DSIWG

SCSC Data Safety Initiative – WG Meeting 73

19th December 2022, Zoom

Minutes

Attendees

Mike Parsons (MP) – AAIP, Nick Hales (NH) – Consultant, Divya Atkins (DA) – MCA, Martin Atkins (MA) – MCA, Jennifer Kracht (JK) – TomTom, Roland Rosier (RR) - TomTom, Paul Hampton (PH) – CGI, Mike Standish (MS) – Dstl, Mark Templeton (MT) – Qinetiq, Brent Kimberley (BK) – Durham, Tim Rowe (TR) – Consultant, Gordon Hurwitz (GH) – Thales, Michael Green (MG) – Ecomergy, Arch McKinlay (AM) – NGA.

Apologies

Richard Garrett (RG) – SQEP, Andy Williams (AW) – Consultant, Paolo Giuliani (PG) – Atkins, Paul McKernan (PMcK) – Consultant, Sam Robinson (SR) – EDF Energy, Mark Nicholson (MN) – University of York, Carl Tipton (CT) – JohnsonMatthey, Richard Clarkson-Webb (RCW) – Atkins, Dave Murray (DM) - BAE Systems, Lee Glazier (LG) - Rolls-Royce, Daniel Clegg (DC) - BAE Systems, Ali Hessami (AH) – Vega, Graham Sutherland, Oscar Slotosch (OS) – Validas, Paul Ensor (PE) – Boeing, Graham Sutherland (GS) - Consultant

Agenda

1. Welcome
2. New Guidance Version
3. Poster for SSS WG
4. Abstract for SSS Poster on NHS Work
5. New Data Risk Types – Update
6. DSITN (Data Safety in the News)
7. Update on Tooling
8. SCSC SSS'23
9. Actions
10. Next meeting
11. AOB

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/73rd_DSIWG_Slides_v1-1486.pptx

1. Welcome

MP opened the meeting and welcomed those attending.

2. Next Guidance Version

MP had received a set of suggestions from JK and RR, motivated by a discussion about whether to investigate turning the Data Safety Guidance into a Standard. There was a discussion about the content of the next version of the guidance starting with TomTom's inputs. There were two categories (A – more important, and B – nice-to-have). The lists of suggestions were annotated in the meeting: https://scsc.uk/file/gd/TomTom_Update-1487.docx

There was lots of discussion about these points and it was agreed that they would go forward as candidate updates to the next version of the guidance.

It was also felt that TomTom had built up experience in applying the guidance to a real problem which would be valuable to include in the guidance as a short appendix[1].

Action 73.1 (JK, RR) – Consider production of a short note which could be used as an appendix to the guidance on lessons learnt using the guidance at TomTom

There was a discussion about which development lifecycles the guidance applied to. It was noted that the guidance refers to 'V' and a more continuous 'service' lifecycle. MP noted that it doesn't have to be used with a development lifecycle and there are organisations which operate as a 'data pipeline' where there is nothing developed, but input data is processed and data products produced. This doesn't have to involve computer systems and could be completely manual. TR noted that this sort of data production model is used extensively for aeronautical data, and MP said he had been involved in a company producing critical drug information this way.

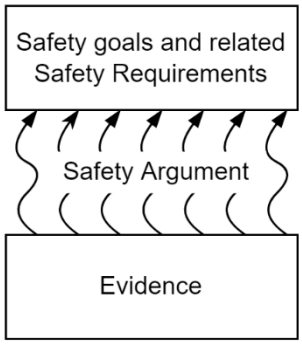
Action 73.2 (MT) – Consider how the guidance fits with different lifecycles considering 'V', Continuous Service, Agile and 'Data Pipeline'

It was noted that the guidance doesn't cover estimating, sizing and costing of the data safety effort required on a particular project, and something on this would be useful.

[1] JK said that she initially struggled with the difference between Data Quality and Safety. The solution is to take some existing numeric data (quantity) that can be used to define thresholds for categories like 'good', 'moderate', 'bad' (quality) and with the addition of context, process, time (or whatever is specific for your data in a system ('3D view')) it is possible to identify the requirements for safety.

There was a discussion about hazards flow up and down from system level when data is involved. There was a consensus that hazard information should flow downwards from system level to components dealing with data, and that those components should be able to flow any identified (data-related) hazards upwards.

RR stressed that there is a need to follow a safety case approach with the guidance where work products are the evidence to support the safety argument, and therefore the top-level safety claims.



This was agreed: there have been various discussions about producing templates or exemplars for data safety arguments over the years, and this was considered to be important.

MP said there are many possible additions to the guidance (see last minutes) but that it would be good to get Thor Myklebust's additional properties in, viz. 'Analysability', 'Explainability' and 'Verifiability (see action 66.6).

An editorial team comprising MP, MT, PH and TR will produce the next version of the guidance. [MP will contact Oscar Slotosch to see if he is still interested.]

3. SSS Poster

MP said that a WG poster was required for SSS in February. Some suggestions for the poster included:

- Use of the DSIWG 'elephant' image as a central focal point
- W3W as example as to why data a problem
- Picture of guidance book cover
- Rationale: why data safety is an issue
- 3 assurance elements: HW / SW / Data
- Some example accidents: e.g. Soyuz, Covid data loss in Excel
- Data properties (sample)
- Excerpts from mitigation tables
- Add SCSC logo + possibly QR code to take to WG area

Action 73.3 (DA) – Produce initial outline of DSIWG poster by 23rd December 2022

4. Abstract for SSS Poster on NHS Work

An abstract was produced in the meeting for the SSS presentation to be given by DA and MA:

"Analysing Data Risks in the NHS Response to the Covid-19 Pandemic"

This work looks at some of the Covid-19 data flows and analyses what happened and the impacts from a data safety point of view (including loss of data increasing transmission). It includes output of workshops held with the NHS and looks at how this can be used to enhance both the new Data Safety Guidance document and the functionality of the data safety tool (RADISH).

5. New Data Risk Types

MP outlined the latest list of Dead Data Types, https://scsc.uk/file/gd/Dead_Data_Types_v4.3-1488.docx . PH presented the diagram showing the taxonomy / ontology for these.

New types discussed were:

- Lazarus Data - Data that should not be defunct, i.e. it is currently unused due to a mistake (e.g. incorrect change applied), but actually should be currently used, incorporated or available within the system.
- Treasure Island Data - Data to which the method of access (e.g. index, map or hash table) has become lost or corrupt so that, although the data still exists, there is no easy way of finding it (or getting to it).
- Enigma Data - Data to which the encryption key or keys have become lost so that, although the data still exists, there is no way of decrypting it.
- Steganographic Data (Long John Silver Data) - Data that is concealed intentionally within other data
- Square Peg Data[2] - This is data where on migration or on translation between systems data doesn't map to a sensible new category or type, or doesn't fit in a field width, so it is mapped either to something which is not quite right, truncated or set to a default or null value [or lost/rejected]. I.e. you are trying to squeeze it into something that it doesn't quite fit, and losing information in the process

There was some interest in these and it was suggested these could be added as an appendix to the new guidance, however it might be better to wait for a SCSC Newsletter article in June 2023.

6. Data Safety in the News (DSITN)

The following were discussed from a data safety point of view:

- Data-based technology in the Ukraine war: https://www.bbc.co.uk/news/technology-63109532
- Elon Musk taking legal action over Twitter account that tracks his private jet: https://www.bbc.co.uk/news/world-us-canada-63978323
- Malicious Microsoft-signed Windows drivers wielded in cyberattacks: https://www.theregister.com/2022/12/14/microsoft_drivers_ransomware_attacks

7. Update on Tooling

DA explained some of the recent contract wins for MCA Ltd.
MA and DA then ran a short workshop for meeting participants on creating logins and using the RADISH tool and then creating some entries in the tool.

[2] BK said that LIMS (Laboratory Information Management System) data is a good example of square peg data. Some of the data may be analysts notes. For a given LIMS measurement (over time), some of the measurements will be numeric, some will be limit statements, some will be analyst comments.
Issues which can occur can include over-diluting the sample (below the limit of detection). Too many colonies for the analyst to count (beyond the limit of quantification), also using the wrong scale or range, etc. Trying to measure too many things in a simple go. You could do something similar for radar, minimum separation, pilot and ATC (air traffic controller) comments, etc.

The https://data-safety.tech website was demonstrated and the accidents viewed. It was thought that this list of accidents could be enhanced using the DSITN entries from the minutes.

Action 73.4 (DA) – See whether any of the DSITN entries from previous minutes and slides could be used to enhance the list of accidents

8. SSS'23

MP mentioned that SSS'23 takes place in February 2023 in York:
https://scsc.uk/e898 - bookings are open and there are still spaces for posters.
If you would like to present a poster, please contact mike.parsons@scsc.uk

9. Actions

Actions 72.1 and 72.2 are now closed.

10. AOB

None

11. Next Meeting

Next meeting will be held 24th January by Zoom from 15:00-17:00.

12. Thanks

Thanks to DA and MA for hosting in Bath and the Christmas drinks and snacks.
Thanks to MP for taking the minutes.
Thanks to MP for chairing.

Summary of Open Actions

Actions greyed out are considered closed and will be removed from the list at next issue.

Ref	Owner	Description	Target Guidance Version
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
44.2	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
53.1	MP	To talk to Kevin King about what we need to do in the guidance for digital twins.	4.0
61.2	AW	Research the relevance of digital currencies and report back to the group (with MA and MT)	N/A
63.1	CT	Look at both Dark Data and Dazzle Data for sensors (e.g. when a sensor is saturated, in noisy environment or when readings are below the detection level floor)	4.0
64.1	MP	Contact Thor and establish the details of the guidance proposals in the paper.	4.0

Ref	Owner	Description	Target Guidance Version
66.6	MT	Add these three properties ['Analysability', 'Explainability', 'Verifiability'] to the user-visible further work section. If time allows then develop into the guidance further.	4.0
68.2	MP/MT	Develop the migration work further and present at next meeting	4.0
69.2	RR	Explore the issue of data / software compatibility issues and to what extent data can impose requirements on software	4.0
69.3	PMcK	Develop a scoping diagram that shows how the DSG fits into the overall lifecycle process and other standards	4.0
69.4	MA	Write a short note on the issues of aggregation	4.0
69.6	MA/DA	Update the data safety tool to use the latest version of the guidance document	-
70.1	MA/DA	Investigate feasibility of creating searchable web database of data safety-related accidents.	-
71.1	MP	Add Homophones/Homonyms explicitly to the guidance.	4.0
71.3	PH/DA/RO	Develop security properties thinking further for next DSIWG	4.0
71.4	PH/DA/RO	Present security properties work to next SISWG meeting	-
71.5	AM	(i) Establish if any of this can be published within the DSIWG and (ii) Consider a structuring similar to that used in security standards or ISO26262	-
71.7	MP/CT	Consider impact of FAIR data on the guidance	4.0
72.1	JK	Send MP the suggestions and ideas from TomTom application of the DSG for consideration in the new version	4.0
72.2	RR	Send MP the suggestions for improvements related to the typical content of standards documents	4.0
73.1	JK, RR	Consider production of a short note which could be used as an appendix to the guidance on lessons learnt using the guidance at TomTom	4.0
73.2	MT	Consider how the guidance fits with different lifecycles considering 'V', Continuous Service, Agile and 'Data Pipeline'	4.0
73.3	DA	Produce initial outline of DSIWG poster by 23rd December 2022	-
73.4	DA	See whether any of the DSITN entries from previous minutes and slides could be used to enhance the list of accidents	-