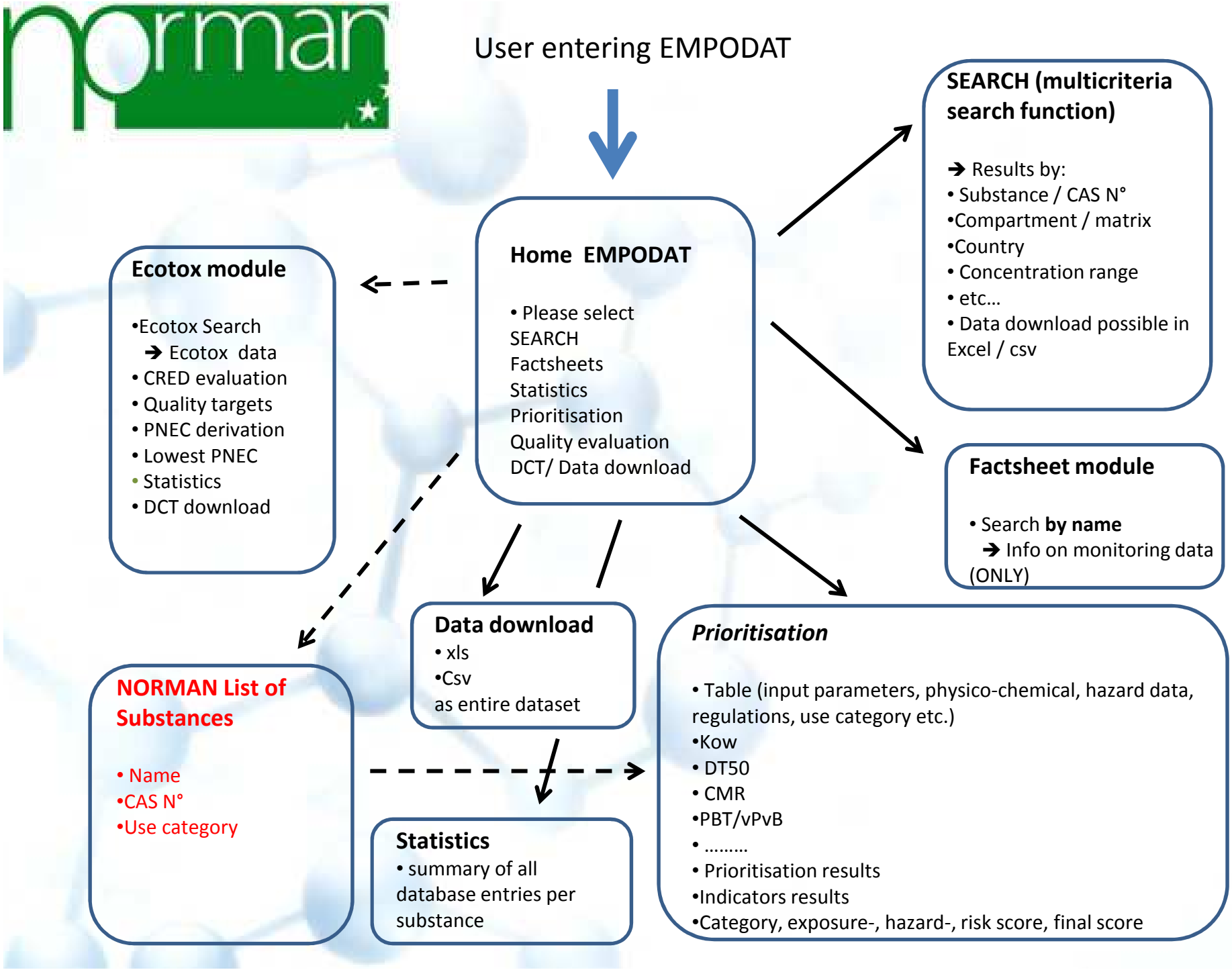
NORMAN Databases workshop

EMPODAT - present status and options for improvement

Ildiko Ipolyi, Environmental Institute
ipolyi@ei.sk

NORMAN databases current status

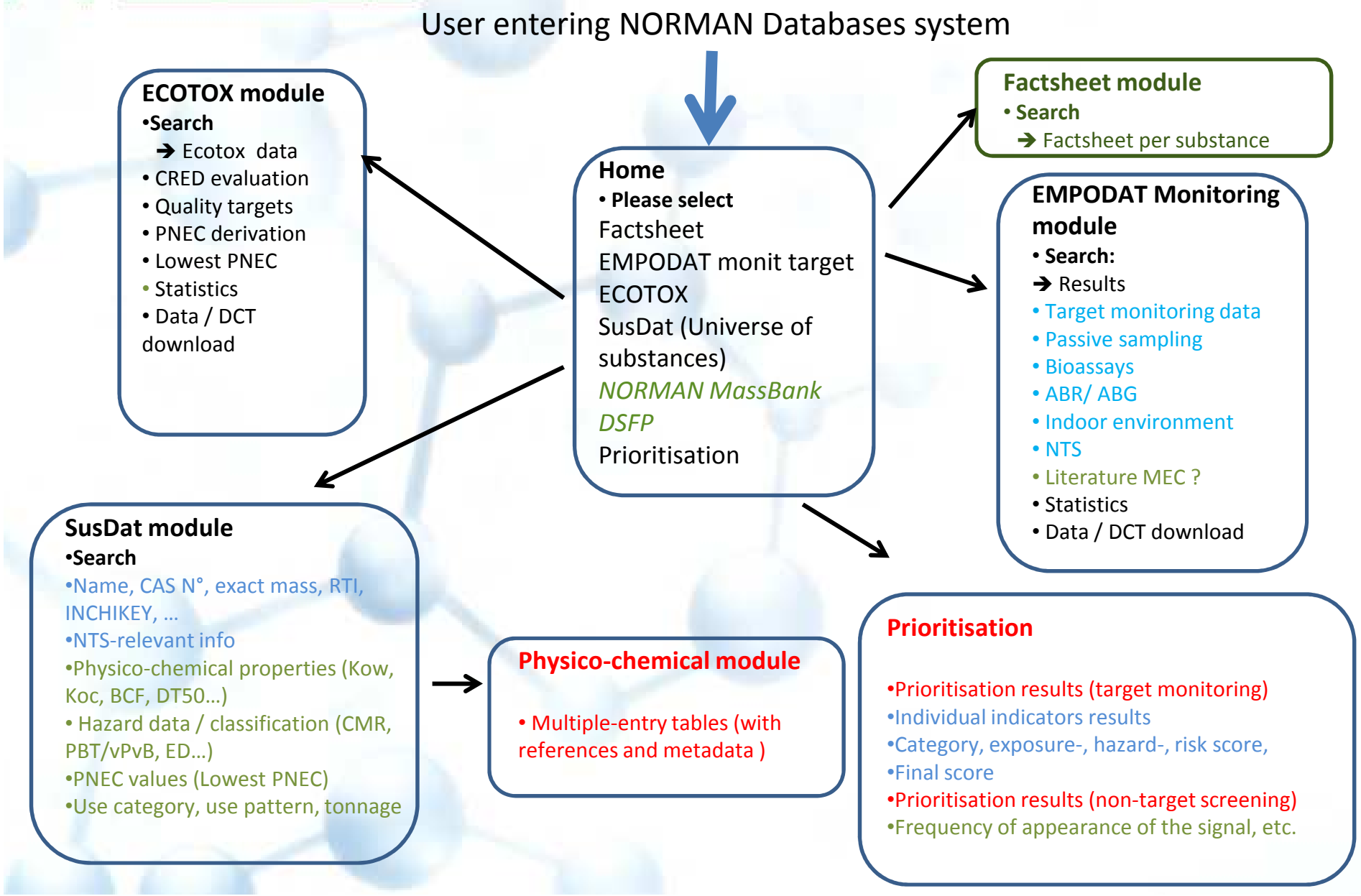

User entering EMPODAT
Home EMPODAT
• Please select
SEARCH
Factsheets
Statistics
Prioritisation
Quality evaluation
DCT/ Data download
SEARCH (multicriteria search function)
➔ Results by:
• Substance / CAS N°
•Compartment / matrix
•Country
• Concentration range
• etc…
• Data download possible in Excel / csv
Ecotox module
•Ecotox Search
➔ Ecotox data
• CRED evaluation
• Quality targets
• PNEC derivation
• Lowest PNEC
• Statistics
• DCT download
Factsheet module
• Search by name
➔ Info on monitoring data (ONLY)
Data download
• xls
•Csv
as entire dataset
NORMAN List of Substances
• Name
•CAS N°
•Use category
Statistics
• summary of all database entries per substance
Prioritisation
• Table (input parameters, physico-chemical, hazard data, regulations, use category etc.)
•Kow
• DT50
• CMR
•PBT/vPvB
• ……
• Prioritisation results
•Indicators results
•Category, exposure-, hazard-, risk score, final score

Main limitations of the current database system

- **Designed for "classical" monitoring data** => with the New modules ……the "Search" interface needs to be reviewed
- Designed **for a limited number of substances** => the system needs to be upgraded to allow smooth integration of large batches of new substances
- **Static system**
  - Factsheets / statistics - no dynamic link to the data source
  - Prioritisation module - need to ask the IT for each new prioritization run, static table with results of prioritization is provided but it is not possible to view / interact on the choice of the datasets
  - No dynamic links for retrieval of data from other databases
- **Single-entry master table for substance info**, no info on data references and no metadata (prioritization)
- **No visual presentation** of the data

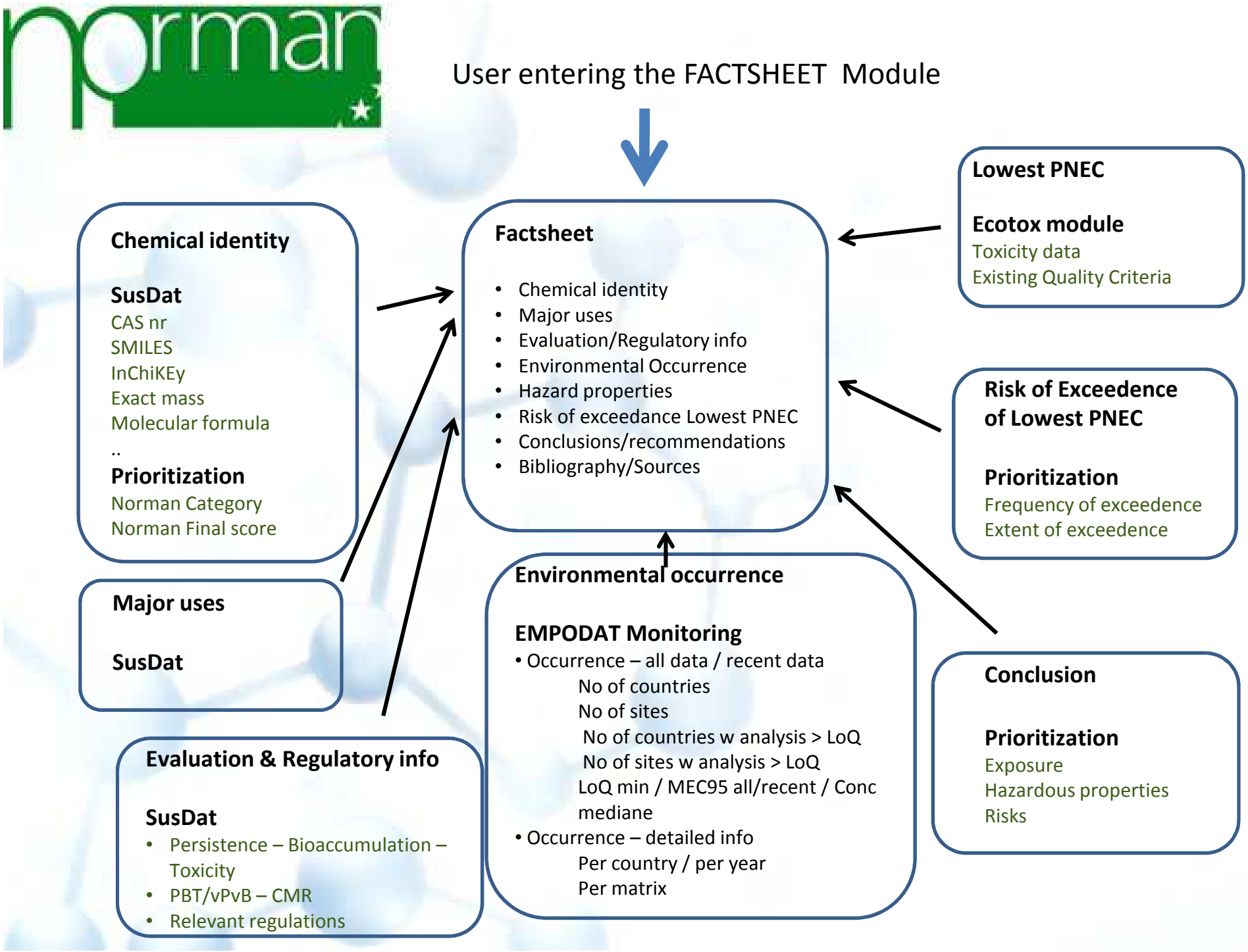
New NORMAN databases structure
our proposal

All new modules integrated in the overall structure



NORMAN databases structure: our proposal

- **SusDat:** central unit taking the place of the NORMAN List of substances
- **Monitoring module:** search function to be revised
  - To expand to all new modules
  - New search parameters needed
- **Prioritisation module**: will be revised
  - Direct interaction of the user on the datasets?
  - New Exposure index (?)
  - New indicators derived from NTS data (e.g. frequency of appearance) to prioritise compounds from NTS data
- **Substance Factsheet**: will be the recipient of all info available in the DB for a substance
  - Info retrieved from all different modules of the DB
  - For each data reported in the factsheet an active link will show the reference + metadata


User entering the FACTSHEET Module
Chemical identity
SusDat
CAS nr
SMILES
InChiKEy
Exact mass
Molecular formula
..
Prioritization
Norman Category
Norman Final score
Major uses
SusDat
Evaluation & Regulatory info
SusDat
Persistence – Bioaccumulation – Toxicity
PBT/vPvB – CMR
Relevant regulations
Factsheet
Chemical identity
Major uses
Evaluation/Regulatory info
Environmental Occurrence
Hazard properties
Risk of exceedance Lowest PNEC
Conclusions/recommendations
Bibliography/Sources
Environmental occurrence
EMPODAT Monitoring
Occurrence – all data / recent data
No of countries
No of sites
No of countries w analysis > LoQ
No of sites w analysis > LoQ
LoQ min / MEC95 all/recent / Conc mediane
Occurrence – detailed info
Per country / per year
Per matrix
Lowest PNEC
Ecotox module
Toxicity data
Existing Quality Criteria
Risk of Exceedence of Lowest PNEC
Prioritization
Frequency of exceedence
Extent of exceedence
Conclusion
Prioritization
Exposure
Hazardous properties
Risks

Issues to be discussed (1) – New Factsheet module

- Factsheets template - additional info, amendments needed?
- From the current single-entry master table to multiple-entry modules for BCF, Kow, Koc, T50 …
- Links with external databases: dynamic link/ updated on the fly or static format + regular update?
- « Search » function of the new Factsheet module: what pre-defined search fields?

Search function of the “Factsheet module” - v_mockup

Search for Emerging Substances

Phys-Chem:

	>=	<=
Log K_{ow}		
$logD_{(pH7)}$		
$logD_{(pH9)}$		
Log K_{oc}		
Water solubility [mg/L]		
Henry constant		

Uses & production:

	Appr.	Not appr.
Biocide		
PPP	☑	
Industrial chemicals		
Pharmaceutical		
Cosmetics		
Tonnage		

>=	<=

Monitoring data:

Monitored	
Quantified	☑
Passive sampler	
Non-Target	
Literature	

Monitoring Matrix:

Freshwater	
Sediment	☑
Biota	
Marine Water	
Groundwater	
Indoor Wastewater	

Ecotox & hazard:

Acute data	
Chronic data	☑
PNEC/EQS	☑
P,B or T	
C, M or R	
ED	☑

Ecotox Matrix:

Freshwater	
Marine Water	☑
Sediment	

Substance (title or CAS no.)

Title of substance

(Â±)-1-(.beta.-allyloxy-2,4-dichlorophenylethyl)imidazole / Technical grade Imazalil / Enilconazole []
(benzothiazol-2-ylthio)methyl thiocyanate []
1-Decanol [112-30-1]
1-Octanol [111-87-5]
1,1,2-Trichloroethane [79-00-5]
1,2-benzisothiazol-3(2H)-one []
1,2,3-Benzotriazole [95-14-7]
1,2,3,4,5,6,7,8,12,12,13,13-Dodecachloro-1,4,4a,5,8,8a,9,9a,10,10a-decahydro-1,4:5,8:9,10-Trimethanoanthracene []
1,2,5,6,9,10-Hexabromocyclododecane (HBCD) (5 isomers - alpha to epsilon) [3194-55-6]
1,3-dichloro-5,5-dimethylhydantoin []
1,3,4,6,7,8-hexahydro-4,6,6,7,8,8-hexamethylin-deno[5,6-c]pyran (Galaxolide) [1222-05-5]
1,4-Dichlorobenzene [106-46-7]
1,4,5,6,7,7-Hexachlorobicyclo[2.2.1]hept-5-ene-2,3-dicarboxylic anhydride []
17-alpha-Estradiol [57-91-0]
17-alpha-Ethinylestradiol [57-63-6]

Search

Risk:

	>=	<=
NORMAN Category		
Priority Score		
Exposure Score		

Structure:

InChi Code	
InChi Key	
Smiles	
Summary	

Save

Issues to be discussed (2) – Data quality and treatment of outliers – *Monitoring module*

1, Identification and treatment of outliers

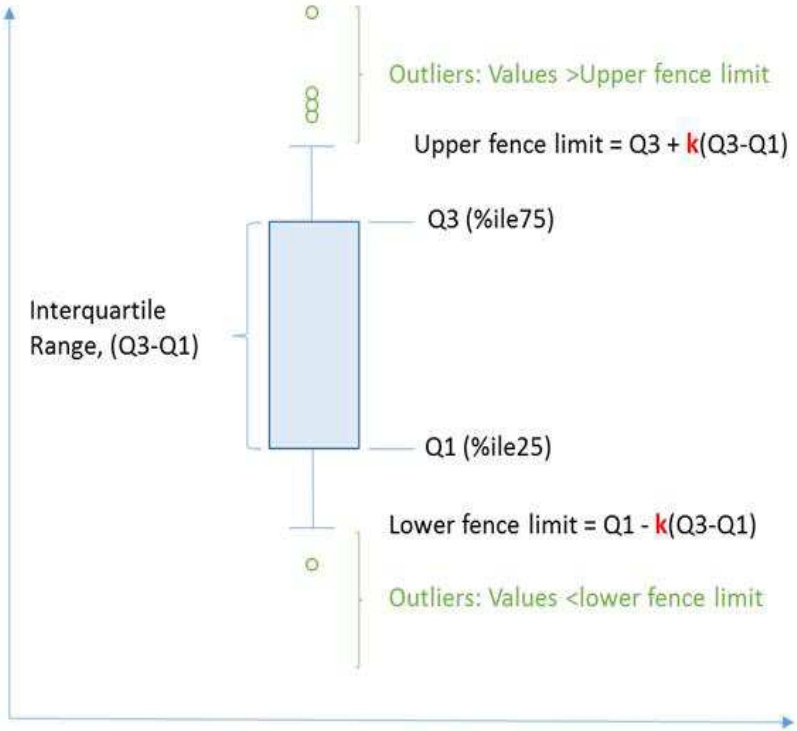
- **option 1 -** to adopt the JRC guidelines w modification

  Advantage

  based on the intrinsic distribution of the statistical sample, hence no need to define threshold for each matrix or (different families of) substance(s)

  k definition – at 2-levels

  - Elimination k=1000?
  - Warning k= 5 – 50?



Issues to be discussed (2) – Data quality and treatment of outliers - *Monitoring module*

- **option 2**

  Alert–based model w thresholds

  Different threshold per substance / families of substance

  HOW to decide what to do with data flagged as outliers? Input from the NORMAN experts to decide on a case-by-case basis whether to discard or leave and park the identified outliers?

  OR application of a standard automated procedure?

- 2, Data quality and selection of datasets for prioritisation
  - Direct interaction of the provider on the datasets?

Issues to be discussed (3) – Data Collection Templates

- New data collection templates

  - Indoor environment (Indoor WG / EBL)
  - Passive sampling (Recetox /IA)
  - Bioassays (EDA-Emerge / ZR)
  - Non-target screening (EI / JS)
  - ABR / ABG (ANSWER / IVM)

  Format

  - Original 3-sheet (Data source – Analysis – Analytical method) format suitable? Or a more user-friendly version needed?

  - Obligatory parameter-set to be defined for all data collection templates!

  - general fields for all DCTs / module-specific fields