SCSC Data Safety Initiative – WG Meeting 54

2nd September 2020, Teams

Minutes

Attendees

Mike Parsons (MP) – CGI, Paul Hampton (PH) – CGI, Alastair Faulkner (AF) – Abbeymeade, Divya Atkins (DA) – MCA, Martin Atkins (MA) – MCA, Paul McKernan (PMK) - DSTL, Dave Banham (DB) – Blackberry, Dale Callicott (DC) – BAE, David Hand (DH) – Imperial College, Paul Mayo (PM) – SQEP Ltd, Jim Mateer (JM) – SQEP Ltd.

Apologies

Louise Harney (LH) – Leonardo, Mike Ainsworth (MA) – Ricardo, Any Williams (AW) – NewTechNo, Ali Hessami (AH) – Vega, Paul Ensor (PE) – Boeing, Paolo Giuliani (PG) – EDF, Janette Baldwin (JB) – Thales, Tim Rowe (TR) – Consultant, Bill Blackburn (BB) – Process Renewal, Brent Kimberley (BK) - Durham

Agenda

- COVID-19
- Dark Data Talk – David J. Hand
- Alastair's Book - hardcopy
- Update on Guidance Document v3.2 (download/sales)
- Update on SSS'21
- Aim for 2021
- Future Events
- Minutes and action status
- AOB, etc.
- Data Safety in the News
- Next Meeting

NOTE: All comments or opinions in these notes are attributed only to individual attendees of the meeting, not to their respective organisations.

*[Note that actions are presented in the form **N.Mx** where **N** is the meeting number, **M** a reference number for the action raised in that meeting and **x** is an optional letter that differentiates related actions arising from the same discussion point].*

The meeting slides are available at [1].

1. COVID-19

MP discussed the Covid-19 crisis and the volume of data that is being published. He thought the data being published has been misleading at times and has not provided a complete picture. The selected

data that has been shared with the public is very poor, as illustrated from the material published on the government website https://coronavirus.data.gov.uk/. MP also noted that infection rates seem to be increasing but deaths are not. PM said there is a particular issue with Covid-related deaths where individuals had other underlying health issues; there is difficulty in assessing whether the death can actually be ascribed to Covid or not, which could end up being a highly subjective judgement. AJ said the reduced number of deaths may be because we understand the disease a lot better now and so treatments are getting better.

2. Dark Data Talk – David J. Hand

DH gave a presentation [2] on Dark Data – based on his publication: *Dark Data - why what you don't know matters* [3]. He noted that the Data Safety Guidance (DSG) has some similarities with his work but there are some differences: he said his main objective is to extract understanding from data whereas the DSIWG take a more systems approach in DSG. However, both recognise how important data is and his work on Dark Data argues that missing data can be even more important than the data you do have.

DH provided examples of Dark Data, such as "elephant powder", measles eradication, bullet holes in aircraft and Hurricane Sandy where insight is gained from the absence of data. Having large data sets is also not a solution as Big Data has all the same associated problems as smaller data sets.

DH provided a specific example of where Dark Data contributed to the 1986 Challenger Space Shuttle Disaster relating to the failure of a seal (O-ring), which caused a catastrophic failure of the vehicle. DH said there was pressure on the organisation to launch the night before. A graph with previous O-ring failures was presented, which seemingly disconnected the relationship between temperature and O-ring problems. DH noted that it was odd that there were no entries with zero values for previous flights, and when these data points were added, it was clear there was a direct relation between failure and cold temperatures.

DH gave another example relating to the comparison of death rates between Japan and US, where there is a much higher rate of heart disease in the US compared to Japan, but the percentage of individuals with risk factors such as smoking are actually higher in Japan than in US. He said this could be diet-related but DH noted that the data may be misreported due to cultural issues: strokes are considered more culturally acceptable than heart attacks and this theory is supported by the much higher reported stroke rates in Japan compared to the US.

DH went through the 15 different types of Dark Data types and noted a number of pitfalls with the interpretation of data such as Simpson's Paradox, the Reproducibly Crisis and bias arising from data-driven (rather than theoretical) models.

DH said the solution is to "Shine a Light" and "Use what you do know about what you don't know". This involved the use of specific strategies for dealing with dark data. For example, by restricting access to data it can allow more faithful reporting of other data such as in controlled trials (blinding). It is also possible to generate new data points that could have arisen using techniques such as simulation, smoothing, bootstrap, boosting, etc.

DH concluded by saying strategies for use of dark data can be a powerful tool, but dark data can have serious consequences.

The relevance to machine learning in image detection was noted and edge and corner cases that are not part of the machine learning training set are good examples of dark data. DC said there will now be AI-based training to review hydrographic charts based on satellite images to address the large amounts of data. It was thought that this could be prone to Dark Data types of issue.

MA asked whether statistical data needs specific treatment in the Data Safety Guidance, although he noted that adding it as a new category may not be appropriate as it seems to cross many of the existing types. PH suggested that statistical data could be treated in the same way as, say, an acceptably noisy analogue sensor but this might not then recognise the value in the data itself. It was agreed that the group needs to consider how Dark Data should be addressed in the Data Safety Guidance.

Action 54.1 [MP] Investigation what changes are required to the Guidance to accommodate issues associated with Dark Data.

3. Alastair's Book - hardcopy
MP reminded the group that AF has produced a hard copy book [4] on data safety and is available for purchase.

4. Update on Guidance Document v3.2 (download/sales)
MP noted that there have been 1349 hits and 567 downloads since Feb 2020.

There were no physical copy purchases in August 2020.

5. Update on SSS'21
MP Note that Nick Hales will be presenting a paper at SSS'21 discussing data safety in virus outbreaks.

6. Aims for 2021
MP asked whether we should update the guidance to produce an update for SSS'21 to include topics such as:
- Dark Data
- More accident case studies
- Tidy up the text
- Addressing comments arising from Thor Myklebust's comments
- Covid-19
- Dissemination / communication (NB: Emma Taylor is giving a talk at SSS'21 in this area).

However, it was agreed that we need to work with the Overleaf editor [MT] to lead the implementation of any changes.

7. Future Events
MP noted the upcoming SCSC Events:
- New Safety Analysis Techniques on 12th November 2020.
- Management and Oversight of Complex Systems on 3rd December 2020;
- SSS'21 in February 2021.

MP showed the draft programme for SSS'21 and said there has been a very good response to the event. MP also showed the proposed pricing structure for the online streaming for SSS'21 and there were no issues raised. [This will now be free to members and completely online.]

8. Minutes and action status

See table at end.

52.1: MP said LRF are keen for any SCSC Working Group to submit proposals on communication and dissemination of issues that are typically impacting industry. The funding is relatively small but the proposal process is lightweight so ideas are welcome from the DSIWG.

9. AOB, etc.

None.

10. Data Safety in the News

AF reported an error in a spreadsheet delayed the opening of a new £150m hospital.
https://www.bbc.co.uk/news/uk-scotland-edinburgh-east-fife-53893101:

MP discussed Aeronautical Data that has its own standard and noted that the updated edition, EUROCAE ED-76A / DO-200B [5] is now available. He also said that there is a training course "Aviation Data-driven Decision Making (AD3M) – Part 1 Online Course" [6] dealing with how to use and interpret Aeronautical Data that may be of interest.

The following reports were received after the meeting:

From MA: A B-2 Spirit crashed when heavy rains before takeoff caused fuel to get into data sensors that were responsible for calculating speed and altitude for the aircraft. This led to maintenance crew calibrating the sensors incorrectly making the aircraft unstable.
https://worldwarwings.com/most-expensive-usaf-crash-in-history-because-of-condensation/

From MA: Crew members of an Easyjet flight at Lisbon airport used the wrong calculations for the runway length, and were 1.3 seconds away from reaching the end of the runway during takeoff.
https://news.sky.com/story/easyjet-flight-was-one-second-from-crashing-after-pilots-misjudged-runway-by-nearly-a-mile-report-reveals-12048260

Modification to ICAO Annex 15 to include Data Quality Requirements are discussed in the Boeing presentation:
https://www.icao.int/MID/Documents/2018/Interregional_EUR-MID_Workshop_on_PANS-AIM/2.9b ANNEX 15 PANS-AIM Data Quality Requirements - VMEYER.pdf

From TR: Ambulance won't find mislocated addresses
https://catless.ncl.ac.uk/Risks/32/24#subj7

11. Next Meeting

A Microsoft Teams meeting in October 2020 was proposed. [Now set for 7th October.]

12. Thanks

Thanks to all for taking part. Thanks to PH for minutes and MP for chairing the meeting.

SCSC DSIWG

Summary of Open Actions

Rows have been greyed-out to indicate that the actions were closed during this meeting. Those entries will be deleted from future versions of the action log.

Ref	Owner	Description	Target Guidance Version
42.6	PH	Define the process to publish a document developed in Overleaf via Amazon	4.0
42.9	MP	Work out a matrix of data categories (previously 'types') and data properties (as per DB discussion)	N/A
43.4	MP	Write up a data focussed FMEA approach.	4.0
44.1	MT	Review last 12 months of DSIWG minutes and put any actions referring to v4.0 into Appendix O.	4.0
44.2	MP	To discuss with AK on how to get the Wikipedia article published	N/A
46.1	MP	Review the application of DSALs to higher level forms of aggregation	N/A
49.6	MT	Review Overleaf briefing material and aim to hold a briefing Webex (between 16th and 20th December 2019) in the use of Overleaf in the production of the guidance.	N/A
49.11	DA/MP	Prepare an introductory email to send out to the DSIWG group inviting people to join the tooling subgroup.	N/A
50.4	PG	Arrange meeting at EDF to demonstrate the tool	N/A
50.5	DA	Start up the sub-group and initiate first teleconf	N/A
51.1	MA	To ask Brian Jepson to set up a Kanban board for the tooling	N/A
51.2	DA	provide access to the tool for all registered users.	N/A
52.1	All	Contact MP or PH if you would like to contribute to the LRF proposal or work on activities	N/A
52.2	All	Contact MP or MT with any news accident stories	N/A
52.3	MP	Contact RO of Rolls-Royce to seek further input on some of the new areas for data safety	N/A
52.4	All	If interested in taking part in some of the TSG activities, please contact DA directly	N/A
52.5	MP	Ask MT / PH to see which of the previously identified updates can be done for DSG3.3	3.3
53.1	MP	To talk to Kevin King about what we need to do in the guidance for digital twin.	3.3
54.1	MP	Investigation what changes are required to the Guidance to accommodate Dark Data issues	3.3

13. References

Ref	Title	Location
[1]	Meeting slides	https://scsc.uk/file/gd/54th_DSIWG_Slides_V1-864.pptx
[2]	Dark Data slides	https://scsc.uk/file/gd/DARK_DATA_David_Hand_DSIWG_talk-863.pptx
[3]	Dark Data Book	https://www.amazon.co.uk/Dark-Data-David-J-Hand/dp/069118237X
[4]	Data Centric Safety	https://www.elsevier.com/books/data-centric-safety/faulkner/978-0-12-820790-1
[5]	EUROCAE ED-76A / RTCA DO-200B	https://eshop.eurocae.net/eurocae-documents-and-reports/ed-76a/# https://infostore.saiglobal.com/en-us/Standards/RTCA-DO-200B-2015-958843_SAIG_RTCA_RTCA_2250689/
[6]	Aviation Data Driven Decision Making	https://store.icao.int/en/aviation-data-driven-decision-making-ad3m-part-1-online-course