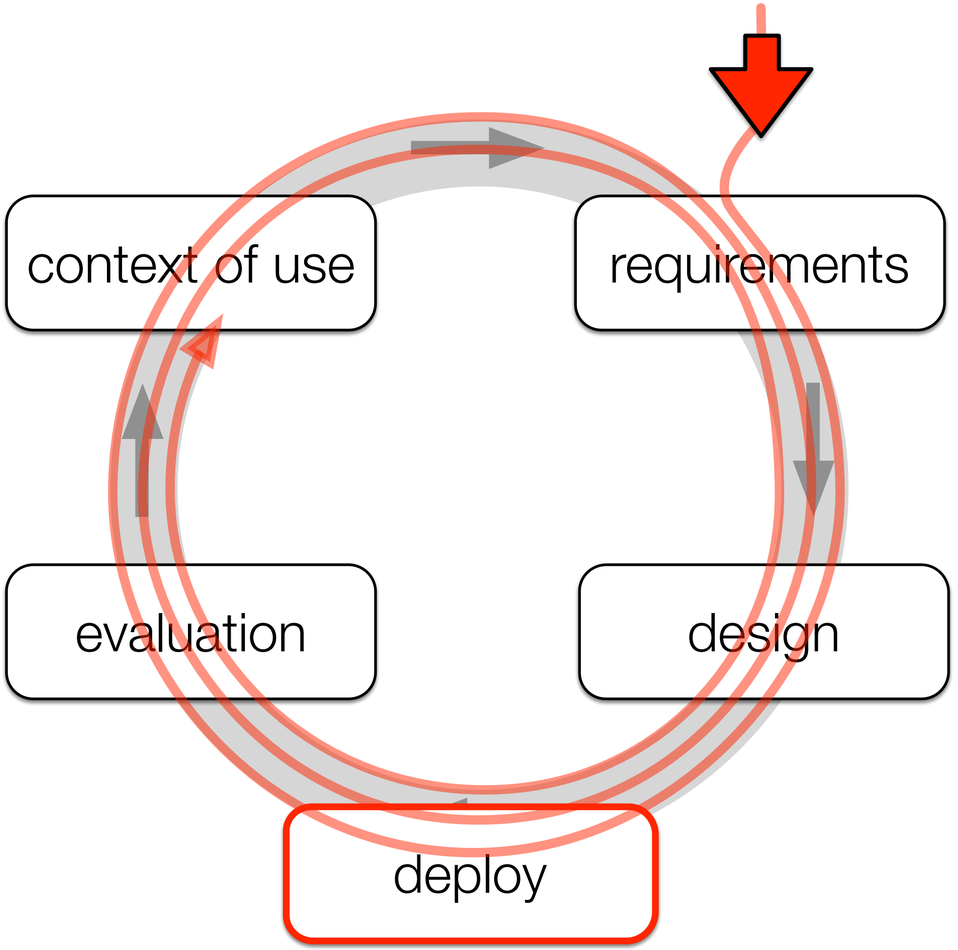
HUMAN-CENTRED VIS DESIGN



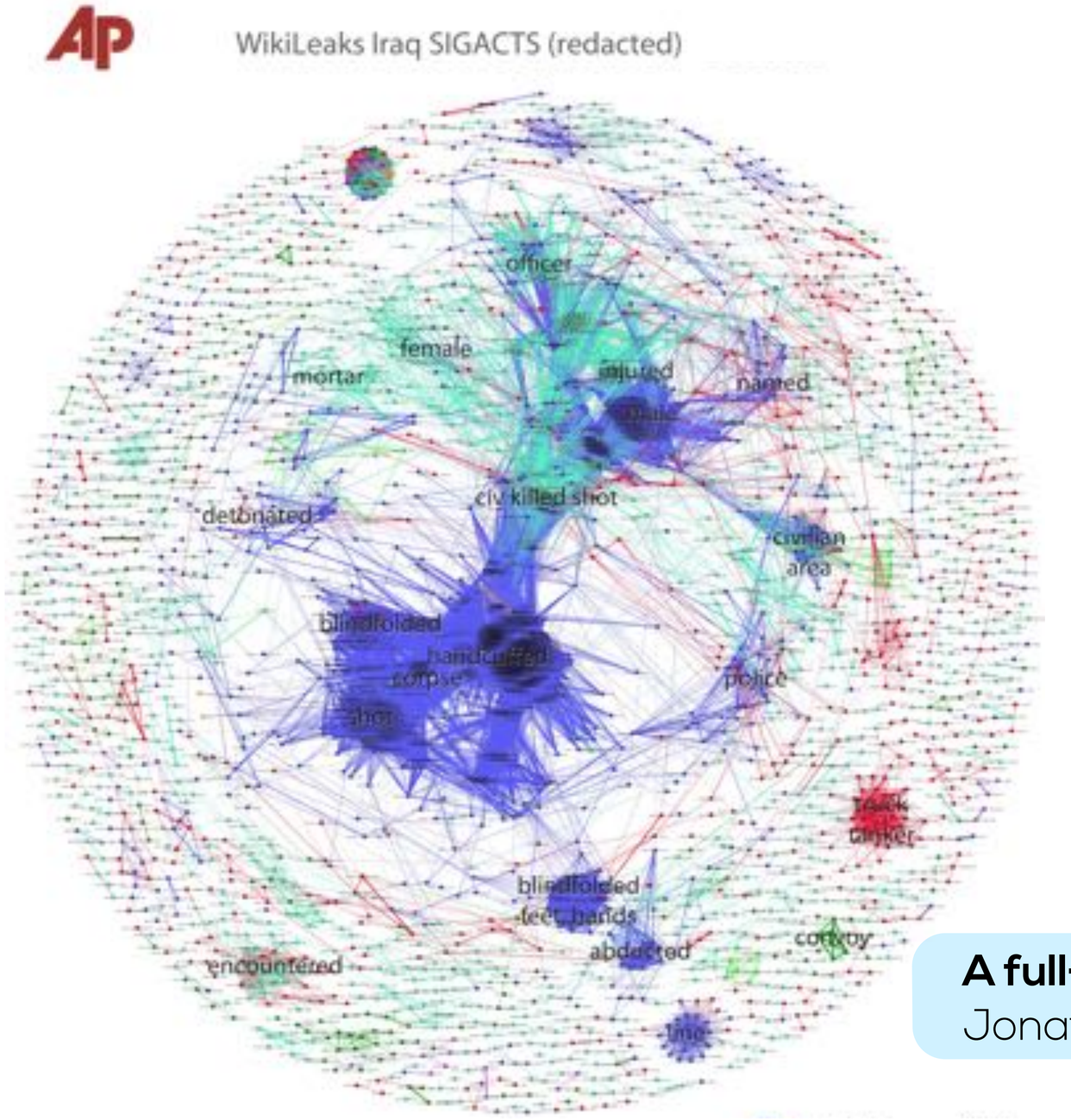
Lloyd & Dykes (IEEE TVCG / InfoVis 2011)

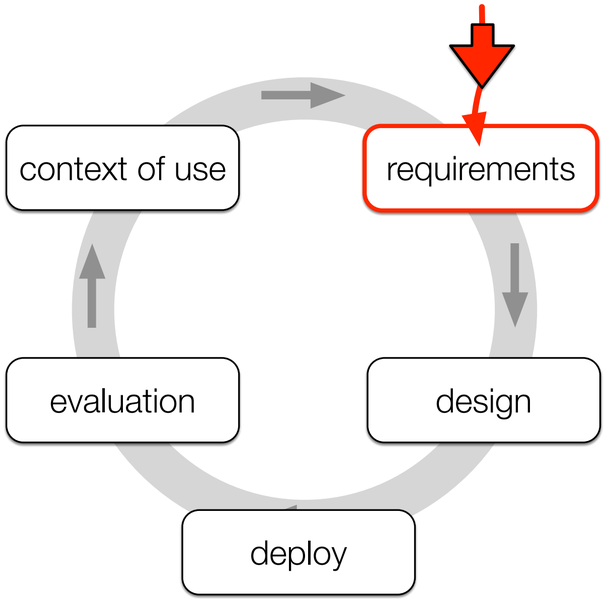
HUMAN-CENTRED VIS DESIGN



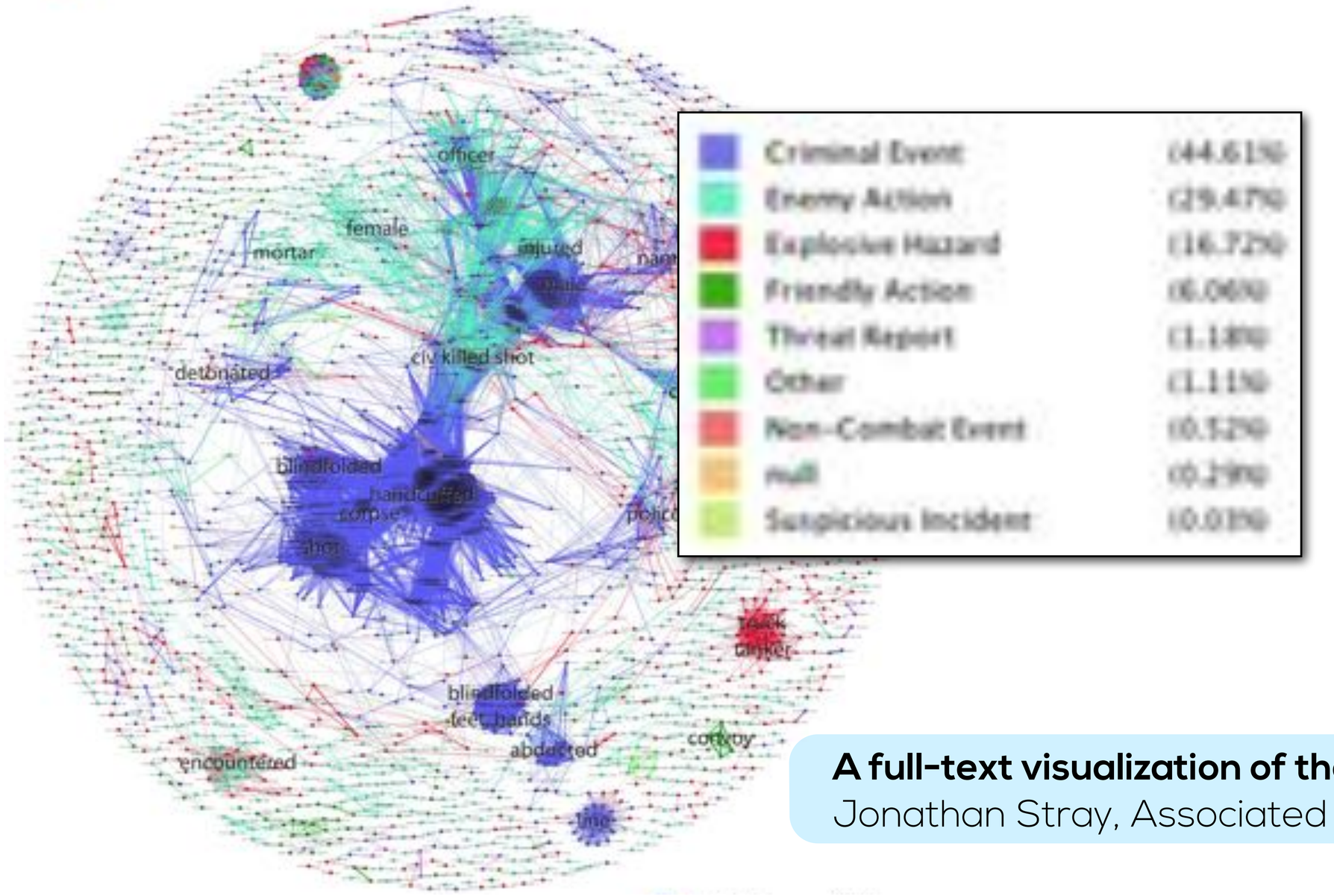


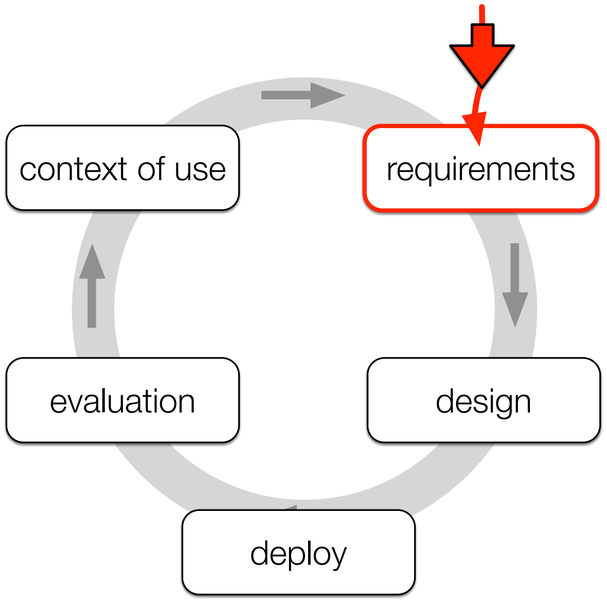
Lloyd & Dykes (IEEE TVCG / InfoVis 2011)



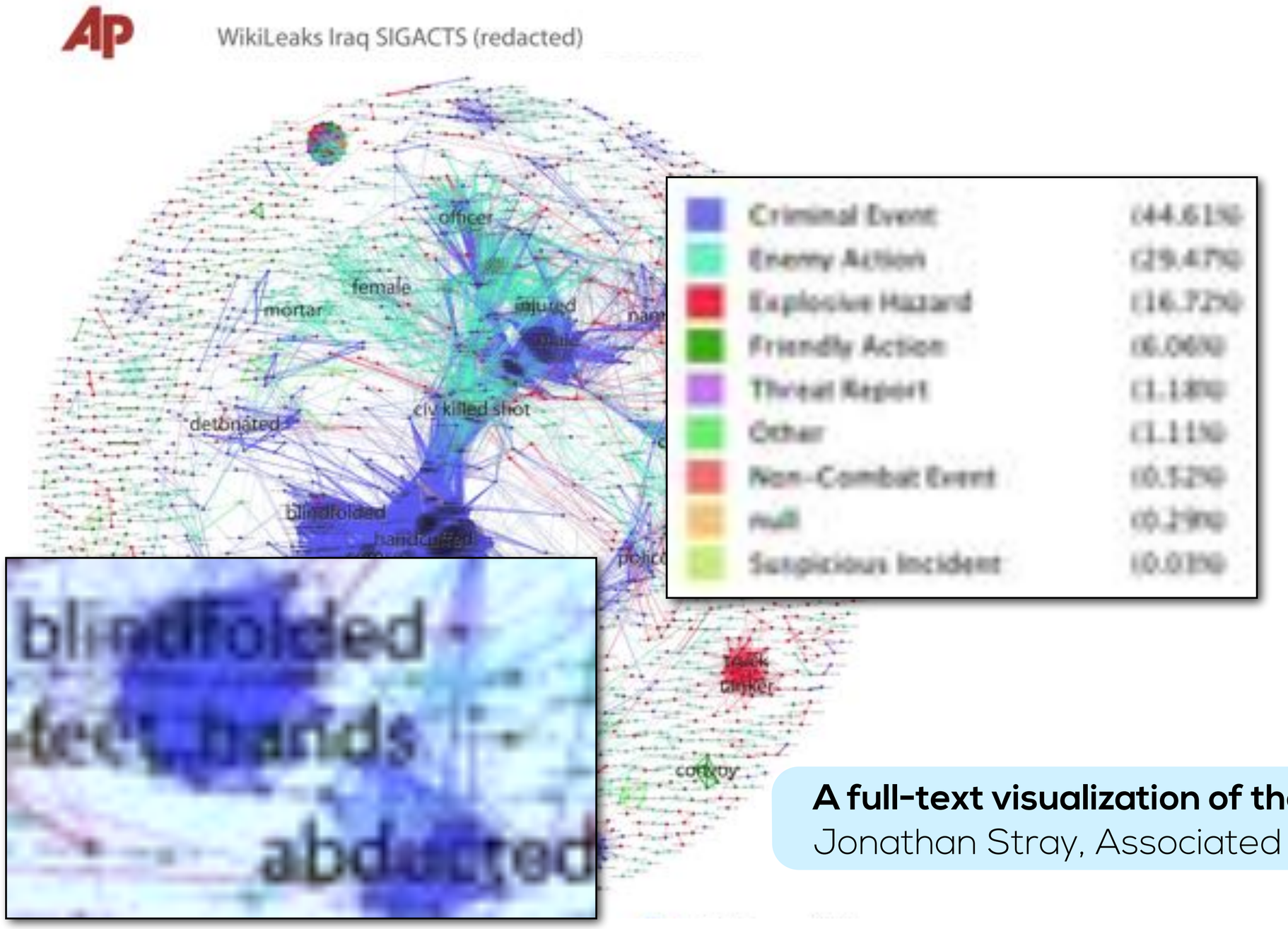


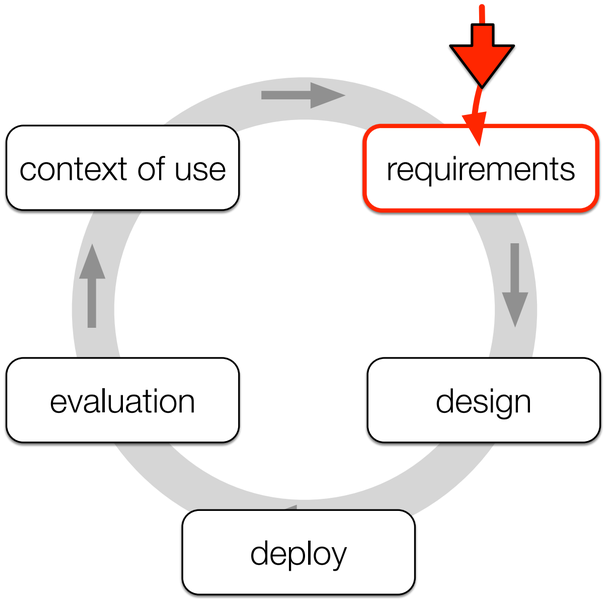
A full-text visualization of the Iraq War Logs.
Jonathan Stray, Associated Press (Dec. 2010)



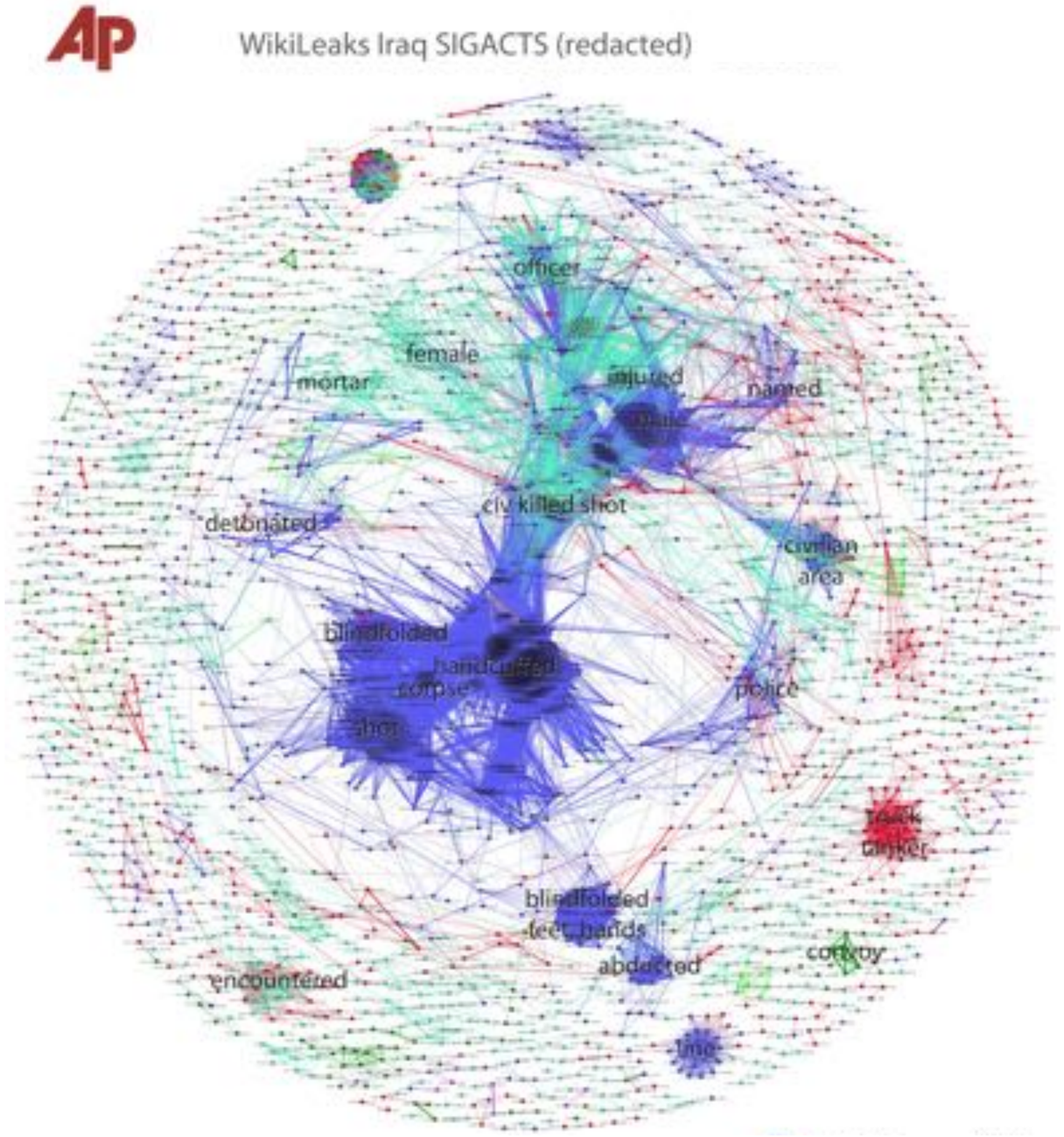


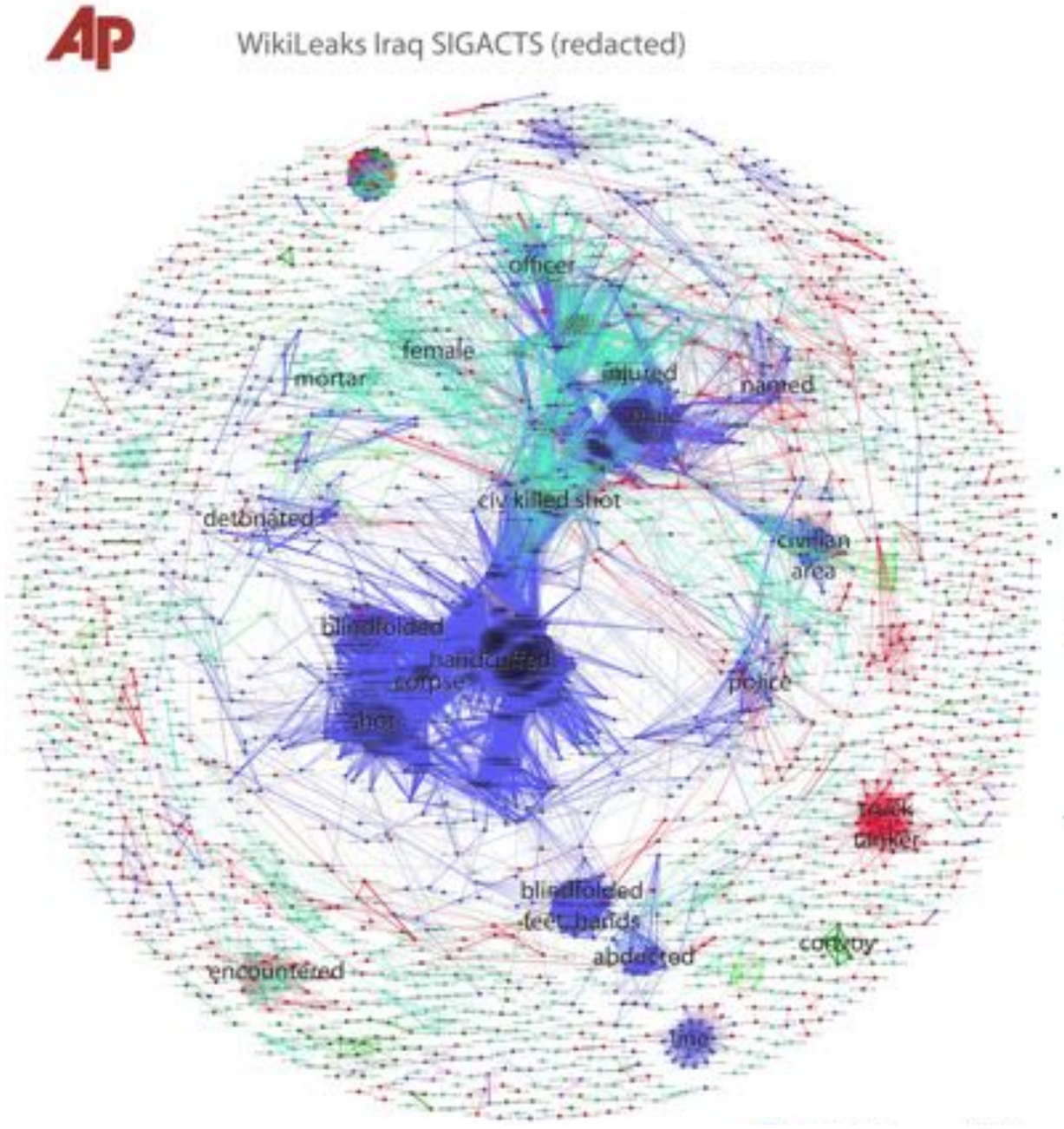
A full-text visualization of the Iraq War Logs.
Jonathan Stray, Associated Press (Dec. 2010)





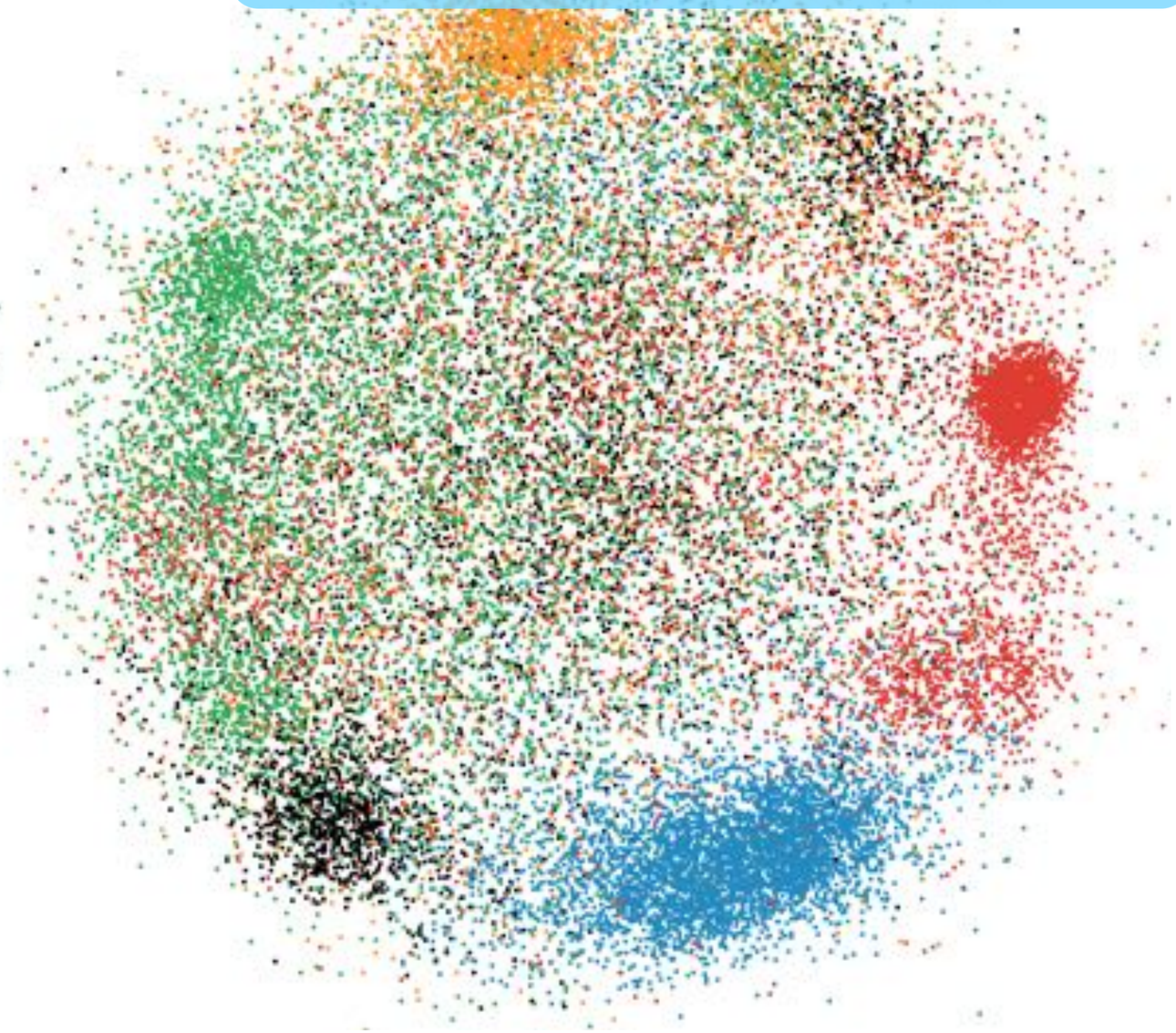
A full-text visualization of the Iraq War Logs.
Jonathan Stray, Associated Press (Dec. 2010)


AP
WikiLeaks Iraq SIGACTS (redacted)
officer
female
mortar
injured
named
civ killed shot
detonated
civilian
area
blindfolded
handcuffed
police
shot
blindfolded
feet, hands
convoy
abducted
encountered

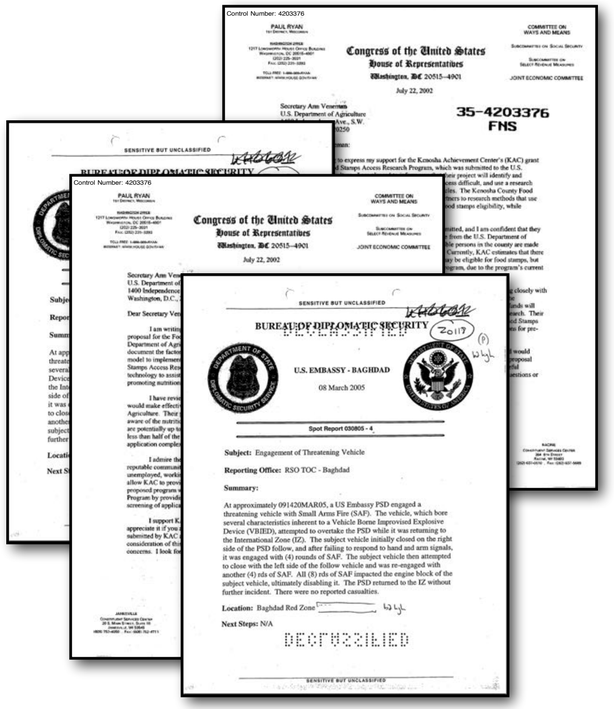


Glimmer MDS: fast & scalable DR

Ingram et al. InfoVis 2009

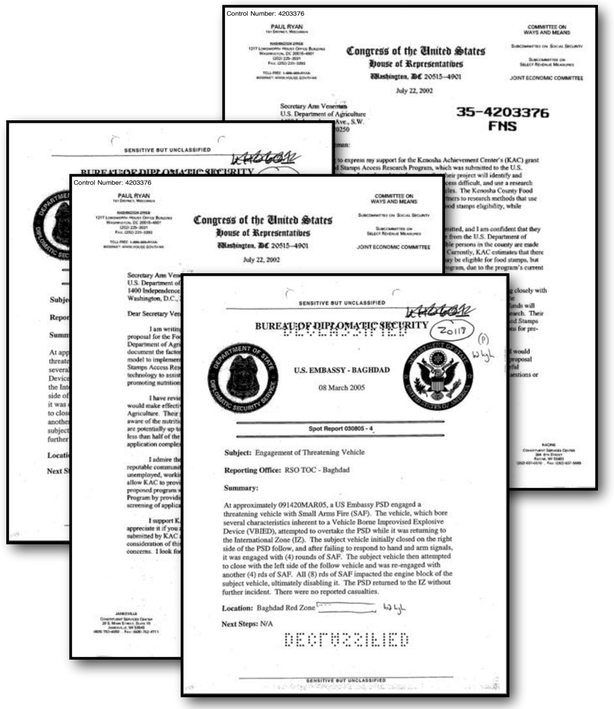


DATA ABSTRACTIONS



TF-IDF vectors

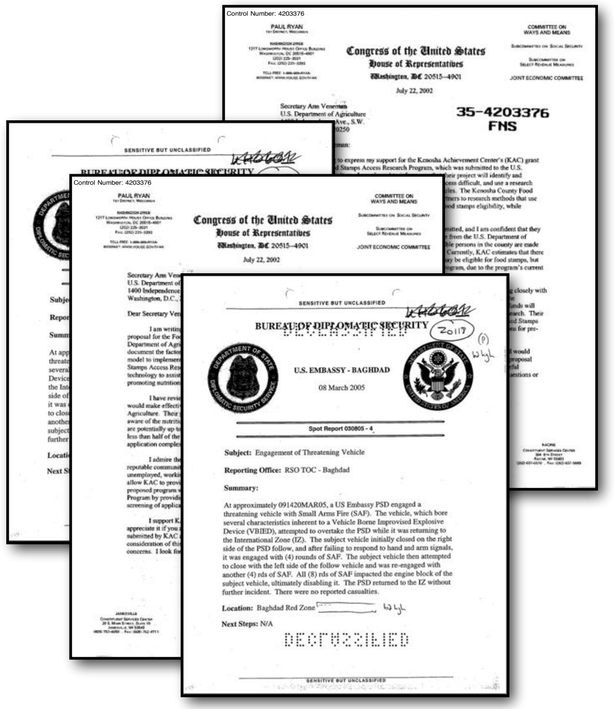
DATA ABSTRACTIONS



TF-IDF vectors

no document metadata!

DATA ABSTRACTIONS



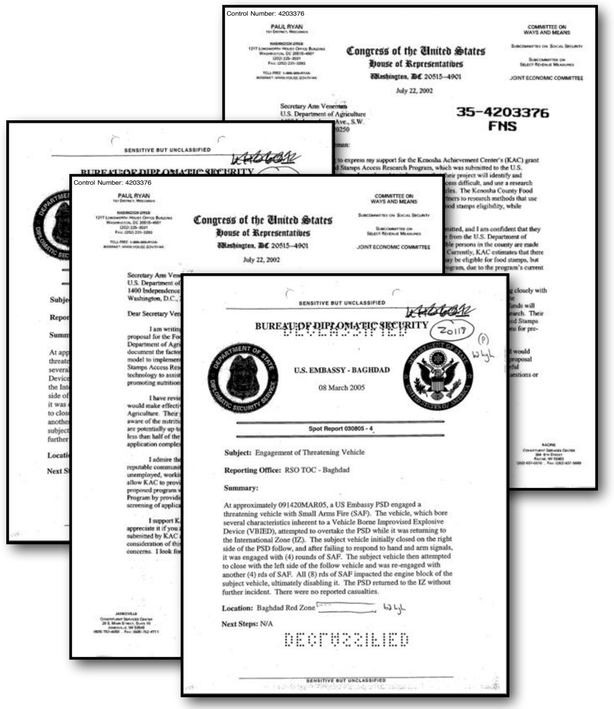
TF-IDF vectors

hierarchical clustering

dimensionality reduction

no document metadata!

DATA ABSTRACTIONS



TF-IDF vectors

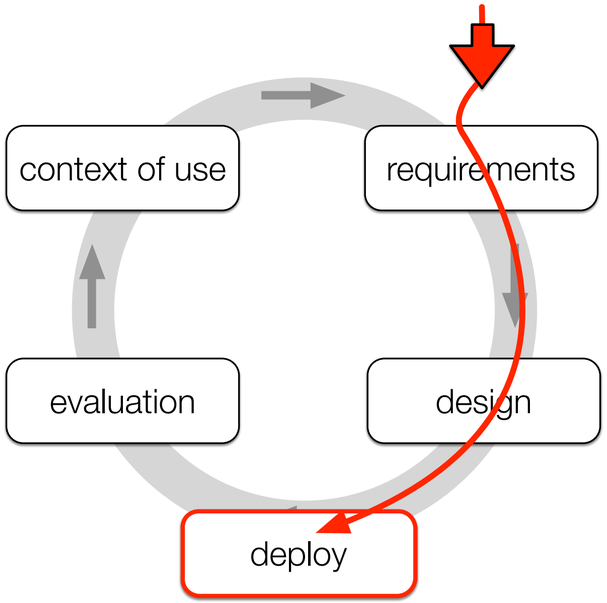
hierarchical clustering

dimensionality reduction

+ user-defined tags

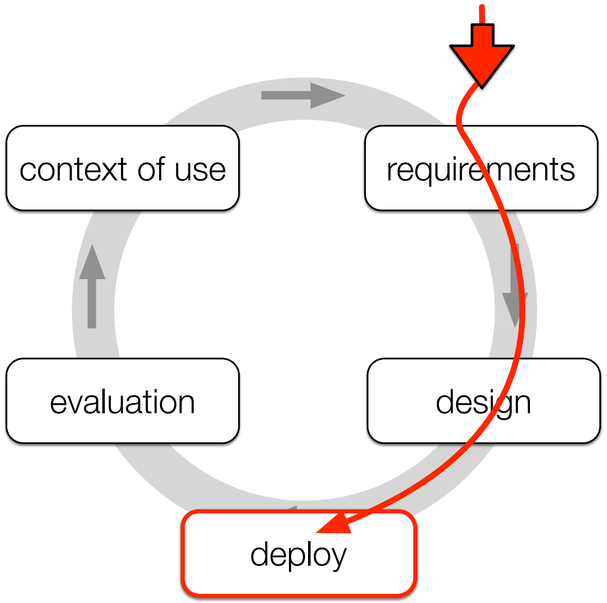
no document metadata!

OVERVIEW V.1 PROTOTYPE, 2011



clockwise from TL:
tree visualization, tag controls, scatterplot, document list, document viewer, cluster list

OVERVIEW V.1 PROTOTYPE, 2011

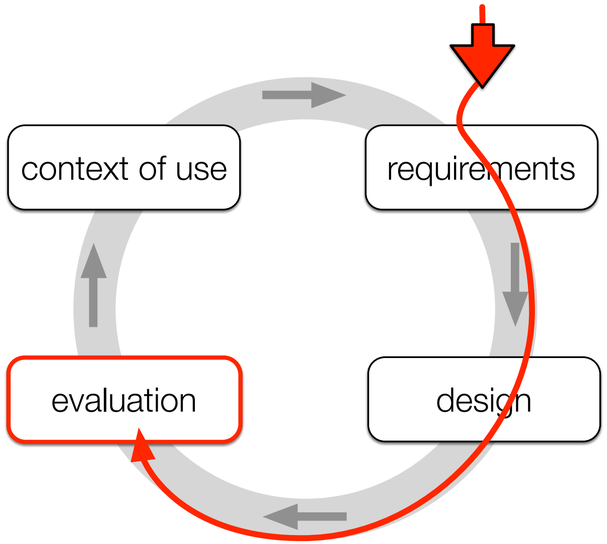


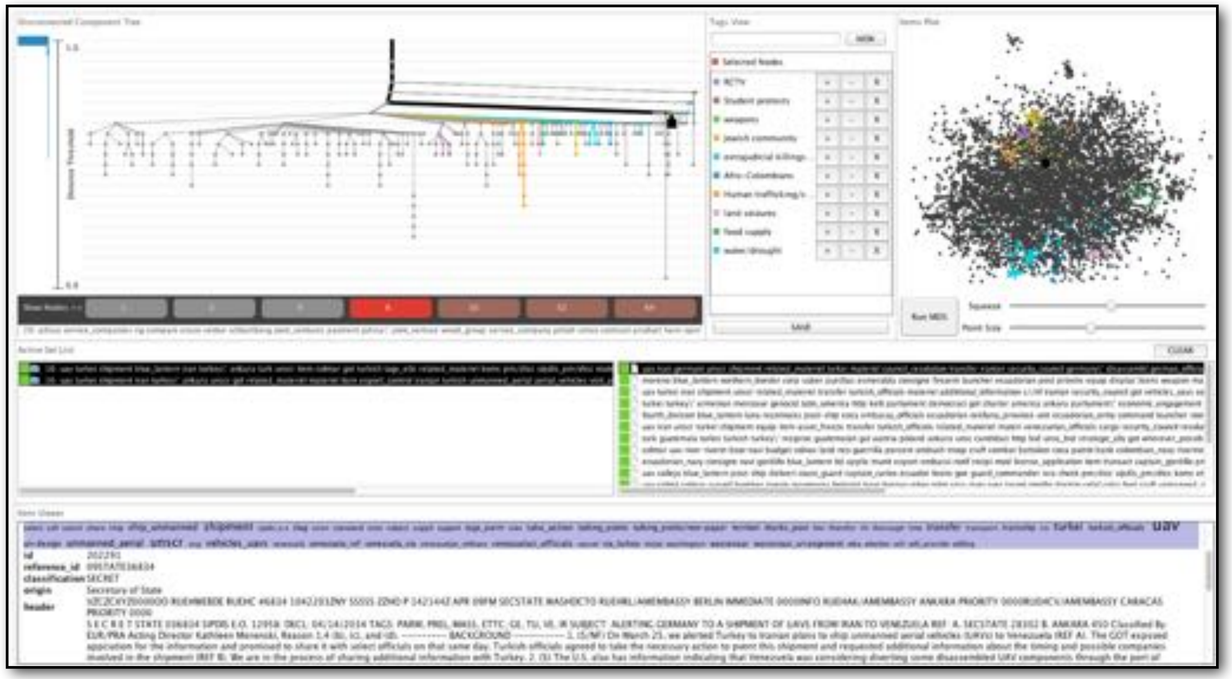
clockwise from TL:
tree visualization, tag controls, scatterplot, document list, document viewer, cluster list

“EXPLORING” THE WIKILEAKS CABLES

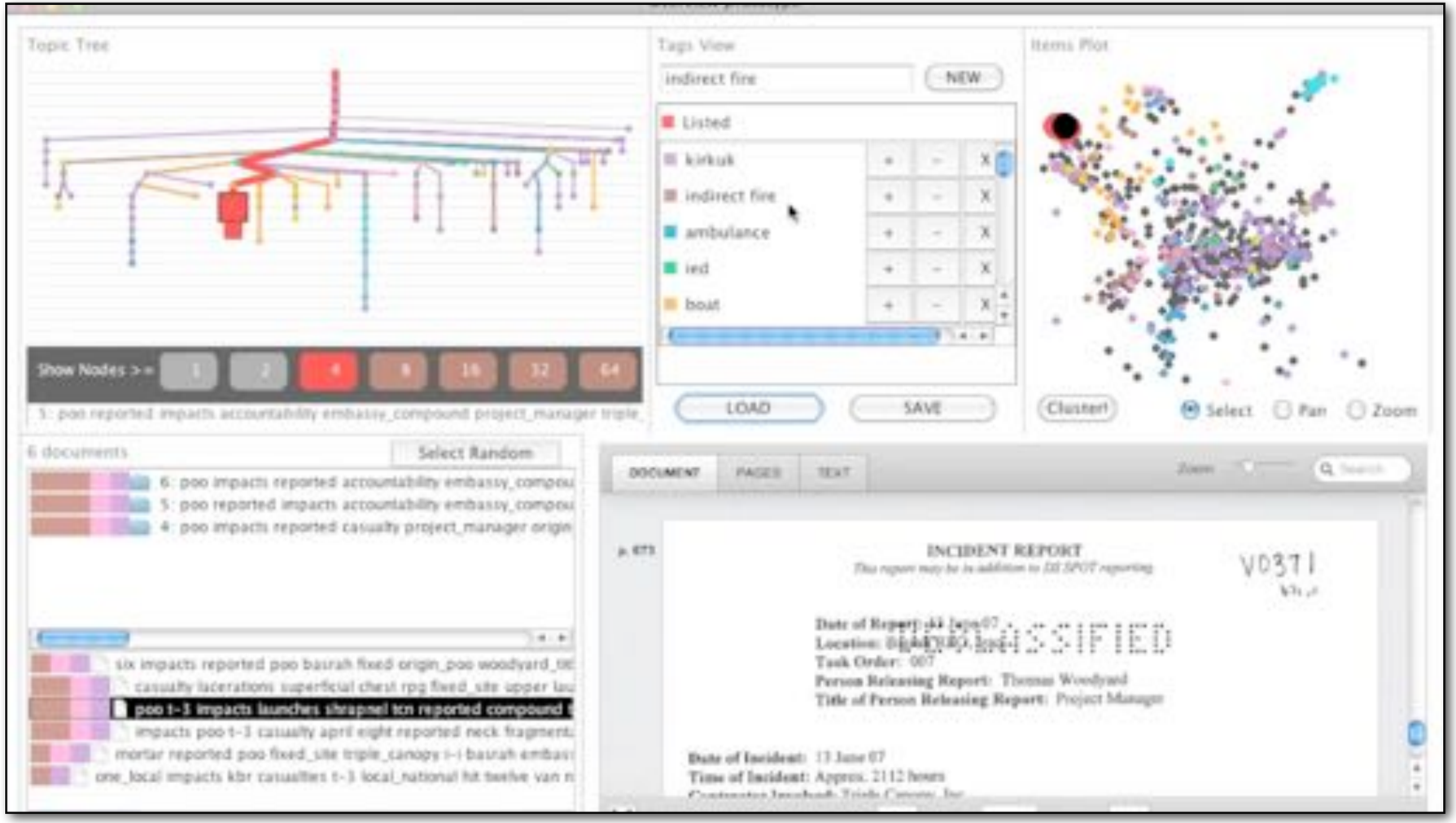
http://bigstory.ap.org/article/venezuela-chavez-era-reporters-view

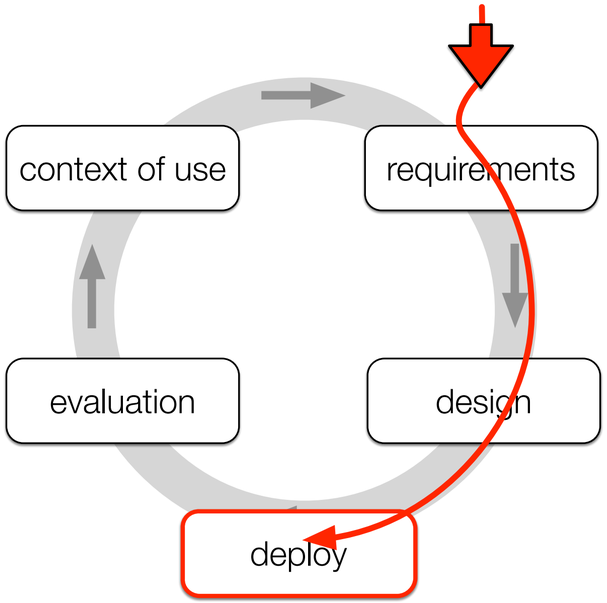
AP Caracas bureau chief
Ian James
(with Hugo Chavez, 2007)



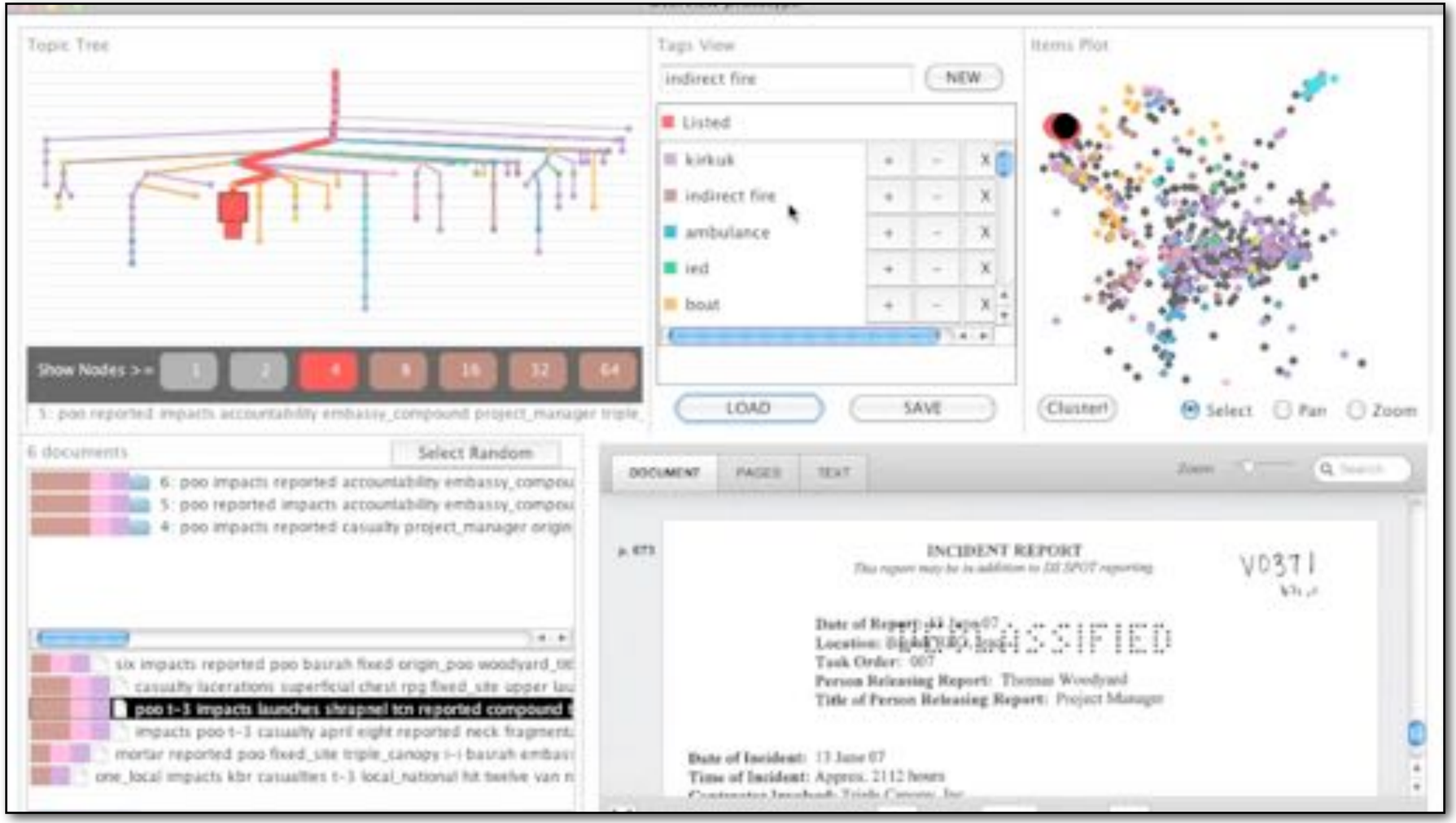


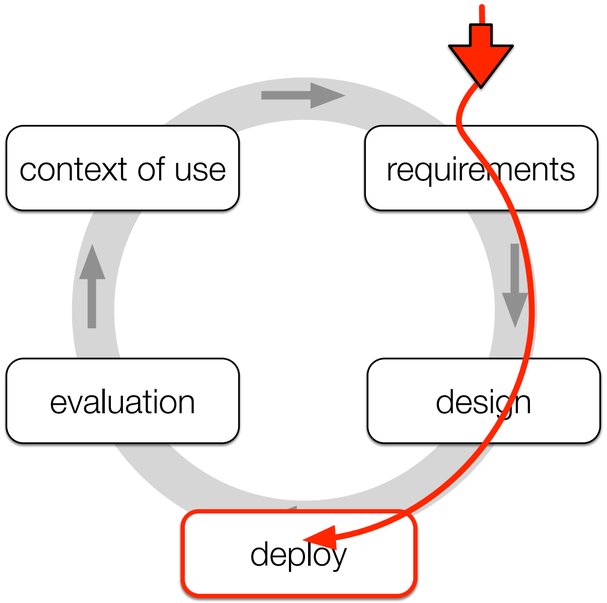
OVERVIEW V.2 (DESKTOP APP), 2012



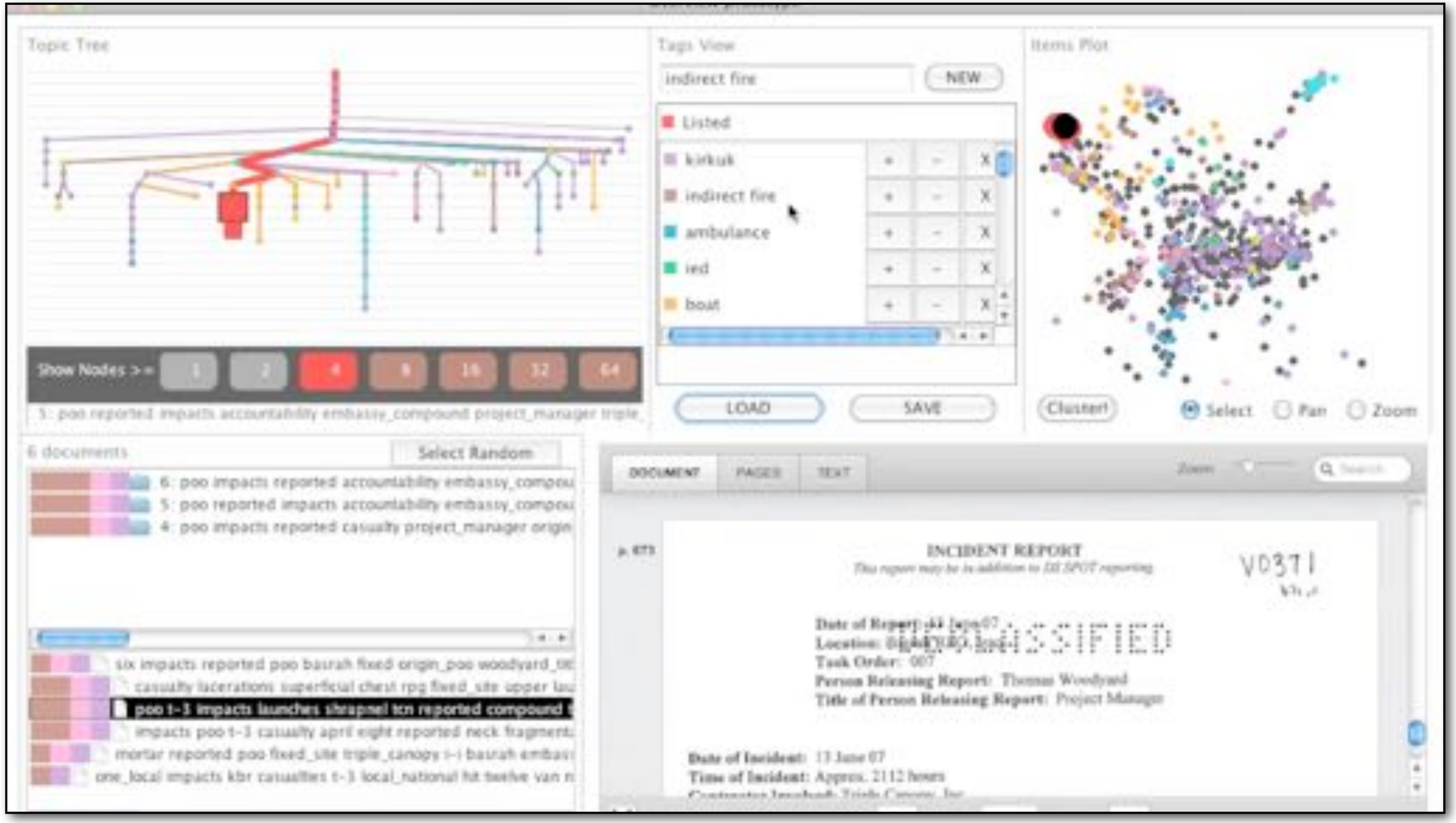


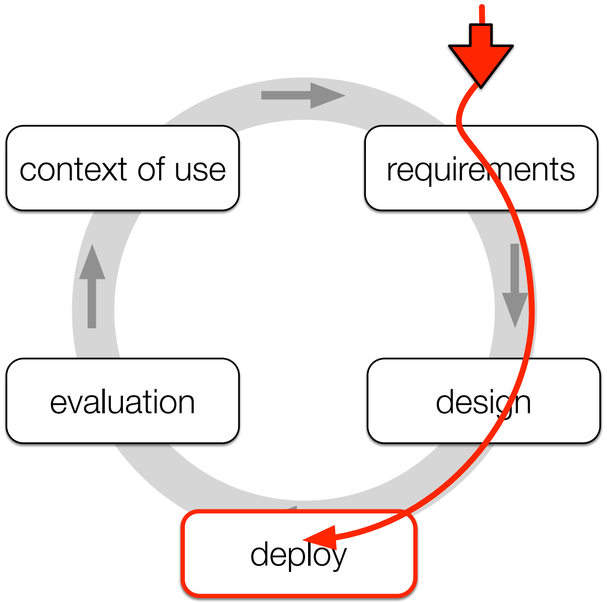
OVERVIEW V.2 (DESKTOP APP), 2012





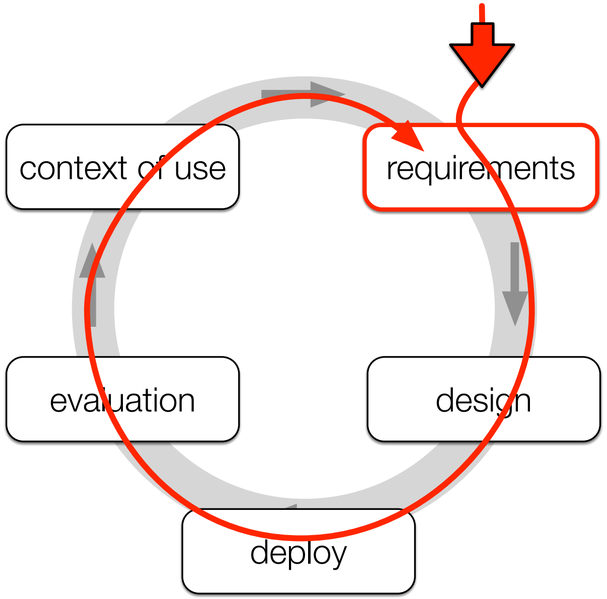
OVERVIEW V.2 (DESKTOP APP), 2012



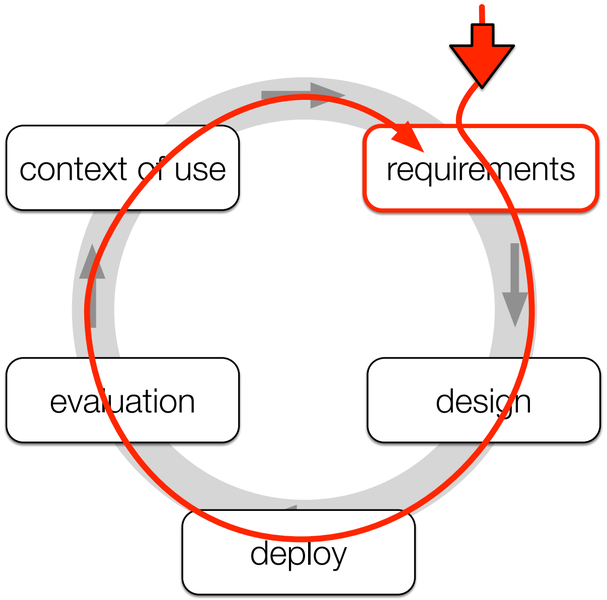


Q: could the use of Overview lead to a published story?

OVERVIEW V.2 ADOPTION, 2012



OVERVIEW V.2 ADOPTION, 2012



What did private security contractors do in Iraq?

by Jonathan Stray on 02/21/2012 | 3 | Edit

AP

TPD working through flawed mobile system

By JARREL WADE World Staff Writer on Jun 3, 2012, at 2:19 AM

TULSA WORLD

RYAN ASKED FOR FEDERAL HELP AS HE CHAMPIONED CUTS

By JACK GILLUM — Oct. 12 7:20 PM EDT

AP

OVERVIEW V.2 CASE STUDIES, 2012

Document Collection

Question

Outcome

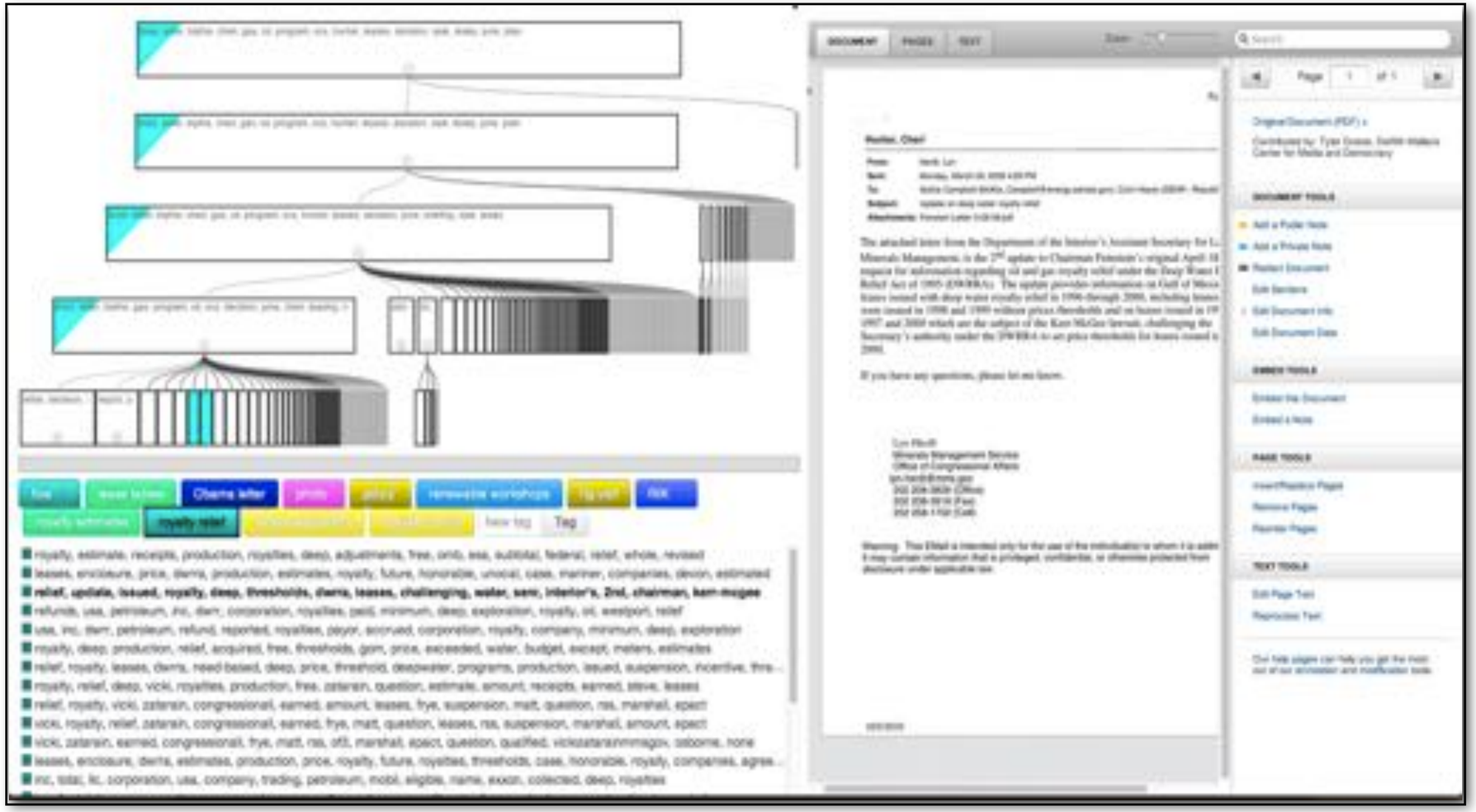
OVERVIEW V.2 CASE STUDIES, 2012

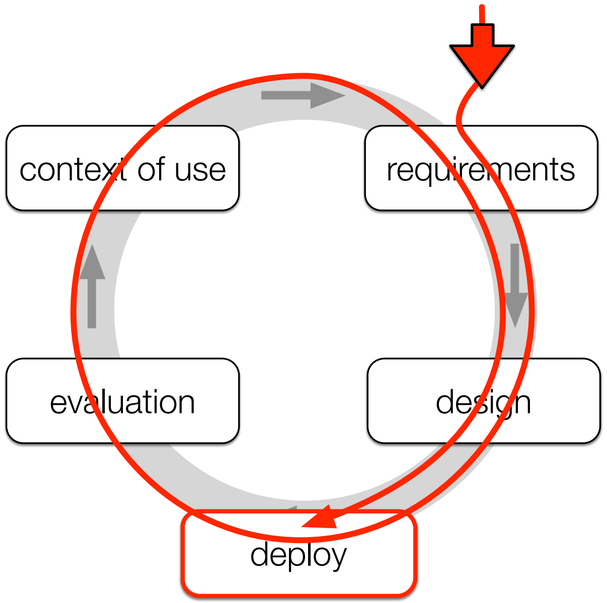
	CASE STUDY #1
Document Collection	666 reports (4,500 pages) from FOIA
Question	*What incidents involved security contractors during Iraq war?*
Outcome	Summarized prevalence of document categories

OVERVIEW V.2 CASE STUDIES, 2012

	CASE STUDY #1	CASE STUDY #2
Document Collection	666 reports (4,500 pages) from FOIA	5,996 emails from FOIA
Question	*What incidents involved security contractors during Iraq war?*	*Were municipal police funds mismanaged?*
Outcome	Summarized prevalence of document categories	Located evidence supporting hypothesis

OVERVIEW V.3 (WEB APP), FALL 2012





clockwise from TL:
tree visualization,
document viewer,
document list,
tag controls

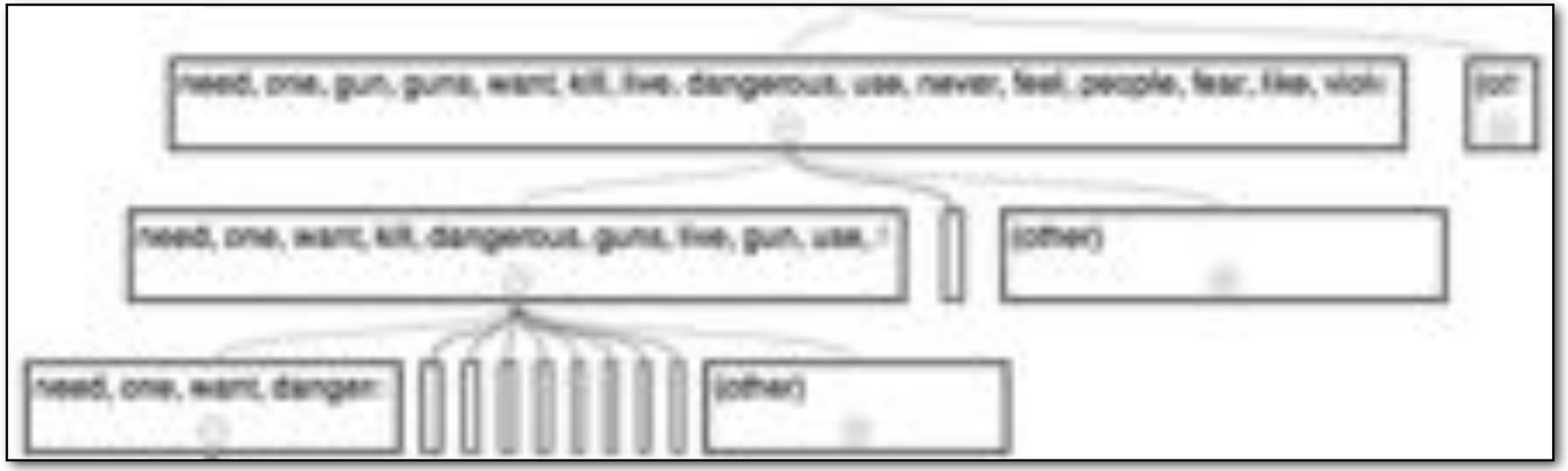
OVERVIEW V.3 EVALUATION, 2013

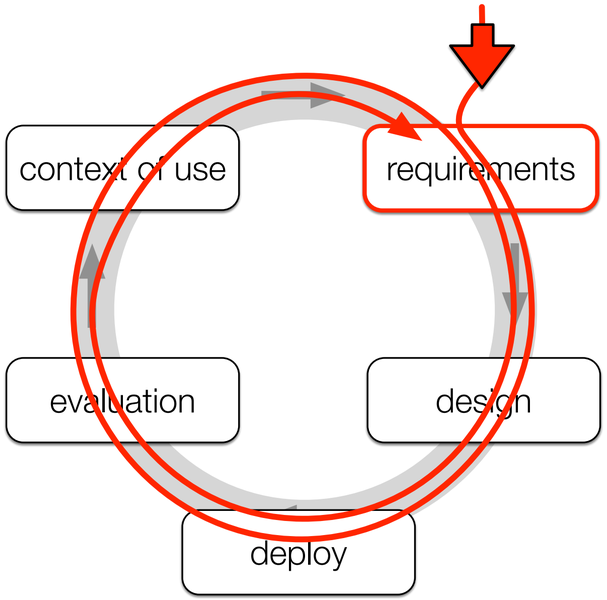

INTERACTIVE 12.22.12

Own a Gun? Tell Us Why

THE DAILY BEAST

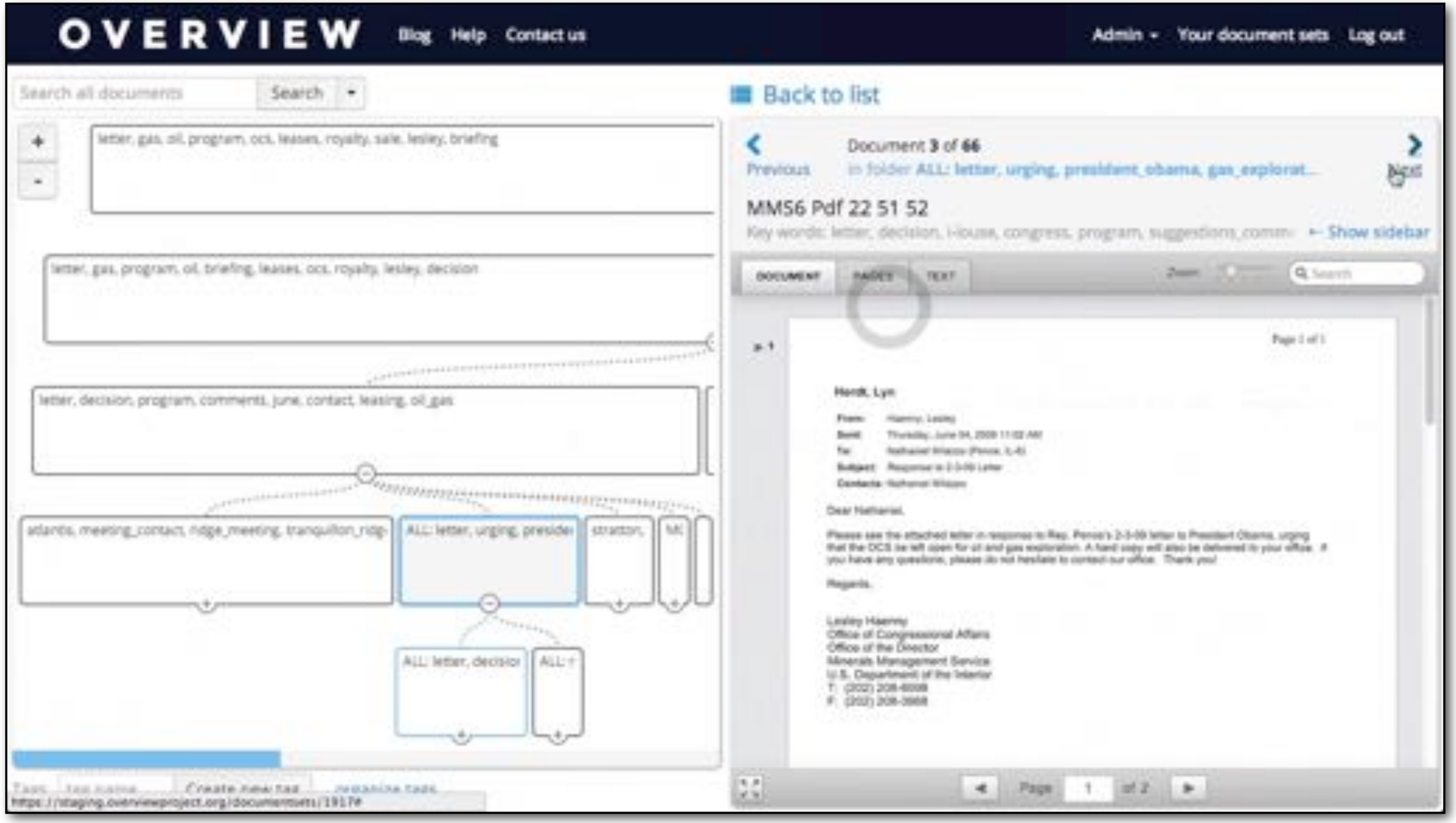
Michael Keller

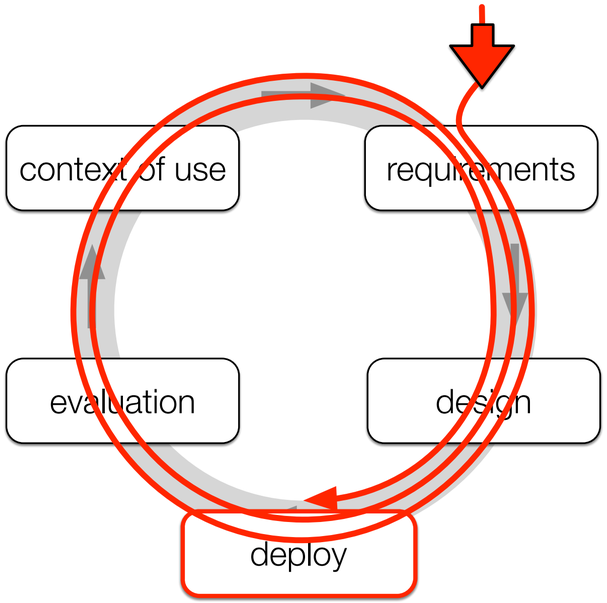




+ think-aloud evaluation with 5 journalists

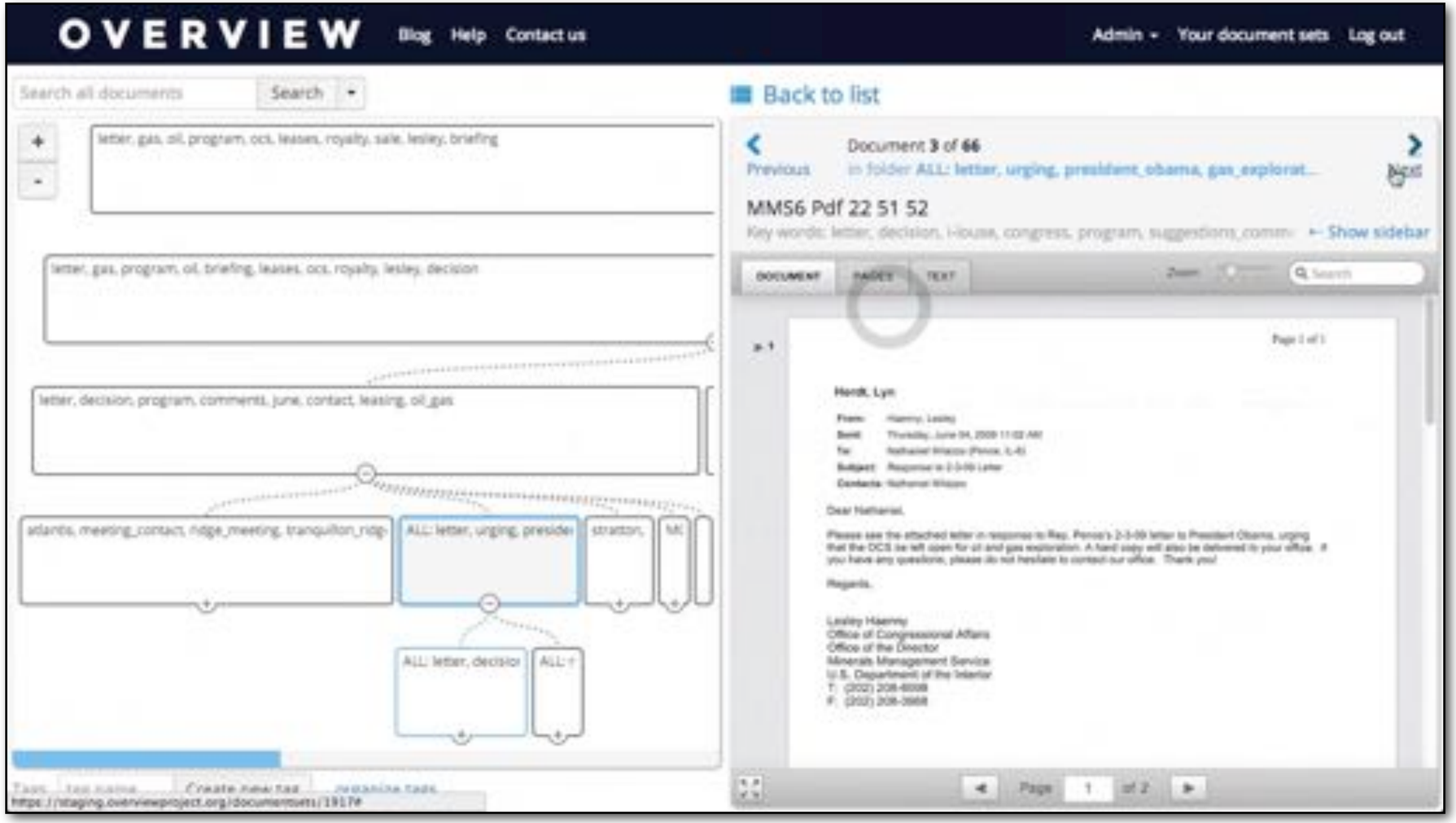
OVERVIEW V.4, FALL 2013

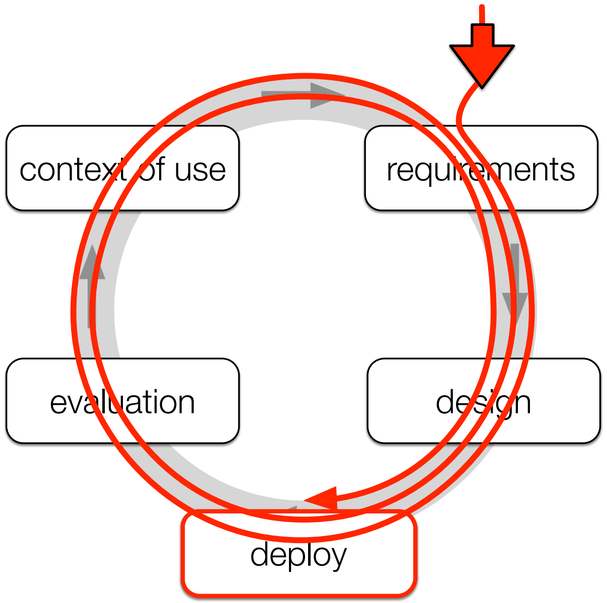




clockwise from TL:
full-text search,
document list / viewer,
tag controls,
tree visualization

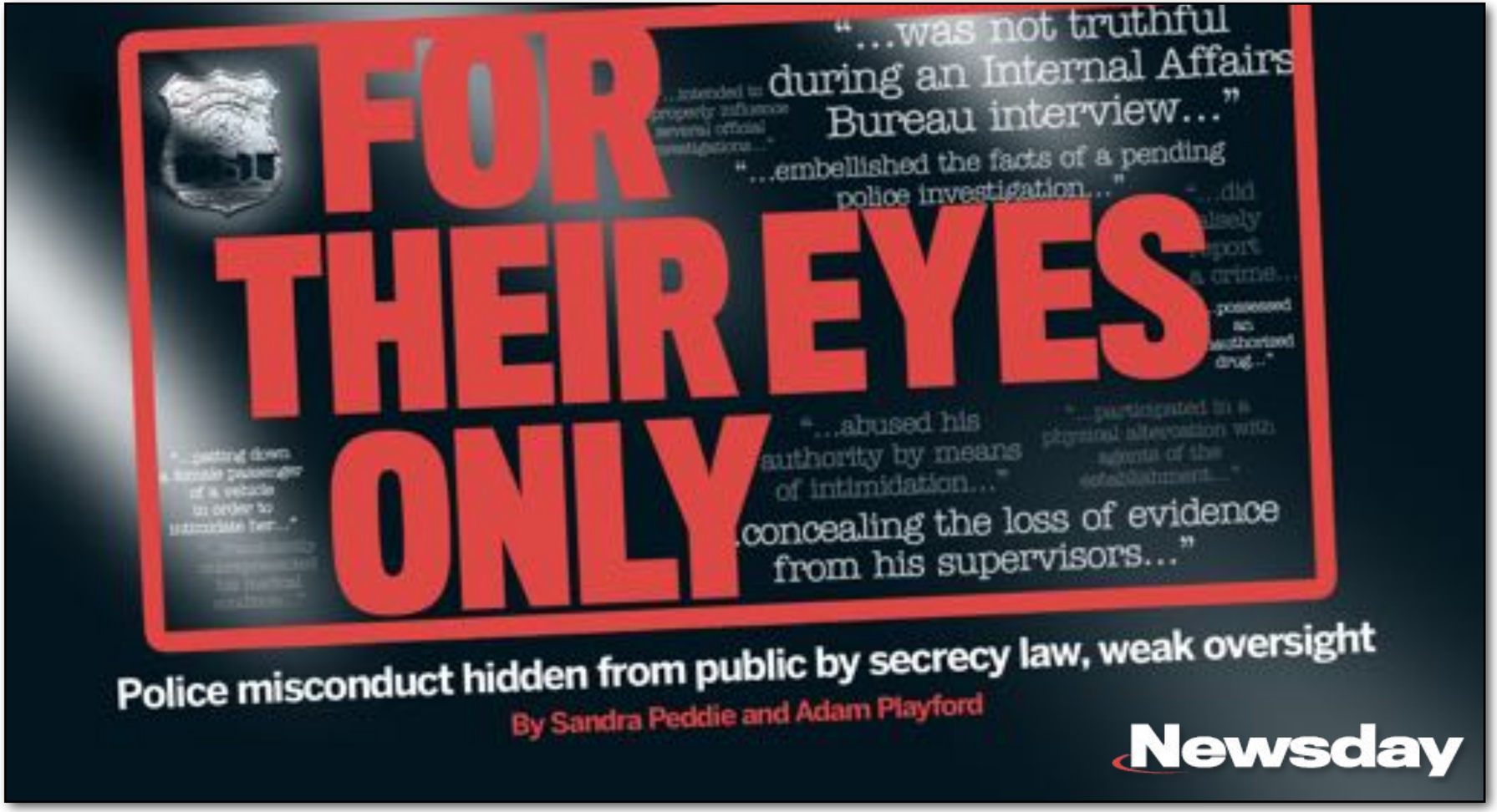
OVERVIEW V.4, FALL 2013

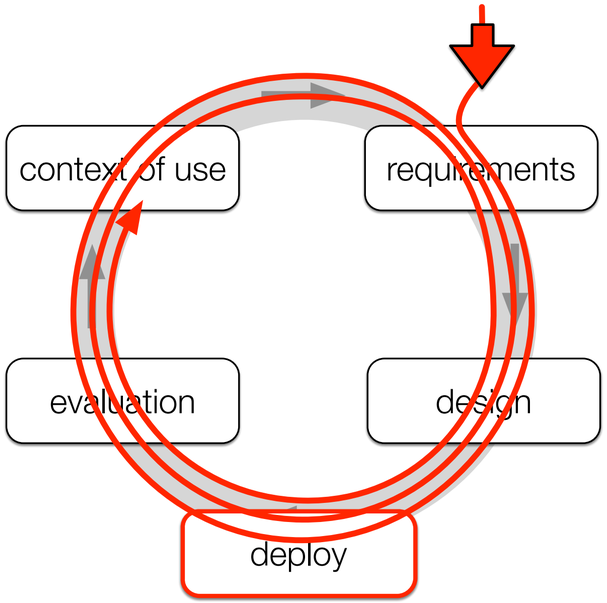




clockwise from TL:
full-text search,
document list / viewer,
tag controls,
tree visualization

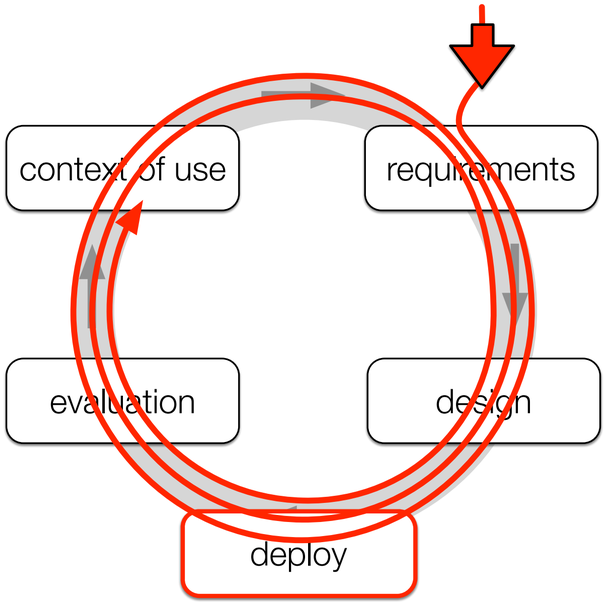
OVERVIEW V.4 ADOPTION





finalist for 2014 Pulitzer Prize in journalism

OVERVIEW V.4 ADOPTION

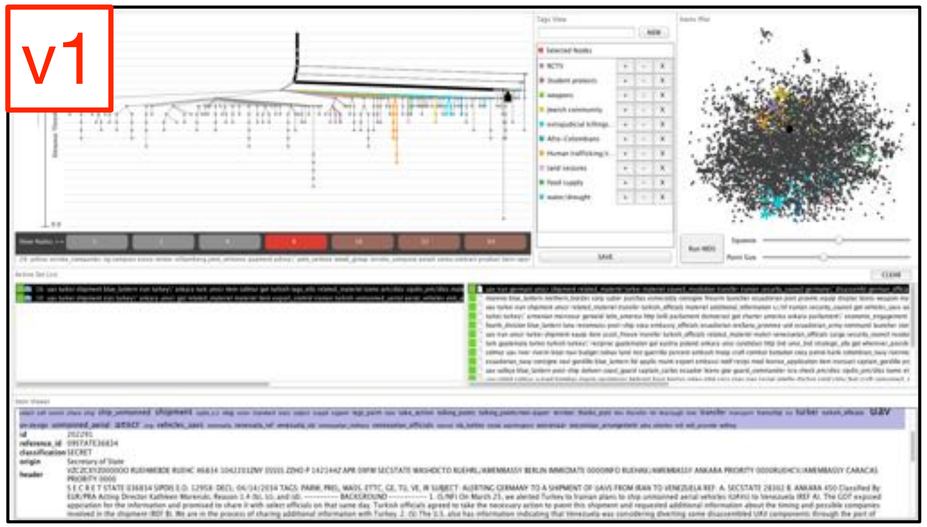


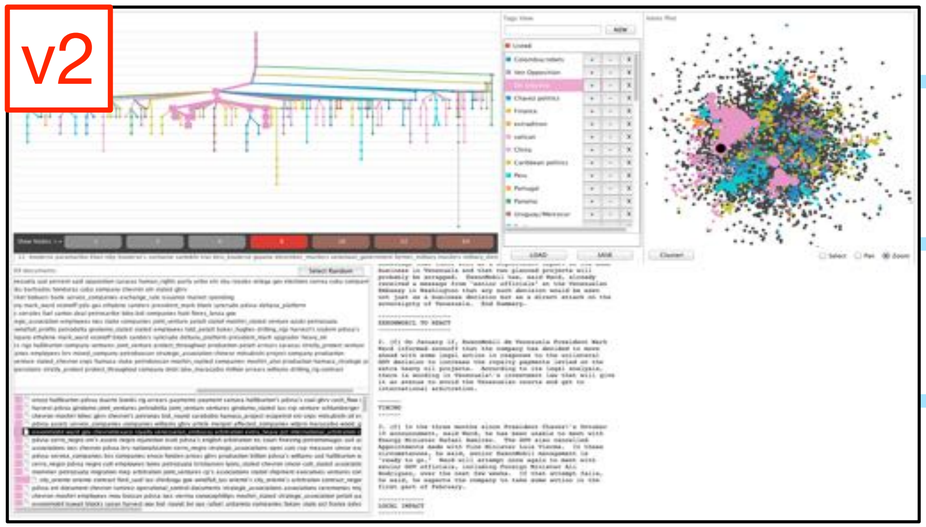
CASE STUDY #6

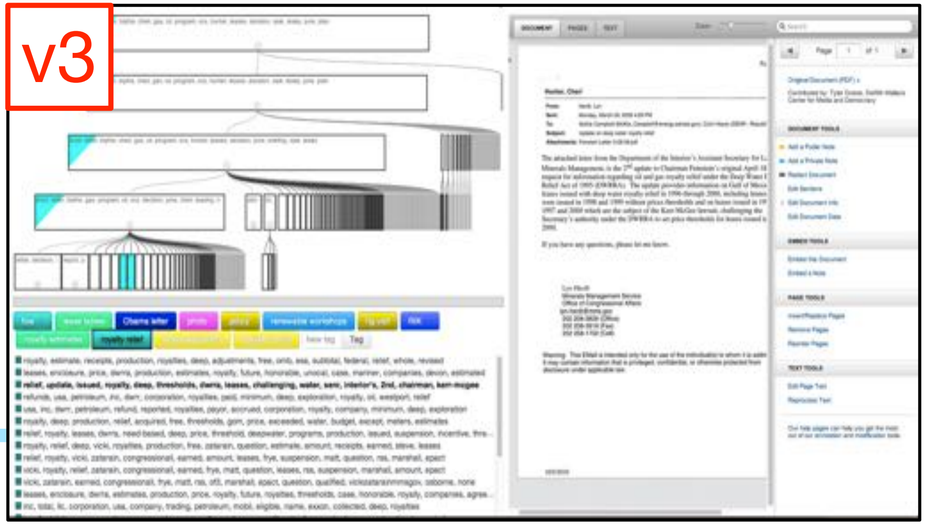
Document Collection	1,680 proposed and passed bills
Question	*Did the NY state legislature fail to pass any bills addressing police misconduct?*
Outcome	Proved non-existence of evidence

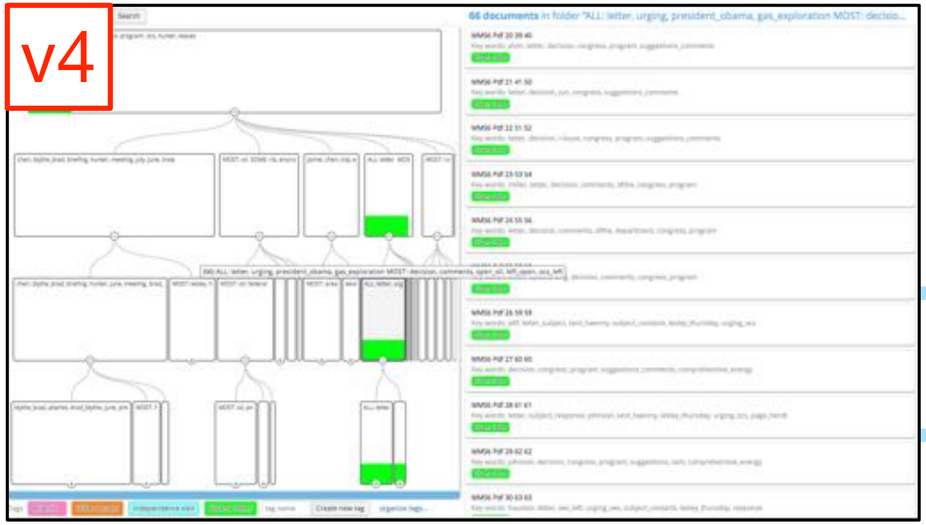
finalist for 2014 Pulitzer Prize in journalism

4 VERSIONS, 6 CASE STUDIES









CS1

CS2

CS3

CS4

CS5

CS6

4 VERSIONS, 6 CASE STUDIES

Case Study	#1	#2	#3	#4	#5	#6
Document Collection	4,500 pages from FOIA	5,996 emails from FOIA	8,680 pages from FOIA	1,278 survey comments	4,653 emails from FOIA	1,680 bills
Question	*What did security contractors do during Iraq war?*	*Were municipal police funds mismanaged?*	*Were Paul Ryan's campaign statements hypocritical?*	*What is the gun ownership debate about?*	*Was gov't response to emergency incident effective?*	*Did gov't pass bills addressing police misconduct?*

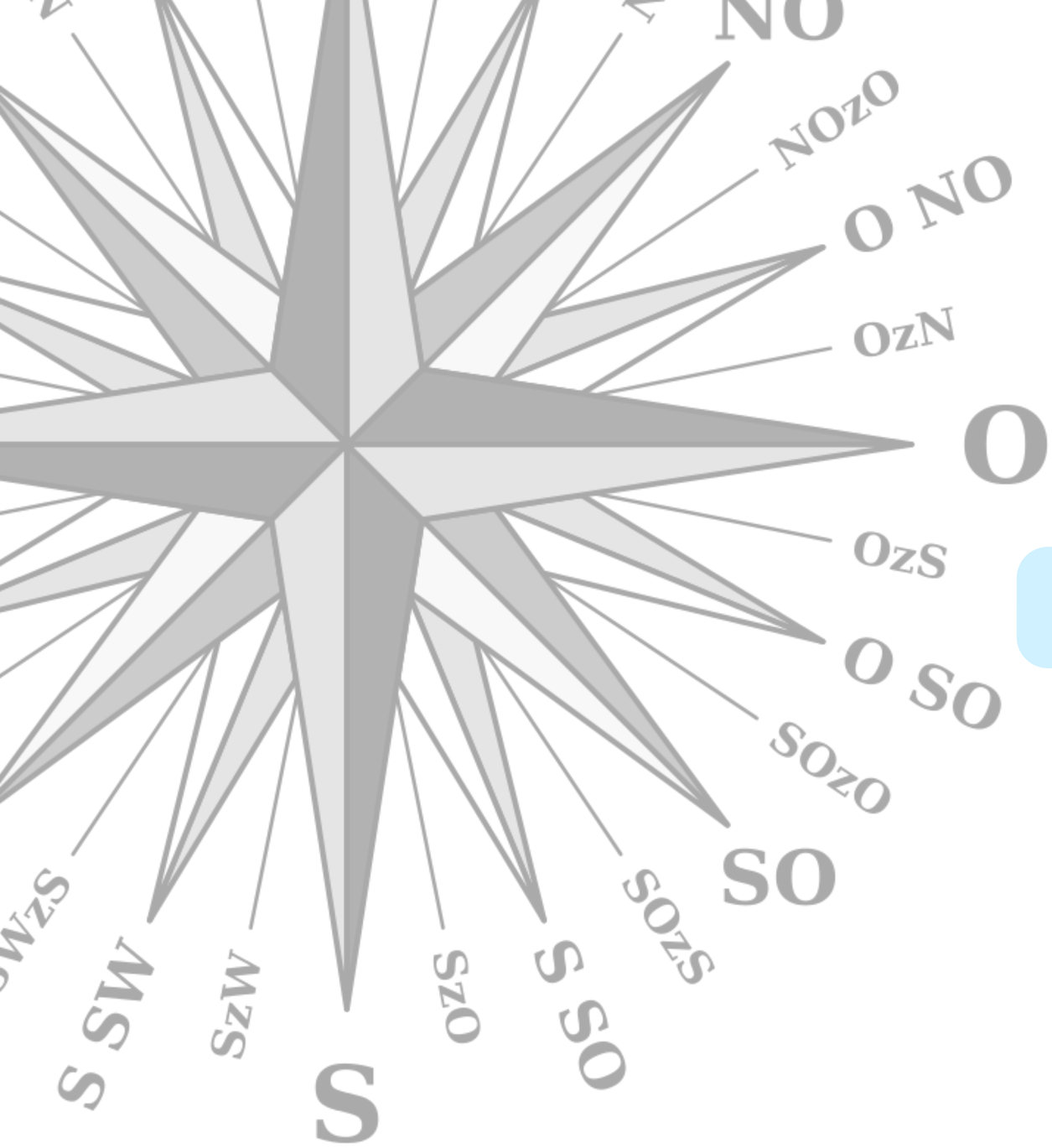
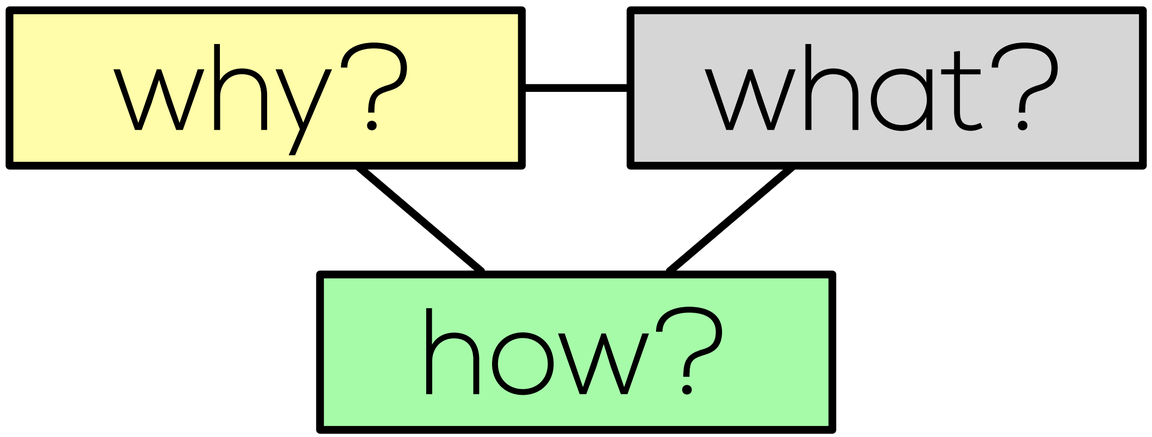


image: commons.wikimedia.org

LESSON #2

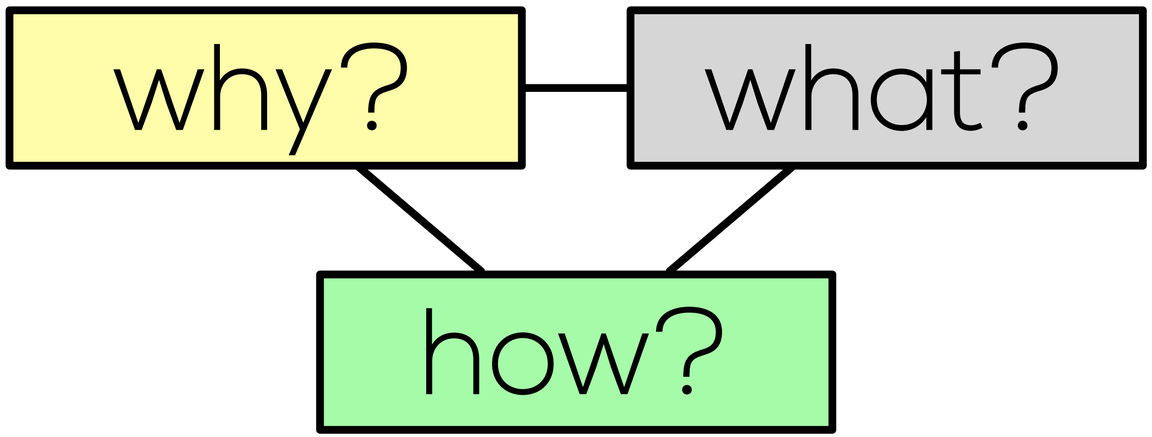
"explore" is not a well-defined task

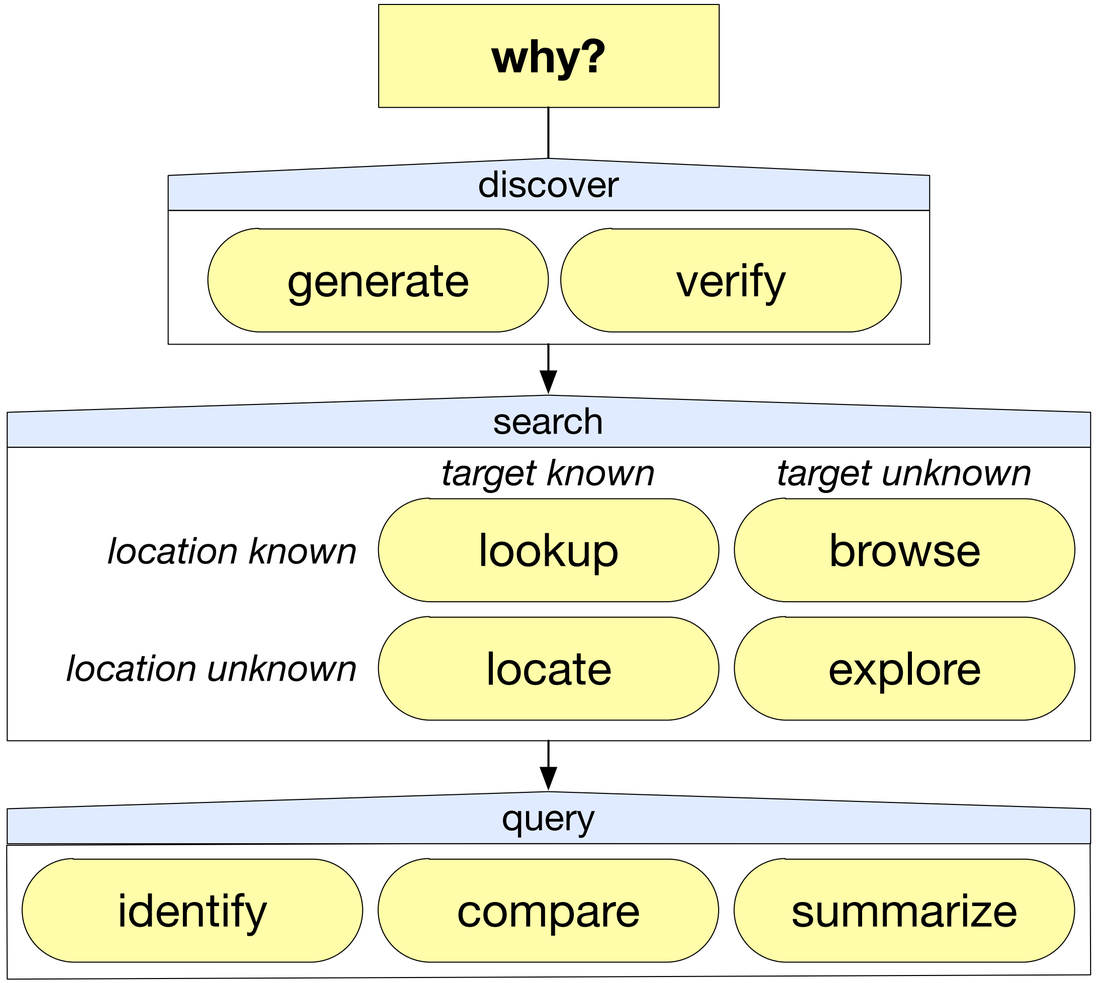
VIS TASK ANALYSIS



Brehmer & Munzner: "A multi-level typology of abstract visualization tasks" (IEEE TVCG / InfoVis 2013).

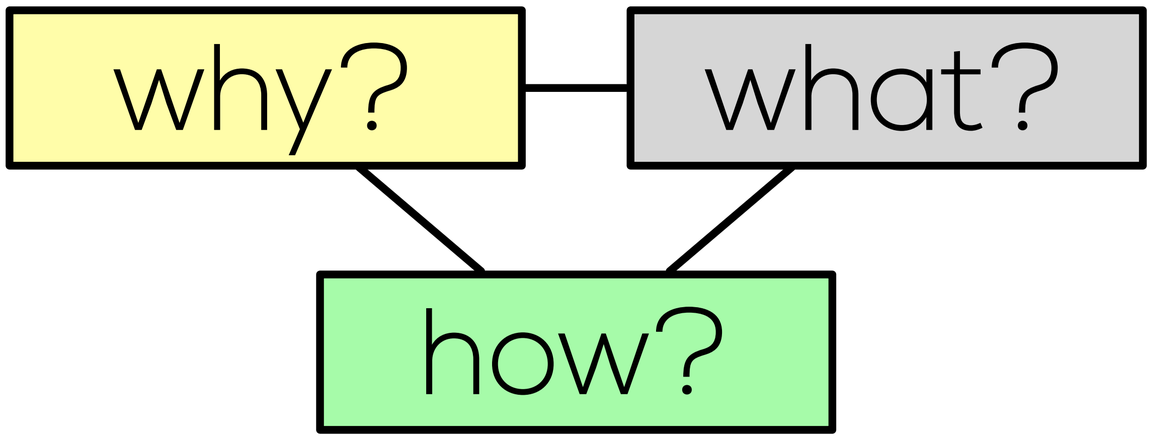
VIS TASK ANALYSIS

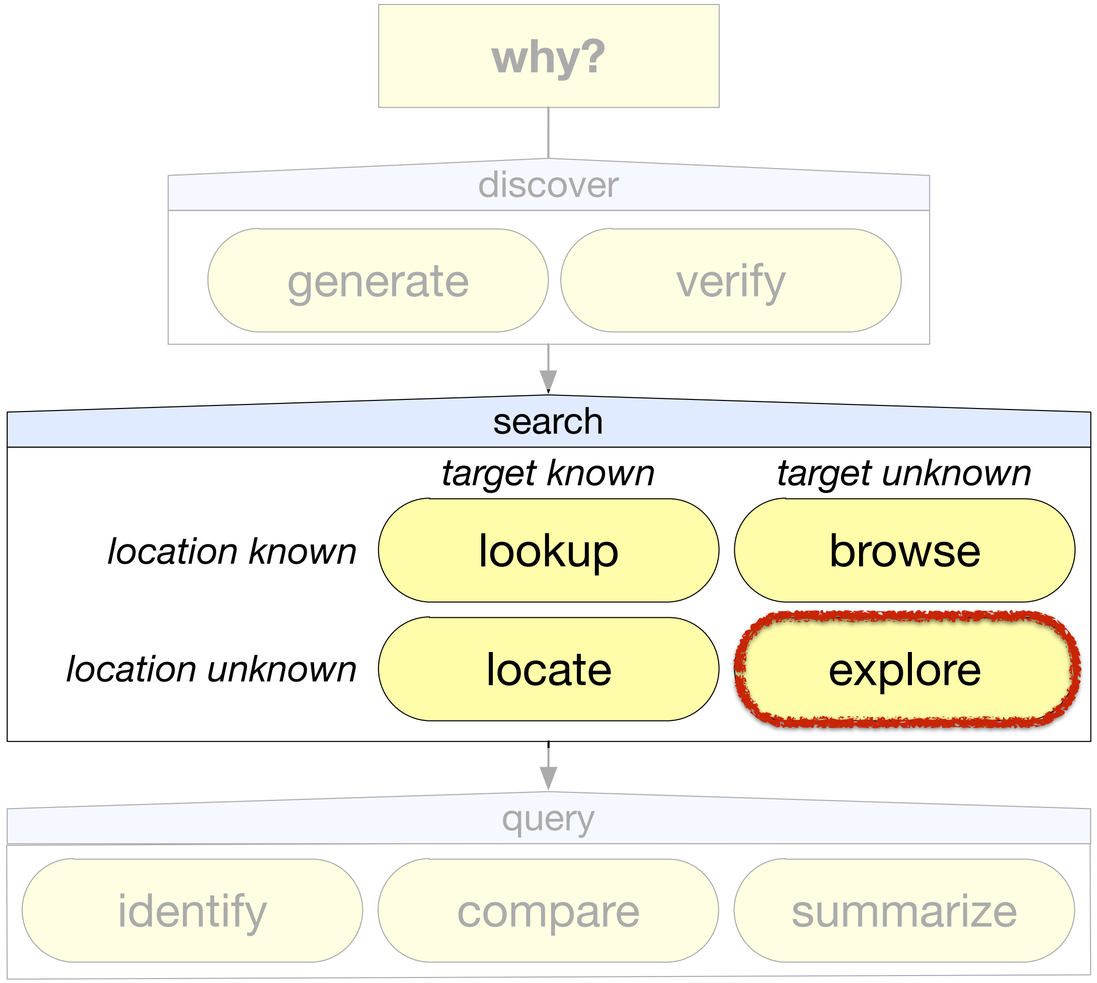




Brehmer & Munzner: "A multi-level typology of abstract visualization tasks" (IEEE TVCG / InfoVis 2013).

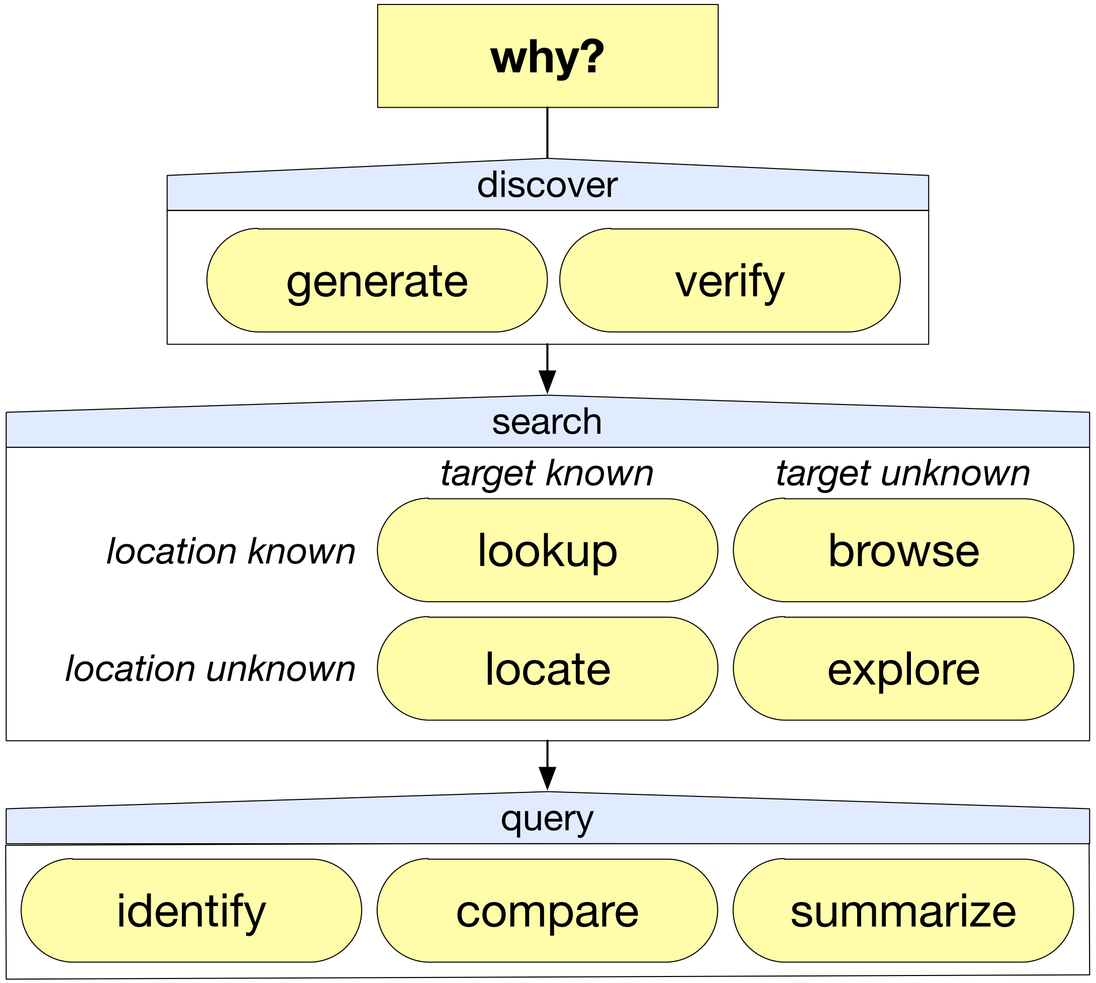
VIS TASK ANALYSIS





Brehmer & Munzner: "A multi-level typology of abstract visualization tasks" (IEEE TVCG / InfoVis 2013).

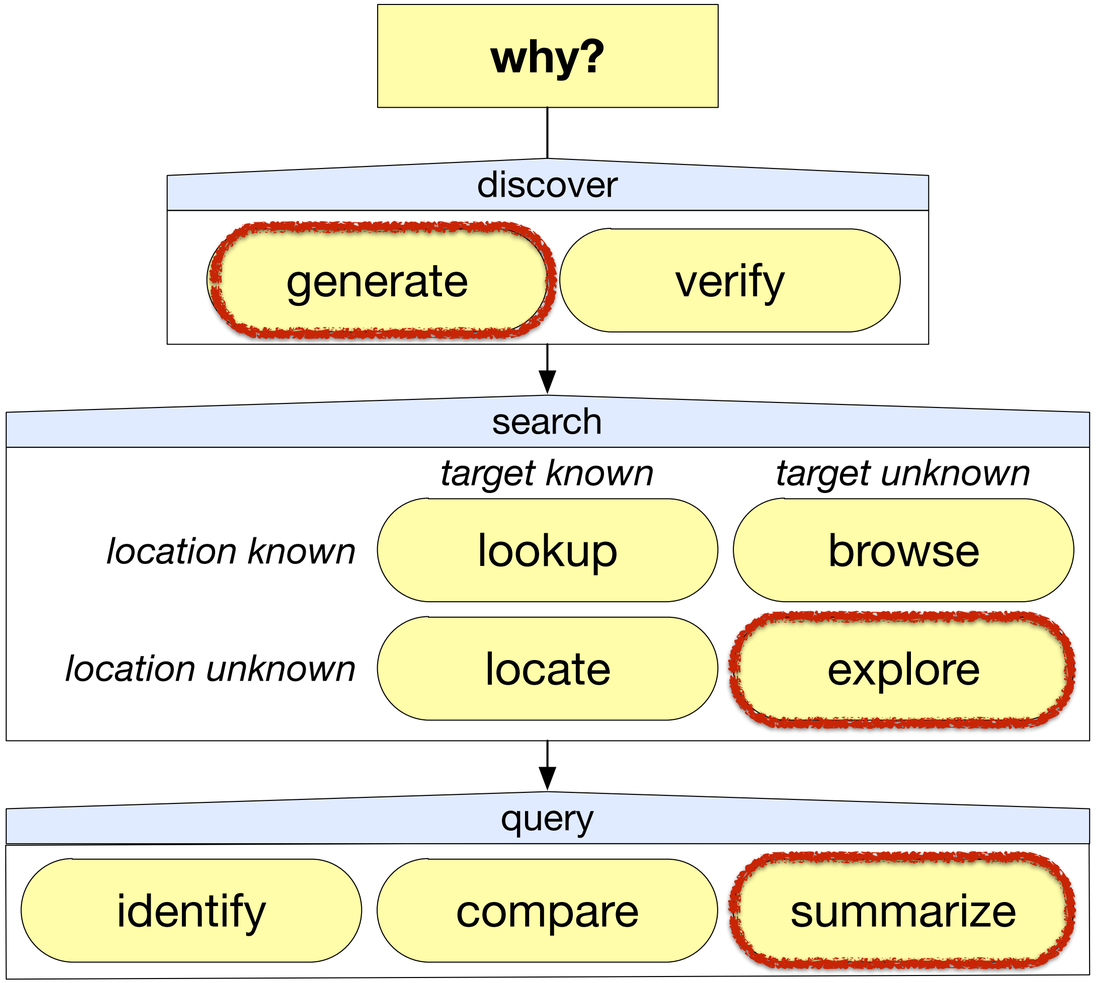
VIS TASK ANALYSIS



VIS TASK ANALYSIS

T1: generate hypotheses → explore → summarize

case studies 1, 4



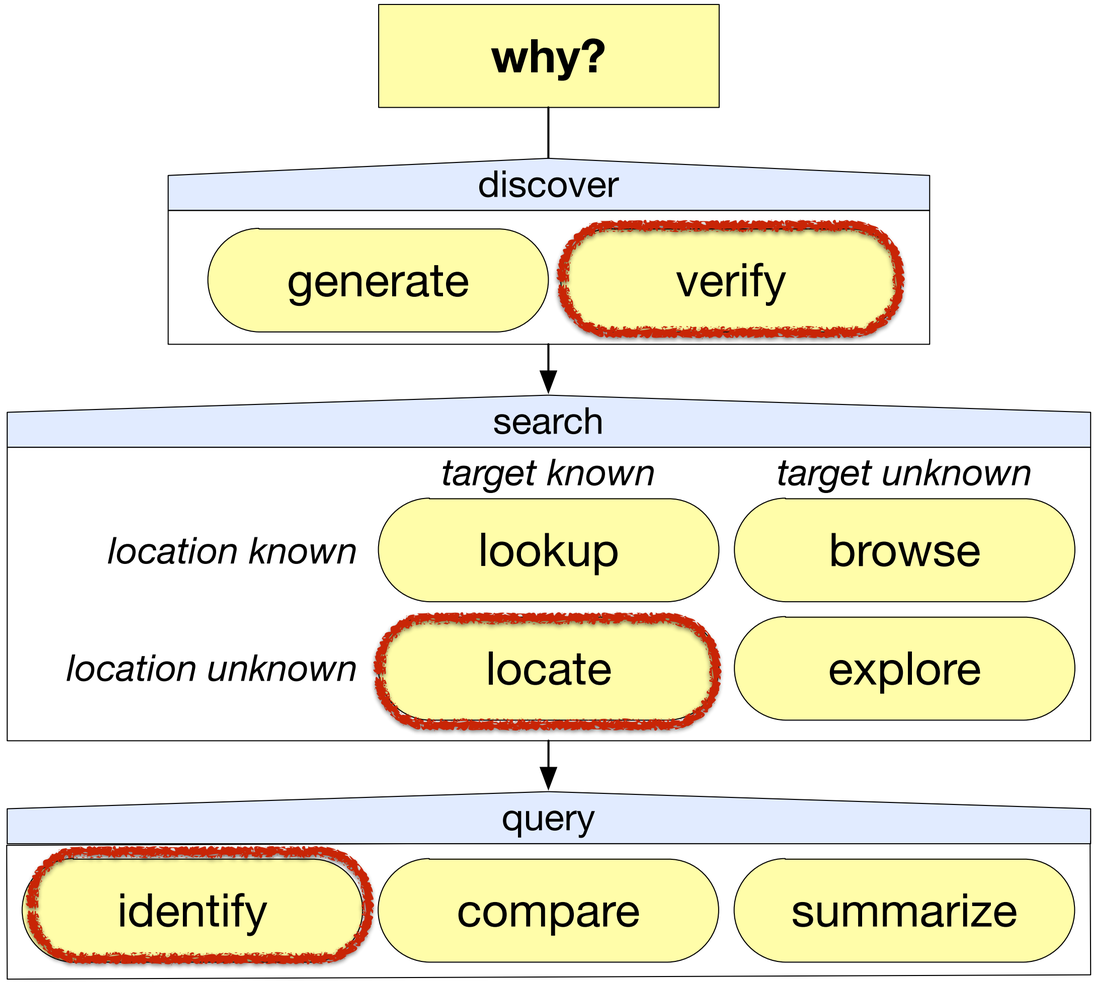
VIS TASK ANALYSIS

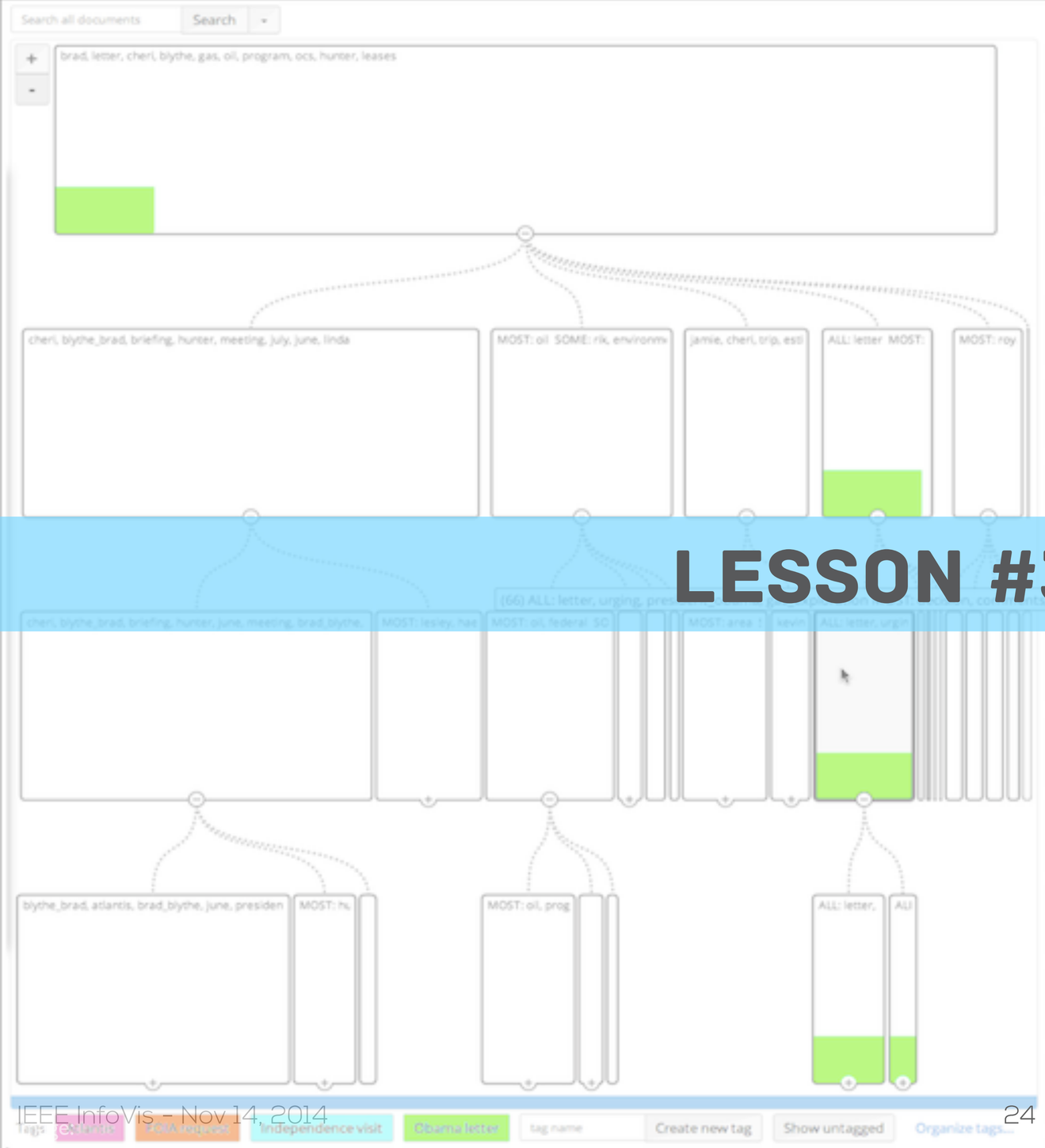
T1: generate hypotheses → explore → summarize

case studies 1, 4

T2: verify hypotheses → locate → identify

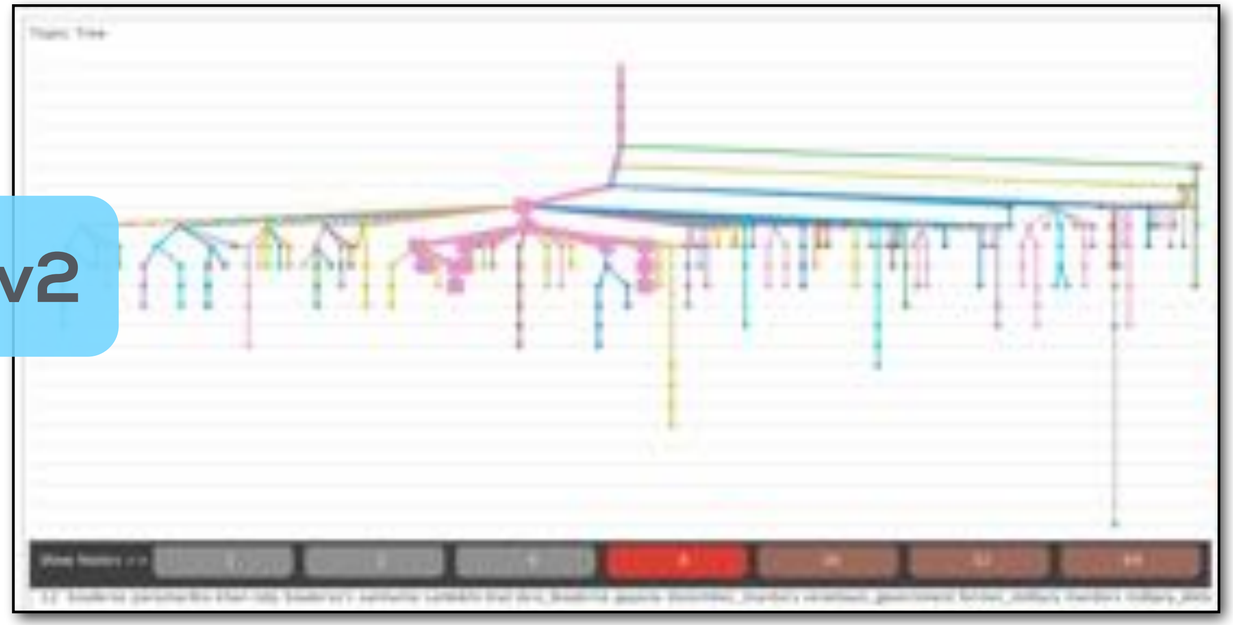
case studies 2, 3, 5, 6



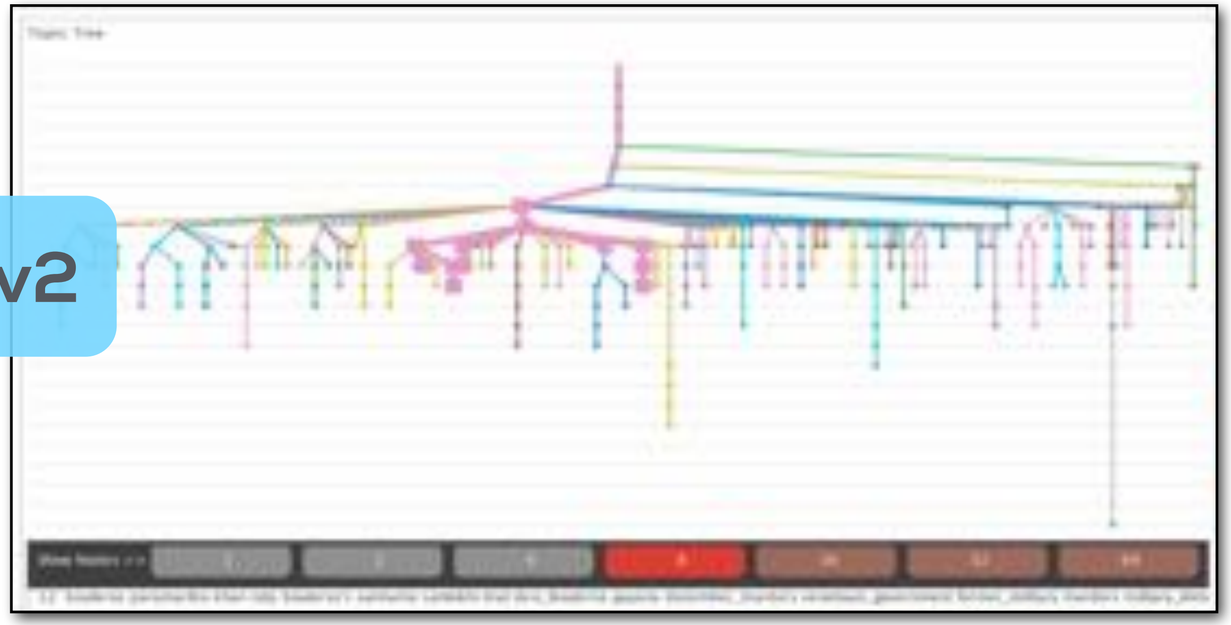


LESSON #3

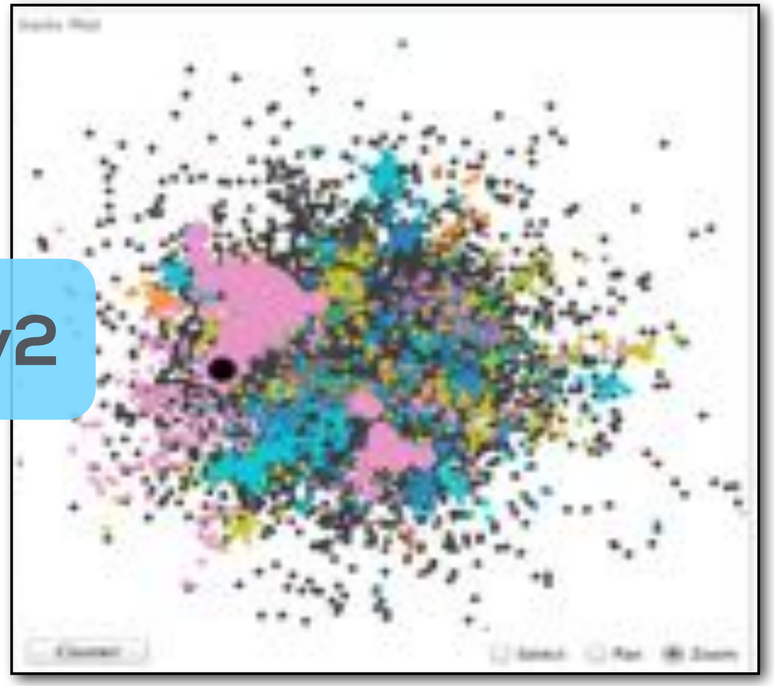
tree vis + tagging
more effective
than scatterplot


v2

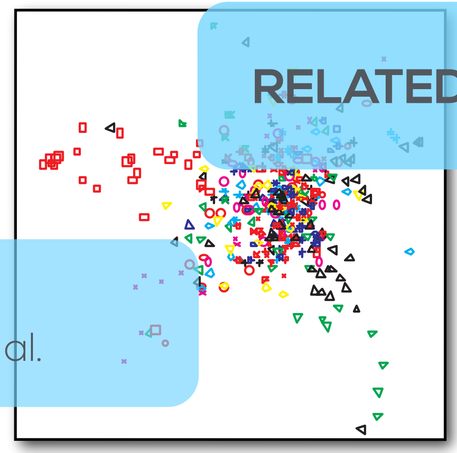
RELATED WORK



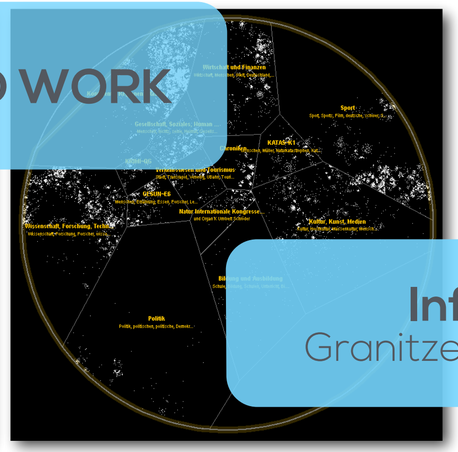
v2



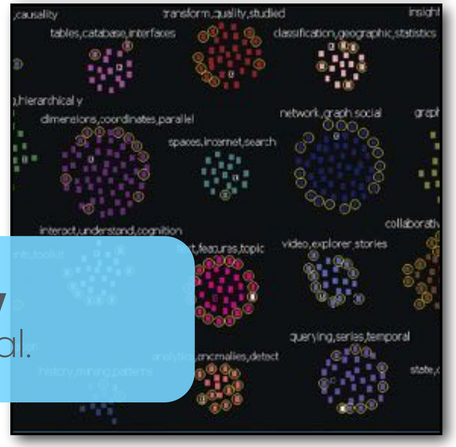
v2



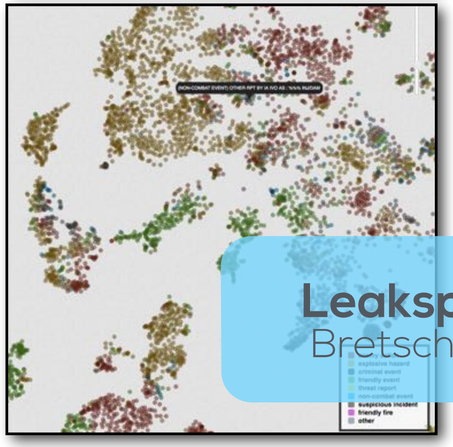
EV
Chen et al.



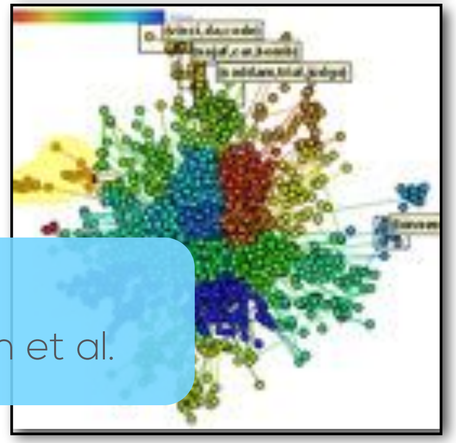
InfoSky
Granitzer et al.



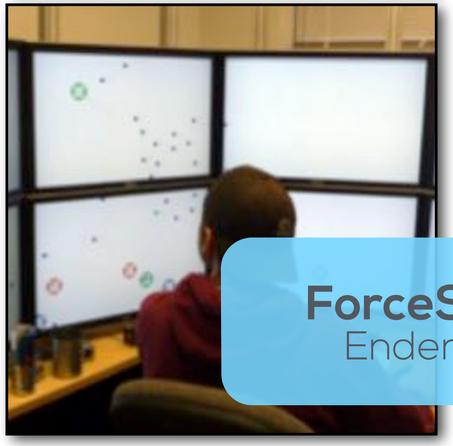
Jigsaw
Görg et al.



Leaksplorer
Bretschneider

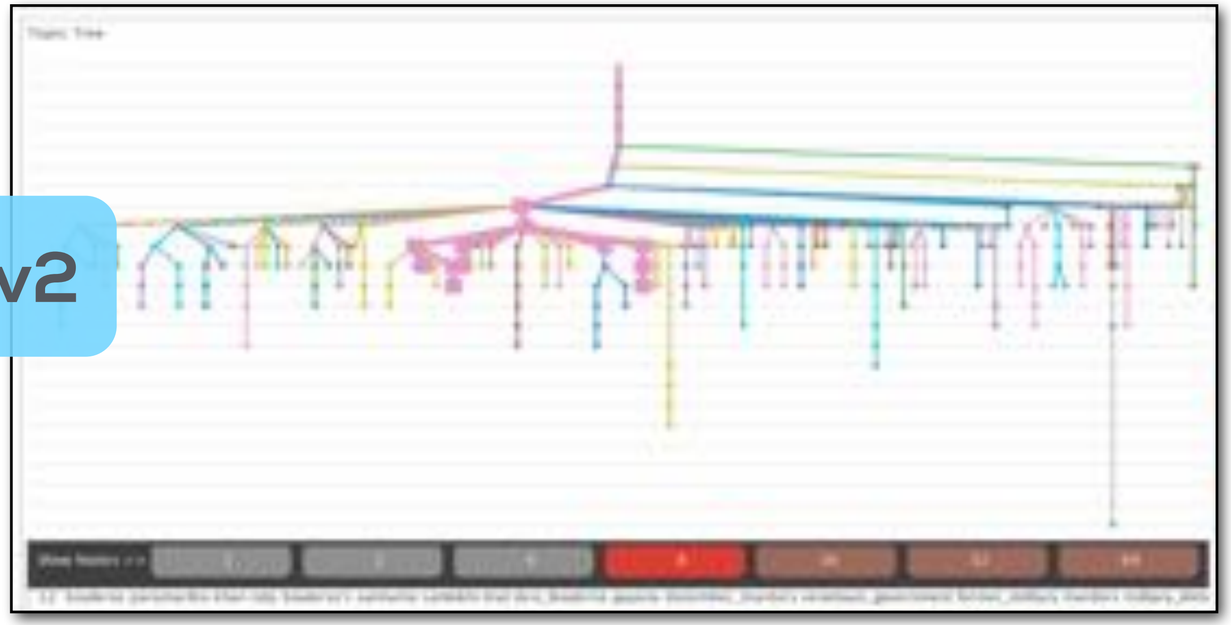


PEx
Paulovich et al.

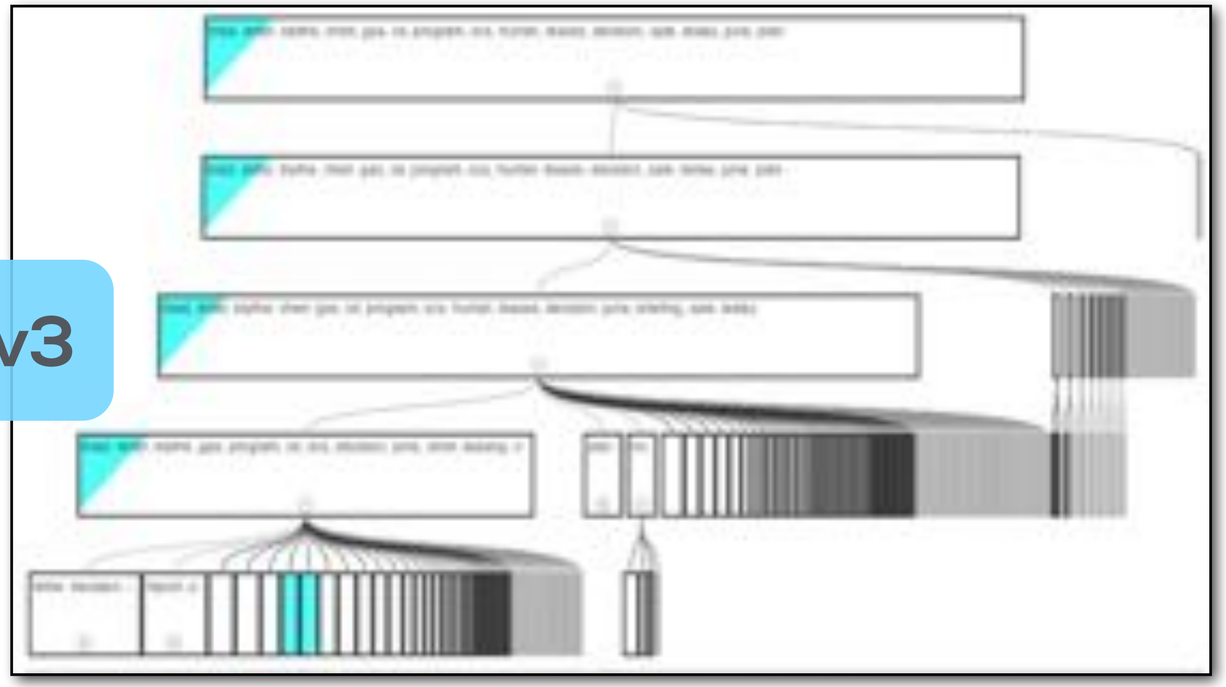


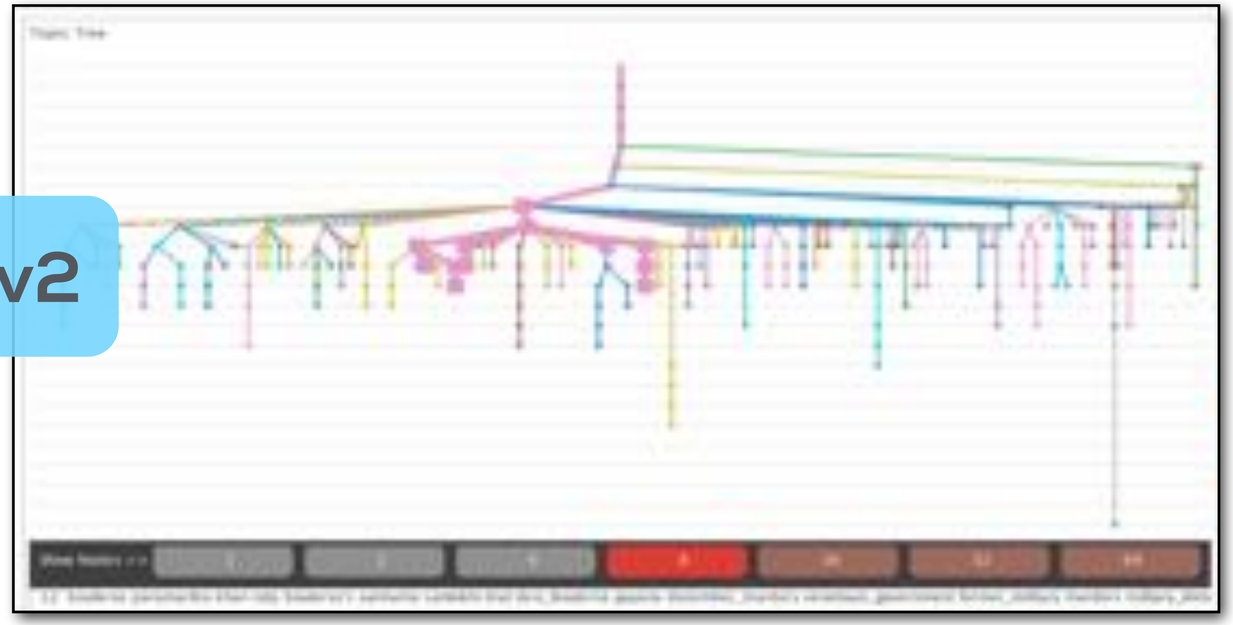
ForceSPIRE
Endert et al.

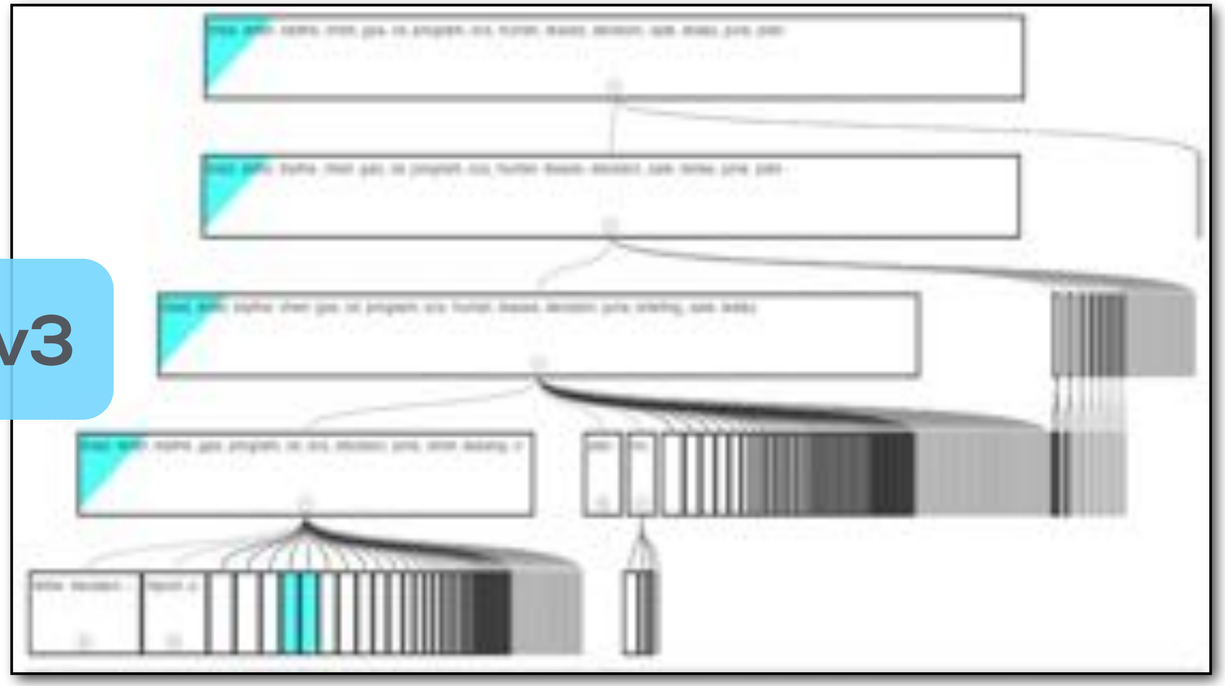
v2

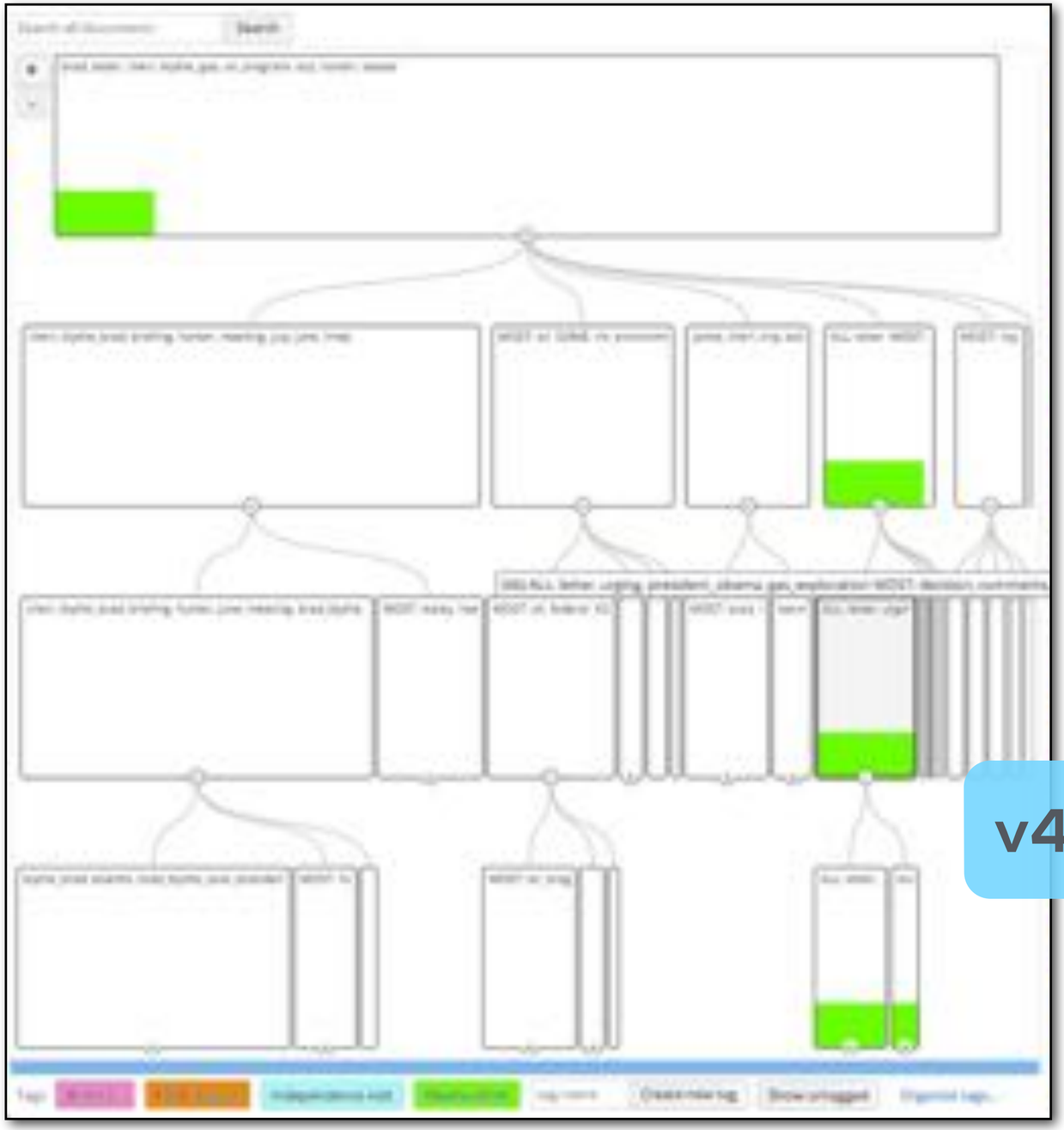


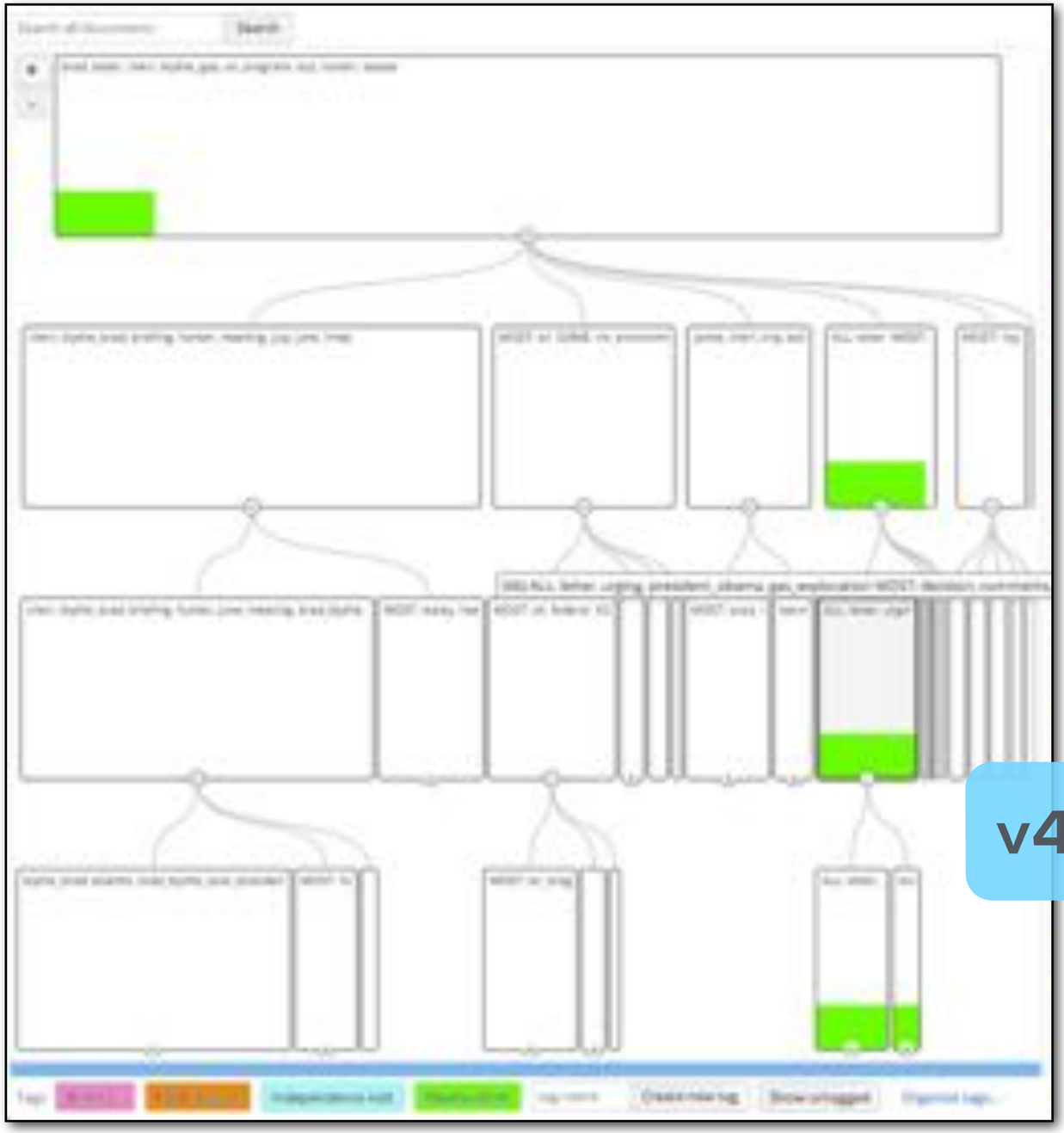
v3




v2


v3


v4


v4

v4 cluster consistency labels

'Show untagged' button

image: commons.wikimedia.org

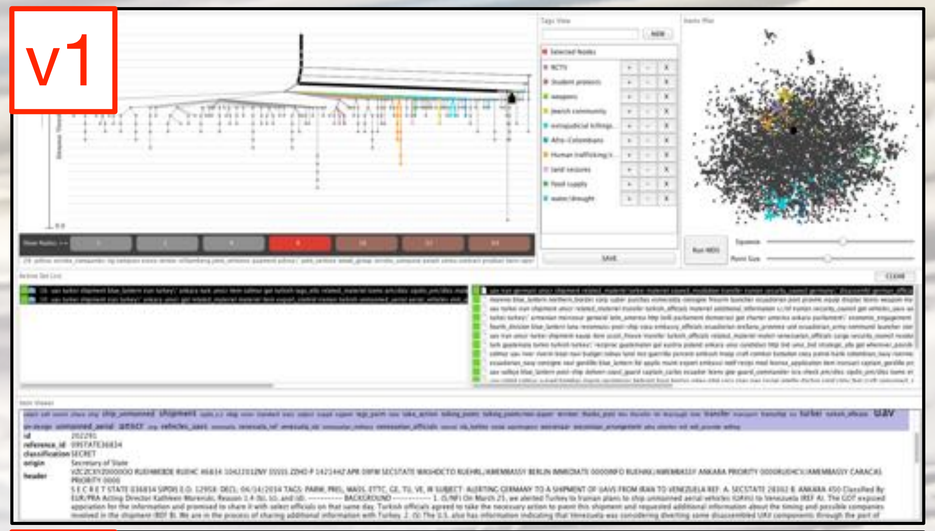
simplify for infrequent use, **reduce** data wrangling

LESSON #4

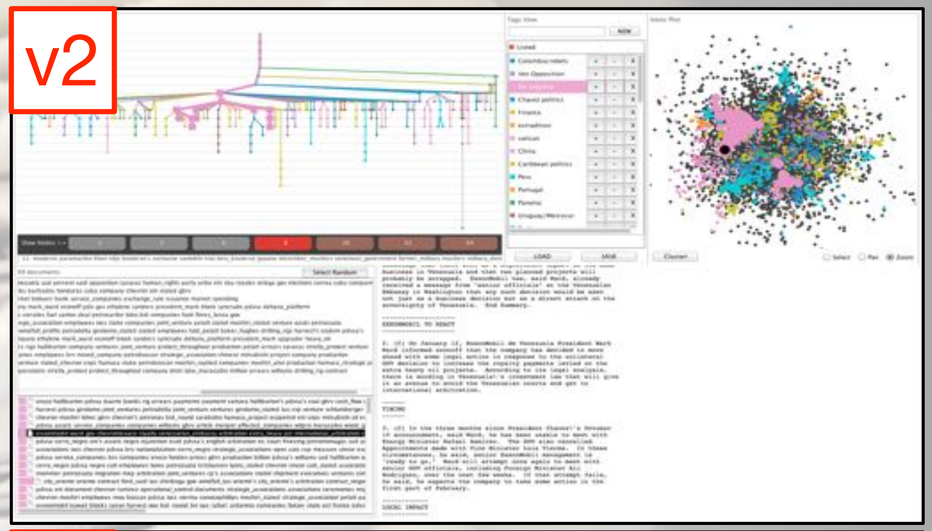
image: commons.wikimedia.org

SIMPLIFY, REDUCE WRANGLING

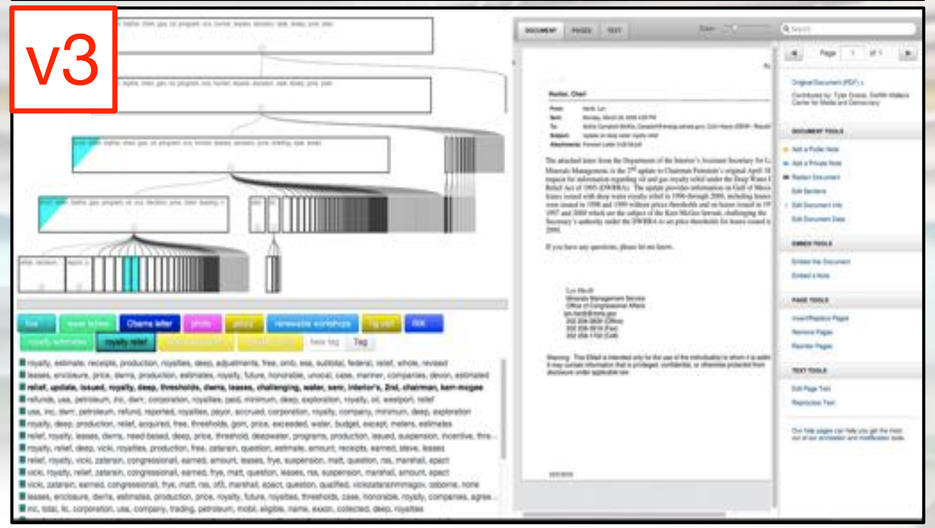
v1



v2



v3



v4

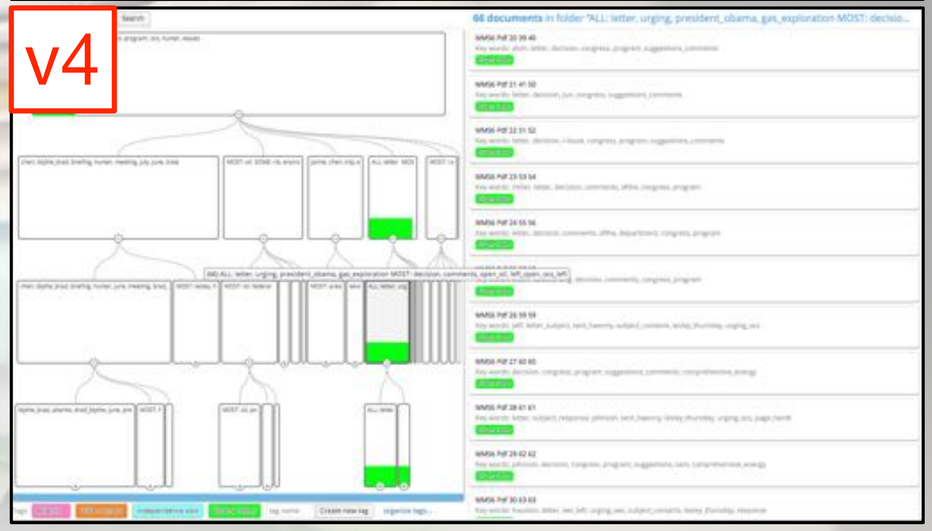


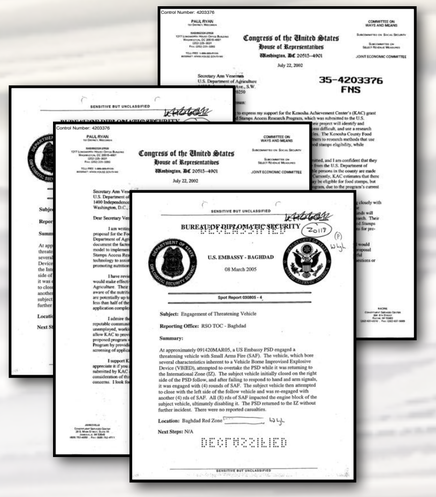
image: commons.wikimedia.org

SIMPLIFY, REDUCE WRANGLING

documentcloud.org

image: commons.wikimedia.org

SIMPLIFY, REDUCE WRANGLING



DocumentCloud

documentcloud.org

image:

LESSONS

image:

LESSONS

#1: study adoption

image:

LESSONS

#1: study adoption

#2: (re)define the tasks

image:

LESSONS

#1: study adoption

#2: (re)define the tasks

#3: trees & tags, not points

image:

LESSONS

#1: study adoption

#2: (re)define the tasks

#3: trees & tags, not points

#4: simplify, less wrangling

ACKNOWLEDGEMENTS

research team

Jonathan

Tamara

Stephen

Matthew

development team

Jonas Karlsson

Adam Hooper

journalists

Jonathan Stray

Jarrel Wade

Jack Gillum

Michael Keller

(anonymous)

Adam Playford

Matthew Brehmer
@mattbrehmer

Stephen Ingram
@FroweFace

Jonathan Stray
@jonathanstray

Tamara Munzner

Private memo reveals winding tale involving John McCain, the NRA and … condors

SUNLIGHT FOUNDATION

by Nancy Watzman | SEPT. 18, 2014, 1:59 P.M.

The Brilliance of Louis C.K.'s Emails: He Writes Like a Politician

Where campaign strategy and comedy marketing collide

the Atlantic

ADRIENNE LAFRANCE | JUL 16 2014, 6:10 AM ET

Surprise! Many credit card agreements allow repossession

Analysis: 'Security interest' clause present on 200 cards

By Fred O. Williams

CreditCards.com™

overviewproject.org **@overviewproject**

blog: **overview.ap.org** **github.com/overview**

thanks: M. Borkin, J. Dawson, J. Ferstay, J. Fulda, S.-H. Kim, H. Lam, J. McGrenere, R. Rensink, M. Sedlmair

UBC AP KF

image: commons.wikimedia.org

SUPPLEMENTAL MATERIAL

Matthew Brehmer

Stephen Ingram

Jonathan Stray

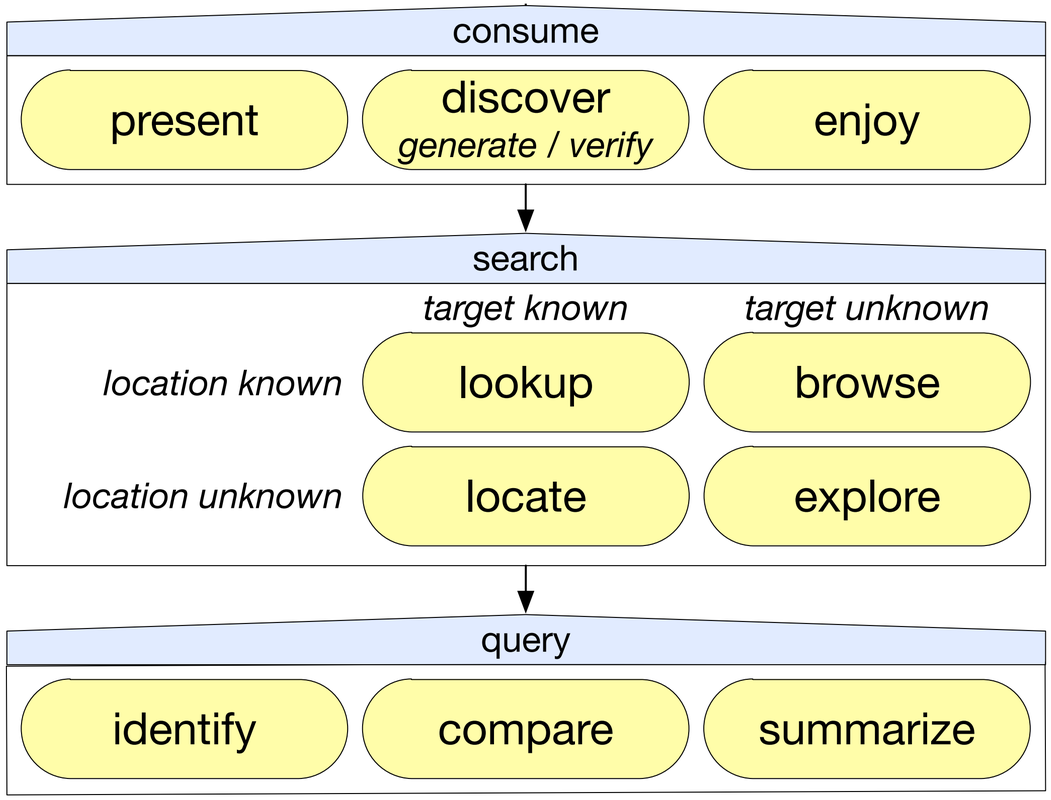
Tamara Munzner

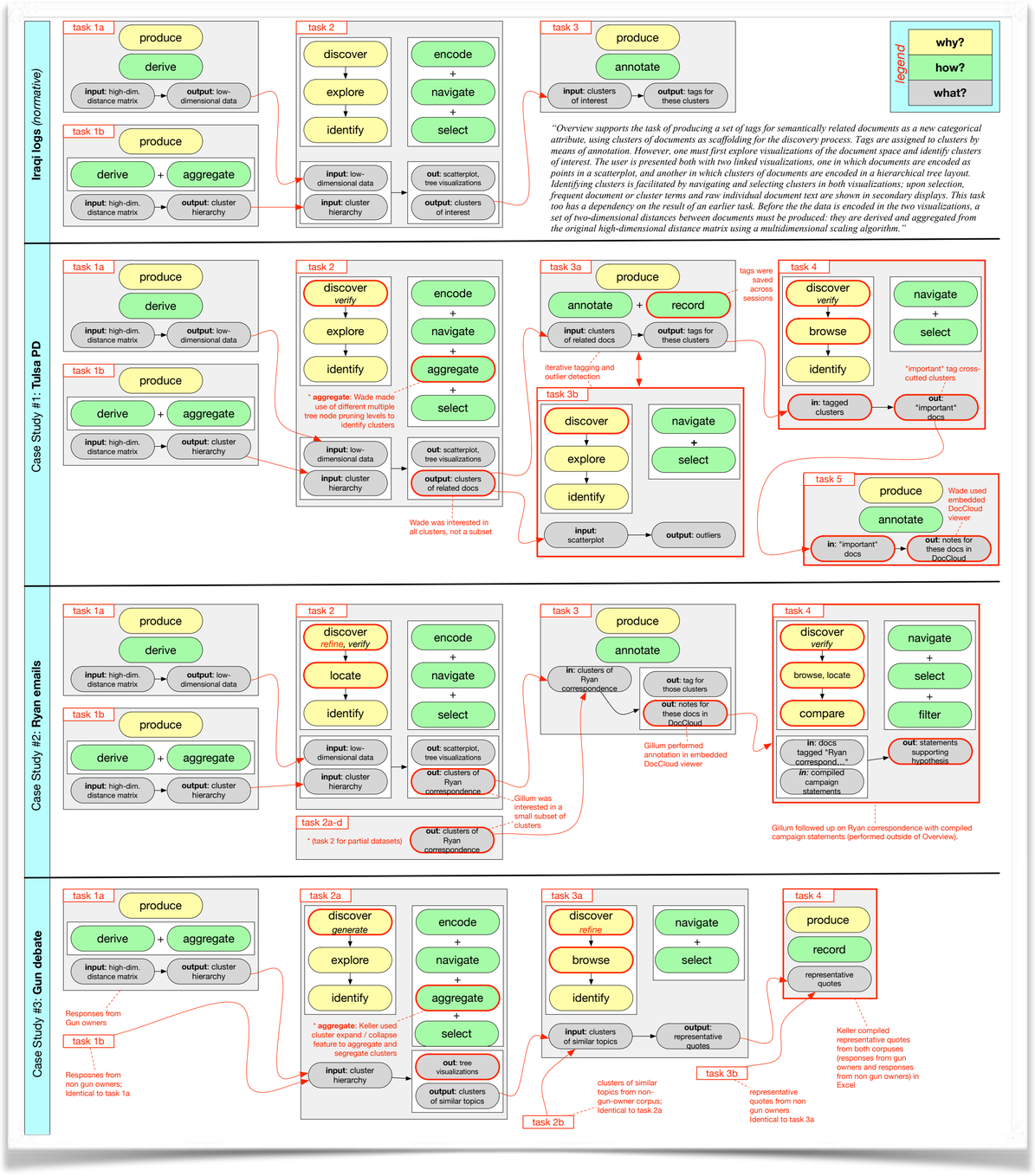
Overview

the design, re-design, deployment, adoption, analysis, re-design, deployment, adoption, analysis, re-design, deployment, adoption, and analysis of a visual document mining tool for investigative journalists: a visualization design study

IEEE InfoVis
Nov 14, 201

VIS TASK ANALYSIS





OVERVIEW VS. JIGSAW

Görg et al. (IEEE TVCG 2013)
Kang & Stasko (IEEE TVCG / VAST 2012)

Domain

Data Abstractions

Scale

Complexity

OVERVIEW VS. JIGSAW

Görg et al. (IEEE TVCG 2013)
Kang & Stasko (IEEE TVCG / VAST 2012)

	OVERVIEW
Domain	Investigative Journalism
Data Abstractions	Hierarchical Clustering of documents (raw text only), tags,
Scale	Tens of thousands of documents
Complexity	Dual view, web app

OVERVIEW VS. JIGSAW

Görg et al. (IEEE TVCG 2013)
Kang & Stasko (IEEE TVCG / VAST 2012)

	OVERVIEW	JIGSAW
Domain	Investigative Journalism	Intelligence Analysis, Law Enforcement, Academic Research
Data Abstractions	Hierarchical Clustering of documents (raw text only), tags,	Documents + document metadata + extracted entities +
Scale	Tens of thousands of documents	Hundreds to thousands
Complexity	Dual view, web app	Multiple view, desktop app

OVERVIEW VS. HIERARCHICAL TOPICS

Dou et al. (IEEE TVCG / VAST 2013)

Domain

Data Abstractions

Scale

Complexity

OVERVIEW VS. HIERARCHICAL TOPICS

Dou et al. (IEEE TVCG / VAST 2013)

	OVERVIEW
Domain	Investigative Journalism
Data Abstractions	Hierarchical Clustering of documents (raw text only), tags,
Scale	Tens of thousands of documents
Complexity	Dual view, web app

OVERVIEW VS. HIERARCHICAL TOPICS

Dou et al. (IEEE TVCG / VAST 2013)

	OVERVIEW	HIERARCHICAL TOPICS
Domain	Investigative Journalism	Academic research (?)
Data Abstractions	Hierarchical Clustering of documents (raw text only), tags,	Hierarchical Clustering of documents + temporal metadata
Scale	Tens of thousands of documents	Hundreds to thousands
Complexity	Dual view, web app	Dual view, deployment details not reported

TIMELINE & ATTRIBUTION

Jonathan

Tamara

Stephen

Matthew

Jonas Karlsson

Adam Hooper

Jonathan Stray

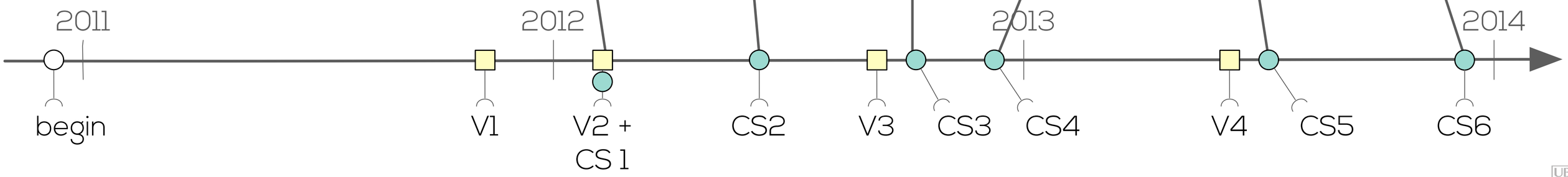
Jarrel Wade

Jack Gillum

Michael Keller

(anonymous)

Adam Playford



overview.ap.org/completed-stories

Private memo reveals winding tale involving John McCain, the NRA and … condors

by Nancy Watzman | SEPT. 18, 2014, 1:59 P.M.

SUNLIGHT FOUNDATION

Records: DHHS downplayed food stamp issues

By Tyler Dukes

Posted: 1:00 p.m. today

WRAL.com
RALEIGH • DURHAM • FAYETTEVILLE

The Brilliance of Louis C.K.'s Emails: He Writes Like a Politician

Where campaign strategy and comedy marketing collide

ADRIENNE LAFRANCE | JUL 16 2014, 6:10 AM ET

the Atlantic

Surprise! Many credit card agreements allow repossession

Analysis: 'Security interest' clause present on 200 cards

By Fred O. Williams

CreditCards.com™