Diving In: A Behavioral Scientist's Guide to Analyzing xAPI-Based Data Lakes

Gregory A. Goodwin and Nick Washburn











- Marksmanship Experiment
- Big data approach to storing/visualizing marksmanship training data (REAPER*)
- Demonstration:
 - Explore and analyze an xAPI data set.

*Range Experience Access Portal for Experimentation and Research (REAPER)





Experimental Setup: Two groups, one gets our experimental training method and the other gets standard training. We want to compare performance of the two groups during two rifle marksmanship training events.



Confirm Zero at 75, 175, and 300m

Record Fire

Position	No. Targets	Range m				
Prone Supported	20	50-300				
Prone Unsupported	10	150-300				
Kneeling	10	50-150				
23-29 hits = marksman 30-35 hits = sharpshooter 36-40 hits = expert						



LOMAH Range



Location of Hits and Misses (LOMAH)



Target





(NOTE: Picture shows 1100 mm BSU)



Record Fire Automated Range



Timed, pop-up targets (single and multiple) Automated detection of hits (not misses)



xAPI Activity Stream





REAPER Overview



BMYNA

RDS





- Relational store data in tables with columns and rows
 - SQL or think spreadsheet
 - Good for data analysis
 - Challenge of referential integrity
- Non-relational store data in a collection or pool
 - JSON (e.g., xAPI)
 - Good for Big Data, flexible
 - Challenge for data analysis



Independent Variable			Cc Va	ontextual ariables		endent riables	
Pa	rticipant #	Group	Target Type	Training Event	Target Dist	Х _{знот}	Y sнo т
1		Experimental	E	BRM5	175	5	-2
2		Experimental	E	BRM5	175	7	3
3		Control	E	BRM5	175	8	-7
4		Control	E	BRM5	175	5	11
	Part	ticipants					

Easily analyzed in statistical analysis software.





```
J,
                                                                                   Contextual Info
"context": {
"revision": "TRACR",
"extensions": {
 "http://www.riptidesoftware.com/products/soldier_tracking_system/step": "175m",
  "http://www.riptidesoftware.com/products/soldier_tracking_system/weapon": "M4",
  "http://www.riptidesoftware.com/products/soldier_tracking_system/sight": "CCO",
  "http://www.riptidesoftware.com/products/soldier_tracking_system/zeroDistance": "300",
  "http://www.riptidesoftware.com/products/soldier_tracking_system/lane": 1,
  "http://www.riptidesoftware.com/products/soldier_tracking_system/target": {
  "type": "E-Type",
  "width": 488.95,
   "height": 1022.35,
  "distance": "175m"
  "http://www.riptidesoftware.com/products/soldier tracking system/killCircle": {
  "diameter": 304.8,
   "offset": 678.180000000001
        Difficult to analyze in statistical analysis software.
```







Accessing and Understanding our Experimental Data in this Data Lake







- Statistical Software (e.g., SPSS, SAS, Statistica, R, and even Excel)
 - Familiar to behavioral researchers
 - Prefer relational databases
 - Are not optimized to handle big data (yet)
- Big Data Tools (e.g., Mongo DB, Google Cloud Platform, Elasticsearch)
 - Unfamiliar to behavioral researchers
 - Don't perform the range of lineal and non-linear inferential statistics that statistical software does
 - Excellent for combining, analyzing, and visualizing big data.









Demo Screenshots





RDECON





View relevant data points

	Liboro	Available Fields 0		
	KIDANA	Popular	3 9000	
0		# X-Position		
	Discover	# Y-Position	2018-01-31 2018-02-38 2018-03-31 2016-04-30 2018-05-31 2018-06-30	
E.	Visualize	8 score	©	
0	Dashboard	* June	Timesource	
	Waterland	t actoriname	May 9th 2018, 14:29:44 Q Q annor move mailto:soldier 1240@reager elemts com annor same 1249 work 14: http://www.mintidesoftmare.com/mroducts/shooter toorking system/	n ste
•	timelion	t object.id	<pre>d verb.display.en-US: missed object.id: http://www.riptidesoftware.com/products/shooter_tracking_system/target/300 object.objectType: Activi</pre>	y
舛	Graph	© timestamp	object.definition.name.en-us: 300 Meter Target result.success: false	50 C
×	Dev Tools	t verb.display.en-US	result.extensions.http://www.riptidesoftware.com/products/shooter_tracking_system/abot/position.xPos: 298	
-			result.extensions.http://www.riptidesoftware.com/products/shooter_tracking_system/shot/position.yPos: 788	
~	Monitoring	t BRM		
۰	Management	t Base	Table JSDN View surrounding documents View single docu	lent.
		O Day	t BRM Q, Q, 🗇 # HTTP://WWW.RIPTIDESOFTWARE.COM/PRODUCTS/SHOOTER,	TRAC
		t Range	KING_SYSTEM/TARGET/300	
		# Shor in kill simle	t Base Q Q 🗇 🕸 Foundry	
			© Doy Q Q □ * Moy 9th 2018, 14:29:44.691	xAPI Data Fields
		t jd	t Ronge Q Q 🗇 🖲 Wulcon	
		t _index	# Shot in kill circle Q Q 🗇 # false	Expanded
		t_type	# X-Position Q Q 🗇 = 295	
		t actor mbox	# Y-Position Q Q II # 788	
			t _ld Q C = * xbyxGRQBejE&THxxH5	
		? context.contextActivities.category	t _index Q Q II = reaper	
		? context.contextActivities.grouping	#_score 0,0 🗆 # -	
		7 context.contextActivities.parent	t _type Q, Q □ # doc	
		# context.extensions.http://www.ri	t actor.mbox @ @ 🗇 🗰 mailto:soldier_1249@reaper.elm/ts.com	
			t actor,name @, Q, () # 1249	
		context.extensions.http://www.n_	7 context.contextActivities.coregory 🔍 🔍 🖾 🛋	
		# context_extensions.http://www.ri	【 【id*: "http://www.rlptidesoftware.com/products	/sho
		t context.extensions.http://www.ri_	oter_tracking_system/position/prone_supported", "objectType": "Activity",	
2	Riptide	t context.extensions.http://www.rl_	"definition": { "nome": {	
a	Logout	# context.extensions.http://www.rl	"en-us": "Shooter firing from the Prone Su ted Position"	.por

ARMY NAS



Check ranges and means

		Visuali	ze Tab			
🖌 kibana	Visualize / New Visualization (up Search (e.g. status /				Save Share	Refresh Reporting C Auto-refresh C O Last 6 months > Options Q
Ø Discover	Add a set					
U Visualize	reaper	Group =	Target ^	Mean X 0	Lower Standard Deviations X 🗄	Upper Standard Deviations X 🗈
O Dashboard	Data Options 💽 🗴	Catalan	175 Meter Target	2.083	-276.408	280.575
Timelion	Metrics	Vulcan	175 Meter Target	0.173	-281.612	281.959
🦄 Graph	Metric Average X.Position C L	Catalan	300 Meter Target	1.88	-292.403	296.163
P Dev Tools	MetricStandard Deviation of X-Postion	Vulcan	300 Meter Target	0.623	-312.317	313.564
	Add metrics	Catalan	75 Meter Target	0.393	-298,578	299.363
Wonitoring		Vulcan	75 Meter Target	7.917	-322.277	338.111
O Management	Buckets Split Rows Range: Descending O I Split uskeyword: Descending O I					
	Add sub-buckets	Export: Raw	▲ Formatted ▲			Mean and Standard
						Deviations of x Axi

BMYN



Select items for csv file



🗾 kibana	203,259 hits Search (e.e	Discover Tab	Share Hepor	Choosii drops	ng data f them into	ields on left panel a table that can
Discover	Add a filter +			be	exported	to a csv file
U Visualize	reaper	3 january 13th 2018, 10:28:50.227 - July 13th 2018, 10:28:50.227 - Auto \$				
 Dushboard Timelion Graph 	Selected Fields t Range # X-Position # Y-Position	000,007				
> Dev Tools	t BRM	0 2018-01-31 2018-02-20 2018-03-31 2018-04-30	2018-0	5-31	2018-06-30	
S Monitoring	Available Fields •	timestamp per day				
Management	Popular	Time – BRM	Range	X-Position	Y-Position	
	actor name	May 9th 2018, 14:29:44.691 HTTP://WWW.RIPTIDESOFTWARE.COM/PRODUCTS/SHOOTER_TRACKING_SYSTEM/TARGET/300	Vulcan	298	788	
	t object id	• Moy 9th 2018, 14:29:44.691 HTTP://WWW.RIPTIDESOFTWARE.COM/PRODUCTS/SHOOTER_TRACKING_SYSTEM/TARGET/175	Vulcan	10	753	
	O timestamp	Moy 9th 2018, 14:29:44.691 HTTP://WWW.RIPTIDESOFTWARE.COM/PRODUCTS/SHOOTER_TRACKING_SYSTEM/TARGET/175	Vulcan	-24	485	
	t verb.display.en-U5	Moy 9th 2018, 14:29:44.691 HTTP://WWW.RIPTIDESOFTWARE.COM/PRODUCTS/SHOOTER_TRACKING_SYSTEM/TARGET/175	Vulcan	-91	688	
	t Base	May 9th 2018, 14:29:44.691 HTTP://WWW.RIPTIDESOFTWARE.COM/PRODUCTS/SHOOTER_TRACKING_SYSTEM/GROUPING/TABLE/175/3	Vulcan	0	8	
	© Day	Moy 9th 2018, 14:29:44.691 HTTP://WWW.RIPTIDESOFTWARE.COM/PRODUCTS/SHOOTER_TRACKING_SYSTEM/TARGET/175	Vulcan	-41	227	
	# Shot in kill circle	Mov 9th 2018, 14:29:44.691 HTTP://WWW.RIPTIDESOFTWARE.COM/PRODUCTS/SHOOTER.TRACKING_SYSTEM/TARGET/175	Vulcan	-9	-41	
	t jd	Mov 9th 2018 14-29-44 691 HTTP://WWW RIPTINESDETWARE COM/PRODUCTS/SHOOTER TRACKING SYSTEM/TARGET/17S	Wilcon	-118	264	
	t _index	May Geb 3018 14-20-44 601 UTTP-//WWW DTDTTDESDETWARE //W/DDDDDI/TTC/CUDDTED TD//YTWC SVCTEM/TAD/ET/125	Wullcon.	207	267	
	t_type		to Loon	107	2.07	
	t actor.mbox	• MGY 3EN 2018, 14:23:44.031 HITF://WWW.NLFTIDESOFTWARE.COMPRODUCTS/SHOUTER_TRACKING_STSTEMPTAROET/1/S	vulcan	-150	1,004	
	? context.contextActivities.category	Moy 9th 2018, 14:29:44.691 MTTP://NWW.RIPTIDESOFTWARE.COM/PRODUCTS/SHOOTER_TRACKING_SYSTEM/TARGET/17S	Vulcan	158	255	
	? context.contextActivities.grouping	 Moy 9th 2018, 14:29:44.691 HTTP://WWW.RIPTIDESOFTWARE.COM/PRODUCTS/SHOOTER_TRACKING_SYSTEM/GROUPING/175 	Vulcan	0	0	
🚨 Riptide	? context.contextActivities.parent	 May 9th 2018, 14:29:44.691 HTTP://WWW.RIPTIDESOFTWARE.COM/PRODUCTS/SHOOTER_TRACKING_SYSTEM/TARGET/300 	Vulcan	128	359	
- Losout	# context.extensions.http://www.rl	May 9th 2018, 14:29:44.691 HTTP://WWW.RIPTIDESOFTWARE.COM/PRODUCTS/SHOOTER_TRACKING_SYSTEM/GROUPING/TABLE/300/1	Vulcan	0	0	

Use statistical software to analyze data...



REAPER Visualizations



REAPER Dashboards



BRM5 Soldier Dashboard







Improved Shot Pattern Analysis - Breathing

- Can be marked on any Grouping or Untimed scenario
- Pattern identified when the bounding box for the shots is largely vertical (3.5 x taller than wide)
- Accounts for "outliers" to mark breathing problem when 80% of the shots fall in the vertical pattern









Improved Shot Pattern Analysis – Trigger Squeeze

- Can be marked on any Grouping or Untimed scenario
- Pattern identified when the bounding box for the shots is largely horizontal (3.5 x wider than tall)
- Accounts for "outliers" to mark breathing problem when 80% of the shots fall in the vertical pattern







- Experiments/Research are "little data" by design
 - IRB restricts participant counts to the minimum needed for inference
 - Inferential stats were developed for small samples
- Will Statistical software like SAS, SPSS, Statistica, and R eventually be able to use web services to analyze petabytes of big data quickly?

- Web services for business intelligence/big data are developing rapidly

- Is there a fundamental philosophical difference between small data research and big data research?
 - Naturalistic vs. laboratory experiments
 - Correlational vs. causal attribution