



perceiving the future



1
00:00:04,550 --> 00:00:02,540
good evening they thank you everybody

2
00:00:07,490 --> 00:00:04,560
for being here especially with the room

3
00:00:10,850 --> 00:00:07,500
being hot so and I like to thank the

4
00:00:15,259 --> 00:00:10,860
society for allowing me to present here

5
00:00:17,330 --> 00:00:15,269
at the conference this evening and I'm a

6
00:00:21,200 --> 00:00:17,340
member of the Society for scientific

7
00:00:24,170 --> 00:00:21,210
exploration and I wanted to talk about

8
00:00:29,060 --> 00:00:24,180
making UFO data more useful for

9
00:00:30,589 --> 00:00:29,070
scientific research and I think and this

10
00:00:39,280 --> 00:00:30,599
could apply to other fields as well

11
00:00:42,760 --> 00:00:39,290
including parapsychology so I think this

12
00:00:46,990 --> 00:00:42,770
symbolizes kind of where the state of

13
00:00:48,830 --> 00:00:47,000

working with UFO data is at the moment

14

00:00:49,880 --> 00:00:48,840

especially for you those in the back of

15

00:00:52,250 --> 00:00:49,890

may not if you can't see it we were

16

00:00:55,010 --> 00:00:52,260

clearly enough it's these alien

17

00:00:56,420 --> 00:00:55,020

spaceship is landed it to aliens get out

18

00:00:58,520 --> 00:00:56,430

once that's the other see I told you

19

00:01:00,349 --> 00:00:58,530

they wouldn't notice this you have all

20

00:01:01,880 --> 00:01:00,359

the there are things got their heads

21

00:01:05,750 --> 00:01:01,890

buried in their smartphones they can't

22

00:01:08,090 --> 00:01:05,760

see them so I think I think that's with

23

00:01:10,609 --> 00:01:08,100

the case in a symbolic way with the UFO

24

00:01:13,700 --> 00:01:10,619

data well there's some things we cannot

25

00:01:16,429 --> 00:01:13,710

see with it because of just the the

26

00:01:20,120 --> 00:01:16,439

nature of the data and also there's just

27

00:01:22,460 --> 00:01:20,130

only fairly recently tools that we have

28

00:01:24,320 --> 00:01:22,470

coming online in which we can address

29

00:01:29,719 --> 00:01:24,330

that but this we just haven't really

30

00:01:31,190 --> 00:01:29,729

been able to employ those tools and what

31

00:01:33,170 --> 00:01:31,200

I'm going to do is just a previews I'm

32

00:01:36,380 --> 00:01:33,180

going to just give a background on UFO

33

00:01:38,420 --> 00:01:36,390

data and then discuss some criticism

34

00:01:40,999 --> 00:01:38,430

about it about it in regard to its

35

00:01:43,580 --> 00:01:41,009

scientific usefulness and also a

36

00:01:45,859 --> 00:01:43,590

proposed method for enhancing it to

37

00:01:47,929 --> 00:01:45,869

dress those criticisms and then going

38

00:01:52,069 --> 00:01:47,939

into a conclusion and then question and

39

00:01:55,219 --> 00:01:52,079

answers so as far as the background

40

00:01:57,770 --> 00:01:55,229

about UFO data some key characteristics

41

00:02:00,620 --> 00:01:57,780

is its there's desparate sources I mean

42

00:02:03,850 --> 00:02:00,630

there's collections of UFO data all over

43

00:02:08,719 --> 00:02:03,860

the world here in the United States and

44

00:02:10,430 --> 00:02:08,729

Britain Europe Asia there's all kind all

45

00:02:13,540 --> 00:02:10,440

sorts all kinds of sources

46

00:02:17,720 --> 00:02:16,070

private researchers non-private

47

00:02:19,699 --> 00:02:17,730

nongovernmental organizations or

48

00:02:21,920 --> 00:02:19,709

non-profit organizations research groups

49

00:02:25,570 --> 00:02:21,930

there's all kinds of sources out there

50

00:02:28,880 --> 00:02:25,580

and also there's active and non-active

51
00:02:31,370 --> 00:02:28,890
collections of data active as in they're

52
00:02:32,900 --> 00:02:31,380
currently accepting data and non act non

53
00:02:35,809 --> 00:02:32,910
active means that's no longer in

54
00:02:38,360 --> 00:02:35,819
operation it's archived and the next

55
00:02:41,930 --> 00:02:38,370
slide that I'll get to will expand upon

56
00:02:43,360 --> 00:02:41,940
that more and then another term and I've

57
00:02:46,340 --> 00:02:43,370
heard this term used today as well

58
00:02:49,039 --> 00:02:46,350
especially in the advent of Indian the

59
00:02:55,240 --> 00:02:49,049
internet and Google and Facebook where

60
00:02:58,610 --> 00:02:55,250
we're talking about lots of data and we

61
00:03:01,400 --> 00:02:58,620
describe big data there's commonly it's

62
00:03:02,960 --> 00:03:01,410
characterized by what is known as the

63
00:03:04,759 --> 00:03:02,970

three V's there's actually more of these

64

00:03:07,910 --> 00:03:04,769

but I here these are three common V's

65

00:03:11,509 --> 00:03:07,920

and that's high volume variety and

66

00:03:14,059 --> 00:03:11,519

velocity and when I talk about volume

67

00:03:16,880 --> 00:03:14,069

terms of computer memory size we're

68

00:03:20,360 --> 00:03:16,890

talking about a zettabyte now that's

69

00:03:23,660 --> 00:03:20,370

equivalent to one trillion gigabytes

70

00:03:24,920 --> 00:03:23,670

that's the warehouse of DVDs and that's

71

00:03:27,530 --> 00:03:24,930

that's especially if you're talking

72

00:03:30,140 --> 00:03:27,540

about trying to pull all this UFO data

73

00:03:31,640 --> 00:03:30,150

together or much of the key databases

74

00:03:36,349 --> 00:03:31,650

out there and trying to bring it

75

00:03:39,110 --> 00:03:36,359

together and work with it and also

76

00:03:40,759 --> 00:03:39,120

variety I'll talk about variety we're

77

00:03:42,740 --> 00:03:40,769

talking about different types of data

78

00:03:47,000 --> 00:03:42,750

some of it struck it can range from

79

00:03:50,569 --> 00:03:47,010

structured to unstructured an example of

80

00:03:53,059 --> 00:03:50,579

us structured data would be a relational

81

00:03:54,140 --> 00:03:53,069

database like maybe like a health

82

00:03:55,819 --> 00:03:54,150

insurance company working with

83

00:03:58,879 --> 00:03:55,829

electronic health records it's all

84

00:04:03,379 --> 00:03:58,889

organized and these nice neat rows and

85

00:04:05,479 --> 00:04:03,389

columns of data to some of your semi

86

00:04:08,590 --> 00:04:05,489

structured which is somewhat looser data

87

00:04:10,910 --> 00:04:08,600

but it might be like some of the

88

00:04:13,460 --> 00:04:10,920

hypertext markup language that's used

89

00:04:15,170 --> 00:04:13,470

for designing web pages runs that run

90

00:04:19,069 --> 00:04:15,180

the internet that's got some structure

91

00:04:22,190 --> 00:04:19,079

to it too to unstructured data just

92

00:04:22,890 --> 00:04:22,200

loose text document like this right here

93

00:04:28,140 --> 00:04:22,900

I mean

94

00:04:29,790 --> 00:04:28,150

it's got some structure to it but if if

95

00:04:32,550 --> 00:04:29,800

you look at some of the UFO cases

96

00:04:34,350 --> 00:04:32,560

especially where people are reporting

97

00:04:37,379 --> 00:04:34,360

they're providing an textual narrative

98

00:04:39,840 --> 00:04:37,389

and a lot of times people are typing and

99

00:04:42,150 --> 00:04:39,850

there are there it's being recorded down

100

00:04:44,340 --> 00:04:42,160

and it's just really loose narrative

101
00:04:48,629 --> 00:04:44,350
it's practically like prose

102
00:04:50,610 --> 00:04:48,639
so it's really unstructured data and and

103
00:04:54,990 --> 00:04:50,620
we talk to another quality is high

104
00:04:57,750 --> 00:04:55,000
velocity big data is where the a batch

105
00:04:59,790 --> 00:04:57,760
of data can be processed in a matter of

106
00:05:02,129 --> 00:04:59,800
seconds or minutes or even fractions of

107
00:05:05,730 --> 00:05:02,139
a second depending on the system that's

108
00:05:08,219 --> 00:05:05,740
used and then also what Big Data you

109
00:05:13,370 --> 00:05:08,229
can't you can't effectively analyze it

110
00:05:20,610 --> 00:05:18,029
and here's a sample of some of key UFO

111
00:05:22,920 --> 00:05:20,620
data's based in the United States some

112
00:05:25,379 --> 00:05:22,930
of you may have heard of these like the

113
00:05:30,230 --> 00:05:25,389

mutual UFO network MUFON case management

114

00:05:33,750 --> 00:05:30,240

system which is active of course and

115

00:05:35,719 --> 00:05:33,760

it's got over 100,000 case files it's

116

00:05:37,950 --> 00:05:35,729

the the type of data it's structured

117

00:05:41,250 --> 00:05:37,960

combined with some semi structured data

118

00:05:44,250 --> 00:05:41,260

it's they're improving that they're

119

00:05:45,740 --> 00:05:44,260

making enhancements to it it's it's

120

00:05:48,719 --> 00:05:45,750

functions like a relational database

121

00:05:51,469 --> 00:05:48,729

since it's gotten into some semi

122

00:05:57,149 --> 00:05:51,479

structured data as textual data links to

123

00:06:00,210 --> 00:05:57,159

photographs even videos and then there's

124

00:06:02,969 --> 00:06:00,220

these other databases and national UFO

125

00:06:04,830 --> 00:06:02,979

Reporting Center which is somewhat

126

00:06:07,409 --> 00:06:04,840

similar to the MUFON one and then you

127

00:06:10,770 --> 00:06:07,419

got Project Blue Book which many of you

128

00:06:15,439 --> 00:06:10,780

may have heard of and that's that's an

129

00:06:18,120 --> 00:06:15,449

example of a non-active database and and

130

00:06:20,580 --> 00:06:18,130

that's a lot those files there they're

131

00:06:25,290 --> 00:06:20,590

even somewhat less structured than the

132

00:06:27,180 --> 00:06:25,300

other type of databases so and

133

00:06:28,950 --> 00:06:27,190

criticisms about the usefulness of the

134

00:06:31,019 --> 00:06:28,960

UFO data and I think this is a key thing

135

00:06:32,159 --> 00:06:31,029

that's a stumbling block in UFO ology or

136

00:06:34,620 --> 00:06:32,169

anybody that has an interest in

137

00:06:36,690 --> 00:06:34,630

researching it is is uh

138

00:06:39,480 --> 00:06:36,700

the analysis is limited

139

00:06:41,160 --> 00:06:39,490

case studies are descriptive statistics

140

00:06:43,490 --> 00:06:41,170

when I talk about limited to case

141

00:06:46,070 --> 00:06:43,500

studies it's usually telling a story

142

00:06:49,050 --> 00:06:46,080

this person saw such-and-such

143

00:06:51,900 --> 00:06:49,060

such-and-such object in the sky on such

144

00:06:54,510 --> 00:06:51,910

and such date like Roswell for instance

145

00:06:57,360 --> 00:06:54,520

the the UFO crash that's claimed to be

146

00:07:01,830 --> 00:06:57,370

at Roswell that's an example of a case

147

00:07:04,620 --> 00:07:01,840

study that's been told I hate I'd say

148

00:07:06,900 --> 00:07:04,630

it's it runs the risk and this it's it's

149

00:07:09,510 --> 00:07:06,910

just my observation from from hearing it

150

00:07:12,180 --> 00:07:09,520

from other people especially those in

151
00:07:15,600 --> 00:07:12,190
the field of UFO ology or UFO research

152
00:07:18,420 --> 00:07:15,610
is that it's it's the risk of becoming

153
00:07:19,980 --> 00:07:18,430
Dade repetitious stated over time you

154
00:07:22,530 --> 00:07:19,990
got the recent one researcher over here

155
00:07:24,780 --> 00:07:22,540
at a conference or a place talking about

156
00:07:26,280 --> 00:07:24,790
this this case study and another one

157
00:07:29,220 --> 00:07:26,290
over here talking about the same case

158
00:07:31,650 --> 00:07:29,230
study so it's like what where we go from

159
00:07:35,550 --> 00:07:31,660
here we've heard this before what what

160
00:07:38,520 --> 00:07:35,560
new can we learn so and what what's

161
00:07:42,000 --> 00:07:38,530
lacking is inferential statistics where

162
00:07:44,610 --> 00:07:42,010
you can we we can actually start to

163
00:07:48,450 --> 00:07:44,620

visualize the data where you can act not

164

00:07:50,550 --> 00:07:48,460

not just read text narratives or looking

165

00:07:53,790 --> 00:07:50,560

just at photographs but you can start to

166

00:07:57,450 --> 00:07:53,800

visualize it you can you can start to

167

00:08:00,900 --> 00:07:57,460

run regression and comparing some means

168

00:08:03,060 --> 00:08:00,910

if you want to compare groups and other

169

00:08:05,340 --> 00:08:03,070

criticisms would be from renowned

170

00:08:08,580 --> 00:08:05,350

computer scientist and UFO researcher

171

00:08:12,090 --> 00:08:08,590

Jacque fillet and he talked about in

172

00:08:14,160 --> 00:08:12,100

debt inadequate data validation no

173

00:08:16,110 --> 00:08:14,170

really missing standards not to say

174

00:08:18,660 --> 00:08:16,120

there's no standards but standards as

175

00:08:23,390 --> 00:08:18,670

far as validating the data before you

176

00:08:25,650 --> 00:08:23,400

before you analyze it there's really no

177

00:08:28,080 --> 00:08:25,660

Universal or formal set of standards

178

00:08:28,650 --> 00:08:28,090

really established out there that I'm

179

00:08:30,780 --> 00:08:28,660

aware of

180

00:08:33,240 --> 00:08:30,790

well according what he's saying as well

181

00:08:35,070 --> 00:08:33,250

and it's a very complex phenomenon I

182

00:08:38,670 --> 00:08:35,080

mean if you you if you have a somebody

183

00:08:41,880 --> 00:08:38,680

saying they've seen something pop out of

184

00:08:44,370 --> 00:08:41,890

midair and vanish D materialized in

185

00:08:46,350 --> 00:08:44,380

midair that's hard to that's hard to

186

00:08:48,330 --> 00:08:46,360

physically gauge you know that's not

187

00:08:50,310 --> 00:08:48,340

following classic physics you know

188

00:08:52,980 --> 00:08:50,320

velocity acceleration that's

189

00:08:55,380 --> 00:08:52,990

so it's that and there's also a lack of

190

00:08:57,750 --> 00:08:55,390

data exchange among many UFO research

191

00:09:01,530 --> 00:08:57,760

groups so that that's also another

192

00:09:03,780 --> 00:09:01,540

challenge so a proposed method to

193

00:09:05,520 --> 00:09:03,790

address this is using what's called a

194

00:09:07,200 --> 00:09:05,530

somebody name may be familiar with this

195

00:09:09,240 --> 00:09:07,210

especially if you're you know working in

196

00:09:11,190 --> 00:09:09,250

data science and so forth is what's

197

00:09:13,860 --> 00:09:11,200

called a data warehouse and that's where

198

00:09:17,510 --> 00:09:13,870

you can pull all these various UFO data

199

00:09:19,470 --> 00:09:17,520

sources together even across various

200

00:09:23,270 --> 00:09:19,480

different structures of data whether

201
00:09:25,920 --> 00:09:23,280
it's photographs radar data recordings

202
00:09:29,220 --> 00:09:25,930
it's I don't see it up there but as far

203
00:09:32,340 --> 00:09:29,230
as even text data if you have documents

204
00:09:34,110 --> 00:09:32,350
that you want to put into the data

205
00:09:36,120 --> 00:09:34,120
warehouse and then you can catalogue it

206
00:09:37,680 --> 00:09:36,130
you can catalogue it by physical

207
00:09:39,960 --> 00:09:37,690
characteristics of the object or

208
00:09:45,890 --> 00:09:39,970
phenomena you can also characterize it

209
00:09:51,780 --> 00:09:47,880
characteristics especially upon the

210
00:09:54,360 --> 00:09:51,790
observer so I just is real quick here

211
00:09:56,640 --> 00:09:54,370
just like 15 seconds sometimes about up

212
00:09:59,220 --> 00:09:56,650
but another thing is the with the data

213
00:10:02,940 --> 00:09:59,230

warehouses come about is using a a

214

00:10:04,950 --> 00:10:02,950

software platform like I do which is you

215

00:10:06,420 --> 00:10:04,960

can help it's especially useful for

216

00:10:08,930 --> 00:10:06,430

working with structured data and you can

217

00:10:11,490 --> 00:10:08,940

use that with the data warehouse and

218

00:10:13,800 --> 00:10:11,500

with the catalog with the with the

219

00:10:16,470 --> 00:10:13,810

analytics sandbox you can do all the

220

00:10:20,460 --> 00:10:16,480

descriptive and inferential statistics

221

00:10:27,560 --> 00:10:20,470

so and that's concludes my presentation

222

00:10:34,560 --> 00:10:27,570

so thank you for your interesting

223

00:10:39,639 --> 00:10:36,639

yeah thank you very much for an

224

00:10:42,550 --> 00:10:39,649

interesting talk there's a ton of data

225

00:10:45,579 --> 00:10:42,560

obviously and I'm I think you can

226

00:10:47,410 --> 00:10:45,589

certainly analyze it and benefit

227

00:10:48,939 --> 00:10:47,420

somewhat but seems to me from what I've

228

00:10:51,460 --> 00:10:48,949

heard it's the quality of the data

229

00:10:53,800 --> 00:10:51,470

that's the problem so how do you get

230

00:10:56,829 --> 00:10:53,810

quality data we have some radar data you

231

00:10:58,420 --> 00:10:56,839

mentioned some psychological data I

232

00:11:01,389 --> 00:10:58,430

think one of those that would be very

233

00:11:03,819 --> 00:11:01,399

interesting but what's being talked

234

00:11:05,889 --> 00:11:03,829

about is like spectrogram the problem is

235

00:11:07,569 --> 00:11:05,899

if you have specialized equipment where

236

00:11:10,689 --> 00:11:07,579

do you put it how do you know when a UFO

237

00:11:12,759 --> 00:11:10,699

is going to appear and so maybe by big

238

00:11:14,949 --> 00:11:12,769

data you can analyze if it there truly

239

00:11:18,910 --> 00:11:14,959

are some hotspots where you might put

240

00:11:20,949 --> 00:11:18,920

rather expensive equipment that can give

241

00:11:25,809 --> 00:11:20,959

you the kind of quality you need to do a

242

00:11:28,300 --> 00:11:25,819

much more thorough analysis yeah yeah

243

00:11:29,530 --> 00:11:28,310

it's also not just the data itself but

244

00:11:33,519 --> 00:11:29,540

the instrumentation that could that

245

00:11:39,170 --> 00:11:33,529

could be a key factor so any other

246

00:11:43,970 --> 00:11:41,510

yeah Russ I'm wondering do you have

247

00:11:46,130 --> 00:11:43,980

you've thought about this a lot what

248

00:11:48,500 --> 00:11:46,140

about the cases where it's inferred that

249

00:11:50,750 --> 00:11:48,510

there's something like a UFO there but

250

00:11:53,600 --> 00:11:50,760

you don't have a direct sighting you

251
00:11:55,370 --> 00:11:53,610
know people who've reported you know

252
00:11:56,840 --> 00:11:55,380
lights coming down on their car the car

253
00:11:59,930 --> 00:11:56,850
stalls and all sudden they're missing

254
00:12:01,250 --> 00:11:59,940
four hours of time or the vehicle is 20

255
00:12:03,380 --> 00:12:01,260
miles away from where it should have

256
00:12:04,850 --> 00:12:03,390
been does that isn't part of the problem

257
00:12:07,430 --> 00:12:04,860
here that we don't even know what

258
00:12:11,060 --> 00:12:07,440
exactly data is because we don't even

259
00:12:12,710 --> 00:12:11,070
know the range of the phenomena what

260
00:12:15,020 --> 00:12:12,720
phenomena we don't know the full range

261
00:12:16,640 --> 00:12:15,030
of the phenomena I mean we know what an

262
00:12:18,230 --> 00:12:16,650
object is that we can't describe but

263
00:12:21,140 --> 00:12:18,240

what about encounters where there's a

264

00:12:22,910 --> 00:12:21,150

lot of missing information yeah and it's

265

00:12:25,700 --> 00:12:22,920

really at the border of what we even

266

00:12:27,920 --> 00:12:25,710

would you know right there to be data

267

00:12:31,130 --> 00:12:27,930

just doesn't fit that that's where you

268

00:12:33,920 --> 00:12:31,140

can look at definitions how how the data

269

00:12:35,510 --> 00:12:33,930

definitions are operationalized and and

270

00:12:38,270 --> 00:12:35,520

how you put like what you're talking

271

00:12:41,120 --> 00:12:38,280

about you you can have a catalog for

272

00:12:42,710 --> 00:12:41,130

that kind of data you know especially if

273

00:12:44,990 --> 00:12:42,720

it needs more validation you need to

274

00:12:47,750 --> 00:12:45,000

follow up and do an additional follow-up

275

00:12:50,090 --> 00:12:47,760

to clarify that special case like that

276

00:12:54,200 --> 00:12:50,100

or it's more more unique physical

277

00:12:55,760 --> 00:12:54,210

phenomena psychic phenomena hi you know

278

00:12:56,930 --> 00:12:55,770

I'm really familiar with the concepts of

279

00:12:58,340 --> 00:12:56,940

data warehouse and I think it's a good

280

00:13:01,520 --> 00:12:58,350

idea to try to bring your data together

281

00:13:03,680 --> 00:13:01,530

and make it accessible all in one place

282

00:13:05,840 --> 00:13:03,690

as someone was saying about cleaning up

283

00:13:07,340 --> 00:13:05,850

the data it's a really good good point

284

00:13:09,620 --> 00:13:07,350

one of the things that you didn't

285

00:13:12,200 --> 00:13:09,630

mention at all is metadata and adding

286

00:13:14,330 --> 00:13:12,210

additional fields for example like who

287

00:13:16,040 --> 00:13:14,340

collected the data or what type of

288

00:13:18,170 --> 00:13:16,050

instrumentation was used to collect the

289

00:13:20,030 --> 00:13:18,180

data so then you can query across your

290

00:13:21,860 --> 00:13:20,040

entire database and see if you get

291

00:13:24,380 --> 00:13:21,870

consistent results from a single person

292

00:13:26,000 --> 00:13:24,390

or using a certain type of equipment and

293

00:13:27,590 --> 00:13:26,010

that can give you more information that

294

00:13:29,660 --> 00:13:27,600

can help to make your data stronger and

295

00:13:34,820 --> 00:13:29,670

easier to analyze I agree that's a valid

296

00:13:38,350 --> 00:13:34,830

point so thank you very much you're