

Data Policy and Data Management – Sample Experiences and Requirements

Silvia Dalla Costa and Nicole Ostlaender

www.jrc.ec.europa.eu

*Serving society
Stimulating innovation
Supporting legislation*



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



The screenshot displays the IPChem website interface. At the top, it features the European Commission logo and the text 'Information Platform for Chemical Monitoring data Enhancing access to chemical data'. Below this, there is a search bar with the text 'Search Chemical:' and a dropdown menu showing search results for 'ben'. The results include: benzene, benzazone, benalaxyl, bendiocarb, benfluralin, benfluracarb, benoxyl, benoxacor, benzamide, 2-amino-n-[1-methylethyl]-, benzene, 1,2-dichloro-3-nitro-. A map of Europe is visible on the right side of the search results. Below the search results, there are four modules: Human Biomonitoring Data, Environmental Monitoring Data, Food and Food Manufacturing Data, and Product and Indoor Air Data. The website also includes a 'Who is participating?' section listing various institutions and a 'Related Information Systems' section listing databases like Indoor Air Quality (IRC), eChemPortal (OECD), EISS (IRC), ERODNet Chemical Portal (DG MARE), and Euronet (EEA). The bottom of the page features a 'Joint Research Centre' logo.

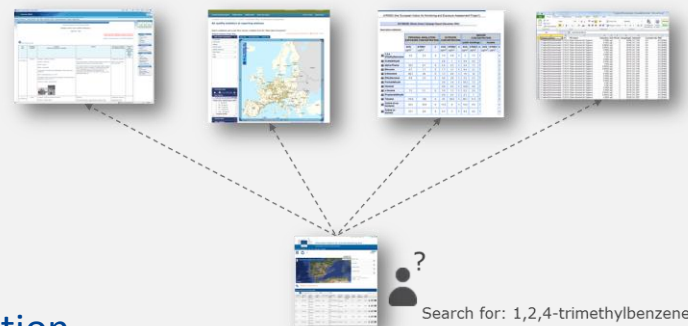
IPChem definition

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

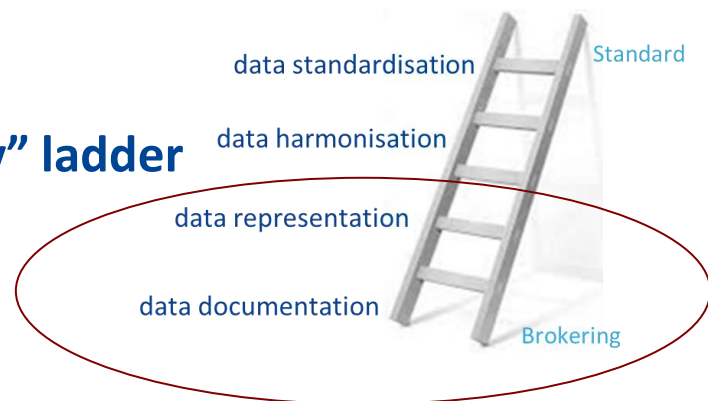
Improve comparability of the data by promoting standardisation of data and metadata and the improvement of quality assurance standards (Stated in the “IPChem Scoping Paper”)

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 exposed 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

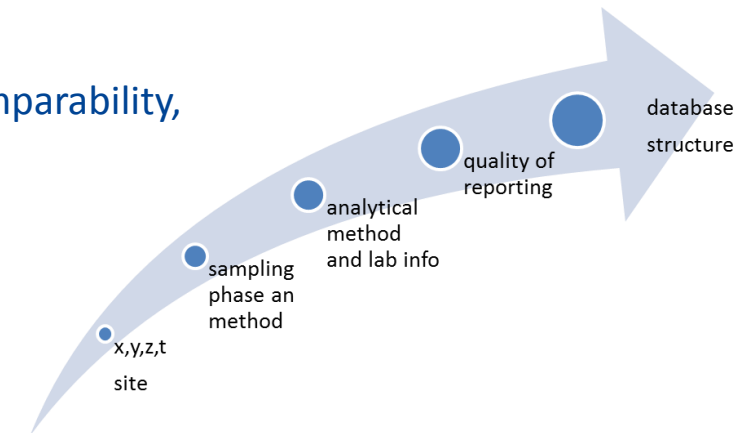
(Sometimes) lack in data documentation and data traceability, lack of QA/QC info

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

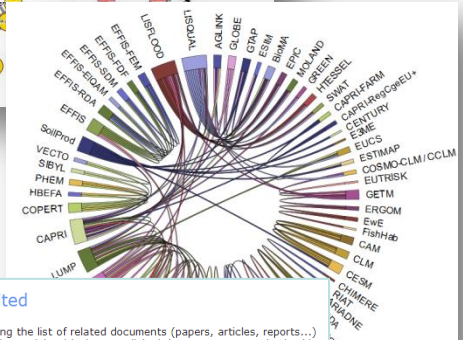
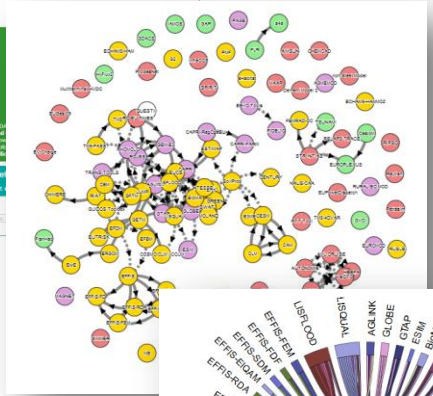
Present: DG RTD is member of both the working groups + supports JRC and DG ENV to promote and inform about IPChem to FP7-FP6 project coordinators

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”

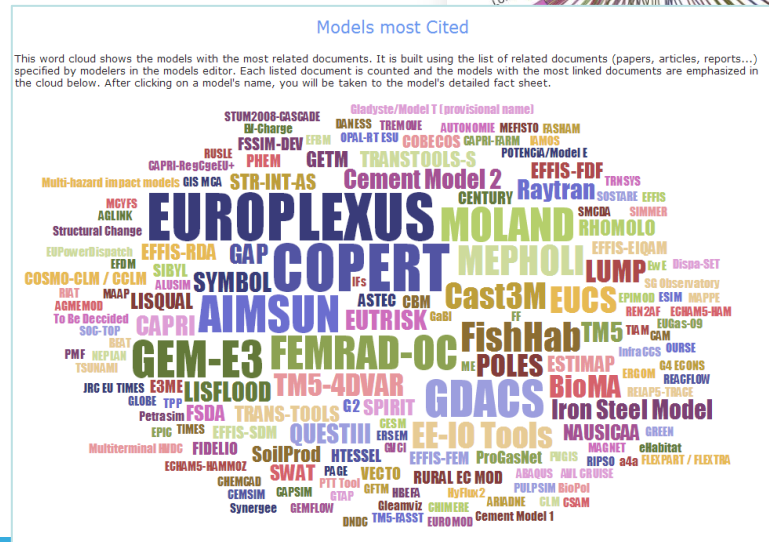


MIDAS

Modelling Inventory Database and Access Services



data policy and management requirements from the perspective of modelling

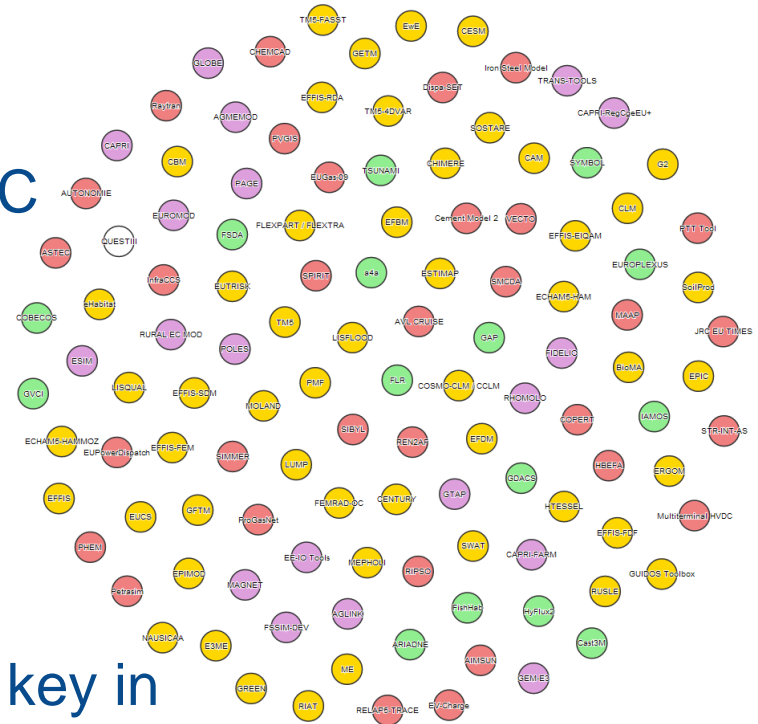


Content

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

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



MIDAS Scope

Online platform accessible from within the Commission Network.

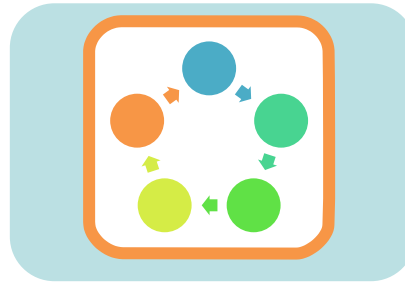
Gives access to an up-to-date database of models that in use in JRC.

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



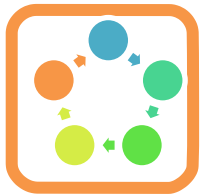
MIDAS Entities

Searchable
entities in the
MIDAS DB

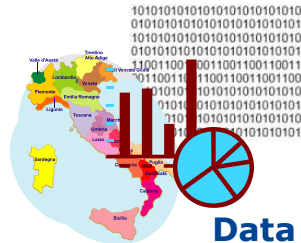


Models

Related entities



Models



Data



People



JRC Scientific
publications

Linka Browser: The graph displays the links between a root Model and the related Models & datasets, as specified in the **models** section.

Refresh

Include external

Exclude external

Only inner circle

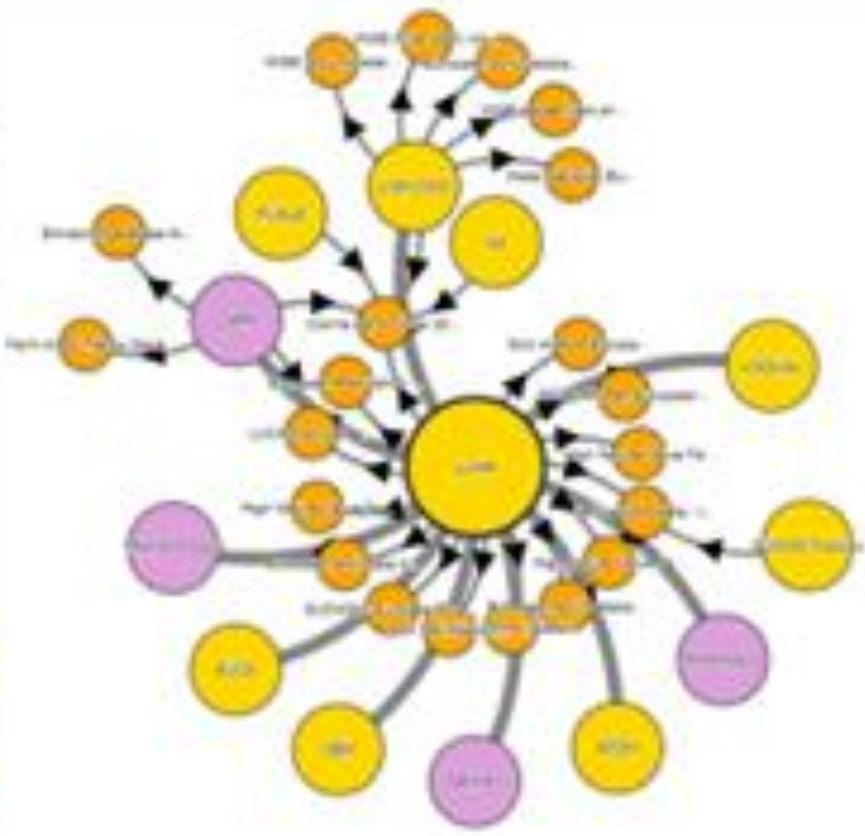
Only string circle

Add datasets

Remove datasets

Reset dataset list

All
 BT
 ITC
 ITR
 Other
 External



Land Use Modeling Platform

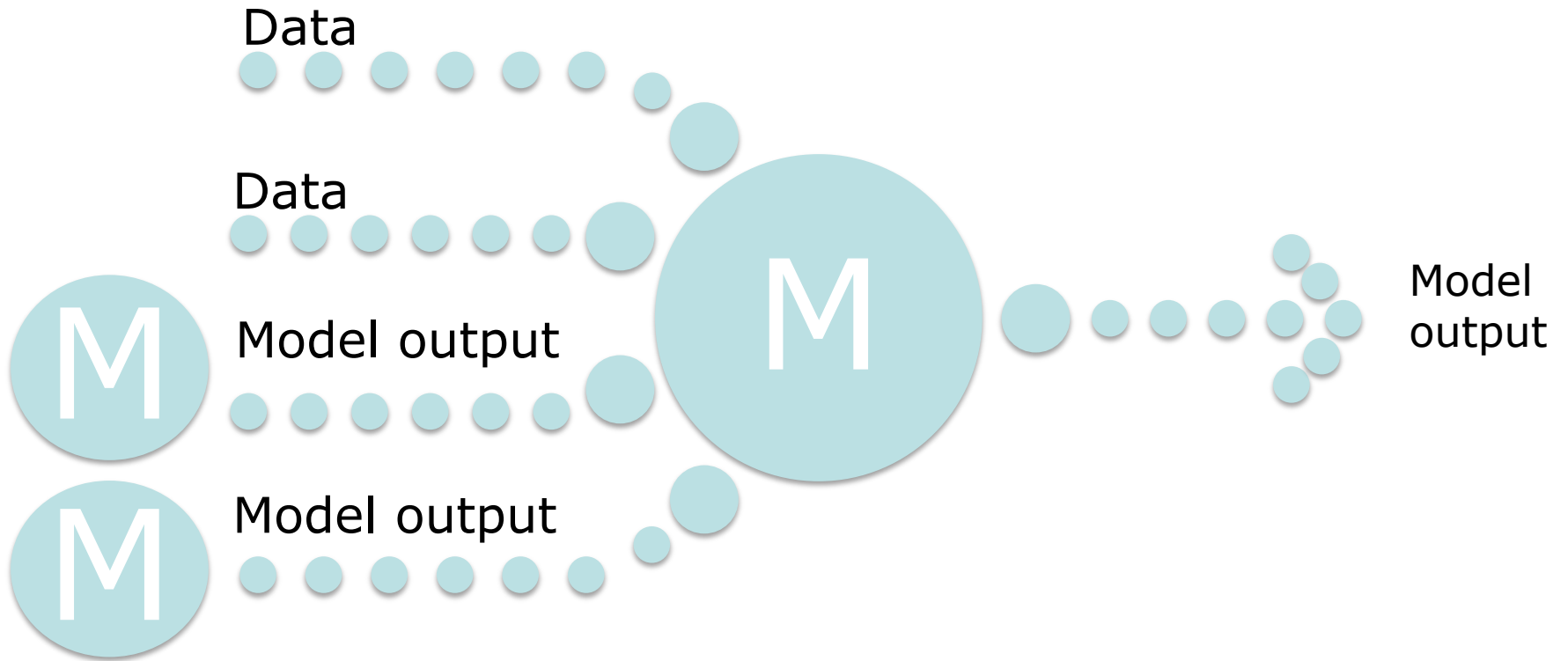
of environmental air-quality, air-land-use and land-use dynamics & related processes. For more information, visit [LINKA Browser](#).

The changes in the cover and use of the surface of the earth depend on natural processes and are, to the same time, affected by demographic, economic, cultural, political, and technological drivers. A land-use/cover model can help to understand and integrate the complex interplay between the bio-physical and human factors that influence land use/cover dynamics. In addition, it can be used as a tool to assess environmental consequences of policies with direct or indirect land-use impacts.

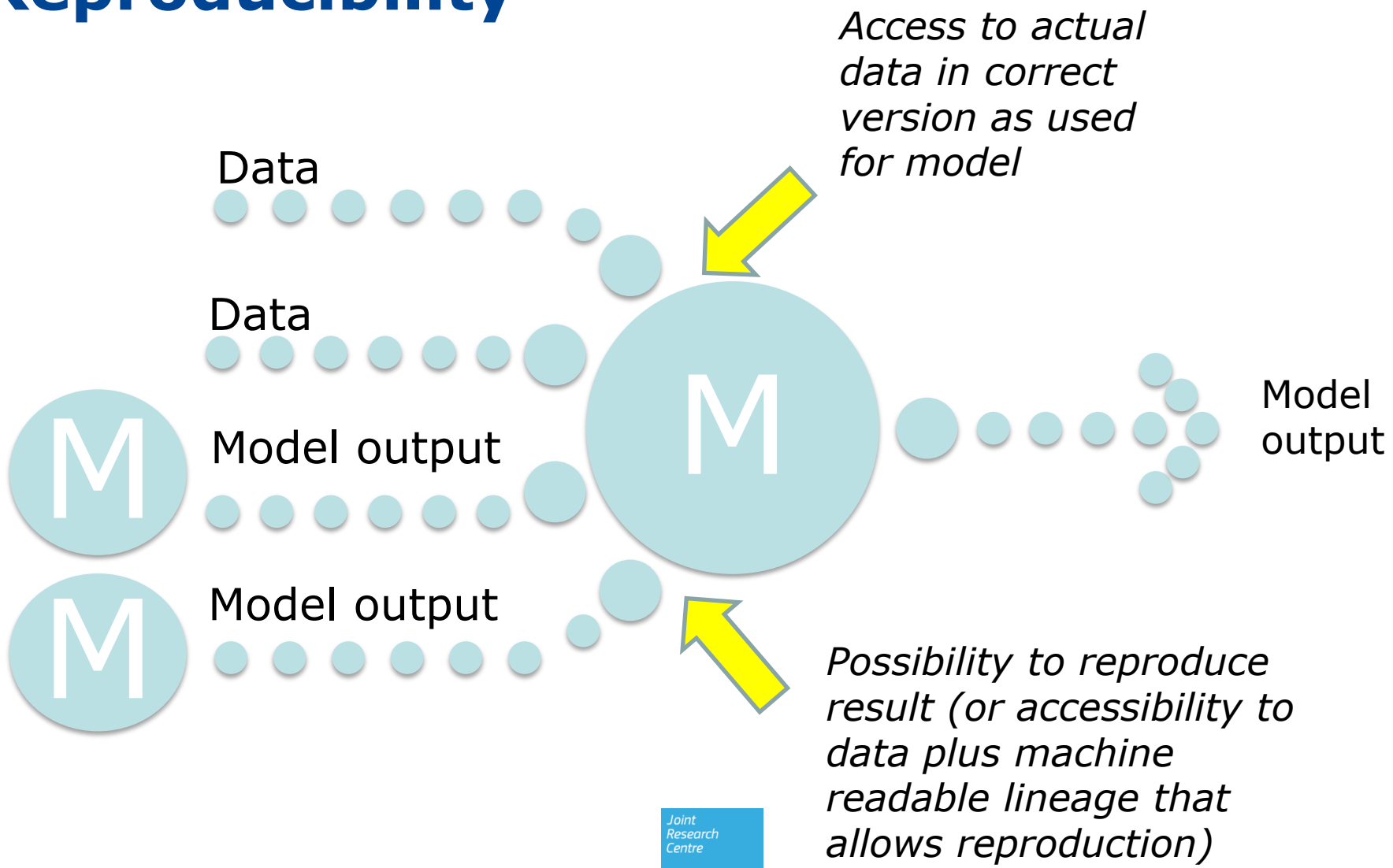
The Land Use Modeling Platform (LUMP) has been developed by the Institute for Environment and Sustainability (IES) of the Joint Research Centre (JRC) to support the policy design of different services of the European Commission. In order to provide a comprehensive, consistent and harmonized analysis of the impacts of policies and/or specific proposals in the context of environmental and socio-economic changes in Europe, LUMP is based upon the combination of a publicly available land use model and to integrate with other modeling services in thematic fields such as forestry, agriculture, economy, forestry, etc.

The Land Use Modeling Platform has a modular structure and is organized in three main components: 1) the land demand module; 2) the land allocation module; 3) the scenario module. The land demand module, which is the core of the platform, is designed for different land uses. A set of different policies and algorithms are used to compute demands for each land use. A range of maximum and minimum demand for each land use, for each year and for each EU/EEA, is entered into the demand module, the land allocation module, which is the core of LUMP. The scenario module is responsible for allocating the yearly projected quantities of land in space (at grid level). This module is also called the *land-use scenario* (LUS) and uses

Areas of impact

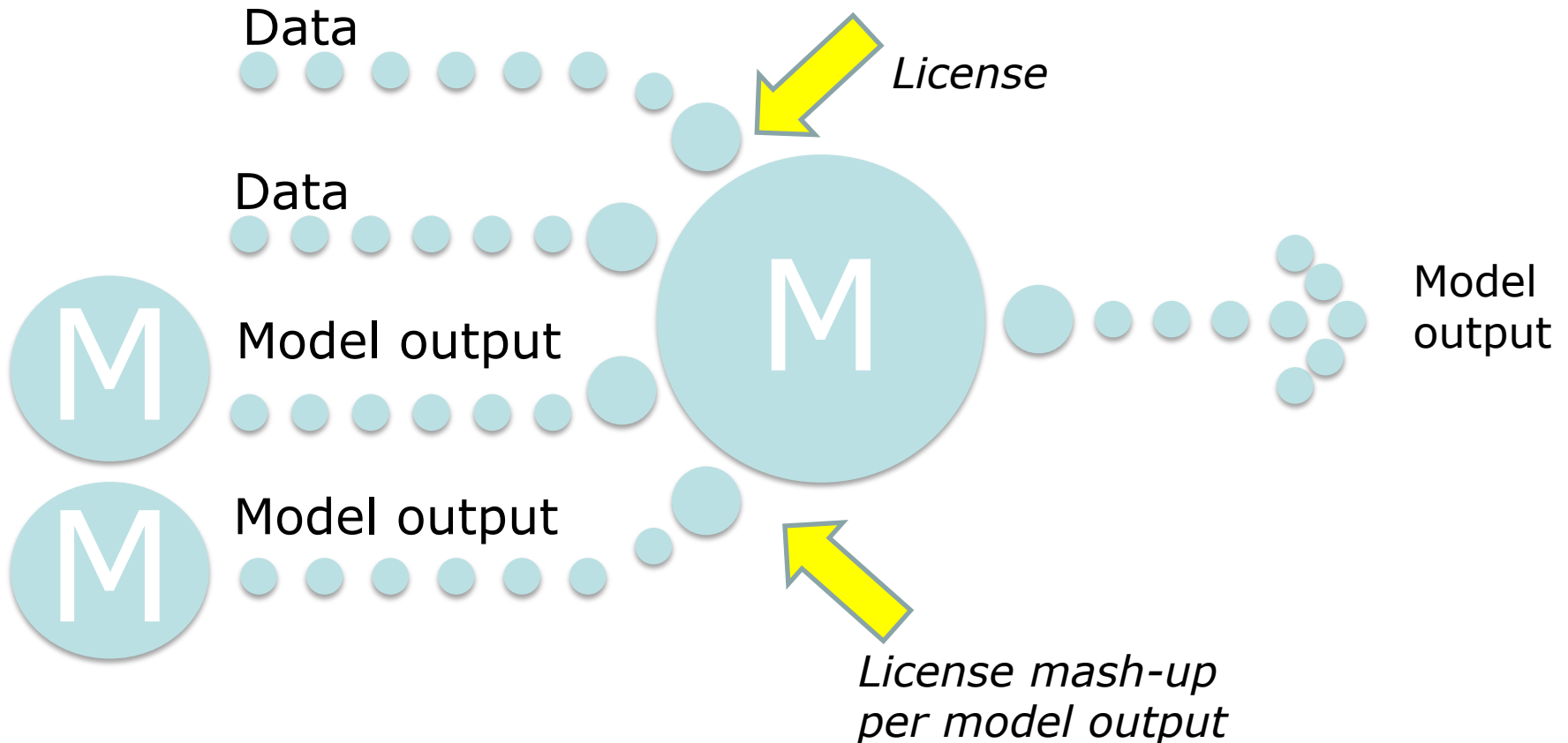


Reproducibility



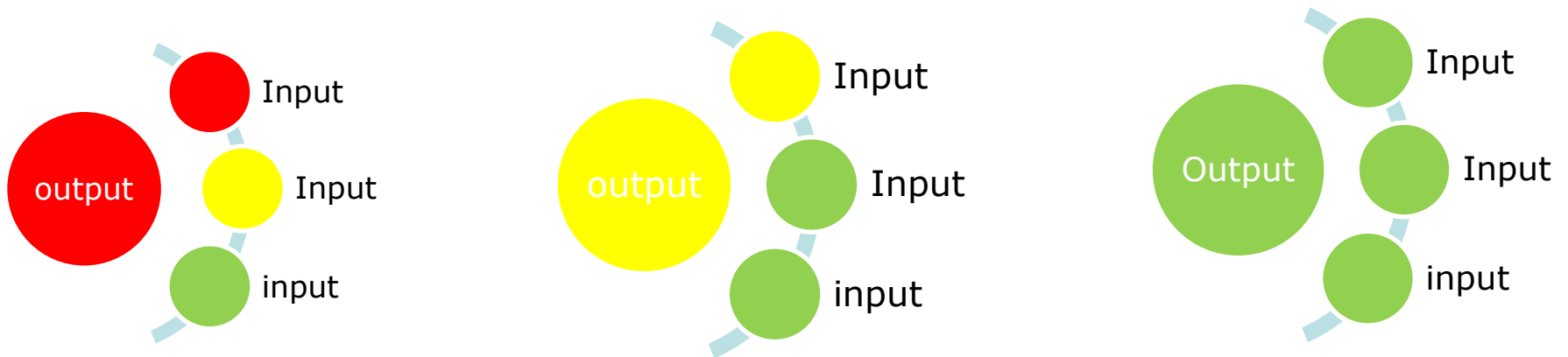
Distribution of results

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-exhaustive) 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 readable licenses
- Machine readable lineage
- ...

