



UNIVERSITY
of York

CREED: An academic's perspective

Dr John Wilkinson

University of York

Outline

- Setting the scene: What got us here?
- The value of CREED
- Is CREED a ‘silver bullet’?
- Looking forward & beyond CREED

**An academic's
perspective**

The Need

- Hundreds of thousands of chemicals in use globally which may reach the env.
- Environmental risk assessment depends on the toxicity of a contaminant & exposure to it.
- Both data sets need to be reliable, relevant and accurate.



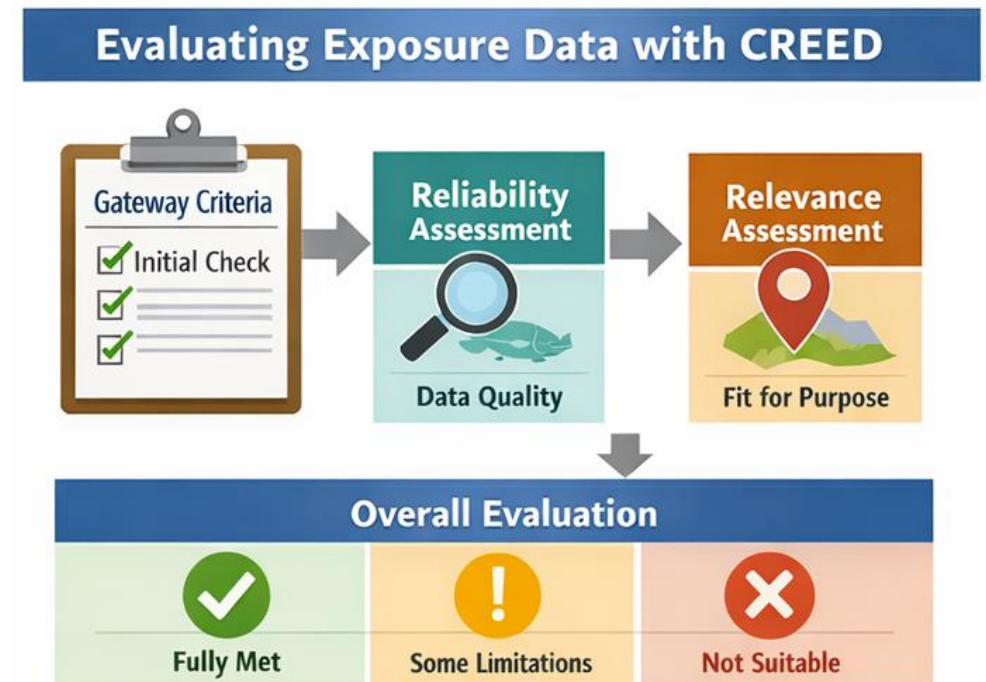
Before CREED

- We've had criteria to evaluate data on the toxicity side of the equation (e.g., Criteria for reporting and evaluating ecotoxicity data: CRED) but not for exposure data.
- Resulted in an ad hoc and inconsistent evaluation of environmental exposure data.
- Specialist skills needed to make informed decisions on data reliability.



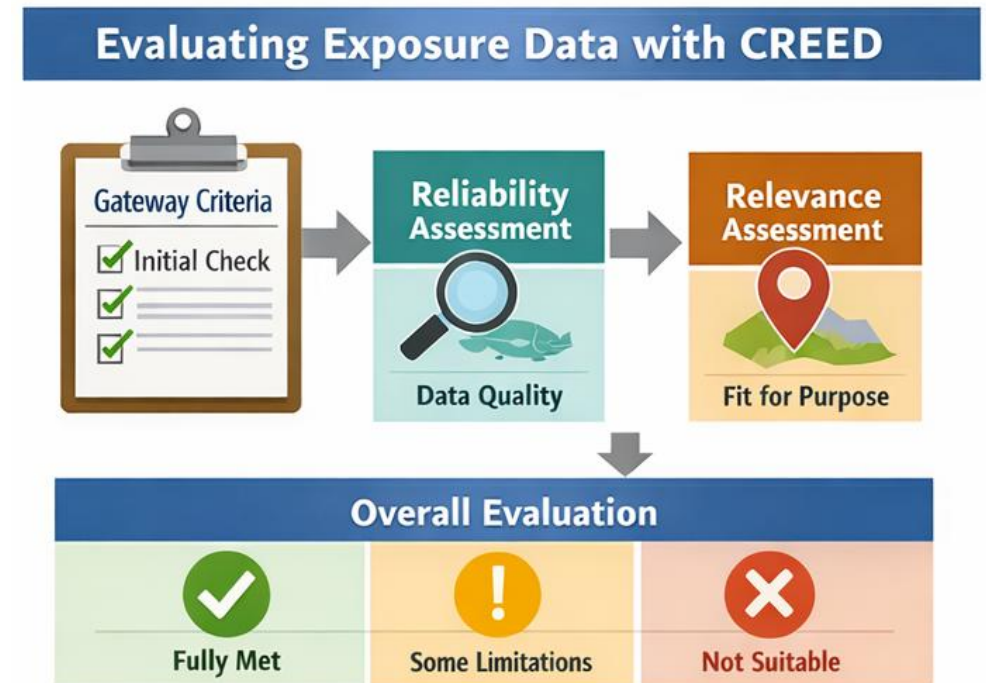
Criteria for Reporting and Evaluating Exposure Datasets

- CREED fills this gap by providing criteria to evaluate exposure data for relevance and reliability.
- Encourages fit-for-purpose data uses (users must first define the assessment purpose).



Criteria for Reporting and Evaluating Exposure Datasets

- Improved transparency and structure (standardization) in how exposure data are evaluated.
- This allows assessors to make informed decisions even when they are not specialists in all aspects of analytical chemistry or monitoring (confidence).

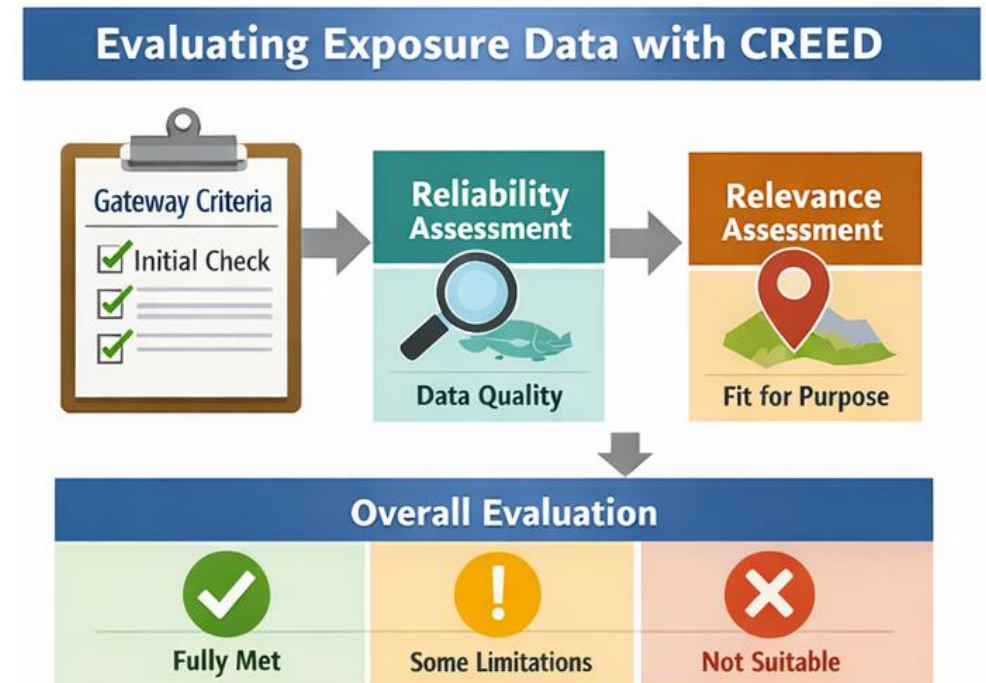


Criteria for Reporting and Evaluating Exposure Datasets

An academic ‘positive’:

CREED lowers the barrier to engaging with exposure data, enabling a wider group of assessors, including students, to evaluate datasets in a structured and transparent way, while still recognising the role of expert judgement.

(my perspective)



Uses for both data generators & users

Criteria for **Reporting** and **Evaluating** Exposure Datasets

How do I know that this dataset is appropriate and reliable for my intended risk assessment?

How can I generate a quality environmental exposure dataset that could be more widely used by the scientific community?



CREED for a data generator

- Most CREED discussion seems to focus on the data assessor, not the data generator.
- Can act as a checklist for reporting environmental monitoring studies:
 - Clarifies what information must be reported for datasets to be usable,
 - Encourages study design that supports exposure assessment,
 - Could encourage better quality and consistency between monitoring studies.



This is all well and good, so...

...is CREED a 'silver bullet'?

Unfortunately, probably not...

(at least from my perspective)



A silver bullet?

Expert Judgment:

- CREED makes (what can be) subjective decisions more transparent and consistent, but it does not remove subjectivity altogether.
 - My MSc students often come to different outcomes even when evaluating the same dataset.
- Disagreements can still happen.



A silver bullet?

Expert Judgment:

- Academically, CREED makes this subjective judgement teachable and defensible.
- When there are disagreements, they can be traced to specific criteria:
 - Something you can point to, explain/ debate.



Expert opinions may vary...

A silver bullet?

Uncertainty:

- CREED makes sources of uncertainty explicit but leaves decisions on how to handle that uncertainty to the assessor.
- Presence of uncertainty, not its magnitude or consequences.



A silver bullet?

Taking a step back:

- Reliability/quality does not start with the data, it starts with the analytical approach to generating it.
- Data use can only be as good as the analytical quality of how data were generated
 - Quantitation quality, instrumentation, chemical standards, QA/QC, validation parameters, etc.
 - Analytical causes of uncertainty can really matter



Expert judgment still essential.

A silver bullet?

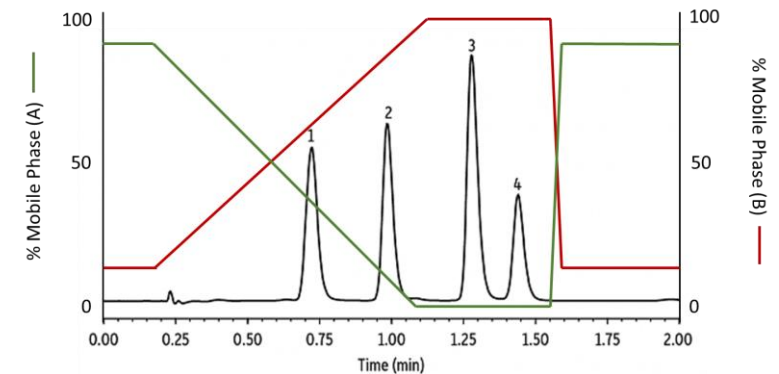
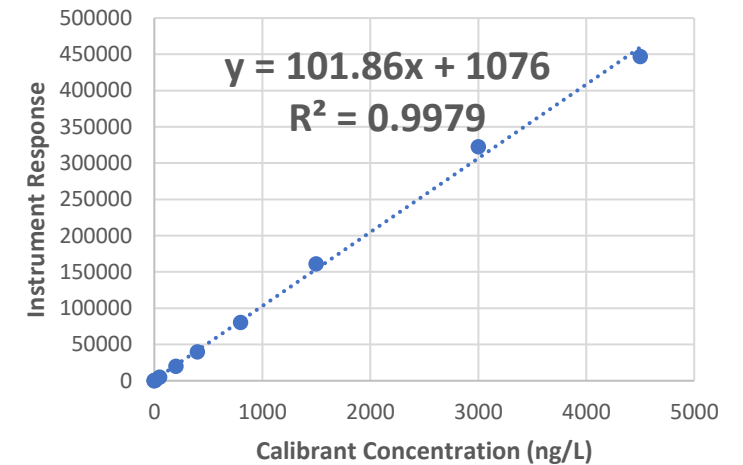
- How can (or should) uncertainty identified through frameworks like CREED be quantitatively translated into risk assessment?
- How does uncertainty in an analytical method (e.g., QA/QC) propagate into data use/ risk assessment?



QA/QC

Looking beyond CREED

- CREED is particularly well suited to evaluating traditionally generated exposure datasets, such as:
 - Targeted monitoring studies with well-defined sampling and analytical methods.
 - Regulatory datasets.
 - HPLC-MS/MS workflows.
- Not all environmental monitoring data fall into these categories.



Looking forward

- How can we take the positives of CREED and apply them to a wider range of environmental monitoring datasets like:
 - Passive sampling
 - High Resolution Mass Spectrometry
 - Semiquantitative/ sensor networks
 - Citizen science?!



Beyond CREED

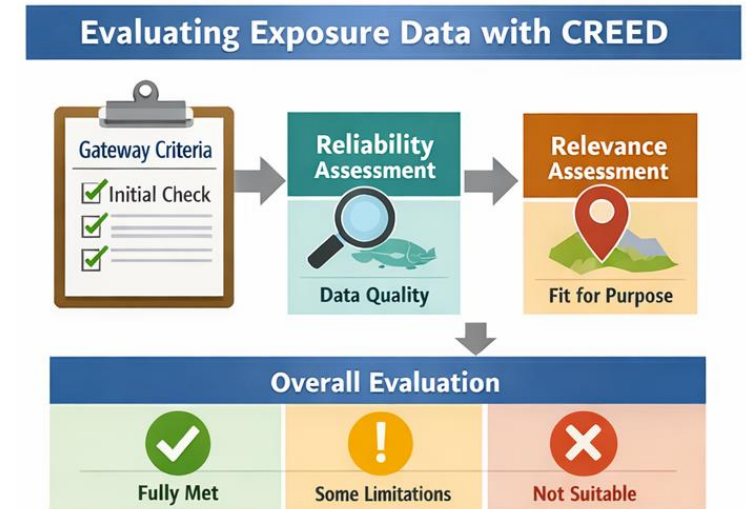
If the goal is reliable data, can (or should) we develop criteria to support the methods/production of environmental exposure data, not just assessment?

- Criteria on sampling design, collection, & analysis for generating exposure datasets
- Quality controls
- Analytical method validation criteria/standards
- Quantifying uncertainty in data analysis



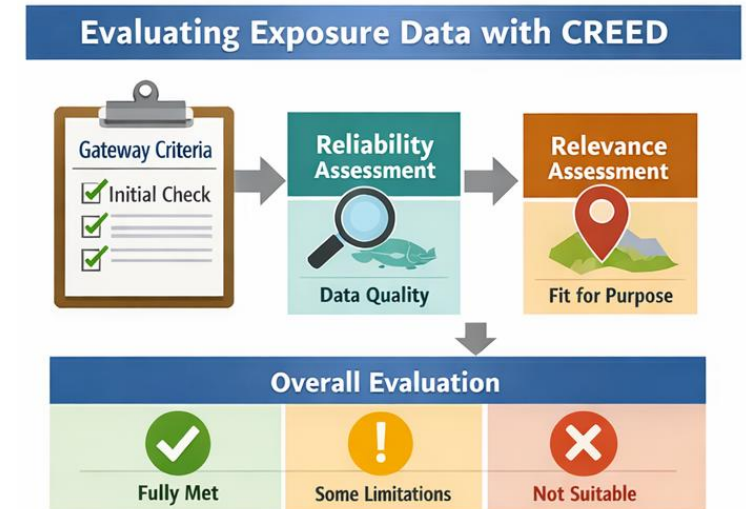
Take aways

- Although unlikely a ‘silver bullet’ CREED is an important step toward harmonising exposure data evaluation.
- It brings exposure assessment closer to the level of rigour already applied to hazard/ecotox data.



Take aways

- It moves subjective expert judgement into something structured and transparent.
- Critically, it provides a platform on or example of which next-generation approaches can built.



Questions?

The CREED framework – a regulator's perspective



Brett Sallach, Kerry Sims, Ian Doyle, Emma Pemberton. Environment Agency

RSC workshop, 24th March 2026



Outline

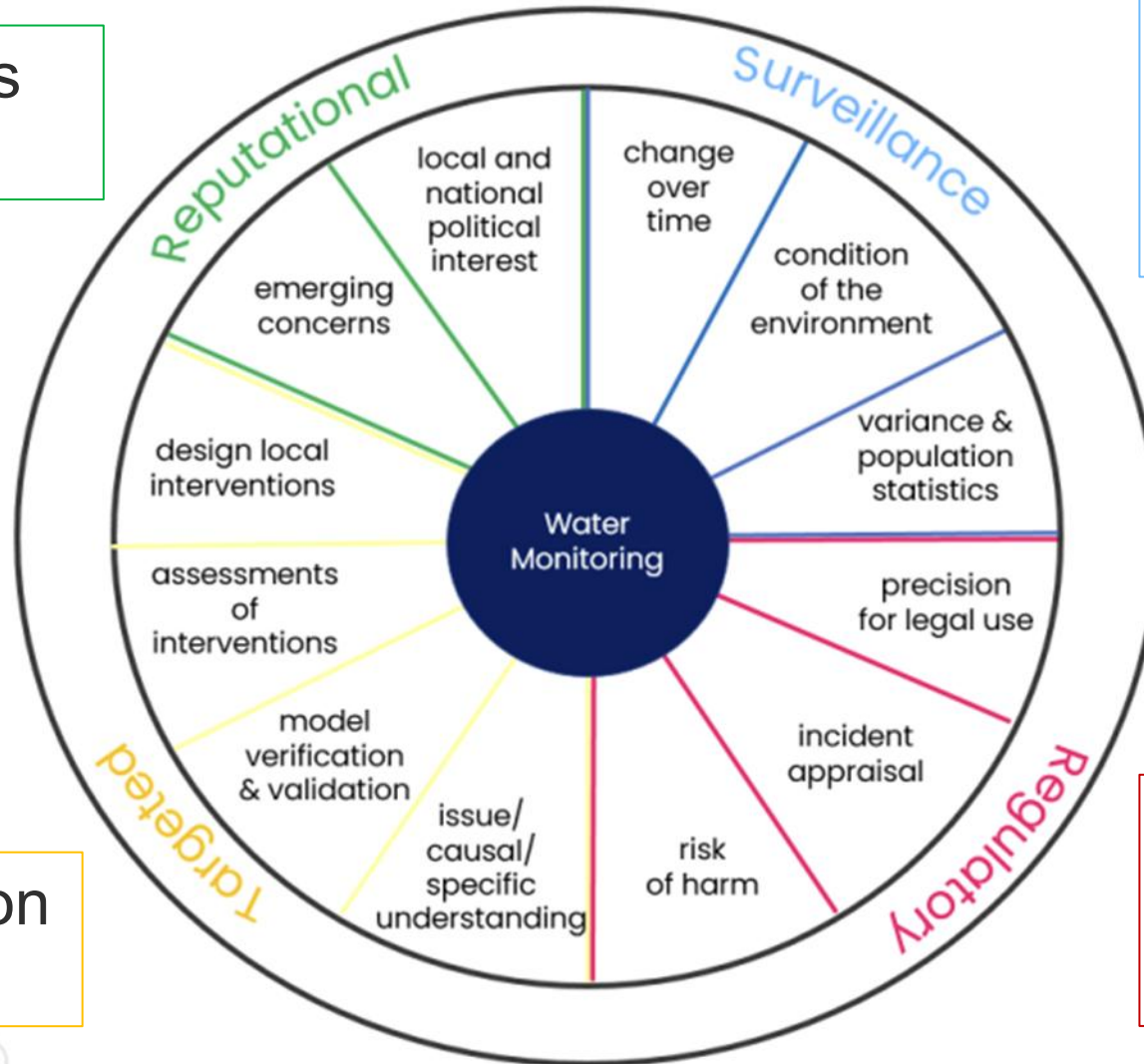
- Why we monitor
- Data quality
- Importance of meta-data
- Making our data available
- Future plans for our monitoring



[Water Data Explorer](#)

Why we monitor?

Emerging Substances Network

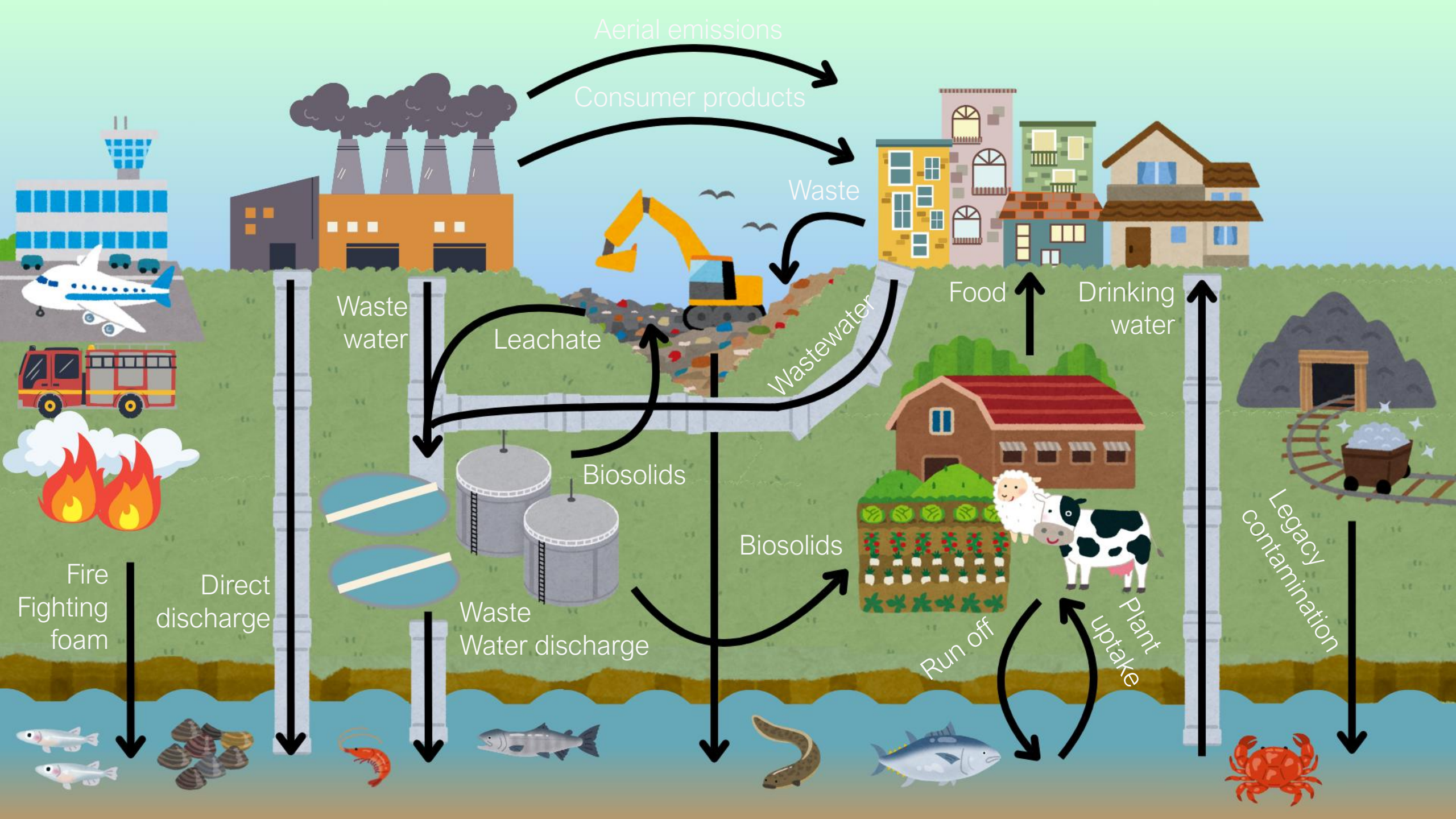


Natural Capital & Ecosystem Assessment Programme- e.g. River Surveillance Network

Chemicals Investigation Programme

Permit Conditions- Discharge compliance/assurance

Water Information Needs framework



Environment Agency Monitoring Networks

Emerging Substances Network

- 21 sites
- Identify emerging chemicals



River Surveillance Network

- 100 sites
- Determine long-term trends

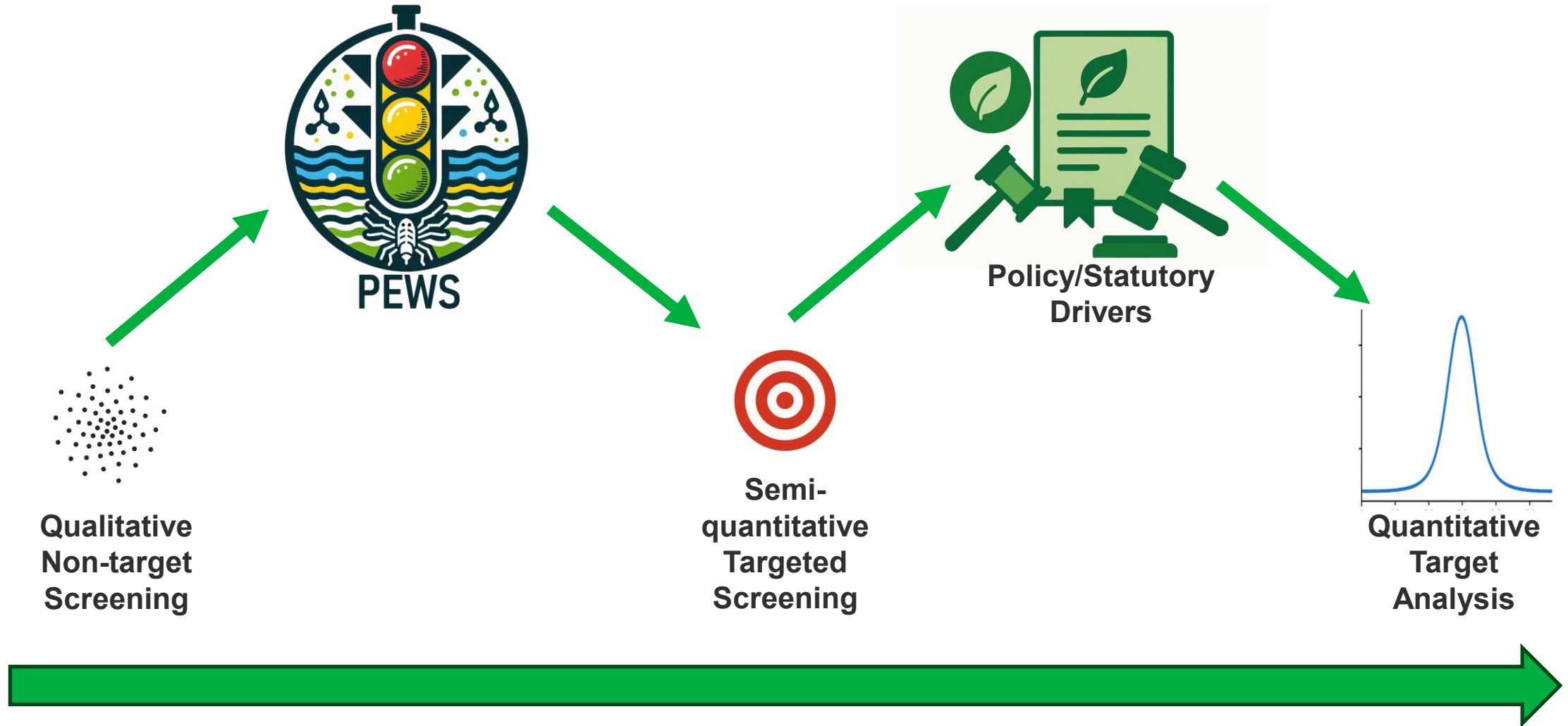


EQSD Surveillance Network

- 83 sites
- Statutory reporting for Priority Substances (WFD)



Smart Monitoring



Reliability : Resource

Where we are headed

- Independent Water Commission Report
- Internal Monitoring Programme Review
 - Scientific assurance assessed
- EA2030 Strategy Document Release



Natural Capital and Ecosystem Assessment

Analysis Ready Data Release

12.12.25

**A Milestone for Environmental
Monitoring**

Benefits of CREED - the regulatory view



Objective framework for evaluation of exposure data



Sets standard for all when generating monitoring data



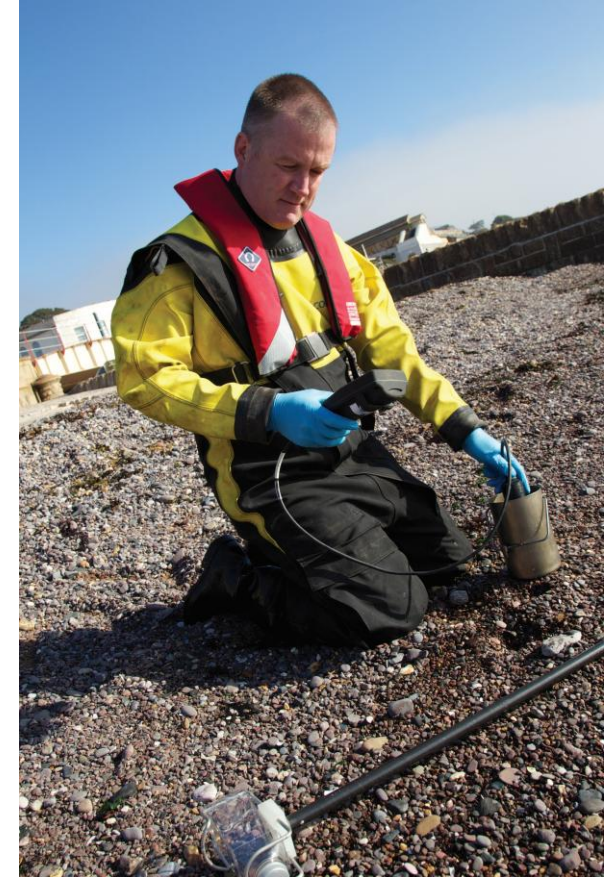
Provide guidance for contracting monitoring



Long term, larger, reliable evidence base

Conclusions

- CREED is a useful framework for fair and transparent evaluation of data
- Our environmental monitoring programmes are vast
- Pragmatic balance between cost and purpose dictates data generated
- A plea for all scientists to report sufficient detail to support a CREED assessment when publishing
- We are committing to make more meta-data available for our data





CREED and a Risk Assessors Perspective

Graham Merrington, Iain Wilson, wca environment Ltd

- Introduction
- Chemical risk assessment
- Is there poor practice regarding exposure data use?
- Would CREED help?
 - » Transparency
 - » Consistency
 - » Reporting
- Conclusions

- Compares effects/hazards to modelled or measured environment data (**PEC or MEC: PNEC**)
- Hazard and effects data mostly laboratory ecotoxicity tests with well studied organisms
- Used to develop PNEC – reliability and relevance assessed by CRED (Klimisch, ToxRTool, national protocols)
- Measured chemicals in the environment – exposure data (PEC or MEC)
- Before CREED, no accepted framework to assess reliability or relevance - best professional judgement – inconsistency



Poor practice (poor environmental outcomes)?



- The Water Framework Directive – risk based chemical regulation – EU wide risk assessment
- Legally binding standard derived if a risk identified
- Approach to compare PNEC with 95th percentile of EU exposure data from surface waters
- Substitution approach (1/2 LoD) or complete removal used for censored data (JRC 2025)

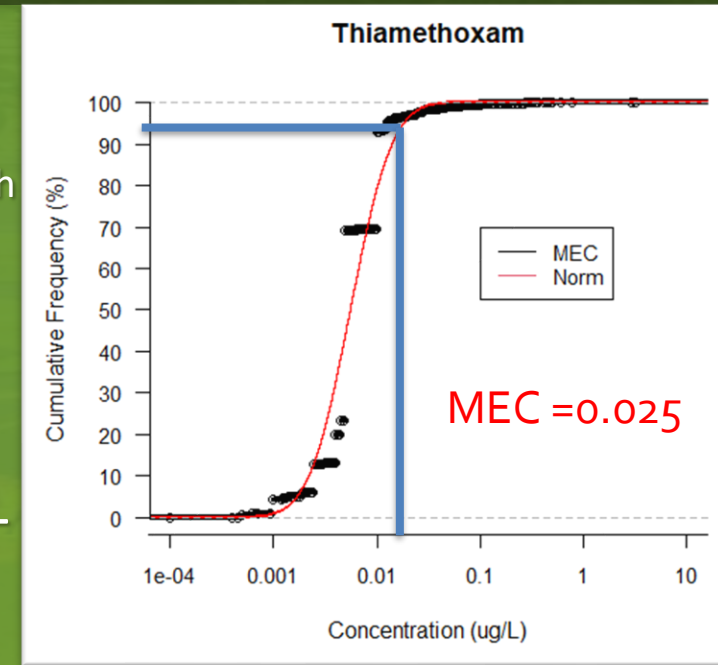
Modified from Merrington et al 2021

Substance	% of unquantified data	Substance	% of unquantified data	Substance	% of unquantified data
Acetamiprid	99	Deltamethrin	0	Triclosan	46
Clothianidin	93	Esfenvalerate	47	Nicosulfuron	52
Imidacloprid	77	Permethrin	75	Diclofenac	27
Thiacloprid	97	Bisphenol A	65	Silver	60
Thiamethoxam	95	Glyphosate	68	Carbamazepine	34

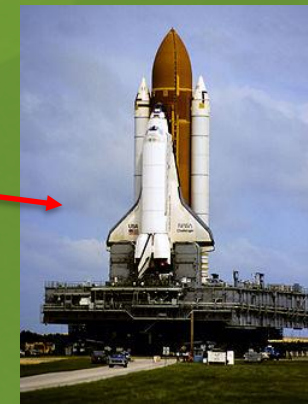
Poor practice (poor environmental outcomes)?



- Substitution approach (especially with multiple LoDs -20 for Estrone) is poor practice for highly censored data (esp. 95th %ile, when data >95% censored)
- There is a practical upper limit in the degree of censoring that nonparametric methods can accommodate (about 80%) – high variability in estimates
- Likely over estimation of 95th percentiles using substitution, removal of censored data means removal of information
- So, do the data really reflect the exposures?
- Is there really an EU-wide risk?



JRC 2022, Final Thiamethoxam Dossier



Poor practice (poor environmental outcomes)?



- NORMAN: (EMPODAT Database - Chemical Occurrence Data)

Category of data	▲	%	No. of Data
Adequately supported by quality-related information		0.07	71,351
Minimal quality-related information		3.59	3,470,805
Not supported by quality-related information		95.61	92,376,780
Supported by limited quality-related information		0.72	700,082

- UBA: (Database - Pharmaceuticals in the environment) “quality flag of each database entry refer to the reliability, plausibility and applied analytical standards of each publication. However, the quality of a publication is difficult to assess and can also be a matter of subjectivity, thus the information on data quality should be seen as an indicator only.”
- Reviewed surface water regulatory monitoring data from 36 countries and states across Europe. Purpose was to use data for bioavailability assessments with metals EQS. 23 of 36 not assignable using CREED
- **If measured data quality not considered important - then risk assessments fundamentally flawed**

Would CREED help?

Transparency



- Purpose is key – for data generators and users
- Regulators can write purpose statement for dischargers to use.....
- Drawbacks and uncertainties are clear – ‘why have you not used my data?’, ‘why have you used these data?’
- Limitations explicitly stated/reported for relevance and the reliability
- Reduce subjectivity, focus BPJ on purpose statement NOT on data evaluation



Would CREED help?

Consistency



- Baseline for data reliability and relevance
- Data generators clear on what is required to enable data to be used for specific purposes – e.g. meet regulatory purposes
- All risk assessors, using reliable and relevant data for the same purpose
- Basis for ensuring data is... 'fit for purpose'



Would CREED help? Reporting



- Most information known, just NOT reported
- Improve usability and utility of all measured chemical data
- Maximise use of finite resources
- Best practice recognised and understood
- Excel based reporting template

SETAC PRESS Environmental Toxicology and Chemistry, Vol. 35, No. 5, pp. 1297-1309, 2016
Published 2015 SETAC
Printed in the USA

Hazard/Risk Assessment

CREED: CRITERIA FOR REPORTING AND EVALUATING ECOTOXICITY DATA

CAROLINE T.A. MOERMOND,*¹ ROBERT KASE,² MURIS KORKARIC,³ and MARLENE ÅGERSTRAND⁴
¹Centre for Safety of Substances and Products, National Institute for Public Health and the Environment RIVM, Bilthoven, The Netherlands
²Swiss Centre for Applied Ecotoxicology, Dübendorf, Switzerland
³Department of Environmental Toxicology, EAWAG, Dübendorf, Switzerland
⁴Department of Environmental Science and Analytical Chemistry, Stockholm University, Stockholm, Sweden

(Submitted 5 June 2015; Returned for Revision 1 September 2015; Accepted 18 September 2015)

Abstract: Predicted no effect concentrations (PNECs) and environmental quality standards (EQSs) are derived in a large number of legal frameworks worldwide. When deriving these safe concentrations, it is necessary to evaluate the reliability and relevance of ecotoxicity studies. Such evaluation is often subject to expert judgment, which may introduce bias and decrease consistency when risk assessors evaluate the same study. The Criteria for Reporting and Evaluating Ecotoxicity Data (CREED) project attempts to address this problem. It aims to improve the reproducibility, transparency, and consistency of reliability and relevance evaluations of aquatic ecotoxicity studies among regulatory frameworks, countries, institutes, and individual assessors. In the present study, the CREED evaluation method is presented. It includes a set of 20 reliability and 13 relevance criteria, accompanied by extensive guidance. Risk assessors who participated in the CREED ring test evaluated the CREED evaluation method to be more accurate, applicable, consistent, and transparent than the often used Klimisch method. The CREED evaluation method is accompanied by reporting recommendations for aquatic ecotoxicity studies, with 50 specific criteria divided into 6 categories: general information, test design, test substance, test organism, exposure conditions, and statistical design and biological response. An ecotoxicity study in which all important information is reported is more likely to be considered for regulatory use, and proper reporting may also help in the peer review process. *Environ Toxicol Chem* 2016,35:1297-1309. © 2015 The Authors. *Environmental Toxicology and Chemistry* Published by Wiley Periodicals, Inc. on behalf of SETAC.

Keywords: Aquatic toxicology Reliability Relevance Study evaluation Reporting recommendation

Purpose Statement		Fully Met (FM)	Partly Met (PM)
Please describe the objective for which the usability of the dataset is assessed, including the required dataset thresholds.			
Assessment of whether atrazine is present at state-of-the-art quantifiable concentrations in surface freshwaters within the Occitanie region, France, between 2017 and 2021, with at least two years of monitoring (fully met) and a minimum of four samples per year at each sampling location (partly met). The dataset should ideally have high-risk agricultural areas included and a sampling density of one site per 100 km ² (28 sites, based on the size of the study area). This assessment is aimed to determine if withdrawing atrazine from use in the European Union has resulted in a reduction in its occurrence across the Occitanie region.			
Purpose Summary Table (optional)		Fully Met (FM)	Partly Met (PM)
Enter thresholds for "Partly Met" (minimum information required) and "Fully Met" (optimal information) to fit the assessment purpose. Thresholds can be specific or general, depending on how important a criterion is to the user for the assessment purpose. Also, users can choose not to enter any threshold(s) for "Partly Met", if they want a yes/no decision.			
RV1	The sampling medium/matrix was appropriate for the given purpose.	Surface freshwater	Not applicable
RV2	The sample collection method (e.g., grab, depth- and width-integrated, discrete, composite, or time-integrated samples, or continuous monitoring) was adequate for the given purpose.	All methods accepted	Not applicable
RV3	The study area and number of locations sampled was suitable for the given purpose.	Occitanie: >1 site per 100 km ²	Occitanie: >1 site per 1000 km ²
RV4	The rationale for selection of sampling locations was provided and it is suitable for the given purpose.	At least 25% of the samples	Any surface freshwater
RV5	The samples were collected over a time scale that was appropriate for the given purpose.	At least 2 years of monitoring	Not applicable
RV6	Over the timespan, the sampling frequency was appropriate for the given purpose.	At least 12 samples a year	At least 4 samples a year
RV7	Conditions during sampling events were documented and relevant for the given purpose (e.g., baseflow, storm events, planned/unplanned discharges, etc.).	Any sampling conditions	Not applicable
RV8	The analyte(s) reported was/were appropriate for the given purpose.	Atrazine	Not applicable
RV9	The method was sensitive enough for the given purpose (i.e., the LOD and/or LOQ were below the benchmarks or metrics to which concentrations in the dataset will be compared).	At least 90% of the LOQs should be < 0.01 µg L ⁻¹	At least 90% of the LOQs should be < 0.6 µg L ⁻¹ (EU standard)
RV10	The summary statistics provided (e.g., median, geometric mean, arithmetic mean, percentiles) were appropriate for the given purpose.	Not applicable	Not applicable
RV11	All supporting parameters that were needed to achieve the given purpose were provided.	Not applicable	Not applicable

RELEVANCE CRITERIA					
Class	No.	Title	Criterion	Conclusion	Status
Media	RV01	Sample Medium/ Matrix	Was the sampling medium/matrix appropriate for the given purpose?	Fully Met	Required
	RV02	Collection Method/ Sample Type	Was the sample collection method (e.g., grab, depth- and width-integrated, discrete, composite, or time-integrated samples, or continuous monitoring) adequate for the given purpose?	Fully Met	Recommended
Spatial	RV03	Study Area	Were the study area and number of locations sampled suitable for the given purpose?	Partly Met	Required
	RV04	Site Type	Was the rationale for selection of sampling locations provided and was it suitable for the given purpose?	Partly Met	Recommended
Temporal	RV05	Sampling Timespan	Were the samples collected over a time scale that was appropriate for the given purpose?	Partly Met	Required
	RV06	Sampling Frequency	Over the timespan, was the sampling frequency appropriate for the given purpose?	Partly Met	Required
	RV07	Temporal Conditions	Were conditions during sampling events documented and relevant for the given purpose (e.g., baseflow, storm events, planned/unplanned discharges, etc.)?	Fully Met	Recommended
Analytical	RV08	Analyte	Was/were the reported analyte(s) appropriate for the given purpose?	Fully Met	Required
	RV09	Sensitivity/ LOD/ LOQ	Was the method sensitive enough for the given purpose (i.e., were the LOD and/or LOQ below the benchmarks or metrics to which concentrations in the dataset will be compared)?	Fully Met	Required
Data Handling & Statistics	RV10	Summary Statistics Type (if dataset contains summary statistics)	Were the summary statistics provided (e.g., median, geometric mean, arithmetic mean, percentiles) appropriate for the given purpose?	Not Applicable	Recommended
Supporting Parameters	RV11	Supporting Parameters (if supporting parameters are required for the purpose)	Were all supporting parameters provided that were needed to achieve the given purpose?	Not Applicable	Required

- Poor practice in collecting, reporting and using chemical monitoring data (by people and organisations who ought to know better). This gives poor environmental outcomes
- CREED - Systematic, transparent, more consistent evaluations
- CREED - Document and characterize uncertainties, identify data gaps
- Guidance on best practice for data reporting (to maximize use of their data by risk assessors)
- Assistance in characterizing uncertainties and data limitations, and fitness of the data for their assessment purpose
- Purpose driven
- The issue is, as with CRED – uptake (OECD: [https://one.oecd.org/document/ENV/CBC/MONO\(2025\)18/en/pdf](https://one.oecd.org/document/ENV/CBC/MONO(2025)18/en/pdf))

Acknowledgments



- Financial support of the initial SETAC workshop was provided by Concawe, GSK, Metals Environmental Research Association (MERA), SETAC Journals (ET&C and IEAM), Syngenta and Unilever.
- All participants in the CREED development process and the beta-testers....
-AND YOU FOR LISTENING