

1
00:00:00,000 --> 00:00:04,559
good evening they thank you everybody

2
00:00:02,549 --> 00:00:07,500
for being here especially with the room

3
00:00:04,559 --> 00:00:10,859
being hot so and I like to thank the

4
00:00:07,500 --> 00:00:15,269
society for allowing me to present here

5
00:00:10,859 --> 00:00:17,339
at the conference this evening and I'm a

6
00:00:15,269 --> 00:00:21,210
member of the Society for scientific

7
00:00:17,339 --> 00:00:24,179
exploration and I wanted to talk about

8
00:00:21,210 --> 00:00:29,070
making UFO data more useful for

9
00:00:24,179 --> 00:00:30,599
scientific research and I think and this

10
00:00:29,070 --> 00:00:39,289
could apply to other fields as well

11
00:00:30,599 --> 00:00:42,769
including parapsychology so I think this

12
00:00:39,289 --> 00:00:47,000
symbolizes kind of where the state of

13
00:00:42,770 --> 00:00:48,840
working with UFO data is at the moment

14
00:00:47,000 --> 00:00:49,890
especially for you those in the back of

15
00:00:48,840 --> 00:00:52,260
may not if you can't see it we were

16
00:00:49,890 --> 00:00:55,020
clearly enough it's these alien

17
00:00:52,259 --> 00:00:56,429
spaceship is landed it to aliens get out

18
00:00:55,020 --> 00:00:58,530
once that's the other see I told you

19
00:00:56,429 --> 00:01:00,359
they wouldn't notice this you have all

20
00:00:58,530 --> 00:01:01,890
the there are things got their heads

21
00:01:00,359 --> 00:01:05,760
buried in their smartphones they can't

22
00:01:01,890 --> 00:01:08,099
see them so I think I think that's with

23
00:01:05,760 --> 00:01:10,618
the case in a symbolic way with the UFO

24
00:01:08,099 --> 00:01:13,709
data well there's some things we cannot

25
00:01:10,618 --> 00:01:16,438
see with it because of just the the

26
00:01:13,709 --> 00:01:20,129
nature of the data and also there's just

27
00:01:16,438 --> 00:01:22,469
only fairly recently tools that we have

28
00:01:20,129 --> 00:01:24,329
coming online in which we can address

29

00:01:22,469 --> 00:01:29,728
that but this we just haven't really

30
00:01:24,329 --> 00:01:31,200
been able to employ those tools and what

31
00:01:29,728 --> 00:01:33,179
I'm going to do is just a previews I'm

32
00:01:31,200 --> 00:01:36,390
going to just give a background on UFO

33
00:01:33,180 --> 00:01:38,430
data and then discuss some criticism

34
00:01:36,390 --> 00:01:41,009
about it about it in regard to its

35
00:01:38,430 --> 00:01:43,590
scientific usefulness and also a

36
00:01:41,009 --> 00:01:45,868
proposed method for enhancing it to

37
00:01:43,590 --> 00:01:47,939
dress those criticisms and then going

38
00:01:45,868 --> 00:01:52,078
into a conclusion and then question and

39
00:01:47,938 --> 00:01:55,228
answers so as far as the background

40
00:01:52,078 --> 00:01:57,779
about UFO data some key characteristics

41
00:01:55,228 --> 00:02:00,629
is its there's desparate sources I mean

42
00:01:57,780 --> 00:02:03,859
there's collections of UFO data all over

43
00:02:00,629 --> 00:02:08,728

the world here in the United States and

44

00:02:03,859 --> 00:02:10,439

Britain Europe Asia there's all kind all

45

00:02:08,729 --> 00:02:13,129

sorts all kinds of sources

46

00:02:10,439 --> 00:02:13,128

government sources

47

00:02:13,550 --> 00:02:17,730

private researchers non-private

48

00:02:16,080 --> 00:02:19,709

nongovernmental organizations or

49

00:02:17,729 --> 00:02:21,929

non-profit organizations research groups

50

00:02:19,709 --> 00:02:25,580

there's all kinds of sources out there

51

00:02:21,930 --> 00:02:28,890

and also there's active and non-active

52

00:02:25,580 --> 00:02:31,380

collections of data active as in they're

53

00:02:28,889 --> 00:02:32,909

currently accepting data and non act non

54

00:02:31,379 --> 00:02:35,818

active means that's no longer in

55

00:02:32,909 --> 00:02:38,370

operation it's archived and the next

56

00:02:35,818 --> 00:02:41,939

slide that I'll get to will expand upon

57

00:02:38,370 --> 00:02:43,370

that more and then another term and I've

58
00:02:41,939 --> 00:02:46,349
heard this term used today as well

59
00:02:43,370 --> 00:02:49,049
especially in the advent of Indian the

60
00:02:46,349 --> 00:02:55,250
internet and Google and Facebook where

61
00:02:49,049 --> 00:02:58,620
we're talking about lots of data and we

62
00:02:55,250 --> 00:03:01,409
describe big data there's commonly it's

63
00:02:58,620 --> 00:03:02,969
characterized by what is known as the

64
00:03:01,409 --> 00:03:04,769
three V's there's actually more of these

65
00:03:02,969 --> 00:03:07,919
but I here these are three common V's

66
00:03:04,769 --> 00:03:11,519
and that's high volume variety and

67
00:03:07,919 --> 00:03:14,068
velocity and when I talk about volume

68
00:03:11,519 --> 00:03:16,890
terms of computer memory size we're

69
00:03:14,068 --> 00:03:20,369
talking about a zettabyte now that's

70
00:03:16,889 --> 00:03:23,669
equivalent to one trillion gigabytes

71
00:03:20,370 --> 00:03:24,930
that's the warehouse of DVDs and that's

72
00:03:23,669 --> 00:03:27,539
that's especially if you're talking

73
00:03:24,930 --> 00:03:30,150
about trying to pull all this UFO data

74
00:03:27,539 --> 00:03:31,650
together or much of the key databases

75
00:03:30,150 --> 00:03:36,359
out there and trying to bring it

76
00:03:31,650 --> 00:03:39,120
together and work with it and also

77
00:03:36,359 --> 00:03:40,769
variety I'll talk about variety we're

78
00:03:39,120 --> 00:03:42,750
talking about different types of data

79
00:03:40,769 --> 00:03:47,009
some of it struck it can range from

80
00:03:42,750 --> 00:03:50,579
structured to unstructured an example of

81
00:03:47,009 --> 00:03:53,068
us structured data would be a relational

82
00:03:50,579 --> 00:03:54,150
database like maybe like a health

83
00:03:53,068 --> 00:03:55,828
insurance company working with

84
00:03:54,150 --> 00:03:58,889
electronic health records it's all

85
00:03:55,829 --> 00:04:03,389
organized and these nice neat rows and

86

00:03:58,889 --> 00:04:05,489
columns of data to some of your semi

87
00:04:03,389 --> 00:04:08,600
structured which is somewhat looser data

88
00:04:05,489 --> 00:04:10,920
but it might be like some of the

89
00:04:08,599 --> 00:04:13,469
hypertext markup language that's used

90
00:04:10,919 --> 00:04:15,179
for designing web pages runs that run

91
00:04:13,469 --> 00:04:19,079
the internet that's got some structure

92
00:04:15,180 --> 00:04:22,199
to it too to unstructured data just

93
00:04:19,079 --> 00:04:22,900
loose text document like this right here

94
00:04:22,199 --> 00:04:25,240
I mean

95
00:04:22,899 --> 00:04:28,149
I mean just the papers like some of the

96
00:04:25,240 --> 00:04:29,800
it's got some structure to it but if if

97
00:04:28,149 --> 00:04:32,560
you look at some of the UFO cases

98
00:04:29,800 --> 00:04:34,360
especially where people are reporting

99
00:04:32,560 --> 00:04:37,389
they're providing an textual narrative

100
00:04:34,360 --> 00:04:39,850

and a lot of times people are typing and

101

00:04:37,389 --> 00:04:42,160

there are there it's being recorded down

102

00:04:39,850 --> 00:04:44,350

and it's just really loose narrative

103

00:04:42,160 --> 00:04:48,639

it's practically like prose

104

00:04:44,350 --> 00:04:50,620

so it's really unstructured data and and

105

00:04:48,639 --> 00:04:55,000

we talk to another quality is high

106

00:04:50,620 --> 00:04:57,759

velocity big data is where the a batch

107

00:04:55,000 --> 00:04:59,800

of data can be processed in a matter of

108

00:04:57,759 --> 00:05:02,139

seconds or minutes or even fractions of

109

00:04:59,800 --> 00:05:05,740

a second depending on the system that's

110

00:05:02,139 --> 00:05:08,228

used and then also what Big Data you

111

00:05:05,740 --> 00:05:12,960

can't you can't effectively analyze it

112

00:05:08,228 --> 00:05:12,959

with traditional databases or methods so

113

00:05:13,379 --> 00:05:20,620

and here's a sample of some of key UFO

114

00:05:18,038 --> 00:05:22,930

data's based in the United States some

115
00:05:20,620 --> 00:05:25,389
of you may have heard of these like the

116
00:05:22,930 --> 00:05:30,240
mutual UFO network MUFON case management

117
00:05:25,389 --> 00:05:33,759
system which is active of course and

118
00:05:30,240 --> 00:05:35,728
it's got over 100,000 case files it's

119
00:05:33,759 --> 00:05:37,959
the the type of data it's structured

120
00:05:35,728 --> 00:05:41,259
combined with some semi structured data

121
00:05:37,959 --> 00:05:44,259
it's they're improving that they're

122
00:05:41,259 --> 00:05:45,750
making enhancements to it it's it's

123
00:05:44,259 --> 00:05:48,728
functions like a relational database

124
00:05:45,750 --> 00:05:51,478
since it's gotten into some semi

125
00:05:48,728 --> 00:05:57,158
structured data as textual data links to

126
00:05:51,478 --> 00:06:00,219
photographs even videos and then there's

127
00:05:57,158 --> 00:06:02,978
these other databases and national UFO

128
00:06:00,220 --> 00:06:04,840
Reporting Center which is somewhat

129
00:06:02,978 --> 00:06:07,418
similar to the MUFON one and then you

130
00:06:04,839 --> 00:06:10,779
got Project Blue Book which many of you

131
00:06:07,418 --> 00:06:15,448
may have heard of and that's that's an

132
00:06:10,779 --> 00:06:18,129
example of a non-active database and and

133
00:06:15,449 --> 00:06:20,590
that's a lot those files there they're

134
00:06:18,129 --> 00:06:25,300
even somewhat less structured than the

135
00:06:20,589 --> 00:06:27,189
other type of databases so and

136
00:06:25,300 --> 00:06:28,960
criticisms about the usefulness of the

137
00:06:27,189 --> 00:06:31,028
UFO data and I think this is a key thing

138
00:06:28,959 --> 00:06:32,168
that's a stumbling block in UFO ology or

139
00:06:31,028 --> 00:06:34,629
anybody that has an interest in

140
00:06:32,168 --> 00:06:36,699
researching it is is uh

141
00:06:34,629 --> 00:06:39,490
the analysis is limited

142
00:06:36,699 --> 00:06:41,170
case studies are descriptive statistics

143

00:06:39,490 --> 00:06:43,500
when I talk about limited to case

144
00:06:41,170 --> 00:06:46,080
studies it's usually telling a story

145
00:06:43,500 --> 00:06:49,060
this person saw such-and-such

146
00:06:46,079 --> 00:06:51,909
such-and-such object in the sky on such

147
00:06:49,060 --> 00:06:54,519
and such date like Roswell for instance

148
00:06:51,910 --> 00:06:57,370
the the UFO crash that's claimed to be

149
00:06:54,519 --> 00:07:01,839
at Roswell that's an example of a case

150
00:06:57,370 --> 00:07:04,629
study that's been told I hate I I'd say

151
00:07:01,839 --> 00:07:06,909
it's it runs the risk and this it's it's

152
00:07:04,629 --> 00:07:09,519
just my observation from from hearing it

153
00:07:06,910 --> 00:07:12,189
from other people especially those in

154
00:07:09,519 --> 00:07:15,609
the field of UFO ology or UFO research

155
00:07:12,189 --> 00:07:18,430
is that it's it's the risk of becoming

156
00:07:15,610 --> 00:07:19,990
Dade repetitious stated over time you

157
00:07:18,430 --> 00:07:22,540

got the recent one researcher over here

158

00:07:19,990 --> 00:07:24,790

at a conference or a place talking about

159

00:07:22,540 --> 00:07:26,290

this this case study and another one

160

00:07:24,790 --> 00:07:29,230

over here talking about the same case

161

00:07:26,290 --> 00:07:31,660

study so it's like what where we go from

162

00:07:29,230 --> 00:07:35,560

here we've heard this before what what

163

00:07:31,660 --> 00:07:38,530

new can we learn so and what what's

164

00:07:35,560 --> 00:07:42,009

lacking is inferential statistics where

165

00:07:38,529 --> 00:07:44,619

you can we we can actually start to

166

00:07:42,009 --> 00:07:48,459

visualize the data where you can act not

167

00:07:44,620 --> 00:07:50,560

not just read text narratives or looking

168

00:07:48,459 --> 00:07:53,799

just at photographs but you can start to

169

00:07:50,560 --> 00:07:57,459

visualize it you can you can start to

170

00:07:53,800 --> 00:08:00,910

run regression and comparing some means

171

00:07:57,459 --> 00:08:03,069

if you want to compare groups and other

172
00:08:00,910 --> 00:08:05,350
criticisms would be from renowned

173
00:08:03,069 --> 00:08:08,589
computer scientist and UFO researcher

174
00:08:05,350 --> 00:08:12,100
Jacque fillet and he talked about in

175
00:08:08,589 --> 00:08:14,169
debt inadequate data validation no

176
00:08:12,100 --> 00:08:16,120
really missing standards not to say

177
00:08:14,170 --> 00:08:18,670
there's no standards but standards as

178
00:08:16,120 --> 00:08:23,399
far as validating the data before you

179
00:08:18,670 --> 00:08:25,660
before you analyze it there's really no

180
00:08:23,399 --> 00:08:28,089
Universal or formal set of standards

181
00:08:25,660 --> 00:08:28,660
really established out there that I'm

182
00:08:28,089 --> 00:08:30,789
aware of

183
00:08:28,660 --> 00:08:33,250
well according what he's saying as well

184
00:08:30,790 --> 00:08:35,080
and it's a very complex phenomenon I

185
00:08:33,250 --> 00:08:38,679
mean if you you if you have a somebody

186
00:08:35,080 --> 00:08:41,889
saying they've seen something pop out of

187
00:08:38,679 --> 00:08:44,379
midair and vanish D materialized in

188
00:08:41,889 --> 00:08:46,360
midair that's hard to that's hard to

189
00:08:44,379 --> 00:08:48,340
physically gauge you know that's not

190
00:08:46,360 --> 00:08:50,320
following classic physics you know

191
00:08:48,340 --> 00:08:52,990
velocity acceleration that's

192
00:08:50,320 --> 00:08:55,390
so it's that and there's also a lack of

193
00:08:52,990 --> 00:08:57,759
data exchange among many UFO research

194
00:08:55,389 --> 00:09:01,539
groups so that that's also another

195
00:08:57,759 --> 00:09:03,789
challenge so a proposed method to

196
00:09:01,539 --> 00:09:05,529
address this is using what's called a

197
00:09:03,789 --> 00:09:07,209
somebody name may be familiar with this

198
00:09:05,529 --> 00:09:09,250
especially if you're you know working in

199
00:09:07,210 --> 00:09:11,200
data science and so forth is what's

200

00:09:09,250 --> 00:09:13,870
called a data warehouse and that's where

201
00:09:11,200 --> 00:09:17,520
you can pull all these various UFO data

202
00:09:13,870 --> 00:09:19,480
sources together even across various

203
00:09:17,519 --> 00:09:23,279
different structures of data whether

204
00:09:19,480 --> 00:09:25,930
it's photographs radar data recordings

205
00:09:23,279 --> 00:09:29,230
it's I don't see it up there but as far

206
00:09:25,929 --> 00:09:32,349
as even text data if you have documents

207
00:09:29,230 --> 00:09:34,120
that you want to put into the data

208
00:09:32,350 --> 00:09:36,129
warehouse and then you can catalogue it

209
00:09:34,120 --> 00:09:37,690
you can catalogue it by physical

210
00:09:36,129 --> 00:09:39,970
characteristics of the object or

211
00:09:37,690 --> 00:09:44,070
phenomena you can also characterize it

212
00:09:39,970 --> 00:09:44,070
by parapsychological or psychic

213
00:09:45,899 --> 00:09:51,789
characteristics especially upon the

214
00:09:47,889 --> 00:09:54,370

observer so I just is real quick here

215

00:09:51,789 --> 00:09:56,649

just like 15 seconds sometimes about up

216

00:09:54,370 --> 00:09:59,230

but another thing is the with the data

217

00:09:56,649 --> 00:10:02,949

warehouses come about is using a a

218

00:09:59,230 --> 00:10:04,960

software platform like I do which is you

219

00:10:02,950 --> 00:10:06,430

can help it's especially useful for

220

00:10:04,960 --> 00:10:08,940

working with structured data and you can

221

00:10:06,429 --> 00:10:11,500

use that with the data warehouse and

222

00:10:08,940 --> 00:10:13,810

with the catalog with the with the

223

00:10:11,500 --> 00:10:16,480

analytics sandbox you can do all the

224

00:10:13,809 --> 00:10:20,469

descriptive and inferential statistics

225

00:10:16,480 --> 00:10:27,570

so and that's concludes my presentation

226

00:10:20,470 --> 00:10:31,290

so thank you for your interesting

227

00:10:27,570 --> 00:10:31,290

presentations are the questions

228

00:10:34,570 --> 00:10:39,649

yeah thank you very much for an

229

00:10:36,649 --> 00:10:42,559

interesting talk there's a ton of data

230

00:10:39,649 --> 00:10:45,589

obviously and I'm I think you can

231

00:10:42,559 --> 00:10:47,419

certainly analyze it and benefit

232

00:10:45,589 --> 00:10:48,949

somewhat but seems to me from what I've

233

00:10:47,419 --> 00:10:51,469

heard it's the quality of the data

234

00:10:48,948 --> 00:10:53,809

that's the problem so how do you get

235

00:10:51,470 --> 00:10:56,839

quality data we have some radar data you

236

00:10:53,809 --> 00:10:58,429

mentioned some psychological data I

237

00:10:56,839 --> 00:11:01,399

think one of those that would be very

238

00:10:58,429 --> 00:11:03,828

interesting but what's being talked

239

00:11:01,399 --> 00:11:05,899

about is like spectrogram the problem is

240

00:11:03,828 --> 00:11:07,578

if you have specialized equipment where

241

00:11:05,899 --> 00:11:10,698

do you put it how do you know when a UFO

242

00:11:07,578 --> 00:11:12,769

is going to appear and so maybe by big

243
00:11:10,698 --> 00:11:14,958
data you can analyze if it there truly

244
00:11:12,769 --> 00:11:18,919
are some hotspots where you might put

245
00:11:14,958 --> 00:11:20,958
rather expensive equipment that can give

246
00:11:18,919 --> 00:11:25,818
you the kind of quality you need to do a

247
00:11:20,958 --> 00:11:28,309
much more thorough analysis yeah yeah

248
00:11:25,818 --> 00:11:29,539
it's also not just the data itself but

249
00:11:28,309 --> 00:11:33,528
the instrumentation that could that

250
00:11:29,539 --> 00:11:36,009
could be a key factor so any other

251
00:11:33,528 --> 00:11:36,009
questions

252
00:11:39,179 --> 00:11:43,979
yeah Russ I'm wondering do you have

253
00:11:41,519 --> 00:11:46,139
you've thought about this a lot what

254
00:11:43,980 --> 00:11:48,509
about the cases where it's inferred that

255
00:11:46,139 --> 00:11:50,759
there's something like a UFO there but

256
00:11:48,509 --> 00:11:53,610
you don't have a direct sighting you

257

00:11:50,759 --> 00:11:55,379
know people who've reported you know

258
00:11:53,610 --> 00:11:56,850
lights coming down on their car the car

259
00:11:55,379 --> 00:11:59,939
stalls and all sudden they're missing

260
00:11:56,850 --> 00:12:01,259
four hours of time or the vehicle is 20

261
00:11:59,940 --> 00:12:03,390
miles away from where it should have

262
00:12:01,259 --> 00:12:04,860
been does that isn't part of the problem

263
00:12:03,389 --> 00:12:07,439
here that we don't even know what

264
00:12:04,860 --> 00:12:11,070
exactly data is because we don't even

265
00:12:07,440 --> 00:12:12,720
know the range of the phenomena what

266
00:12:11,070 --> 00:12:15,030
phenomena we don't know the full range

267
00:12:12,720 --> 00:12:16,649
of the phenomena I mean we know what an

268
00:12:15,029 --> 00:12:18,240
object is that we can't describe but

269
00:12:16,649 --> 00:12:21,149
what about encounters where there's a

270
00:12:18,240 --> 00:12:22,919
lot of missing information yeah and it's

271
00:12:21,149 --> 00:12:25,709

really at the border of what we even

272

00:12:22,919 --> 00:12:27,929

would you know right there to be data

273

00:12:25,710 --> 00:12:31,139

just doesn't fit that that's where you

274

00:12:27,929 --> 00:12:33,929

can look at definitions how how the data

275

00:12:31,139 --> 00:12:35,519

definitions are operationalized and and

276

00:12:33,929 --> 00:12:38,279

how you put like what you're talking

277

00:12:35,519 --> 00:12:41,129

about you you can have a catalog for

278

00:12:38,279 --> 00:12:42,720

that kind of data you know especially if

279

00:12:41,129 --> 00:12:45,000

it needs more validation you need to

280

00:12:42,720 --> 00:12:47,759

follow up and do an additional follow-up

281

00:12:45,000 --> 00:12:50,100

to clarify that special case like that

282

00:12:47,759 --> 00:12:54,210

or it's more more unique physical

283

00:12:50,100 --> 00:12:55,769

phenomena psychic phenomena hi you know

284

00:12:54,210 --> 00:12:56,940

I'm really familiar with the concepts of

285

00:12:55,769 --> 00:12:58,350

data warehouse and I think it's a good

286
00:12:56,940 --> 00:13:01,530
idea to try to bring your data together

287
00:12:58,350 --> 00:13:03,690
and make it accessible all in one place

288
00:13:01,529 --> 00:13:05,850
as someone was saying about cleaning up

289
00:13:03,690 --> 00:13:07,350
the data it's a really good good point

290
00:13:05,850 --> 00:13:09,629
one of the things that you didn't

291
00:13:07,350 --> 00:13:12,210
mention at all is metadata and adding

292
00:13:09,629 --> 00:13:14,340
additional fields for example like who

293
00:13:12,210 --> 00:13:16,050
collected the data or what type of

294
00:13:14,340 --> 00:13:18,180
instrumentation was used to collect the

295
00:13:16,049 --> 00:13:20,039
data so then you can query across your

296
00:13:18,179 --> 00:13:21,870
entire database and see if you get

297
00:13:20,039 --> 00:13:24,389
consistent results from a single person

298
00:13:21,870 --> 00:13:26,009
or using a certain type of equipment and

299
00:13:24,389 --> 00:13:27,600
that can give you more information that

300

00:13:26,009 --> 00:13:29,669

can help to make your data stronger and

301

00:13:27,600 --> 00:13:34,830

easier to analyze I agree that's a valid

302

00:13:29,669 --> 00:13:38,360

point so thank you very much you're

303

00:13:34,830 --> 00:13:44,830

interesting that's recitation

304

00:13:38,360 --> 00:13:44,830

[Applause]