





# FY23 Sigma Data Challenge

October 2022

NTR RN – Detection Programme

[philip.martin@awe.co.uk](mailto:philip.martin@awe.co.uk)



# Talk Outline

- Challenge Objectives
- Overview of pilot/data
- Purpose of the challenge
- Challenge statement
- Challenge output
- How to engage – links,
- Calendar / Timeline of the challenge



# Objectives of the Challenge

The Sigma Challenge is a pilot project to trial the access for academic researchers to large scale data from multi-site radiation detectors. It's an opportunity for researchers to develop machine learning or AI algorithms to analyse this large data set of spectra, and to make conclusions about the distribution of radiation sources that were present in the field.

- The Sigma Data contains up to 1.5 billion gamma spectra, spread across 3 data pilot data sets. The spectra are typically very sparse and contain few counts.
- Academic researchers will be given access to the data sets and invited to develop automated analysis methods that can interpret this data.
- Some geographical context will be given to the data, however the full 'ground truth' of the location and type of radioactive sources present in the field is not available.
- A key outcome of the project will be for the academic teams to give feedback to AWE on the quality of the data, the performance of their analysis algorithms, and recommendations for how the data could be improved in future trials.

The challenge will run from October 2022 to October 2023, and participants will be invited to present their results at a NuSec Technical Workshop in Autumn 2023.

# Overview of SIGMA and pilot data

- What is SIGMA
  - SIGMA, a US Defense Advanced Research Projects Agency (DARPA) programme developed a low-cost, high capability networked radiation detector system for countering nuclear and radiological terrorism through continuous city-scale radiation monitoring.
- Why AWE
  - On behalf of UK HO, AWE lead the UK technical component of a UK-US collaboration effort on SIGMA to a level allowing us to provide expert advice to the customer.
- 3 Pilots 2017-2018 (each detector uploads a spectrum every second to the cloud)
  - Pilot 1
    - 12/06/17 – 07/07/17
    - ~100 D3S, ~10 static.
  - Pilot 2
    - 05/01/2018 – 02/02/2018
    - ~50 D3S, ~20 static.
  - Pilot 3
    - 06/08/2018 – 09/10/2018
    - ~100 D3S, ~10 static.

## Kromek D3S Revision 1.1

2 x 1 x 0.5 inch CsI (TI)  
gamma detector

32 mm x 100 mm LiF:ZnS  
neutron detector



## Physical Sciences Inc. PERM-S (unheated and heated variants)

2 x 4 x 16-inch NaI (TI)  
gamma detector





# Challenge Purpose

- For you
  - Access to a reasonable number of spectra.
    - All three pilots total around 1.5 billion spectra, pilot 2 has been suggested as a starting point with a round 250 million.
  - Sigma Data Challenge Pilot projects.
    - To aid with the generation of projects, special pilot project funding is available. These pilot projects are two months in length with bursaries of £3,000 available.
    - Information on how to apply for funding will be briefed out with the roll out of the data
    - Applications assessed on a quarterly basis (Deadlines 1<sup>st</sup> Dec, 1<sup>st</sup> Mar, 1<sup>st</sup> Jun)
  - Facilitated workshops.
    - This is a real dataset and we are happy to talk those taking part in the challenge through the details of the data provided
    - Spoiler – it is imperfect
- For us
  - A chance to engage the academic community in the area of advanced, real-time, spectroscopic ID algorithm development.
  - Hopefully the start of a regular (yearly) feedback loop where we learn what useable large datasets look like and work towards providing them for the community.
    - We could provide our adjudicated alarm information at a future date
    - We could model relevant threat spectra and inject them into the datasets next year
  - Ultimately we believe we all benefit by understanding the steps involved, and inherent challenges within, this type of algorithm development.



# Challenge Statement

- Actionable, real-time, spectroscopic ID of short-duration spectra in a large distributed-detection dataset

Big Challenge	Specific Challenge
What advanced data analysis concepts can be applied to the dataset?	What specific data analysis approaches are applicable to the dataset and why, also what should not be applied to the dataset and why?
Threat Identification	Accurate identification of any/all isotopes which are detected, regardless of threat-type (medical/industrial/etc).
Optimised performance against threats	Optimising the ability to identify and/or locate a threat spectra whilst not mis-identifying non-threat spectra.
Device health monitoring	Some detectors are co-located – can this provide information on ground truth or individual device performance?
Operational and actionable output	What algorithm outputs must be usable by a responder, i.e. within seconds?  What algorithm outputs can be fused with other data within minutes?  Can a long term (i.e. over days) change in detector behaviour be spotted?
Maximising efficacy of a system of detectors	How much area is adequately covered by the detectors in this data set?  Can we know how many detectors are needed to cover an area?  What deployment techniques coupled with analysis techniques can optimise this coverage?
Learning for future trials/challenges	What additional information is required to investigate such a dataset to high complexity in the future?  What data in the set is superfluous or valueless?



# Challenge output

- R&D Initial and End workshops
  - Kick off event planned where the data format is presented and a questions asked/answered
  - FAQ event planned a few months in to allow any issues to be raised/iterated
  - Workshop planned around September 2023 where participants are expected to present something (even if it is just a list of “what we need next time”)
- Comparison to adjudicated output and other challenge work
  - Not a ground truth but a reasonable first approximation
  - Can look to understand differences
  - Small sigma London Pilot datasets with known moving sources do exist and may be shared at a future date
- Publishing
  - Publishing is welcome however all publications must undergo the AWE permission to publish process as an additional step





# How to engage

- Links
- Universities must agree to the below before being sent the data:
  - They are to use the data for the sole purpose of the activities outlined as part of the challenge(s) and will delete the data from all systems when those activities are finished or upon the request of HSG.
  - HSG remain the data owners so they will not share the data with any other party unless given specific permission from HSG.



# Calendar / Timeline of the challenge

- Launch 11th Oct
- Registration period
- Preparation and introduction event, early Nov
- Release of data, late Nov
- FAQ event, Jan
- Sigma Data Challenge Pilot projects proposals – deadlines 1st Dec, 1st Mar, 1st Jun
- Sigma data workshop, Autumn 2023
- Final NuSec event, Autumn/Winter 2023