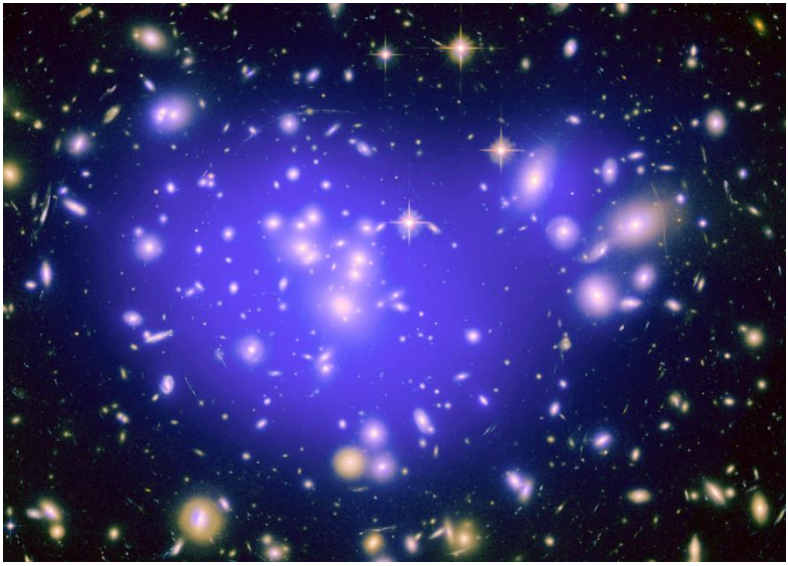
Dark Data and Safety



Mike Parsons discusses the results of a focussed analysis by the Data Safety Initiative Working Group on Dark Data. The assessment looked at the safety implications arising from the various types of Dark Data and the plans to provide more specific guidance to address Dark Data safety risks.

At the last SCSC Data Safety Initiative Working Group (DSIWG) meeting (DSIWG#55, [1, 2]) the group had some initial thoughts on the implications of David Hand's work on Dark Data (see the previous article) for the Data Safety Guidance (DSG) [3]. It was suggested that we need to go through all the data categories in the current guidance 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 a COTS component

In the meantime, the meeting examined the previously identified categories of Dark Data:

Dark Data Categories and Safety Examples

1. **Data We Know Are Missing: "Known unknowns"**

This case is very common in safety justifications where assurance information may be withheld for commercial reasons or does not exist, but we know, or are informed, that it isn't available. Common examples include:

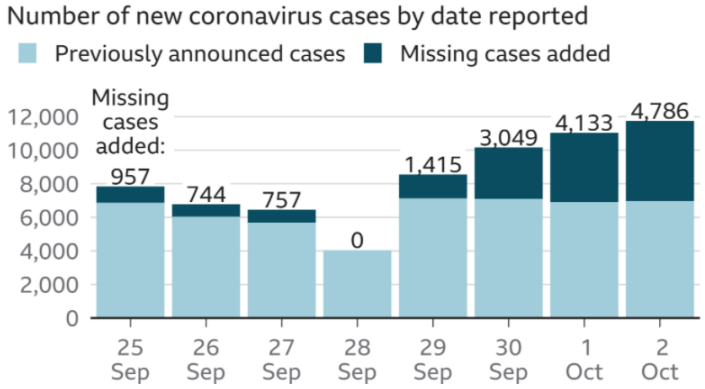
- Assurance information for COTS components
- Technical information about legacy systems

If this is the case we can mitigate in several ways, including use of warnings, training, restrictions of use, etc. We can also substitute with other more indirect assurance e.g. established organisational track-record in the sector, or audit reports.

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

This case is hopefully less common than (1), but we have to acknowledge that it happens. Some examples are:

- The recent Covid-19 Track and Trace data loss [4], where the organisation handling the data was unaware that rows were missing from a spreadsheet for some time



- Somebody knows a problem with a system (i.e. has counter-evidence regarding the safety claim) but does not communicate this to the person creating the safety position

In many cases the loss may be discovered after some period, and it is incumbent on the organisation involved to analyse the impact of the missing data over the time period, including subsequent decisions and actions. Its effects should not be underestimated. This case can fundamentally change the safety picture and is probably in the highest safety risk category.

The DSIWG discussed the issue of data that is discovered to be missing but then subsequently found. It may still be available so what should be done with the "rediscovered data"? Four options were identified: (1) Apply the missing data; (2) Ignore the missing data (3) Mention that the data was missing but not take it into account and (4) perform an impact analysis on the missing data and then act according to the results.

3. **Choosing Just Some Cases**

This is where something or somebody has been selective. Examples might be:

- Selection of test runs that succeeded (ignoring failed runs and their diagnostics)
- Selective sampling from sensors, or where the sampling intervals are chosen badly
- Incorrect filtering of the data, leaving out more cases than intended

Note that with complex or informal criteria the effects could be as case (2), i.e. you don't know what has been left out.

4. **Self-Selection**

This was considered to be similar to case (3), but could be even more informal or ambiguous.

5. **Missing What Matters**

This was considered similar to case (2) in impact terms. Examples might be:

- Measuring the wrong things, e.g. poor safety metrics / indicators
- Too much data to deal with or analyse, so some is ignored
- Too much filtering or processing, so losing information along the way
- Being too close to the data, i.e. the "wood for the trees". This is when the detail masks the overall issue with data, e.g. a slow trend or bias masked by peaks.

6. **Data Which Might Have Been**

This was considered an interesting case, an example might be:

- Inappropriate system architecture e.g. single data channel when a multiple channel approach should have been used, such as in the Boeing 737 MAX accidents [5]

7. **Changes With Time**

This was recognised as a common problem in a safety context. Data in safety systems often becomes obsolete or out-of-date and may still be mistakenly used. Some examples are:

- System configuration data not kept up to date as software or hardware changes
- Medical drug interaction databases
- Software patches require updates to configuration or system data – not always done.

8. **Definitions of Data**

This was considered a common case for systems with databases or those exchanging data with external systems, e.g.

- Data schemas in medical record systems often evolve over time. These can render old data obsolete / subject to misinterpretation, and possibly needing migration/translation.

9. **Summaries of Data**

Often seen in data about safety systems and projects:

- Safety metrics or indicators where data is aggregated to create a composite value
- Could be misleading or cause "boundary reactions", e.g. Red-Amber boundary
- Data fusion across multiple sensors

10. **Measurement Error and Uncertainty**

Sensors can degrade and fail over time, especially in harsh environments such as automotive, marine or aviation:

- Sampling techniques can cause artefacts, interval polling interval can be incorrect, etc.
- Data fusion again

11. **Feedback and Gaming**

This can happen in safety justifications or test case production:

- Early production of a safety argument could lead to only those artefacts which support the claim being generated
- Confirmation bias in safety justifications

12. Information Asymmetry

This is common where there are multiple stores or sources of the same data:

- Multiple / backup databases where they are not kept in sync

13. Intentionally Darkened Data

This can and does happen, e.g.

- Defence, security and government sectors where data is purposefully hidden or destroyed
- This could apply to records deleted after an accident to make things 'look better'

14. Fabricated and Synthetic Data

Fabrication is surprisingly common, e.g. medical, policing and maritime sectors. Synthetic data is often used where there are difficulties in producing enough real data with the right characteristics:

- Data is retrospectively entered / patched to make a "clean" record
- Synthetic autonomous vehicle training databases can have issues with artificial data if not realistic

15. Extrapolating Beyond Your Data

Machine learning systems have to cope with extrapolation outside of their training data, but the outcomes may be unexpected:

- Machine learning data, especially real or recorded data that may not contain edge/corner cases

A decision was taken by the meeting to create a standalone appendix on Dark Data in the next version of the data safety guidance (3.3) to be issued in Feb 2021 at SSS'21 [6].

References

[1] DSIWG, "Minutes of Data Safety Initiative Meeting #55", SCSC, scsc.uk/file/gd-main/SCSC-Data-Safety-Initiative-Meeting-55-FINAL.pdf, Oct 2020

[2] DSIWG, "Slides for Data Safety Initiative Meeting #55", SCSC, scsc.uk/file/gd/55th_DSIWG_Slides_v1-899.pptx, Oct 2020

[3] DSIWG, "Data Safety Guidance (Version 3.2)" SCSC, February 2020, SCSC-127E, ISBN-13: 9798601577359, scsc.uk/scsc-127E

[4] Leo Kelion, "Excel: Why using Microsoft's tool caused Covid-19 results to be lost", BBC, accessed 5th Oct 2020, www.bbc.co.uk/news/technology-54423988

[5] Boeing 737 MAX, en.wikipedia.org/wiki/Boeing_737_MAX_groundings, accessed 22 Oct 2020

[6] SSS'21, "Safety-Critical Systems Symposium (SSS'21)", SCSC, February 2021, https://scsc.uk/e683

Image Attributions
Top image: galaxy cluster Abell 1689, with the mass distribution of the dark matter in the gravitational lens overlaid (in purple). NASA, ESA, E. Jullo (JPL/LAM), P. Natarajan (Yale) and J-P. Kneib (LAM). Licensed under the Creative Commons Attribution 3.0 Unported license.
Covid-19 graph: Gov.uk dashboard, PHE. Contains public sector information licensed under the Open Government Licence v3.0

Mike Parsons, SCSC DSIWG Chair

Mike is the SCSC Director and Events Coordinator. He also leads the SCSC Service Assurance, Data Safety Initiative and Covid-19 Working Groups. He is currently a safety engineer at CGI UK working on various healthcare projects. He has been in the business of safety since 1989.