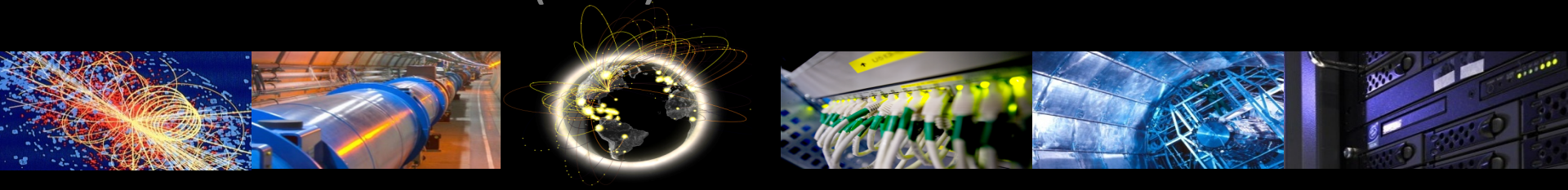
Harnessing the power of threat intelligence for WLCG cybersecurity

HEPiX October 16 2019

David Crooks (STFC/EGI CSIRT)
Liviu Vâlsan (CERN)



WLCG Worldwide LHC Computing Grid

egi CSIRT WISE COMMUNITY

CERN

Background

- The modern threat landscape facing research communities include actors that are sophisticated, motivated and well funded

- An important response to this is the sharing of threat intelligence within our community

Background

- If one site sees an attack they can share indicators to allow others to react
  - so called Indicators of Compromise (IoCs)

- Sites also require appropriate monitoring, storage and alerting software to make best use of this information

Security Operations Centre

- The purpose of a Security Operations Centre (SOC)
  - Gather relevant security monitoring data from different sources
  - Aggregate, enrich and analyse that data
  - Use it in the detection of security events and during any subsequent actions

- A SOC consists of a set of software tools and the processes connecting them

WLCG SOC WG

- Working group started in 2016
- Identified need to monitor cluster environment in a new context which can include virtualised / containerised systems
  - Potentially more opaque than existing grid systems
  - Effective network monitoring is critical

WLCG SOC WG

- Mandate to generate reference designs for deployment of SOC components for a range of site topologies
- Work enhanced by including neighbouring communities
  - NRENs
  - University CSIRTs
  - Other research communities
- Participants include Tier-1s, Tier-2s, NRENS, other academic institutes

SOC initial model

- Follow idea of *minimum viable SOC*

- The goal of this model is to
  - synchronise threat intelligence with a remote source
  - ingest security monitoring data
  - store it in a searchable repository and visualise it
  - enrich this data with threat intelligence
  - alert based on any consequent correlations

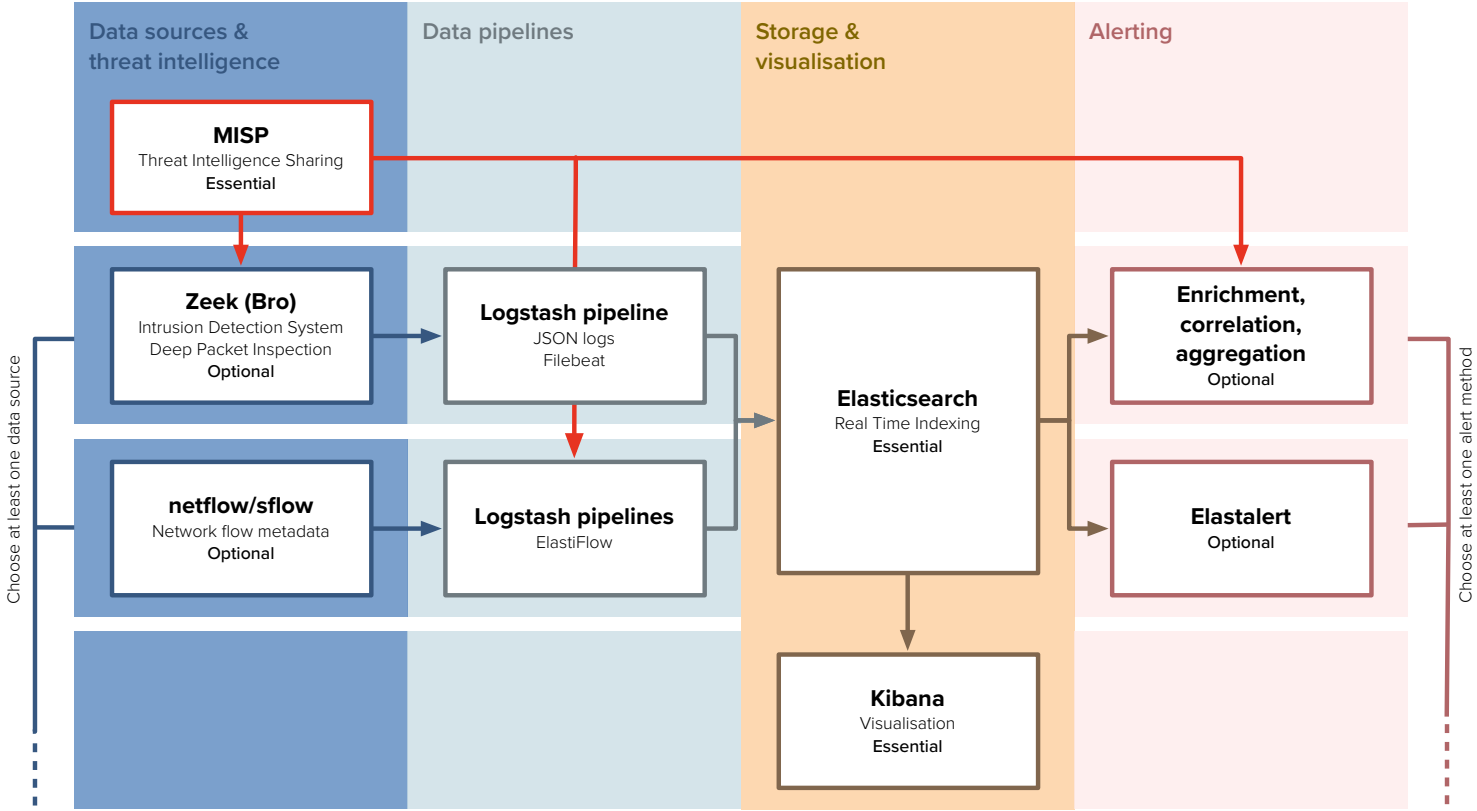
SOC initial model

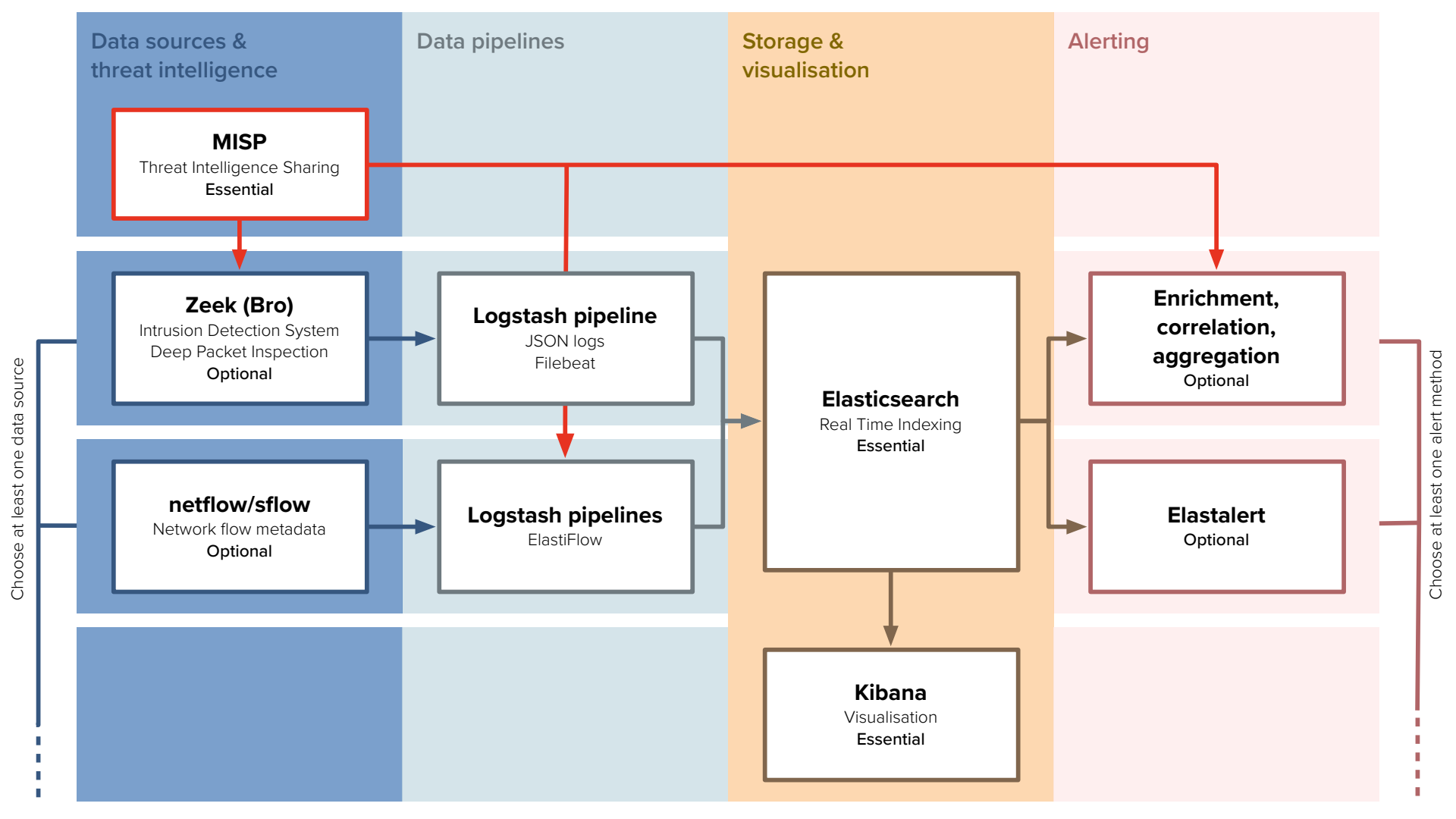
- In February, decided on SOC initial model consisting of the following stages
  - Data sources and threat intelligence
  - Data pipelines
  - Storage & visualisation
  - Alerting

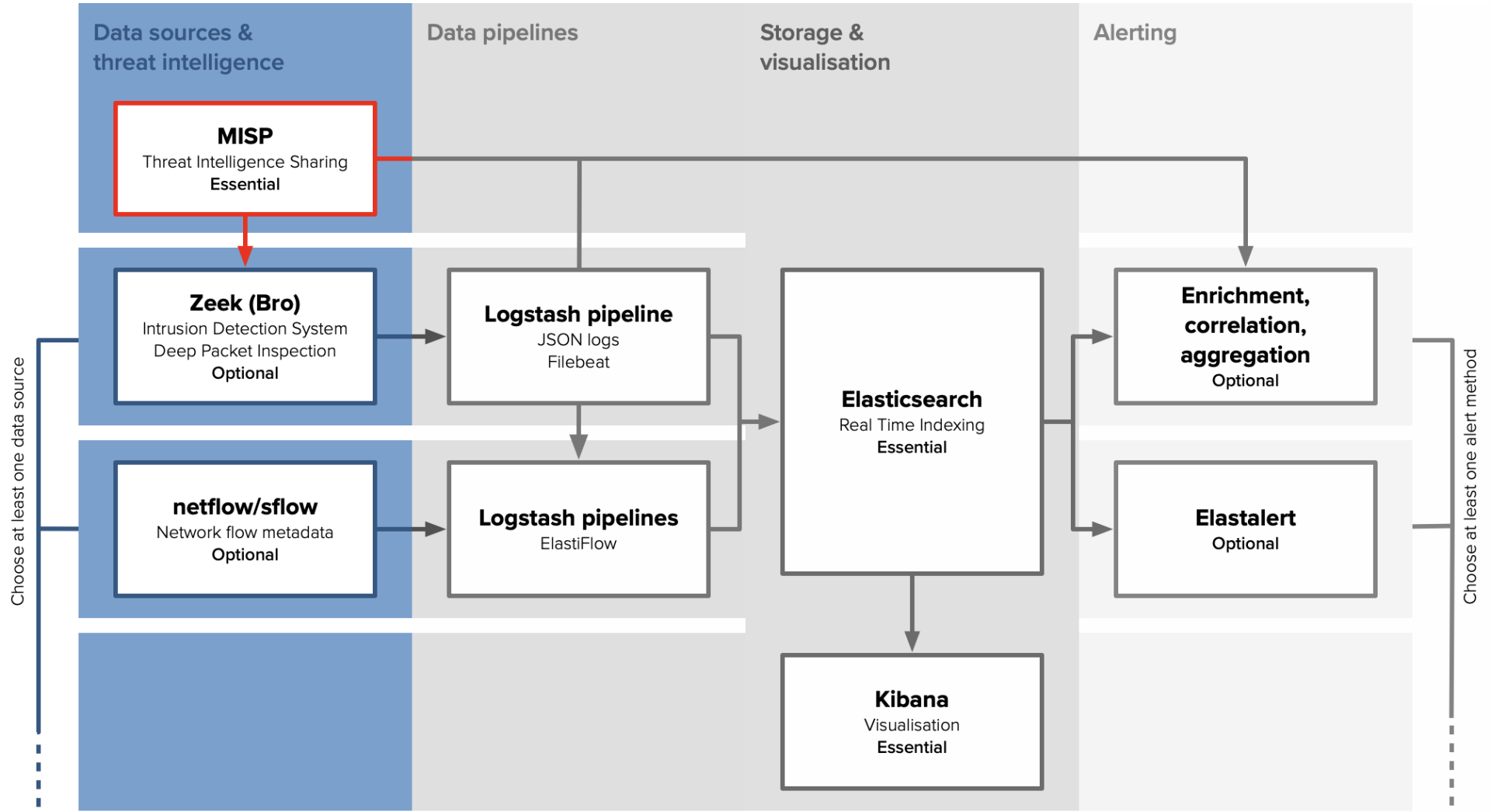
SOC initial model

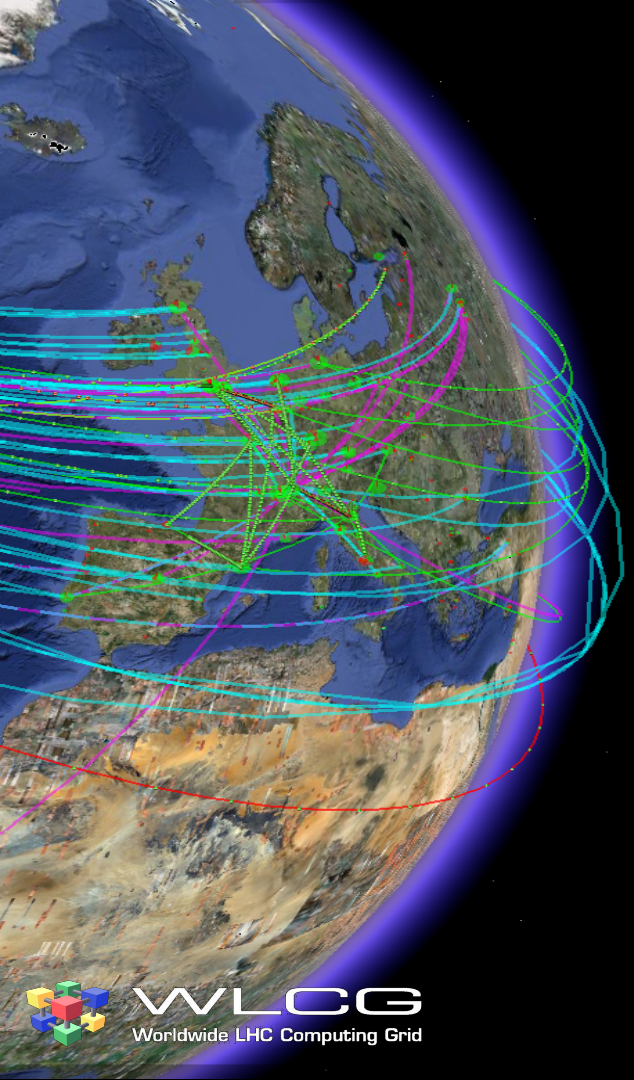
- Define 2 types of component
  - Essential
  - Optional (but require at least one)

SOC initial model




Data sources & threat intelligence
Data pipelines
Storage & visualisation
Alerting
MISP
Threat Intelligence Sharing
Essential
Zeek (Bro)
Intrusion Detection System
Deep Packet Inspection
Optional
netflow/sflow
Network flow metadata
Optional
Logstash pipeline
JSON logs
Filebeat
Logstash pipelines
ElastiFlow
Elasticsearch
Real Time Indexing
Essential
Kibana
Visualisation
Essential
Enrichment, correlation, aggregation
Optional
Elastalert
Optional
Choose at least one data source
Choose at least one alert method


Data sources & threat intelligence
Data pipelines
Storage & visualisation
Alerting
MISP
Threat Intelligence Sharing
Essential
Zeek (Bro)
Intrusion Detection System
Deep Packet Inspection
Optional
netflow/sflow
Network flow metadata
Optional
Logstash pipeline
JSON logs
Filebeat
Logstash pipelines
ElastiFlow
Elasticsearch
Real Time Indexing
Essential
Kibana
Visualisation
Essential
Enrichment, correlation, aggregation
Optional
Elastalert
Optional
Choose at least one data source
Choose at least one alert method



Threat Intelligence

Threat Intelligence

- Malware Information Sharing Platform (MISP) [Essential]
- *"A threat intelligence platform for gathering, sharing, storing and correlating Indicators of Compromise"*
  - *https://misp-project.org*
- Enables development of trust frameworks between sites to allow rapid sharing of threat intelligence
- Cornerstone of this work

WLCG MISP

- Specific external MISP instance hosted at CERN
  - Shares a portion of the intelligence from their trusted partners

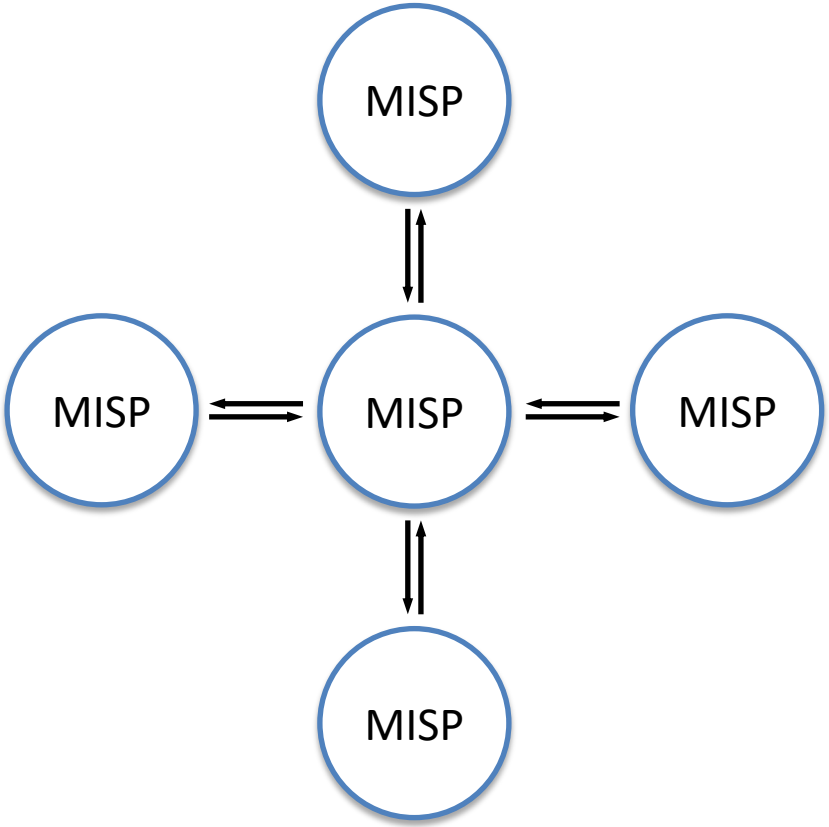
- Available for use by WLCG and neighbouring communities

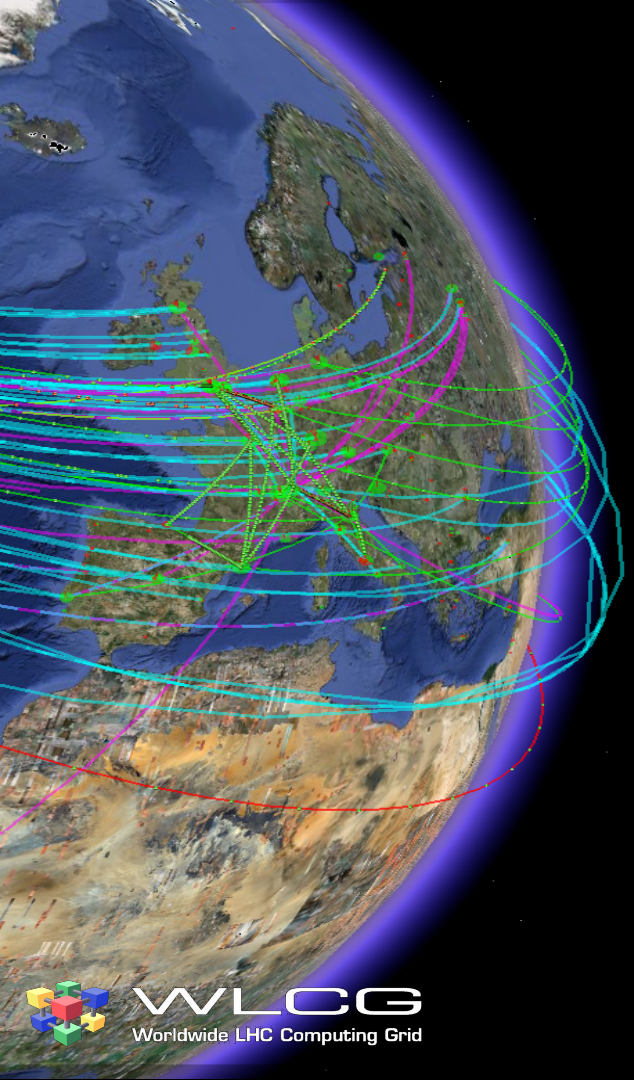
WLCG MISP

- Event types
  - Mostly TLP: GREEN/WHITE
  - TLP: AMBER for events created by CERN

- Access via CERN SSO
  - Focus on federated identity
    - EduGAIN + SIRTFI

WLCG MISP





Data sources

Data sources

- At least one of
  - Zeek (Bro): deep packet inspection
  - Netflow: network metadata

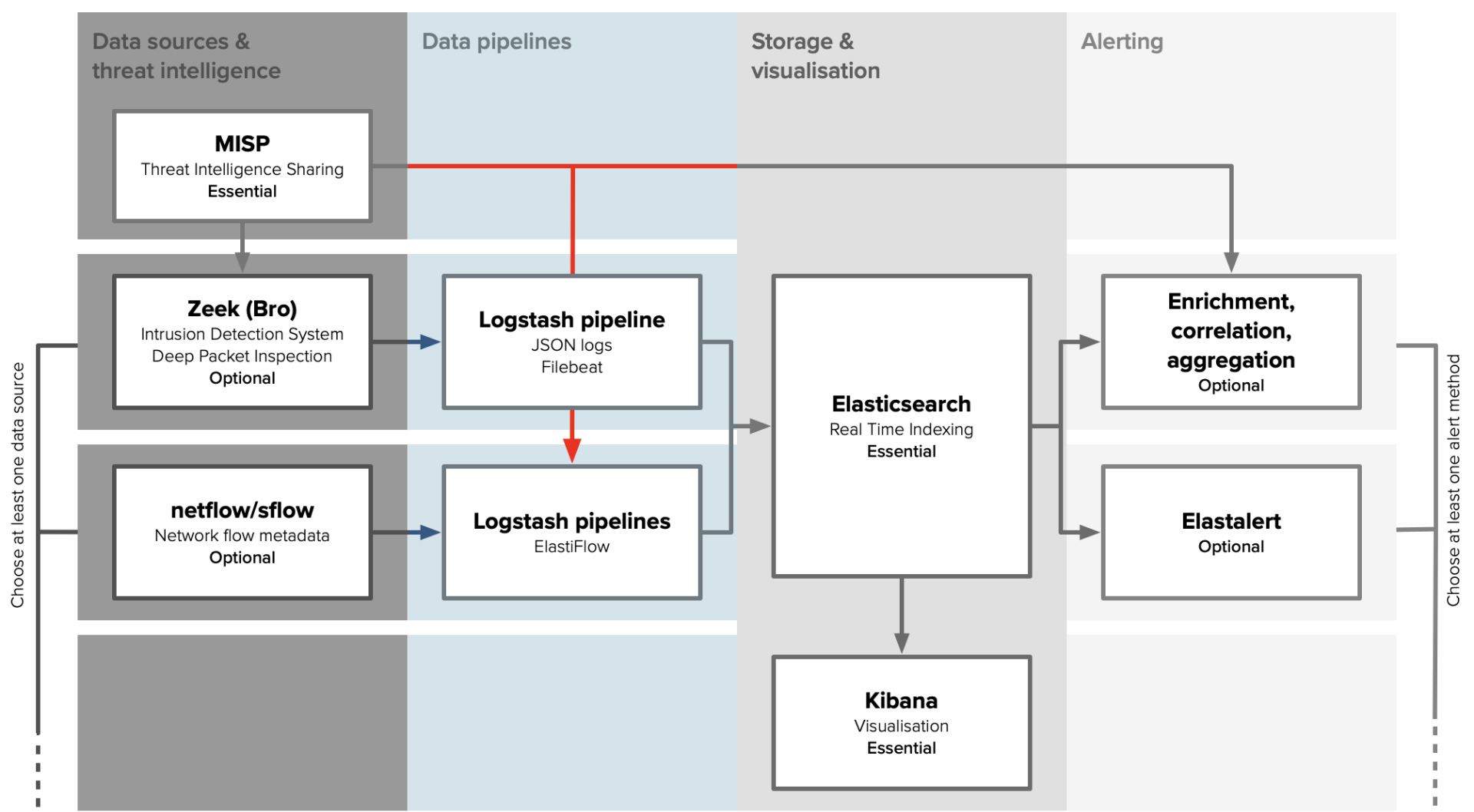
- Provide two options to hopefully cover range of use cases

Data sources

- Zeek
  - High level of information
  - Scalable and flexible
  - Dynamic protocol analysis
- However
  - Hardware implications
- Commercial options available

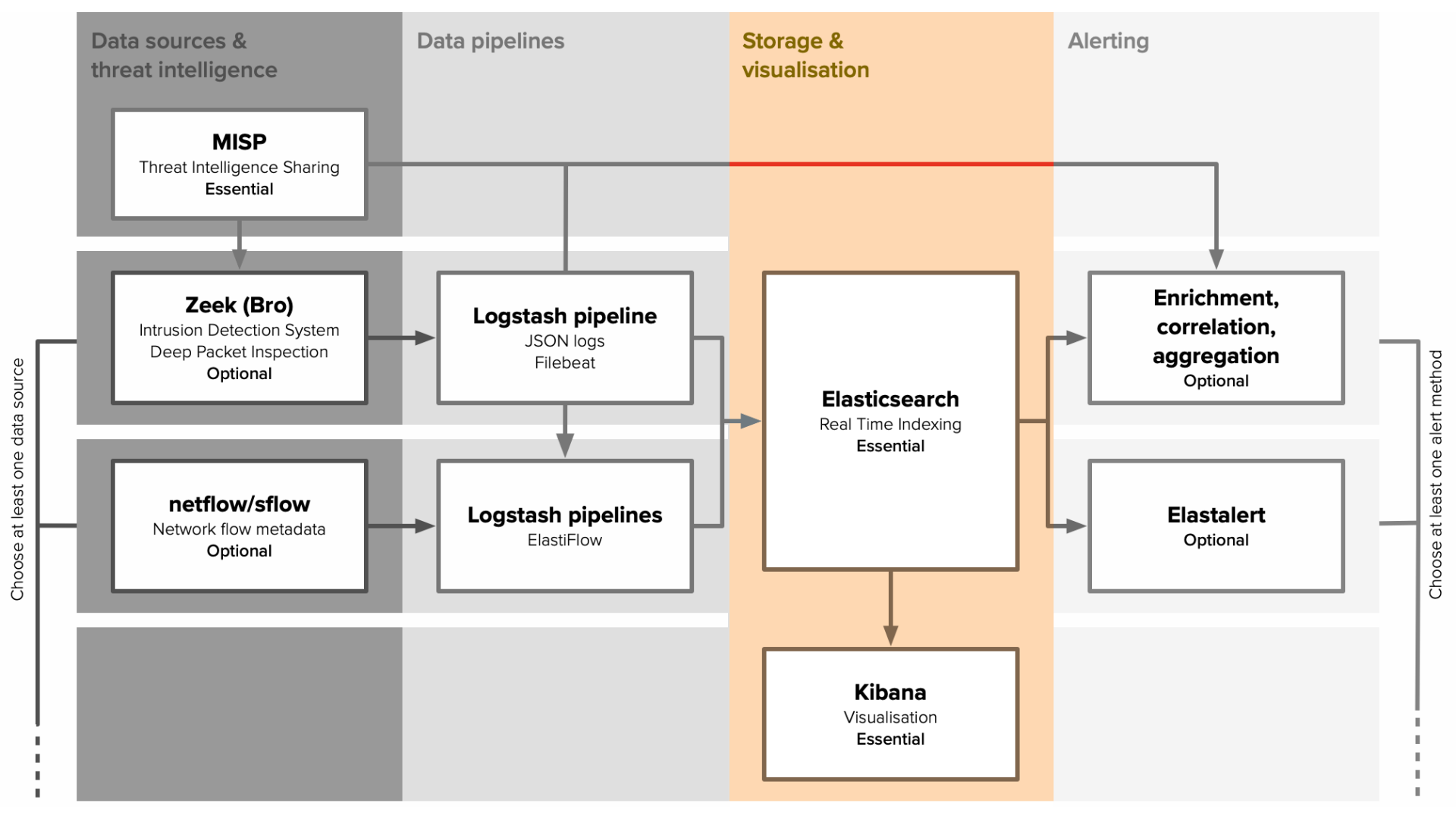
Data sources

- Netflow/Sflow
  - Network metadata
  - Many network vendors provide generators
  - Software clients
- However
  - Less data than Zeek


Data sources & threat intelligence
Data pipelines
Storage & visualisation
Alerting
MISP
Threat Intelligence Sharing
Essential
Zeek (Bro)
Intrusion Detection System
Deep Packet Inspection
Optional
Logstash pipeline
JSON logs
Filebeat
Elasticsearch
Real Time Indexing
Essential
Enrichment, correlation, aggregation
Optional
netflow/sflow
Network flow metadata
Optional
Logstash pipelines
ElastiFlow
Elastalert
Optional
Kibana
Visualisation
Essential
Choose at least one data source
Choose at least one alert method

Data pipelines

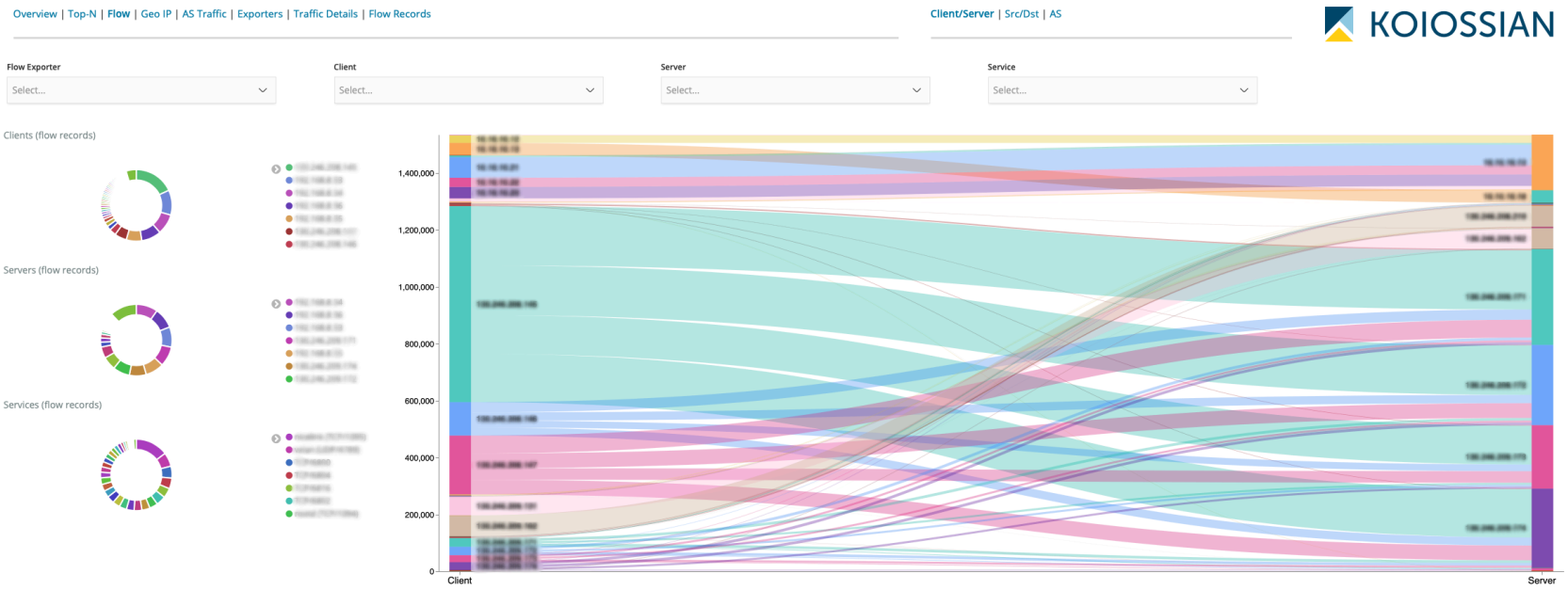
- Logstash [Essential]
- Pipelines to ingest data into Elasticsearch
  - Matched to data sources
  - Provide documentation for Zeek pipeline
  - Suggest use of Elastiflow for netflow pipeline


Data sources & threat intelligence
Data pipelines
Storage & visualisation
Alerting
MISP
Threat Intelligence Sharing
Essential
Zeek (Bro)
Intrusion Detection System
Deep Packet Inspection
Optional
Logstash pipeline
JSON logs
Filebeat
Elasticsearch
Real Time Indexing
Essential
Enrichment, correlation, aggregation
Optional
netflow/sflow
Network flow metadata
Optional
Logstash pipelines
ElastiFlow
Elastalert
Optional
Kibana
Visualisation
Essential
Choose at least one data source
Choose at least one alert method

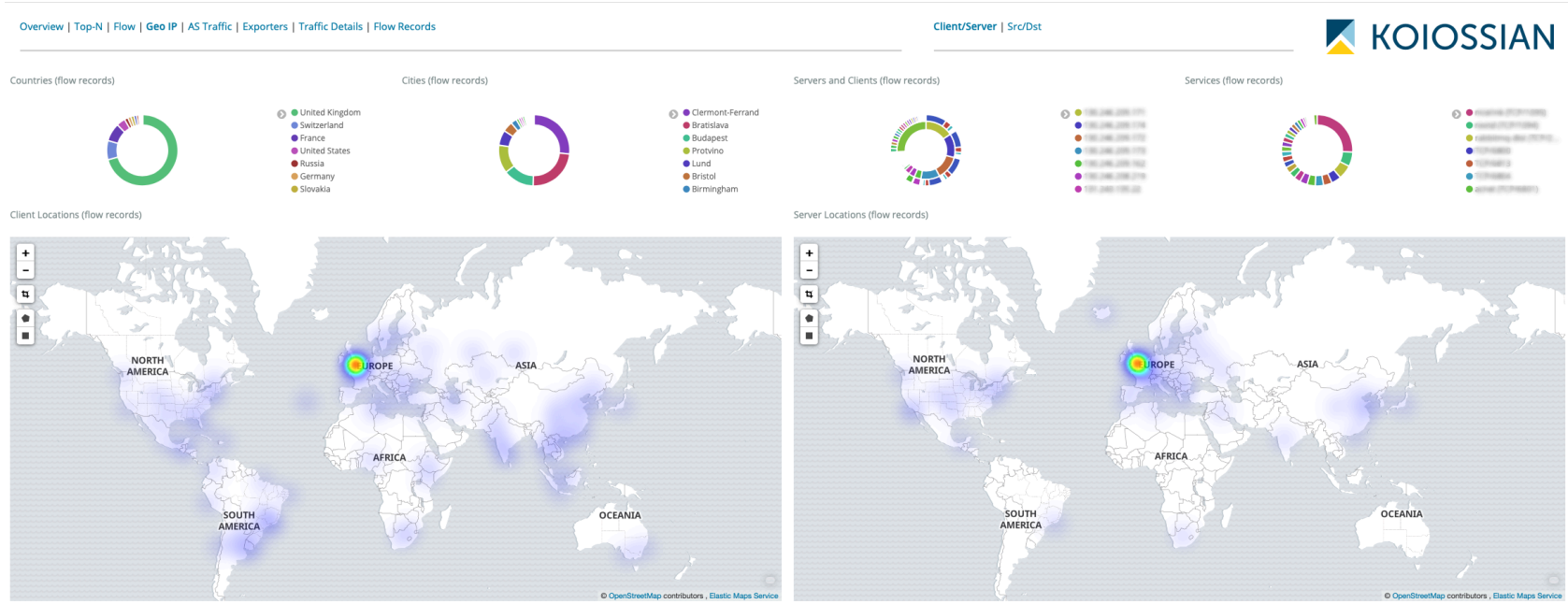
Storage & visualisation

- Elasticsearch [Essential]
  - Provide deployment tips based on experience of group members
- Kibana [Essential]
  - Provide some dashboards based on CERN SOC experience
  - Elastiflow provides dashboards for netflow visualisation

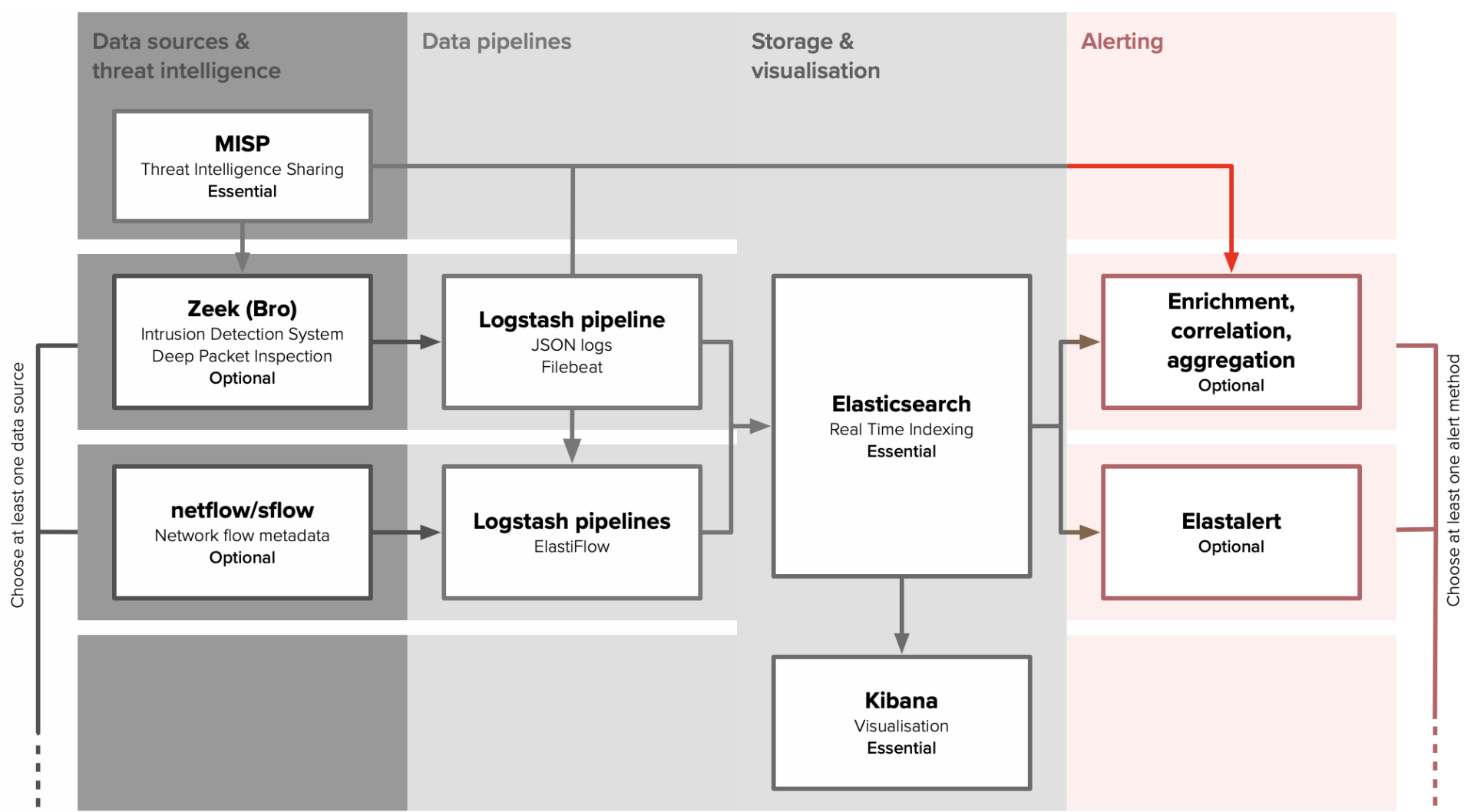
Elastiflow


Overview | Top-N | Flow | Geo IP | AS Traffic | Exporters | Traffic Details | Flow Records
Client/Server | Src/Dst | AS
KOIOSSIAN
Flow Exporter
Client
Server
Service
Select...
Clients (flow records)
Servers (flow records)
Services (flow records)
1,400,000
1,200,000
1,000,000
800,000
600,000
400,000
200,000
0
Client
Server

Elastiflow



HEPiX, October 2019


Data sources & threat intelligence
Data pipelines
Storage & visualisation
Alerting
MISP
Threat Intelligence Sharing
Essential
Zeek (Bro)
Intrusion Detection System
Deep Packet Inspection
Optional
Logstash pipeline
JSON logs
Filebeat
Elasticsearch
Real Time Indexing
Essential
Enrichment, correlation, aggregation
Optional
netflow/sflow
Network flow metadata
Optional
Logstash pipelines
ElastiFlow
Elastalert
Optional
Kibana
Visualisation
Essential
Choose at least one data source
Choose at least one alert method

Alerting

- At least one of
  - Enrichment, correlation and aggregation scripts based on CERN example
  - Elastalert
    - Trigger on Elasticsearch query
    - Spike of events
    - Matching on field content

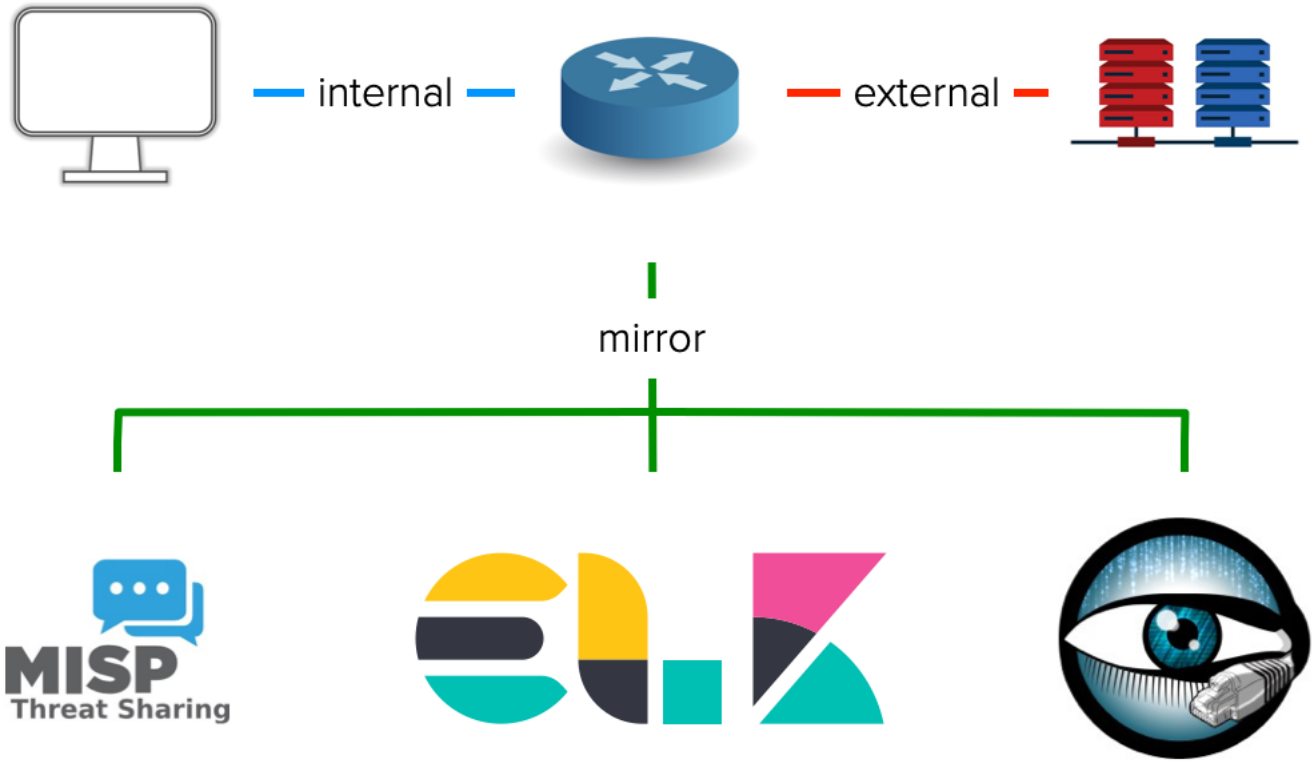
Current implementations

- AGLT2
  - Zeek/Netflow + Elasticsearch + MISP
- STFC Cloud (this summer)
  - Sflow + Elasticsearch + MISP
- Nikhef (this summer)
  - Zeek + Elasticsearch + MISP

PocketSOC

- SOC demonstrator
  - Docker cluster designed to run on a laptop
  - Essential components and network elements
  - Minimal traffic to demonstrate workflow
  - Test new components

PocketSOC

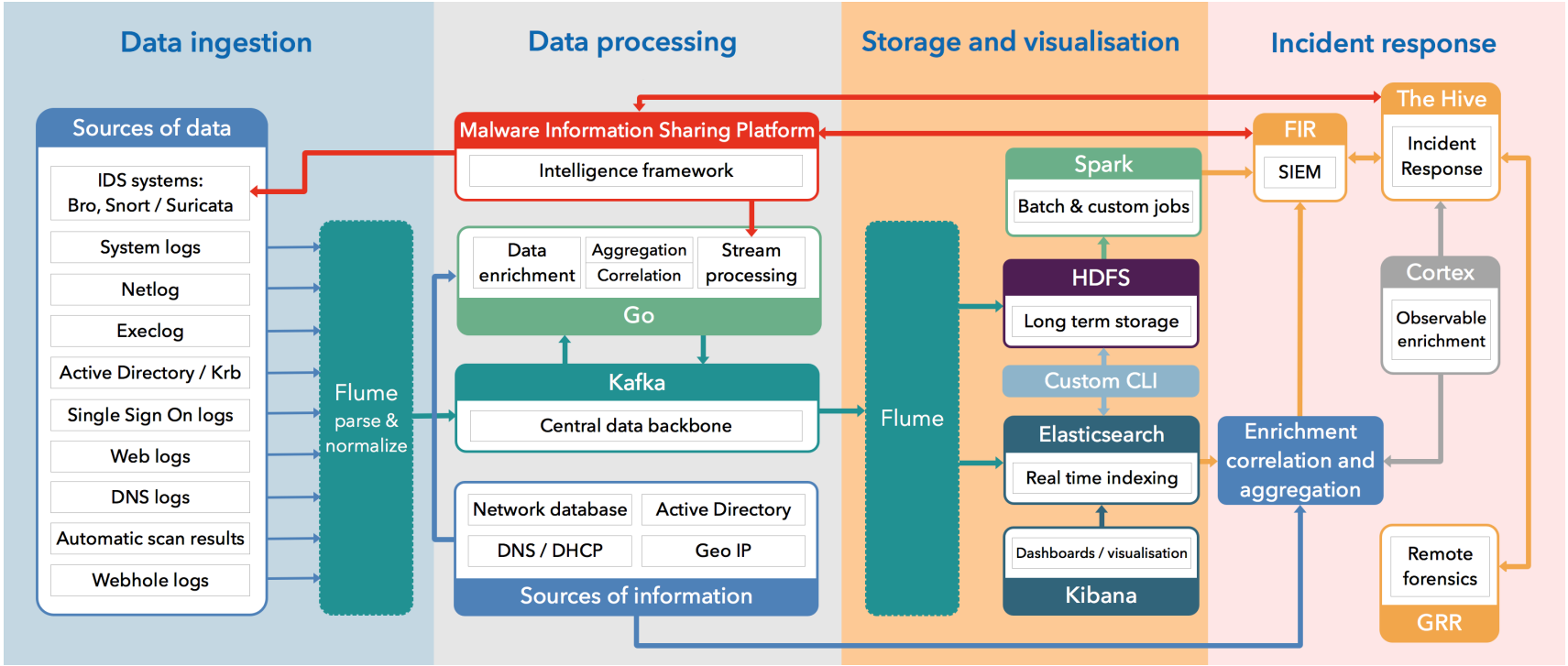


CERN SOC

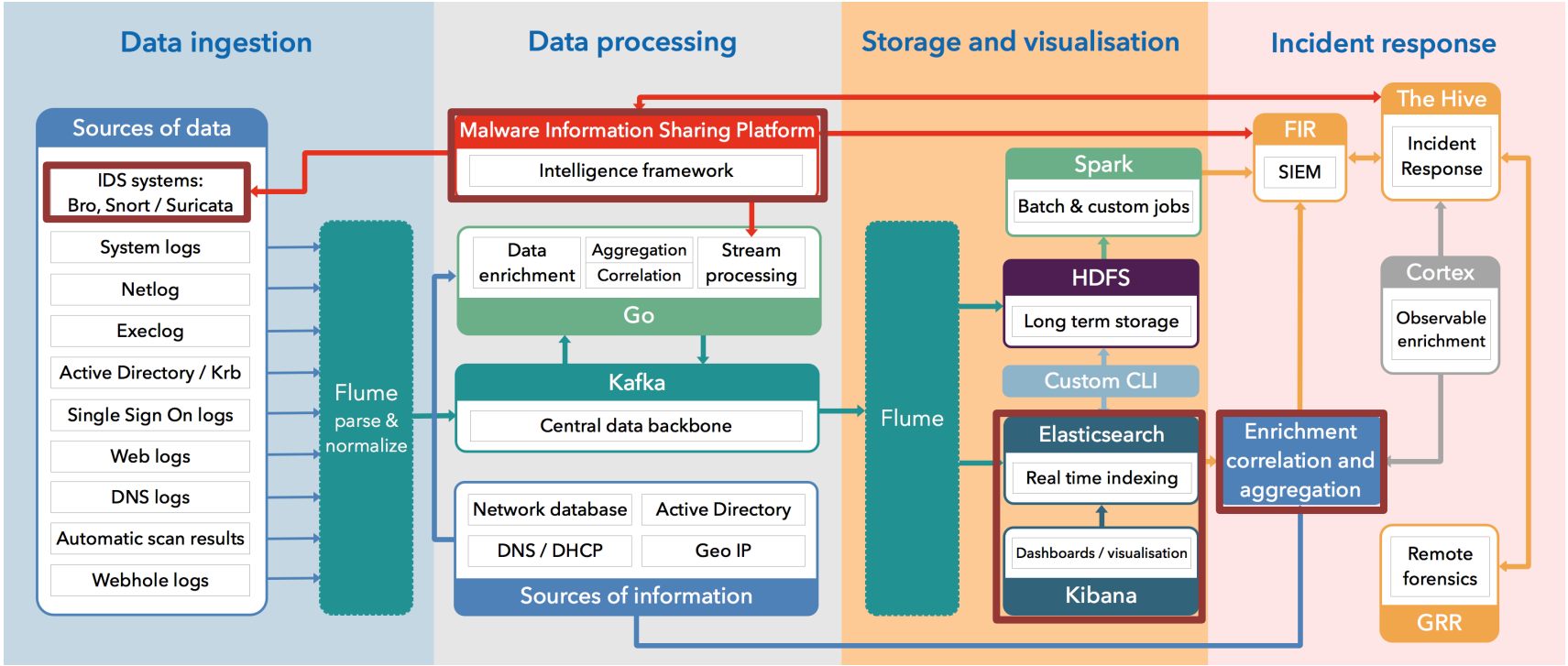
- Closely following and benefitting from work on the CERN SOC

- Gives indicators for future direction of the SOC model

CERN SOC


Data ingestion
Data processing
Storage and visualisation
Incident response
Sources of data
IDS systems: Bro, Snort / Suricata
System logs
Netlog
Execlog
Active Directory / Krb
Single Sign On logs
Web logs
DNS logs
Automatic scan results
Webhole logs
Flume
parse & normalize
Malware Information Sharing Platform
Intelligence framework
Data enrichment
Aggregation
Correlation
Stream processing
Go
Kafka
Central data backbone
Network database
Active Directory
DNS / DHCP
Geo IP
Sources of information
Flume
Spark
Batch & custom jobs
HDFS
Long term storage
Custom CLI
Elasticsearch
Real time indexing
Dashboards / visualisation
Kibana
FIR
SIEM
The Hive
Incident Response
Cortex
Observable enrichment
Enrichment correlation and aggregation
Remote forensics
GRR

CERN SOC


Data ingestion
Data processing
Storage and visualisation
Incident response
Sources of data
IDS systems: Bro, Snort / Suricata
System logs
Netlog
Execlog
Active Directory / Krb
Single Sign On logs
Web logs
DNS logs
Automatic scan results
Webhole logs
Flume
parse & normalize
Malware Information Sharing Platform
Intelligence framework
Data enrichment
Aggregation
Correlation
Stream processing
Go
Kafka
Central data backbone
Network database
Active Directory
DNS / DHCP
Geo IP
Sources of information
Flume
Spark
Batch & custom jobs
HDFS
Long term storage
Custom CLI
Elasticsearch
Real time indexing
Dashboards / visualisation
Kibana
FIR
SIEM
The Hive
Incident Response
Cortex
Observable enrichment
Enrichment correlation and aggregation
Remote forensics
GRR

Next SOC Workshop

- 4th SOC Workshop
  - Next week!
    - 21-23 October @ Nikhef
  - Major focus: threat intelligence sharing
  - Validation of SOC workflow
  - Roundtable discussing social and technical issues of intelligence sharing

Contact

- Main working group page
  - https://wlcg-soc-wg.web.cern.ch
- Documentation
  - https://wlcg-soc-wg-doc.web.cern.ch