



Company Level Environmental Accounting Reporting and Information

A practical guide for regulators on the challenges and potential benefits of integrating environmental data



1

CLEAR Info is a demonstration project lead by the Environment Agency. The project aims to improve how businesses implement environmental legislation by using information on the environmental performance of parent companies.

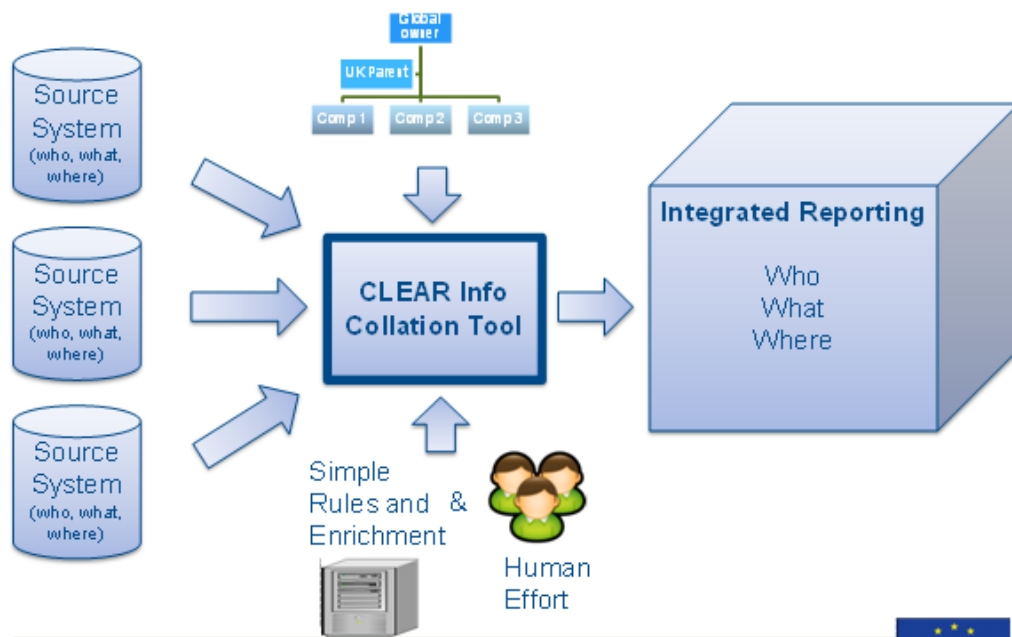
Regulated Industry

- Size
- Activity
- Permits
- Legislation

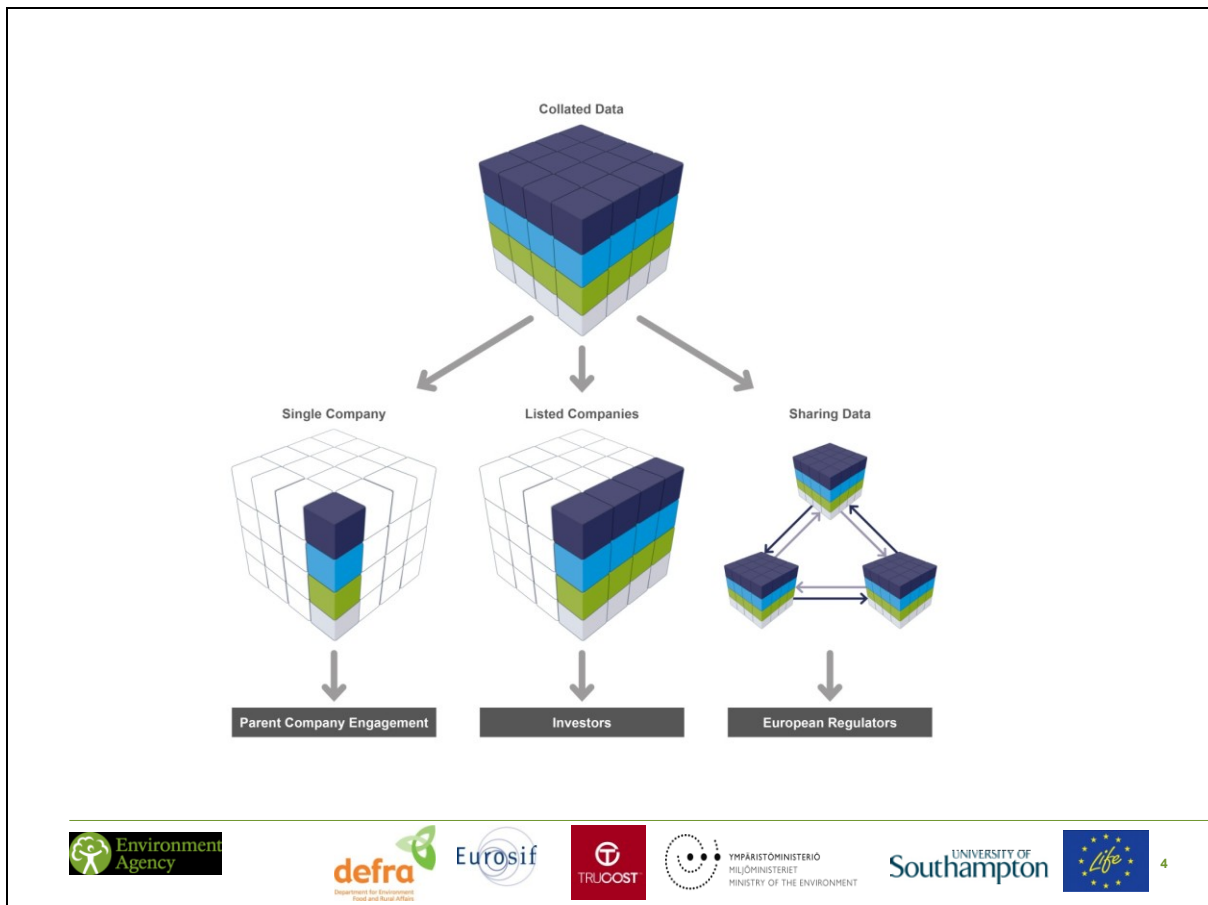


- How much data do we have on a company? Depend on size, activity, what permits and which legislation e.g. relatively large company like BT, very little regulation from us. Waste company with large landfill, constant monitoring and sites visits across most of estate. England ONLY
- How we gather data - site visits, web based submissions, sampling, operator data portals
- How we store data – Regime based, different systems, often using different software or environments. Some spreadsheets, access databases, business object universes or custom built systems. Some Area based, some national – with permits up to 60yrs. 2 broad categories of data systems, those that hold permits, and those that contain monitoring data

What is CLEAR Info doing?

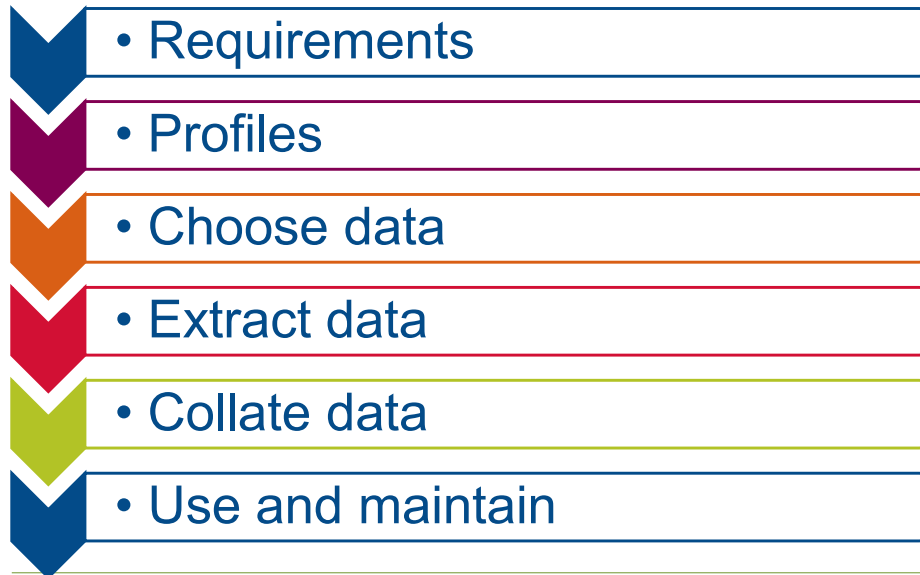


- Source systems, different regulatory regimes
- Feed into a big data warehouse /
- Rules and enrichment – company names fixing – datasets are at best 80% match, at worst 10% match. Not the role of the project to fix the EAs underlying data issues Therefore focused on identifying permits for our 3 parent companies.
- Added company hierarchy data as a structure to aggregate against
- Decide on things might want to search for – geographic region, Sector, type of permit breach. And attach flags to data 'dimensions'.



- Use dimensions to interrogate data.
- Use parent company to search for all the permits a company had.
- Look for listed companies to engage with investors.
- Look for specific permits or legislative links to share data with other agencies.
- Benefit of cube – multiple reporting functions that use this as base. Pre-processes combinations of dimensions so you get instant answers. Our cube has 170K companies in it.

Producing a collated dataset

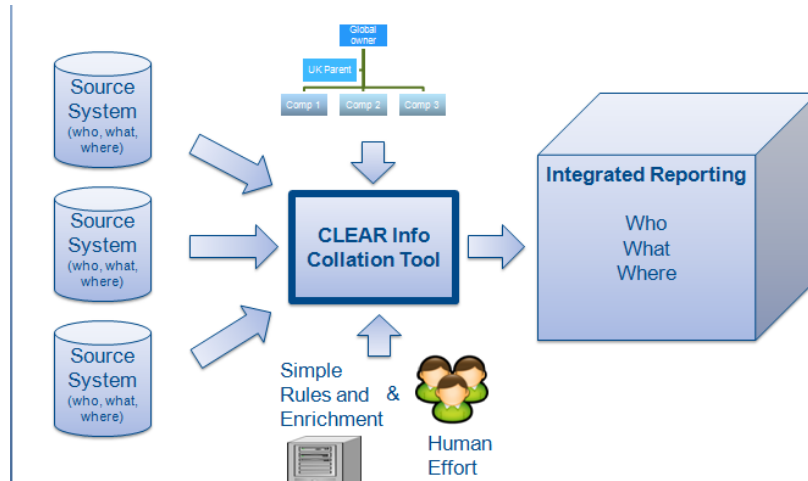


5

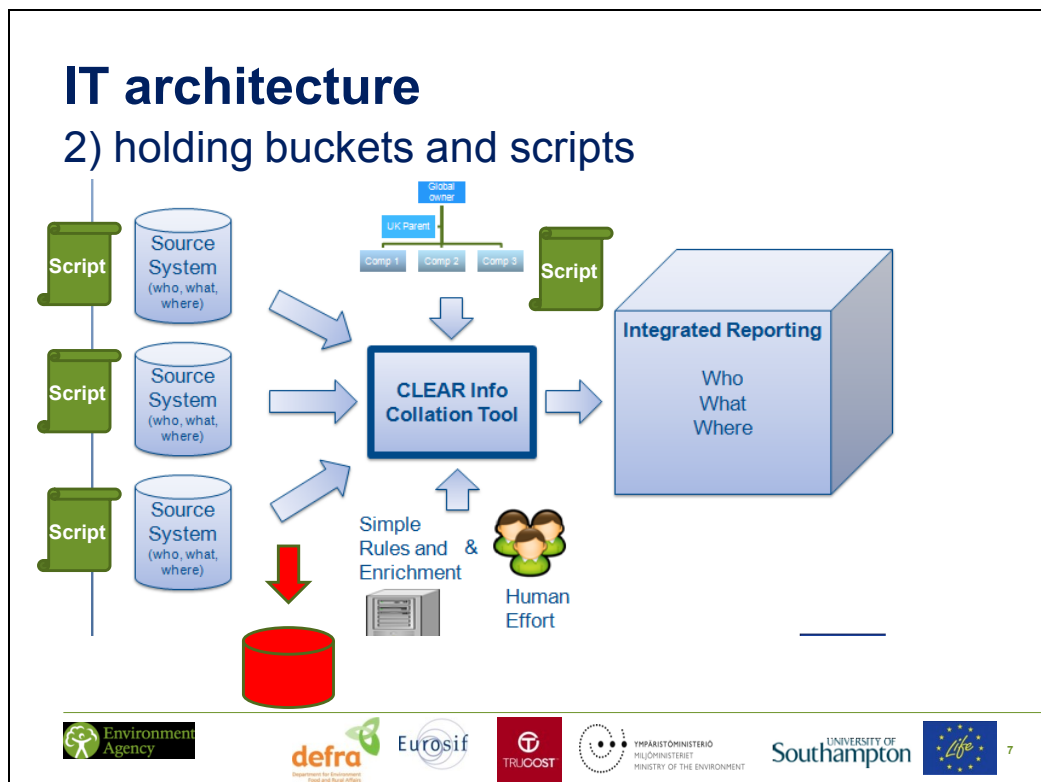
- Requirements – what data is available, any external data that might be of interest, what is required for implementing legislation / monitoring compliance? How will you use the data, who will have access?
- Profiles – speaking to the day-to-day system managers, information do the contain / data fields, what could be used as keys to link, security issues, frequency of data change.
- Choose data – speak to user groups and stakeholders, what info useful, does it need special authorisation, usable format (national, limited free txt, data quality, linking factors). Objective datasets, do they need narrative to aggregate to parent level, are there issues of double counting?
- Extract data – from source systems into flat files e.g. CSV, excel, access. Define what you want to search (dimensions) region / type of breach and count (facts) Litres of water / tonnes of carbon. Apply any data cleansing or enrichment.
- collate data – use linking factors and common fields to populate big relational database. Apply aggregation structure (comp hierarchy, organisational string). Populate cube or other MDM software.
- Maintain – who will have access, establish set reports, governance structure, data refreshes (inputs, frequency and formats).

IT architecture

1) Proof of concept



Over 18 month, 3 main iterations of background IT architecture

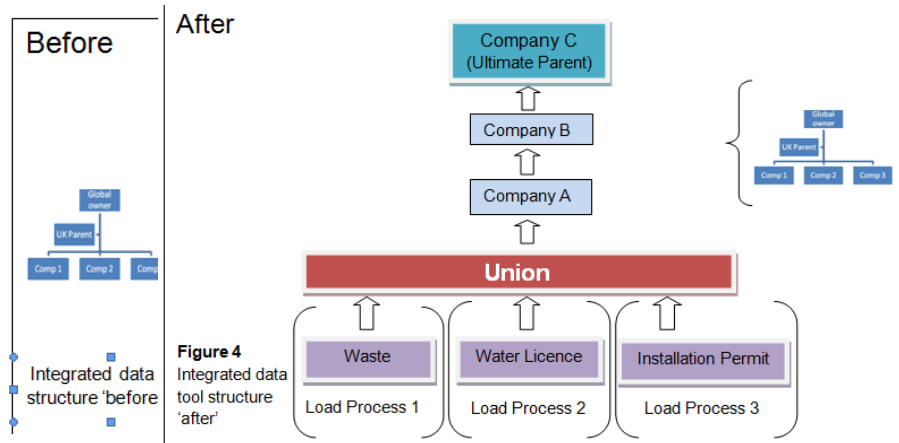


2nd iteration enabled us to see how much of the data was being loaded and evaluate problems and fix accordingly.

- Capture all data that wouldn't load with standard business rules in a 'holding bucket'. Further analysis was then done with system owners to assess if there were data quality issues or we needed to expand out field lists and business rules.
- Adding scripts was something that saved us considerable time during development. These detail business rules at strategic points in the data flows. If you wish to change a small detail, you can do it here and the script will help identify where the item is referenced throughout the system and cascade the change through.

IT architecture

3) Processing structure



3rd iteration

- Streamlined data processing. Significantly improved performance, work on separate parts (or refresh data) without re-running whole cube.

- LNK – by third structure we were confident we were capturing the maximum amount of data and able to aggregate it to a higher level. However, still some significant problems that meant data could not be aggregated. Sure no surprise, stem from lack of data standards, format and quality.

What's in the dataset?



Permitting information, breaches, waste treated and produced, water abstraction, EAs Site based Operational Risk Assessment and emissions from the pollution inventory

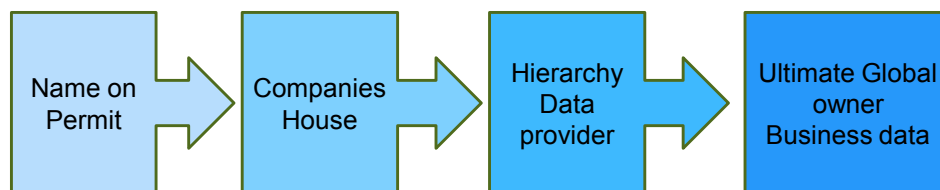
What's in the dataset?



- Geographical reference data
- Packaging, batteries and WEEE
- Hazardous waste
- Data on pollution incidents and enforcements
- Carbon data = CRC, EUETS,

Common data problems

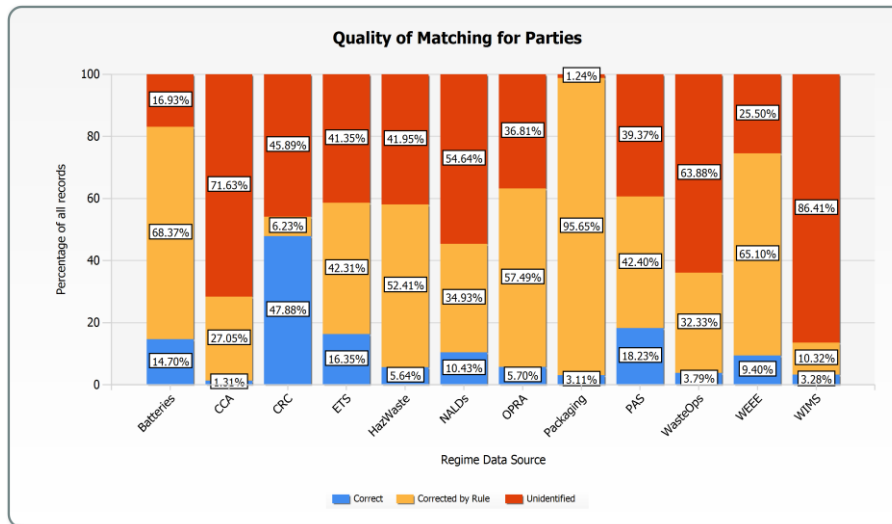
- ➔ Security and confidentiality
- ➔ Free text fields
- ➔ Field definition
- ➔ Company names



11

- Security and confidentiality – different user groups (parent company's see their own data). Get feed files of data, careful to exclude certain sites of national security. Added sensitivity flags to certain fields or date ranges of data so they could be easily filtered. If going live into production also set care levels of access for different users.
- Free text fields (type into box instead of select from list) mean can't be 'read' by a computer. Good example = water abstraction licence. Have maximum abstraction limit per year say 10,000m³. one company, 2 abstraction licences of 10,000m³ limit. However, combined abstraction limit of 15K over both licence. In dataset the field which indicates if there is a combined licence limit is free text. = no automatic way of searching based on returns if they have breached their limits without case by case review. Feed back for future system development not to have any free text fields if possible
- Field definition – several of our datasets would have apparently common fields such as geographical region or sector. Whilst appear the same – Region = local council region, water catchment region, EA internal region (not same geographical boundaries). Sector = Env Perm Regulations, Stand Industry Classification, EA's Future Approach to Regulation or FAR sector. Working with data standards team to find commonly used but different field definitions. In conjunction with the EA data teams to outline all definitions.
- Company name – underlying principles are (boxes). If person / site manager applying for permits hasn't written legal entity name of company – start to run into issues. E.g. Large companies in the UK = British Telecom Ltd. Most permits say 'BT'. Several things we can do. Automatic fixes that strip out limited, ltd, Ltd. Manual fixes e.g. BT, several 1000 permits with same 'error'. Not uncommon error, we have 24 ways of spelling 'The Environment Agency'!

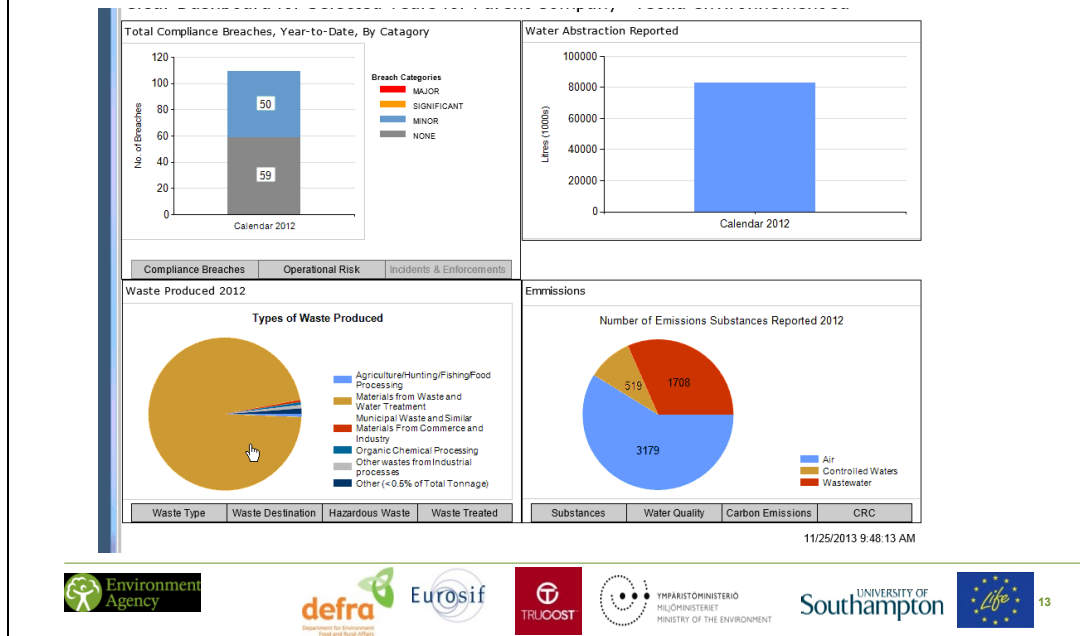
Company names matching



- Wont be able to see names, don't worry.
- Blue = direct match with Companies House – newer system (CRC 5years) better hit rate (lowest 1.3%).
- Orange = automatic cleansing OR manual fixes combined.
- Red = unmatched. Highest 98%, lowest 10% match. Some maybe not required to register with Companies house, domestic, sole traders, (UK public sector) NHS or University. Cannot separate these out.
- Often cannot change in source permitting system. Altering permits could result in charge for company or void any current enforcement notices or action that's ongoing.
- Solution = corrected name field held in collation tool corrects the problem.
- As can see, still a lot of red on these graphs – plan to work with the business to correct this over time. Including asking for company number when issuing new permits.
- Pilot, - not fix the world, focus on parent companies we are actively engaging with at board level.

Benefits

➔ Help identify priority issues.



Priority issues – Businesses, investors or regulators - new insight either looking at parent company, sector or legislative level.

Parent companies – Example of parent company dashboard we developed to interact with parent company's at board level.

- Using the collated dataset helped us target which sectors and which companies within those sectors we would target for engagement. We had a mix of companies that we had existing account management relationships with, UK based, large international companies with a variety of level of regulation by the EA in England.
- New bring so much data together in one place, well received 'provided new insight'. Fully interactive, click on graphs to drill into more detail down to site / day. Much more constructive for engagements than static PDF of previous reports. Highlight areas where investment or action is needed at company level.
- We are taking this forward to agree compliance improvement plans and monitor effectiveness of this type of intervention with companies in a range of sectors, size and activities.



Trucost presented work on environmental factors as market / share indicators (incl. Carbon and OPRA scores)

The Environment Agency presented range of data available: compliance, Risk assessment, Emissions, waste, water

Discussion:

- What they do within the investment or finance sector
- what types of data do they use and where do they get it from
- how widely used is environmental data
- what challenges do they face with accuracy and data integration.

Regulators

➔ Target poor performers



Reviewing the data at a higher level has enabled us to look for trends in performance more easily. We could use sector analysis or review types of breaches to indicate areas to target in future initiatives.

Future potential

Compliance Breach Reasons (By Category) for Selected Year 2012

Major & Significant
No Data Available

Minor & None

Category	Count
Permitted activities	1
Infrastructure	7
Operational Management	18
Incident management	13
Emissions	12
Energy	6
Monitoring and records, maintenance and reporting	4

- Use it to produce our Sustainable Business Report – Significant data processing investment by the business which could be dramatically improved. Sector drill downs. Many areas of the business that could benefit from efficiencies in reporting and data processing.
- Power pivot reporting platforms could be used for slicker management reporting or instantaneous answers to FOI requests that are a significant cost the EA.
- Use it for a common customer list, one version of the truth for business use. This could free more data processors up for data analytic providing more insight from our existing data. One example....
- Looking into permit breaches and pollution incidents at a company level over time. Is there a threshold of lower level breaches that indicates a more damaging environmental incident could be about to occur?
- Other regulators - target poor performers across the EU if we can better share and link up data.

Discussion

- ➔ Does this sound familiar to things you have done within your own organisations?
- ➔ Is there anything different that you, as a regulator, have used integrated data to achieve?

