DSIWG

SCSC Data Safety Initiative – WG Meeting 55

7th October 2020, Teams

Minutes

Attendees

Mike Parsons (MP) – CGI, Louise Harney (LH) – Leonardo, Alastair Faulkner (AF) – Abbeymeade, Divya Atkins (DA) – MCA, Martin Atkins (MA) – MCA, Paul McKernan (PMK) - DSTL, Dale Callicott (DC) – BAE, Jim Mateer (JM) – SQEP Ltd, Andy Williams (AW) – NewTechNo, Paolo Giuliani (PG) – EDF, Tim Rowe (TR) – Consultant, Mark Templeton (MT) – Qinetiq, Carl Tipton (CT) - Johnson Matthey.

Apologies

Paul Hampton (PH) – CGI, Mike Ainsworth (MA) – Ricardo, Ali Hessami (AH) – Vega, Paul Ensor (PE) – Boeing, Janette Baldwin (JB) – Thales, Bill Blackburn (BB) – Process Renewal, Brent Kimberley (BK) – Durham, Dave Banham (DB) – Blackberry, Fan Ye (FY) – ESC, Sam Robinson (SR) - EDF

Agenda

1. Covid-19 and PHE data error
2. Dark Data Implications
3. Data Safety in the News
4. Update on SSS'21 papers: WG presentation and papers
5. Update on Guidance Document v3.2 (downloads/sales)
6. Aims for 2021
7. SCSC Events
8. Data Safety Tooling
9. AOB, etc.

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 https://scsc.uk/file/gd/55th_DSIWG_Slides_v1-899.pptx

The group welcomed new attendee CT who gave a brief background to his interest in the group.

1. COVID-19 and PHE Data Error

MP discussed data in the Covid-19 crisis and two items in particular:

(i) Excel: Why using Microsoft's tool caused Covid-19 results to be lost, https://www.bbc.com/news/technology-54423988

(ii) Disappearing Covid-19 app alerts cause alarm, https://www.bbc.com/news/technology-54389083

There was a lot of discussion about the first item. MP felt that, although Excel was blamed, in fact the real issue was much deeper with no sensible criticality analysis of the data or the data processing architecture (i.e. what data is safety-critical, how is it protected, etc.).

There was a suspicion that the whole data chain had been constructed in a rushed manner, with no serious analysis[1]. It was thought likely that the people who designed the processing were not engineers, with little understanding of systems safety (and no safety engineers were likely involved). Hence it was probably a "un-engineered" system. It may also have been a legacy architecture, which had been expanded way beyond its original capacity design. There is also likely to be a culture problem at PHE with lack of safety engineering capability, lack of safety awareness, and probably a lack of safety and data governance. There was a concern that it may have been a result of Agile development[2].

The apparent fix of splitting into smaller files was felt to be a poor short-term patch. It was also noted that PHE is about to be replaced.

ACTION 55.1 (MP/MT): Create additional 'war story' for the Excel data loss for version 3.3 of the guidance

2. Dark Data Implications

MP described his initial thoughts on the implications of David Hand's work https://darkdata.website/ on Dark Data for the data safety guidance. He suggested that we need to go through all the data categories and look for common "dark" examples, for example:

DSG Category 3: Requirements data may not be formally written down
DSG Category 13: Staff and training data may be missing or may be falsified
DSG Category 23: Justification data may be missing e.g. for COTS component

He then explained some suggested examples for each of the Dark Data types:

1. Data We Know Are Missing: "Known unknowns"
    - Typical for missing assurance for COTS components
    - Can add warnings, etc. to mitigate
    - Could change the safety position
    - Can try to "fill-in" with other assurance e.g. established organisation track-record.

2. Data We Don't Know Are Missing: "Unknown unknowns"

[1] PMK: It is highly unlikely that the system was 'designed' it appears to have 'evolved' therefore there was almost certainly no national approach to the design, formal requirements, scaling and scalability. In addition, there was probably no clear-cut understanding of what to do with the metrics – this is common across COVID metrics…

[2] PMK: It is more likely to have been rushed rather than Agile. An Agile development still plans to deliver a working solution , but quickly, however, if the requirements are wrong it does not matter which development lifecycle you choose, you are unlikely to get what you intended out of the system

- **PHE Track and Trace data!**
- May be discovered after some period
- Can fundamentally change the (safety) picture
- Probably the biggest safety risk[3]

(An example may be: somebody knows a problem with a system but not the person trying to make the safety position.)
(There was a question about data discovered missing after some time - the issue is then what to do with the "rediscovered data": (i) Apply it? (ii) Ignore It? or (iii) Mention it but not take into account? (iv) Do an impact analysis on the missing data and then act according to the results?

3. Choosing Just Some Cases
    - Sampling from sensors / sampling intervals chosen badly
    - With complex or informal criteria could be as case (2) - you don't know what has been left out

4. Self Selection
    - Probably as (3), but could be more informal / ambiguous

5. Missing What Matters
    - Probably as (2). Are we actually measuring the right things, e.g. safety metrics[4]

    (Also: The case of being too close to the data, i.e. the "wood for the trees" case i.e. the detail masks the overall pattern of the data)

6. Data Which Might Have Been
    - Possibly could be applied to architectures, e.g. single channel / multiple channel situations (e.g. Boeing 737 MAX[5]) where a second channel would give a much better set of data to compare with the first

7. Changes With Time
    - Data in safety systems often becomes obsolete or out of date and may still be mistakenly used. E.g. system configuration data or medical drug databases. New versions of software often require updates to data, which are not always done.

8. Definitions of Data
    - Data schemas often evolve over time. These can render old data obsolete / subject to misinterpretation, and possibly needing migration/translation[6].

9. Summaries of Data

[3] PMK: The provision of a 'safety net' or some other provision is essential. Key factors in these safety nets is they must not mask/ignore dark data but they should also allow for it to be considered before it adversely affects the system, as it should for all data
[4] PMK: Also need to be aware of 'optimism bias' and the tendency to accept data that supports your position (your para 11 (Confirmation Bias)– sometimes this bias is not obvious to the person involved as they are not aware of their personal biases
[5] PMK: The 737 MAX MCAS was probably a design flaw from the outset. The fact that this design flaw led to the platform being reliant on a single data source led to the accidents
[6] PMK: There is also a challenge with systems where the data can be input in free text as classification can become very complex

- Often seen in safety metrics where data is aggregated. Could be misleading or cause "boundary reactions".
- Data fusion

10. Measurement Error and Uncertainty
    - Often happens (Boeing 737 MAX again). Sampling techniques, interval polling, etc.
    - Data fusion again
    - Lossy models

11. Feedback and Gaming
    - Confirmation bias in safety justifications

12. Information Asymmetry
    - Can be many sources of data – different databases out of sync

13. Intentionally Darkened Data
    - Can and does happen, e.g. medical, maritime sectors
    - Potentially huge safety impact

14. Fabricated and Synthetic Data
    - Can and does happen, e.g. medical and maritime sectors where data is sometimes retrospectively entered / patched to make a "clean" record
    - Autonomous vehicle training databases
    - Potentially huge safety impact

15. Extrapolating Beyond Your Data
    - Machine learning data in spades!
    - Data take out of scope

There was a useful discussion on these implications. A decision was taken to create a standalone appendix on Dark Data in the next version of the guidance (3.3) to be issued in Feb 2021.

ACTION 55.2 (MP/MT): Create appendix on Dark Data for version 3.3 of the guidance

3. Update on Guidance Document v3.2 (download/sales)

MP noted that there have been 1587 hits and 669 downloads since Feb 2020.

4. Update on SSS'21

MP noted that SSS'21 is now online, spread over 3 afternoons and is free to SCSC members https://scsc.uk/e683 It was noted that Nick Hales is giving a Covid-19 Data Safety paper at the symposium.

5. Aims for 2021

It was agreed that the version 3.3 to be produced for SSS'21 will have:

- Dark Data Appendix
- More accident case studies
- Tidy up the text (as needed)
- Addressing comments arising from Thor Myklebust's comments

6. Future Events

MP noted the upcoming SCSC Events:

Seminar: New Safety Analysis Techniques
November 12, 2020 - Online, Free to Members
This seminar is relevant to safety engineers and safety consultants who have to perform analysis of systems. It will also be useful for safety auditors and assessors who may have to interpret or ... »

Seminar: Management and Oversight of Complex Systems
December 3, 2020 - Online, Free to Members
This online seminar will be useful for all those involved in running a complex operation that involves safety. It is aimed at Managers, Operators, Regulators and Assurance staff. If you operate a ... »

Symposium: Safety-Critical Systems Symposium (SSS'21)
February 9 - 11, 2021 - Online, Free to Members
The 29th SCSC Safety Critical Systems Symposium 2021 (SSS'21) will be held from 9-11th February 2021 in an online format. It will be split into three sessions, held in the afternoons over the ... »

7. Minutes and action status

See table at end.

8. Tooling

DA and MA did a demo of the latest version of the Data Safety Tool by MCS which was well received.
DA will set up a subgroup to look at the wider issues of tooling.
Some of the meeting attendees will be trialling the use of the web based tool.

9. AOB, etc.

MA noted the data aspect to this aviation incident: https://www.flightglobal.com/safety/wizz-a321-left-out-of-balance-by-seat-allocation-mishap/140542.article

10. Next Meeting

A Teams meeting in mid-November 2020 was proposed [Now set for 11th November 2020].

11. Thanks

Thanks to all for taking part. Thanks to MP for chairing the meeting.

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	3.3
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.	3.3
44.1	MT	Review last 12 months of DSIWG minutes and put any actions referring to v3.3 or v4.0 into Appendix O.	3.3
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 before end of March 2021 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/DA	Arrange remote meeting for EDF staff 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
52.3	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
55.1	MP/MT	Create additional 'war story' for the Excel data loss for version 3.3 of the guidance	3.3
55.2	MP/MT	Create appendix on Dark Data for version 3.3 of the guidance	3.3