European Commission

Data Policy and Data Management – Sample Experiences and Requirements

Silvia Dalla Costa and Nicole Ostlaender

Joint Research Centre

www.jrc.ec.europa.eu

Serving society
Stimulating innovation
Supporting legislation

European Commission

IPCheM

Information Platform for Chemical Monitoring (IPCheM) requirements in data policy and data management

Contents

- IPCheM introduction and definition
- Objectives
- Users
- Challenges and issues meet so far
- Ongoing activities

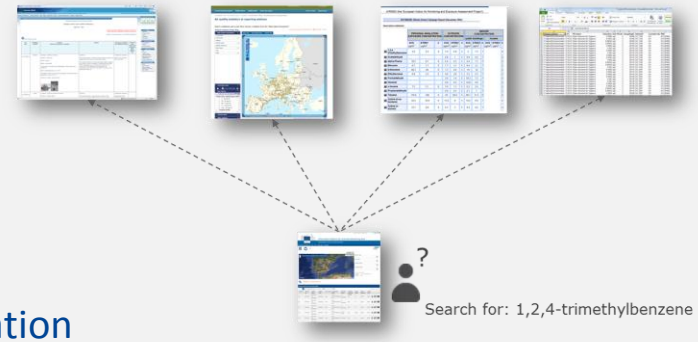
Joint Research Centre

IPCheM definition

A single access point for discovering and accessing chemical monitoring data of Europe

EU Communication "The combination effects of chemicals – Chemical mixtures" (COM/2012/0252 final)

IPCheM is:



- a distributed infrastructure
- avoiding data duplication and information systems replication
- respecting any condition of data access and use defined by Data Providers
- strengthening collaboration between EC Services, Agencies, Research Centres, international and national bodies.
- facilitating links with info systems in the same domain

Main Objectives

Promotion of a coherent and harmonised approach to the generation, collection, storage and use of chemical monitoring data (requested by the "Communication on Chemical mixtures")

Short term objectives focused on **data access**, by:

- Implementing searching facilities to discovery and access chemical monitoring data
- Implementing hosting facilities for data currently not easily accessible or orphan data
- Providing chemical monitoring documentation (Metadata and data docs.) of defined quality

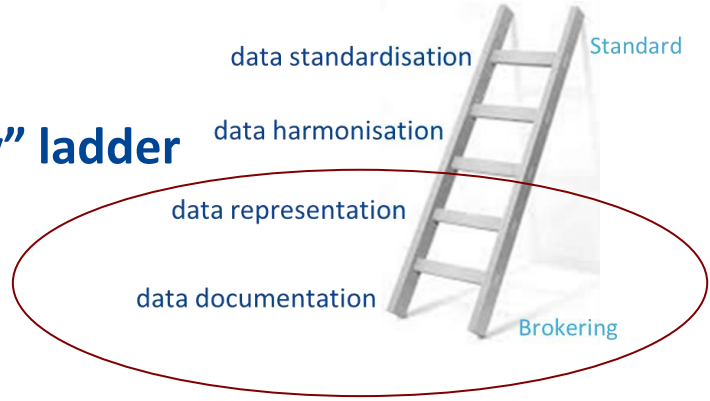
Improve comparability of the data by promoting standardisation of data and metadata and the improvement of quality assurance standards" (Stated in the "IPCheM Scoping Paper")

From data access to data comparability

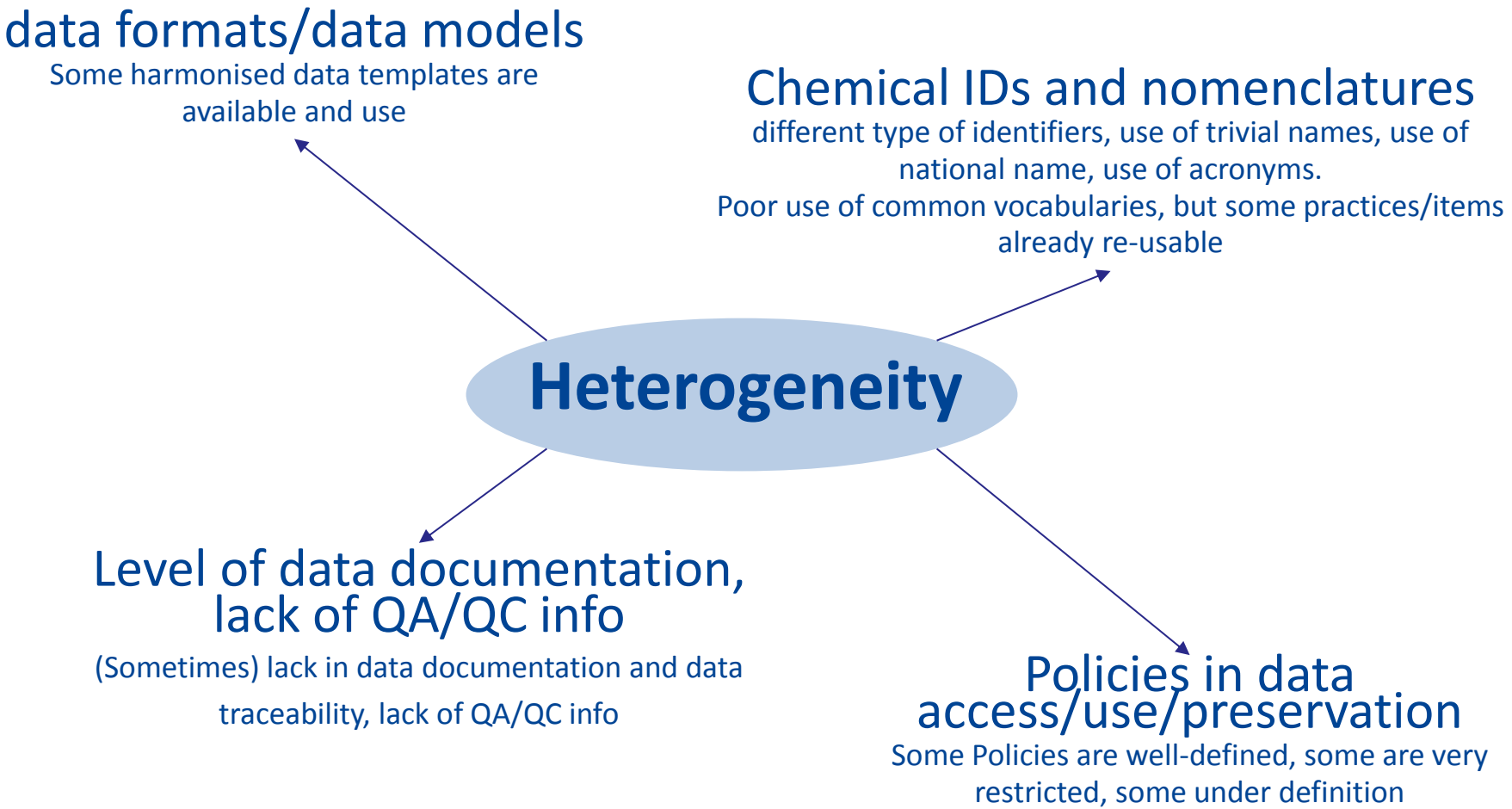
In chemical monitoring domain:

- evaluate multi-media and multi - pathway exposure for human risk assessment
- correlate increasing levels of known and emerging substance in humans with the occurrence in food and feed, products and the environment
- evaluate the simultaneous occurrence of important exposured media (drinking water, products, food)
- facilitate the evaluation of the effectiveness of chemicals and related legislations/policies

IPCheM is oriented to climb the "data comparability" ladder



Issues

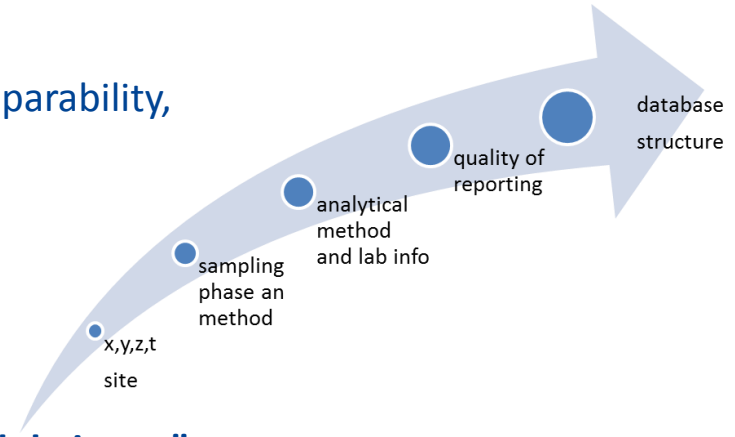

data formats/data models
Some harmonised data templates are available and use
Chemical IDs and nomenclatures
different type of identifiers, use of trivial names, use of national name, use of acronyms.
Poor use of common vocabularies, but some practices/items already re-usable
Heterogeneity
Level of data documentation, lack of QA/QC info
(Sometimes) lack in data documentation and data traceability, lack of QA/QC info
Policies in data access/use/preservation
Some Policies are well-defined, some are very restricted, some under definition

and challenges

- Taking into account important aspects of the data policies for federated data collections
  data sensitiveness, respect of the privacy, ethical considerations → human biomonitoring data

- Working on **metadata and data traceability** ensuring the connection, alignment and integrity with connected data

- Because IPCheM promotes QA/QC statements and data comparability,
  metadata must describe **different phases**



different entities (e.g. station, sample, laboratory, etc.) and **"module items"** (food&feed, environment, product&indoor air, human biomonitoring) complementing chemical monitoring data collections

Ongoing activities

2 working groups (users + project coordinators)

Preliminary contributions of the wg on IPCheM Terms and data Policy

- Glossary of definitions and collection of policies, used by wg members, are the starting point for the future legal interoperability

- IPCheM data policy should consider these different policies in place and outline a general part in which the scope until where IPCheM policy can go is clarified

- Inside the same legislation there are different approaches, to be considered as challenges rather than issues. Wg is providing information on the practices in which these policies are implemented

Ongoing activities

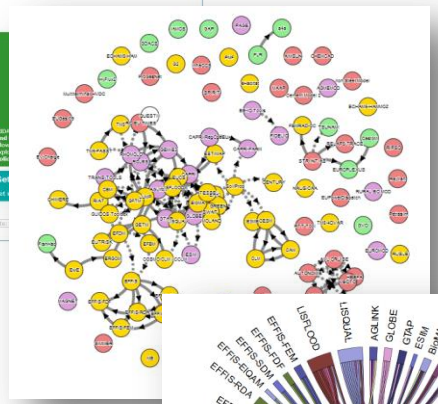
How to incorporate obligation on reporting to IPCheM into the contract agreements under framework programs?

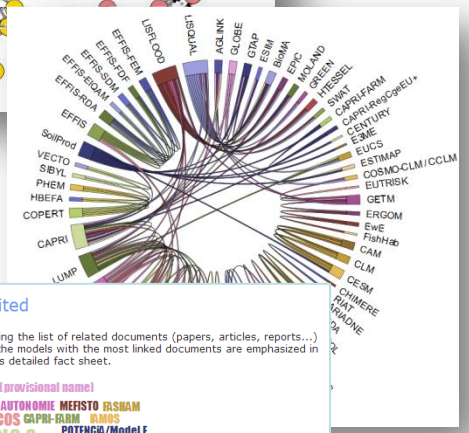
Discussion with DG RTD on how research data could be "by default" part of IPCheM

Present: DG RTD is member of both the working groups + supports JRC and DG ENV to promote and inform about IPCheM to FP7-FP6 project coordinators

Future: Possibility to strive data access using the H2020 Grant agreement, art. 31.5 "Access rights for the EU institutions, bodies, offices or agencies and EU Member"

European Commission



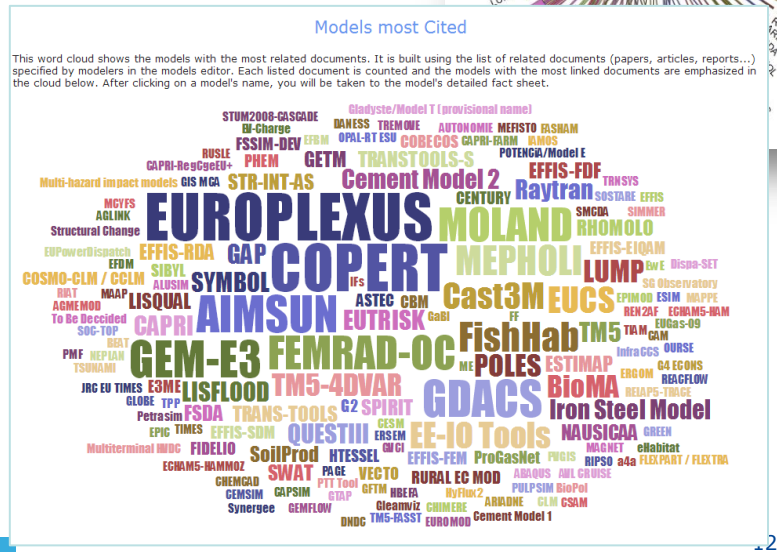


MIDAS

Modelling Inventory Database and Access Services

-

data policy and management requirements from the perspective of modelling



Joint Research Centre

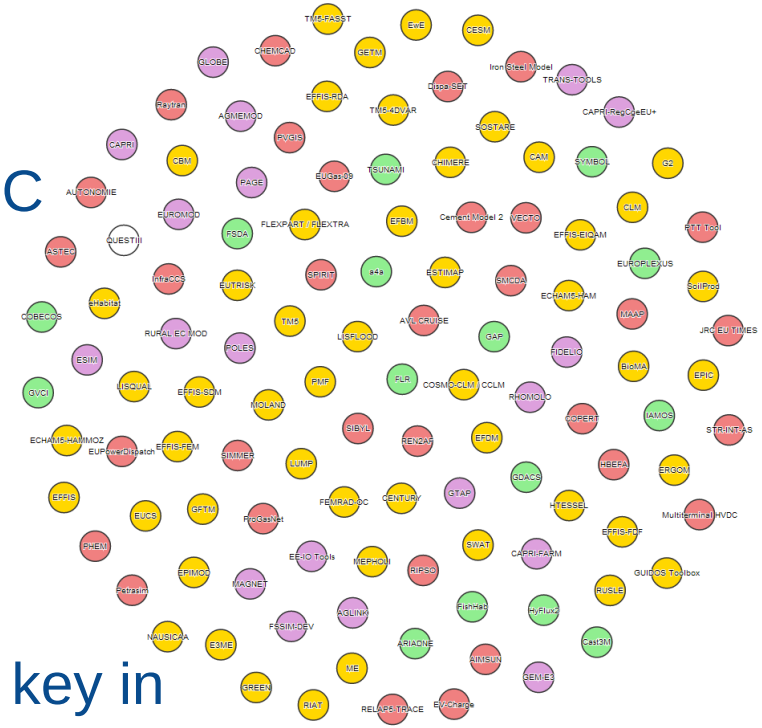
Content

- Rationale
- MIDAS Scope and represented entities
- Data management and policy: Areas of Impact
- A (non-exaustive) shopping list

European Commission

Rational

- **Modelling as key expertise** of JRC
- Support to Impact Assessment as a major contribution to **policy making**
- **Transparency** of models and **reproducibility of their results** is key in science for policy making
- **Access to data** as one of the key issues



Joint Research Centre

MIDAS Scope

Online platform accessible from within the Commission Network.

Gives access to an up-to-date database of models that in use in JRC.

Represent models in context: Links models to models, data policies and people

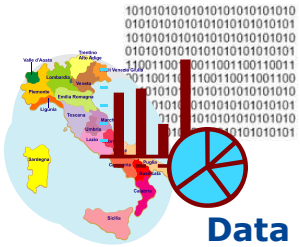
MIDAS Entities

Searchable entities in the MIDAS DB

Models

Related entities

Models

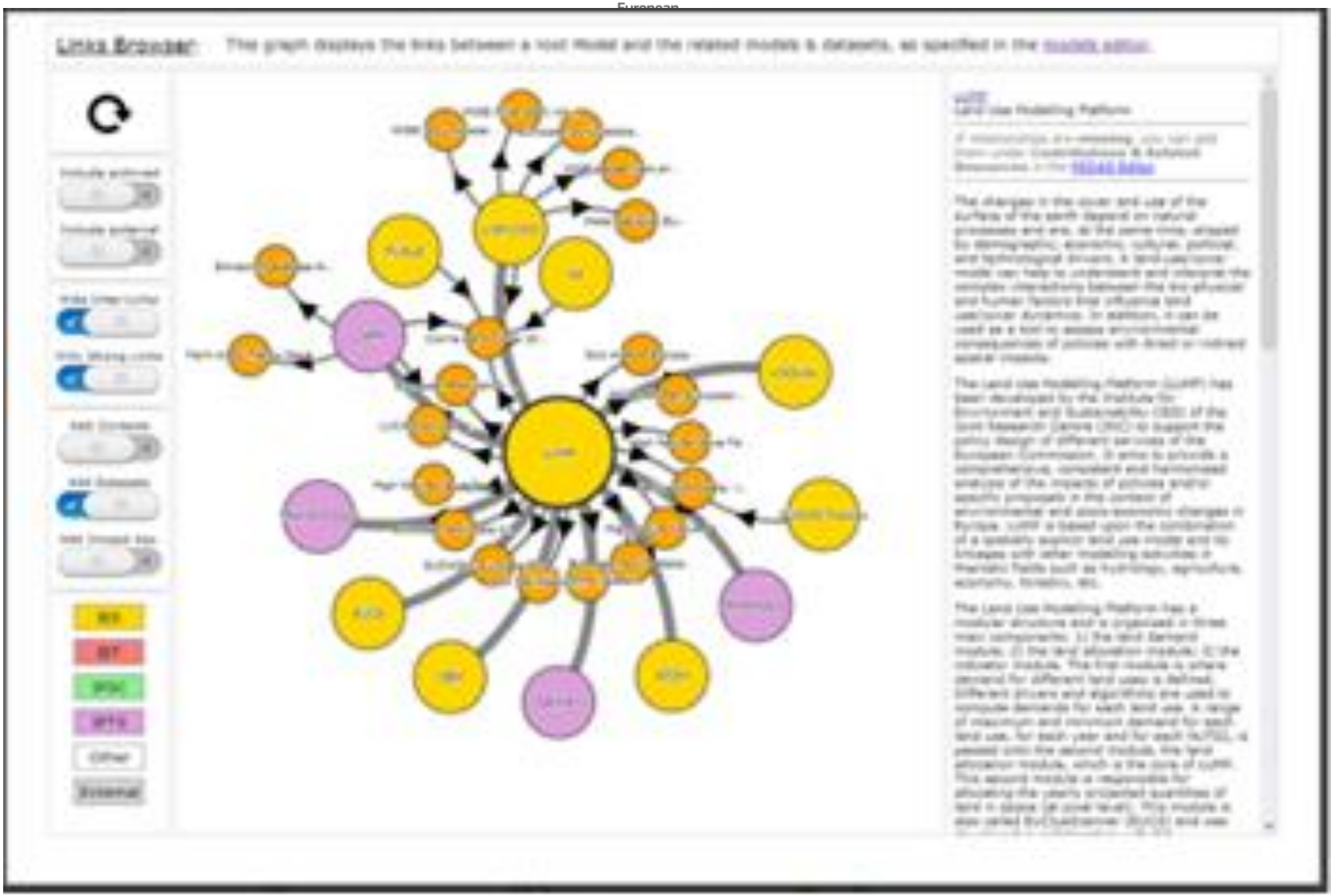


Data

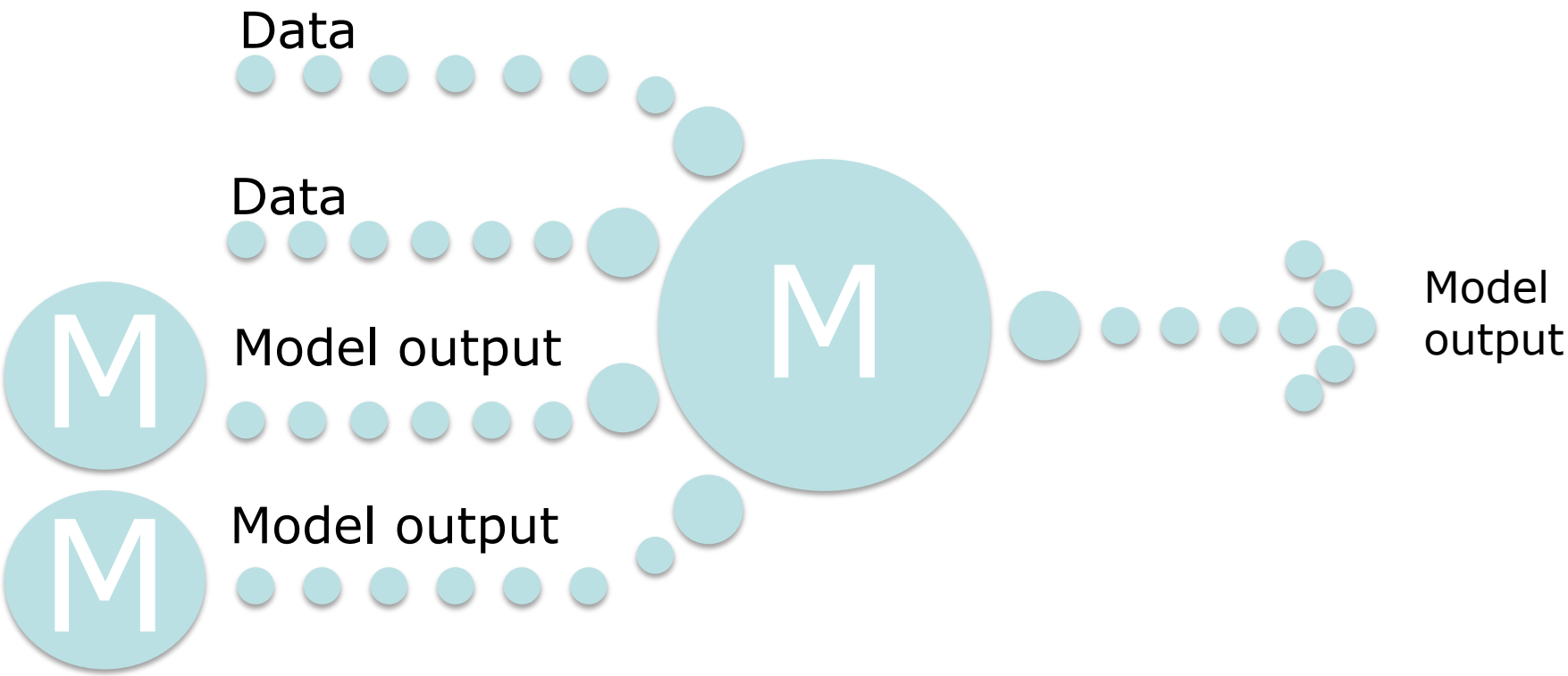
People

Impact Assessment

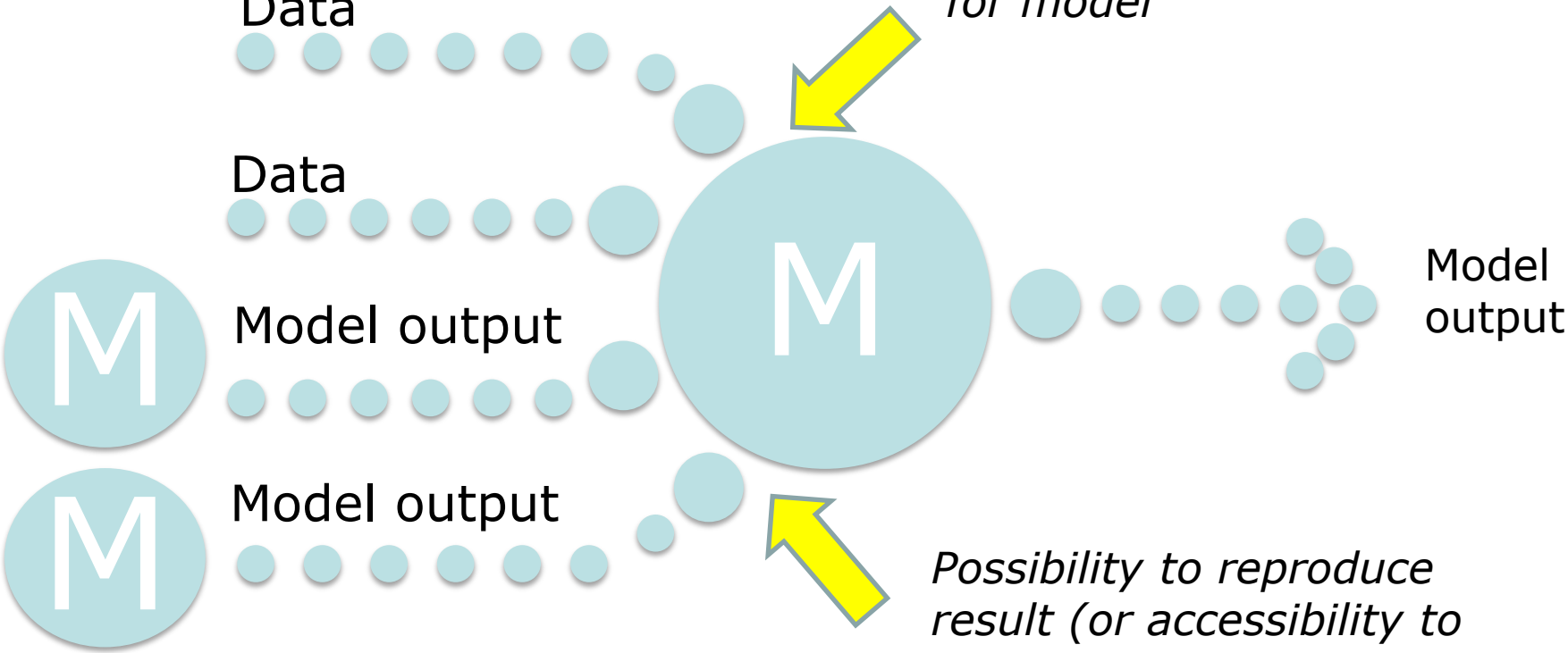
JRC Scientific publications


Links Browser
This graph displays the links between a root Model and the related models & datasets, as specified in the

Areas of impact

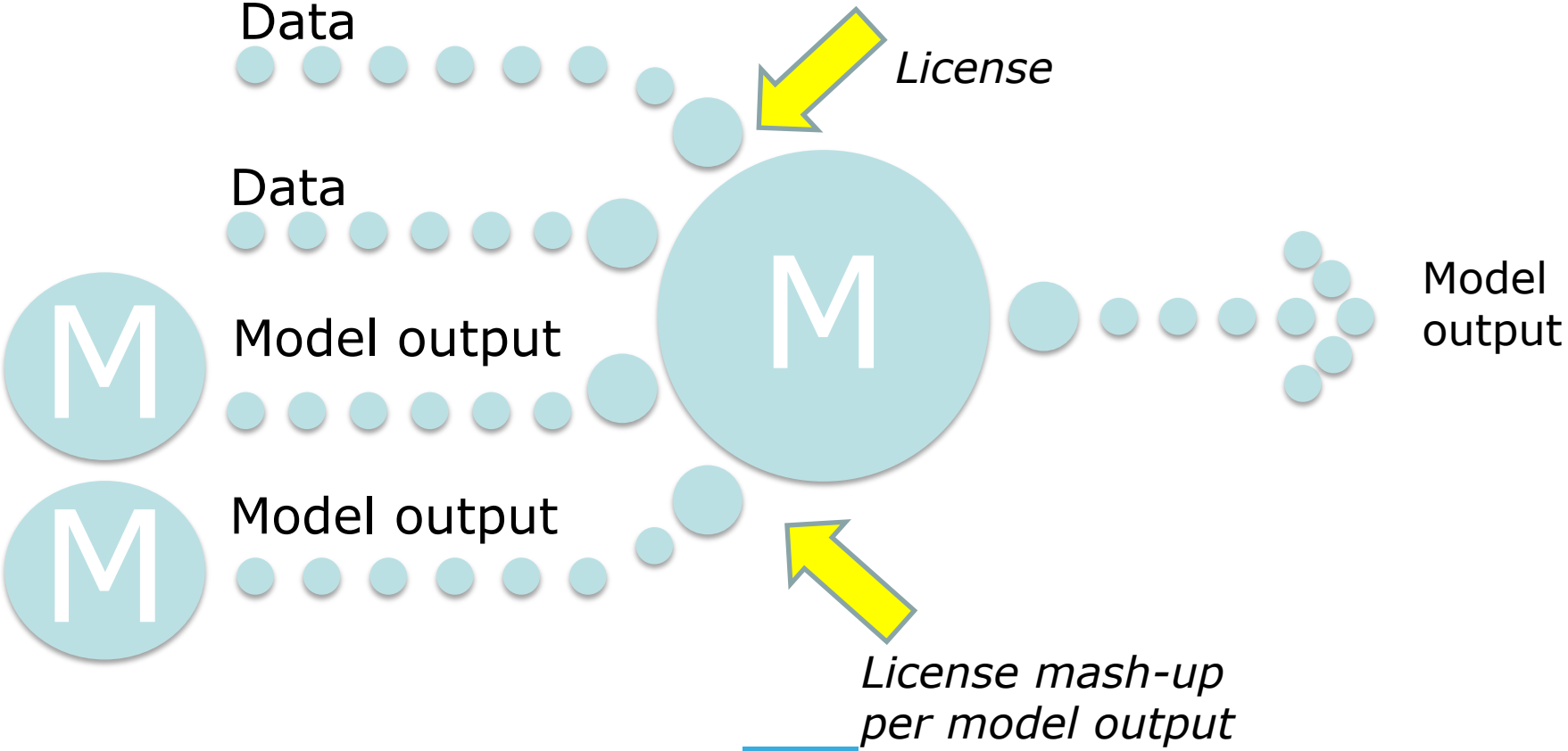


Reproducibility



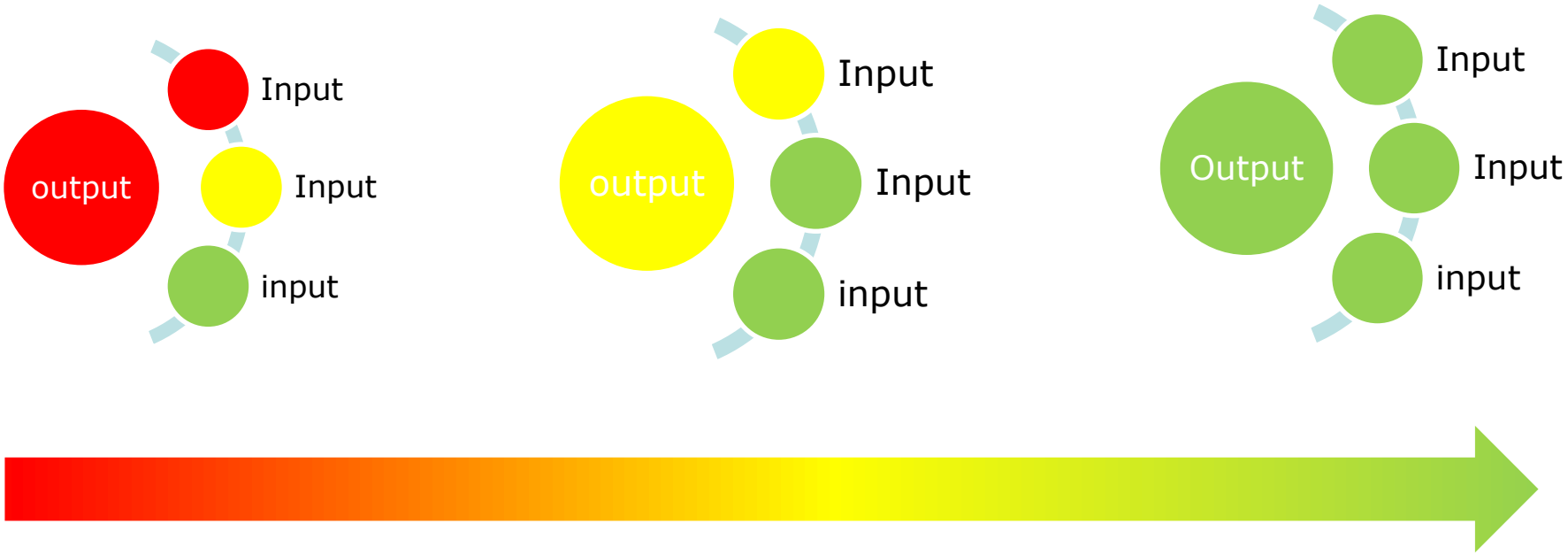
Distribution of results

Unclear or restrictive licensing of inputs prevents re-distribution of result



Traffic light system

to choose input data based on desired output distribution



Level of desired accessibility of model output

A (non-exaustive) data policy and management shopping list …

- Unique persistent identifiers (incl vers.) for internal & external data
- Data should be as much as possible open and accessible by default
- Hierarchical representation of data and licensing
- Long term preservation of data (both internal and external)
- Finite set of machine readible licenses
- Machine readible lineage
- …