# Unified Materials Information System (UMIS): An Integrated Material Stocks and Flows Data Structure

**Citation for published version:**
Myers, RJ, Fishman, T, Reck, B & Graedel, TE 2018, 'Unified Materials Information System (UMIS): An Integrated Material Stocks and Flows Data Structure', *Journal of Industrial Ecology*.
https://doi.org/10.1111/jiec.12730

**Digital Object Identifier (DOI):**
10.1111/jiec.12730

**Link:**
Link to publication record in Edinburgh Research Explorer

**Document Version:**
Peer reviewed version

**Published In:**
Journal of Industrial Ecology

# Unified Materials Information System (UMIS): An Integrated Material Stocks and Flows Data Structure

Rupert J. Myers [1,a,*], Tomer Fishman [1,b], Barbara K. Reck [1,c], T. E. Graedel [1,d]

[1] Yale School of Forestry & Environmental Studies, Yale University, 195 Prospect St, New Haven 06511 Connecticut, United States

[*] Corresponding author. Email: [a] rupert.myers@gmail.com; [b] tomer.fishman@yale.edu; [c] barbara.reck@yale.edu; [d] thomas.graedel@yale.edu.

## Abstract <heading level 1>

Modern society depends on the use of many diverse materials. Effectively managing these materials is becoming increasingly important and complex, from the analysis of supply chains, to quantifying their environmental impacts, to understanding future resource availability. Material stocks and flows data enable such analyses but currently exist mainly as discrete packages, with highly varied type, scope, and structure. These factors constitute a powerful barrier to holistic integration and thus universal analysis of existing and yet to be published material stocks and flows data. We present the Unified Materials Information System (UMIS) to overcome this barrier by enabling material stocks and flows data to be comprehensively integrated across space, time, materials, and data type independent of their disaggregation, without loss of information, and avoiding double counting. UMIS can therefore be applied to structure diverse material stocks and flows data and their metadata across material systems analysis methods such as material flow analysis (MFA), input-output (I/O) analysis, and life cycle assessment (LCA). UMIS uniquely labels and visualizes processes and flows in UMIS diagrams; therefore, material stocks and flows data visualized in UMIS diagrams can be individually referenced in databases and computational models. Applications of UMIS to restructure existing material stocks and flows data represented by block flow diagrams, system dynamics diagrams, Sankey diagrams, matrices, and derived using the ‘economy-wide’ MFA classification system are presented to exemplify use. UMIS advances the capabilities with which complex quantitative material systems analysis, archiving, and computation of material stocks and flows data can be performed.

# Introduction <heading level 1>

A wealth of material stocks and flows data has been compiled and analyzed since the emergence of material systems analysis and materials management practices in the 20th century and the industrial ecology field in the late 1980s (Frosch and Gallopoulos, 1989; Ayres, 1992). These data are diverse in scope, were generated using various analytical approaches, and are published at different levels of detail in various tabular and graphical formats. They cover various topics, e.g., environmental pollutant flows in river basins (Ayres et al., 1988), material use in cities (Hoekman and von Blottnitz, 2016), anthropogenic systems (Graedel et al., 2004), coupled anthropogenic and natural systems (Rauch and Graedel, 2007), and the (life) cycles of materials and their constituent substances (e.g., electrical wire and copper (Wang et al., 2015)).

Material systems analysis fundamentally involves the analysis of the type and quantity of existing materials, how and to what extent they get transformed in and distributed among (enter and leave) processes such as production, use, and recycling in anthropogenic systems, and their associated impacts on economic and natural systems (i.e., environmental impacts). The natural system is constituted by natural processes such as nutrient cycling among organisms in marine ecosystems excluding humans, which is depicted in food webs (Polis and Winemiller, 1996), whereas the anthropogenic system is constituted by anthropogenic processes such as manufacturing, construction, transportation etc. (Ayres, 1994) typically along industrial supply and value chains. Therefore, a process such as fishing represents a linkage, possibly the transformation (e.g., from alive to dead fish), distribution (e.g., from the ocean to boat), and/or storage (e.g., withdrawal from the ocean and deposition into a bucket), of material between anthropogenic and natural systems (Figure 1). It is notable that material stocks and flows data are treated similarly in the analysis of natural (e.g., food webs) and anthropogenic (e.g., supply and value chains) systems, and that material processing changes the location but not the cumulative mass of material in the combined anthropogenic and natural system (excluding nuclear reactions). These data can thus be reconciled into a single unified structure. Consideration of both natural and anthropogenic processes is essential to the holistic analysis of material systems.

Figure 1. Relationships between material stocks and flows in anthropogenic and natural systems. Material stored in a particular reservoir undergoes processing, storage, distribution, and transformation, to again become stored in another (one or more) reservoir(s). Total mass is conserved but the location of the material changes. These relationships between reservoirs and processes provide a basis upon which a unified structure for material stocks and flows data can be built.

Material stocks and flows data have been individually compiled and published for decades in diverse and seemingly inconsistent formats that typically serve small sections of the material systems analysis research community. Although these data have proliferated in recent years, it is challenging to synthesize, build on, and enhance them due to their diverse and inconsistent formatting. For example, the combined use of material stocks and flows data in monetary and mass units can provide a greatly enhanced description of anthropogenic systems relative to what can be accomplished using only one of these data types, and there is an abundance of both types of data (Chen and Graedel, 2012; Lenzen et al., 2014), however these data are relatively infrequently used together in holistic material cycle investigations (Nakajima et al., 2013; Chen et al., 2016). This effort of combining multiple data types is hampered by the absence of a single flexible, universally applicable, standardized, and generic machine readable data structure that can be applied without loss of information. Reconciliation of material stocks and flows data into such a structure has not yet been achieved but would provide a foundation to develop substantially more functional, holistic, and higher complexity databases and quantitative computational models of anthropogenic and natural systems. It would therefore improve data availability, increase the reproducibility of research results, eliminate repetition of work, integrate research efforts to advance our understanding of material systems issues such as the sustainability and resilience of industrial supply chains, and increase the effectiveness of the material systems analysis research community.

Industrial ecology and material systems analysis research occurs to a significant extent through applications of the three following methods, the choice depending on the scope of the investigation and thus also on the level of disaggregation of the available relevant data:

1. Materials flow analysis (MFA), which is described as “a systematic assessment of the flows and stocks of materials within a system defined in space and time” (Brunner and Rechberger, 2005). The level of data disaggregation used in a MFA investigation varies significantly depending on its scope and data availability; it can be relatively low (Graedel et al., 2005; Hoekman and von Blottnitz, 2016) (describing very aggregate processes and

materials, e.g., production and biomass, respectively) or rather high (Meylan and Reck, 2017) (e.g., ‘copper; strip, of a thickness exceeding 0.15 mm, of copper-zinc base alloys (brass), in coils’). MFA data often describe partial or complete material cycles (Graedel et al., 2004), but also frequently describe more aggregate data and indicators such as domestic extraction in ‘economy wide’ MFA (EW-MFA); such data can exist on the firm level and sub-national (e.g., river basins and cities), country, international, and global scales (EUROSTAT, 2001; Fischer-Kowalski et al., 2011).

2. Life cycle assessment (LCA), which has as its objective to “[compile] and [evaluate] the inputs, outputs, and potential environmental impacts of a product system throughout its life cycle” (Hellweg and i Canals, 2014). LCA data are normally relatively highly disaggregated and refer to multiple materials, owing to the need to describe the full ensemble of environmental inputs and outputs relevant to a product system, yet often use generic or non-process specific data.
3. Input-output (I/O) analysis, which differs from LCA and MFA in that it tracks monetary flows through the economy in matrices that are “generally constructed from observed economic data for a specific geographic region” (Miller and Blair, 2009), to e.g., allocate environmental impacts to products and services. More aggregated descriptions of the economy are typically investigated using I/O analysis rather than LCA, consistent with economic data published by e.g., national statistical offices. I/O analysis and LCA data have been harmonized in multi-regional I/O tables (Lenzen et al., 2014) and I/O-LCA models (Hawkins et al., 2007) by reconciling differences in data (dis)aggregation. However, I/O analysis and MFA data, despite sharing some key concepts (e.g., accounting of material flows), are often disaggregated differently. The former normally describe

multiple materials in individual industries and products (i.e., not material specific), whereas the latter typically describe a single material across a small number of products and industries (i.e., material specific).

Pauliuk et al. (2015) recently showed that material stocks and flows data can be unified across MFA, I/O analysis, and LCA by employing the make and use table approach used to compile I/O tables (EUROSTAT, 2008). Consistency with this approach can be achieved by transforming material stocks and flows data into the bipartite directed graph structure (i.e., a graph representing a system containing two types of processes and only flows between processes of different type). In practice, the bipartite directed graph structure can be attained by ensuring that transformative processes are always followed by one or more flows that each terminate at distributive processes, and vice versa. This representation is realistic because transformed materials are typically distributed to locations different from where they were produced. We build on these insights and address the challenge of unifying material stocks and flows data across MFA, I/O analysis, and LCA methods by:

1. Using a substantially more visual approach and nomenclature more closely aligned with MFA rather than I/O analysis;
2. Establishing a labeling system that facilitates referencing between the visualized data, databases, and computational models;
3. Emphasizing connections between different material cycles;
4. Discussing how diverse and differently disaggregated data are harmonized without double counting; and by

5. Demonstrating how to transform different types of material stocks and flows data into a unified structure.

Key MFA concepts are now introduced to establish a foundation upon which a unified structure for material stocks and flows data is developed.

# Material Flow Analysis Data Organization: The Existing State of the Art <heading level 1>

The basic attributes of MFA are that the mass conservation principle is respected and that the investigated system is represented by processes, stocks, and flows. The investigated system is specified using a ‘system boundary’ defined in terms of space (reference space), time (reference timeframe), and one or more materials (reference material) (Brunner and Rechberger, 2005). The reference timeframe can be a time period, e.g., a year, or a specific point in time, e.g., the end of a year. Exemplary block flow type diagrams (Figure 2) depict this information by differentiating among transformative, distributive, and storage processes. They also differentiate among flows that are internal to (hereafter termed ‘flows’) and cross the system boundaries (hereafter termed ‘cross boundary flows’, or ‘trade flows’ if the reference spaces represent independent economic entities e.g., countries in Figure 2) (Pauliuk et al., 2015; Müller et al., 2006). Transformative, distributive, and storage processes transform process inputs to outputs, distribute process outputs to inputs, and produce or release stocks, respectively. It is typical to assign processes to each major stage in anthropogenic material cycles (these are often production, fabrication & manufacturing, use, and waste management) (Graedel et al., 2002). MFA diagrams sometimes display uncertainty (Rauch and Pacyna, 2009) and also differences that result from applications of the mass conservation principle (i.e., a ‘mass balance’) when compared to the observed data (leading to ‘mass balance residuals’) (Graedel et al., 2004). However, MFA diagrams that incompletely distinguish among the aforementioned types of processes and flows dominate (the distinction is often either implied or unnecessary if the system boundary coincides with a single transformative process) (Hendriks et al., 2000; Tanimoto et al., 2010; Uihlein et al., 2006; Davis et al., 2007; Müller, 2006). Material stocks and flows data are also visualized using other types of diagrams, e.g., Sankey (Schmidt, 2008) and system dynamics (Ford, 1999) diagrams, which share some of these attributes.

Figure 2. Exemplary block flow type diagram for the iron cycle, the year 2000, and the United States, adapted from (Müller et al., 2006). Mass quantities in Tg/year are displayed adjacent to each respective flow. Mass balance residuals are not shown (e.g., around the ‘Blast Furnace’ transformative process). Note that some distributive processes needed to avoid material flowing between two processes of the same type and thus to ensure consistency with the bipartite directed graph structure are omitted, e.g., between the ‘Manuf.’ and ‘Scrap Process. & Waste Manag.’ transformative. Production (dashed green box), engineering materials (dashed yellow box), fabrication & manufacturing (dashed purple box), use (dashed orange box), waste management (dashed red box), and environment (dashed blue box) subsystems are added to illustrate the subsystem concept (see Development of the Unified Materials Information System (UMIS)).

However, most MFA diagrams are used to communicate key messages and quantitative results rather than to place and show data in complete detail and in their exact context within material systems. Therefore, the formatting of these MFA ‘communication diagrams’ changes greatly depending on the number of processes displayed, data availability, and investigation scope (Lupton and Allwood, 2017). Consequently, most are significantly mismatched with one another in style and detail even when describing similar systems (Wang et al., 2007; Pauliuk et al., 2013; Cullen et al., 2012; Müller et al., 2006). MFA communication diagrams also typically do not normally use explicit, standardized labeling systems to annotate processes and flows. These attributes hinder their utility to illustrate the kind of highly structured and detailed (meta)data that are used in databases and computational models of complex material systems (e.g., the exact positions of material stocks and flows data in highly and differently disaggregated material cycles). Explicitly and comprehensively indexing material stocks and flows data visualizations is beneficial in computational modeling of complex material systems because it allows visualized information to be precisely referenced. Therefore, the increasing complexity of data analysis and availability of data in industrial ecology is creating a growing need to develop ‘elicitation diagrams’ that can visualize fully detailed material stocks and flows data in their exact systems context within a standardized and labeled structure.

The goal of this paper is thus to develop a Unified Materials Information System (UMIS) to structure, label, and visualize diverse material stocks and flows data and their metadata (e.g., uncertainty, system boundary properties) into a single standardized format. UMIS could then consolidate datasets across the major material systems analysis methods, e.g., MFA, I/O analysis, and LCA. Here, the ‘whole system’ describes the entire system in its most general sense, including the anthroposphere and nature, for all references spaces, reference timeframes, and reference materials. UMIS is visualized in terms of matrix type ‘UMIS diagrams’ showing material inputs, outputs, and processing. The UMIS diagram for each reference material is unique because the processes, stocks, and flows that comprise each material cycle are unique. For example, UMIS diagrams for iron in the United States in the year 2000 and for iron in Australia in the year 2017 are equivalent, but both are different from the UMIS diagram for copper in the United States in the year 2017. This representation means that any irrelevant (e.g., obsolete) processes and flows for a reference material in a particular reference timeframe or reference space remain in UMIS diagrams and are associated with zero material mass. The effort focuses on materials and mass, two fundamental foci of material systems analysis research. Such an approach is naturally aligned with MFA methodology although we show that it can be readily applied to other data types (e.g., monetary and energy) and methods (e.g., I/O analysis and LCA). This paper also aims to develop UMIS so that data visualized in UMIS diagrams can be readily referenced in databases and computational models. Another aim of this paper is to demonstrate how UMIS is used to transform and visualize material stocks and flows data into its standardized structure (these demonstrations are presented as Supporting Information, SI).

# Development of UMIS <heading level 1>

In the sections that follow, the UMIS is developed by: (1) defining concepts and notation needed to define (2) a comprehensive data structure and elicitation diagrams for material stocks and flows data; (3) strategies to facilitate flexible data disaggregation and also (4) to avoid double counting in computational models utilizing the data structure; (5) implementation of multiple reference spaces, reference timeframes, and reference materials into the data structure; and (6) the treatment of metadata in the data structure, including units and uncertainty.

## Reconciling Data across MFA, I/O Analysis, and LCA <heading level 2>

Development of UMIS begins by applying the aforementioned MFA concepts to reconcile MFA, I/O analysis, and LCA data using their common ability to quantitatively analyze flows of materials along their cycles. The architecture for such an effort involves connecting material stocks and flows data into a single structure with a flexible level of disaggregation. It is desirable if this effort structures data independent of its units so that it can be applied to many data types, e.g., radioactivity (Bq), energy (kJ), and monetary ($).

Figure 3. Relationships between (A) I/O analysis (make and use tables), (B) MFA (block flow type diagram), and (C) LCA (inventory) data. Transformative and distributive processes are shown as darker grey filled squares and lighter grey filled circles, respectively. Flows are displayed as arrows. Colored bold arrows (B-C) are flows that are entered into the make and use tables here (A). Subsystem, aggregate subsystem module, and system boundaries are shown as dashed, bold dashed, and alternating dashed double dotted lines, respectively. Process and flow labels are used to reference data between the respective methodologies; their formulation, and also labeling of subsystems, are described in the text. The environment subsystem is included in (C) to demonstrate the compilation of an inventory table, which is done by disaggregating the aggregate production of engineering materials subsystem module *(PEM.1*) (shaded green boxes in B and C) to account for all inflows to and outflows from the *aggregate environment subsystem module* (*ENV.5*) (black bold arrows).

Figure 3 highlights commonalities and linkages between MFA, I/O analysis, and LCA data using standardized UMIS notation. The purpose of this notation is to label the visualized material stocks and flows data so that they can be uniquely referenced in databases and computational models.

### A Prescriptive Condition <heading level 3>

UMIS prescribes one outflow per transformative process. Transformative processes are disaggregated if additional outflows are needed to fully describe it. These disaggregated transformative processes again specify one outflow. A prescriptive condition such as prescribing one outflow from each transformative process defines the UMIS diagram structure so that it is machine readable and can be computationally generated. This condition enables the production of elicitation (UMIS) diagrams for highly disaggregated and complex systems like the global physical economy to be automated, which would be infeasible to do manually, and so is of major benefit in the analysis of high complexity material systems analysis data.

### Subsystems <heading level 3>

UMIS structures data using ‘subsystems’. The subsystem concept facilitates flexible structuring of data at any level and type of disaggregation. A ‘subsystem boundary’ (dashed lines, Figure 3) defines a subsystem, analogous to how a system boundary (dashed double dotted lines, Figure 3) defines a material system. Each subsystem contains a non-zero even number of processes, of which half are transformative and half are their associated distributive processes, because processes occur in pairs in UMIS to ensure consistency with the bipartite directed graph structure (and thus also the make and use table approach) (Pauliuk et al., 2015). For example, the *production subsystem* (*PEM.1;1;1*) in Figure 3B contains one transformative process (*mining*) and one associated distributive process (*mining output*). Procedures to name and label subsystems are discussed below. Subsystems are defined so that their boundaries do not intersect one another. This condition helps to avoid double counting of data (see Avoiding Double Counting of Data).

Subsystems can be infinitely disaggregated to describe more specific material stocks and flows data. Subsystem disaggregation is shown in Figures 3B and 3C, where the *production subsystem* (*PEM.1;1;1*) (Figure 3B) is disaggregated into a *mining subsystem* (*PEM.1;1;1;1*) (Figure 3C). The most aggregated subsystem represents the whole system. If a subsystem is defined to represent a stage in a material cycle (e.g., the ‘fabrication & manufacturing’ stage in Figure 1) and where cumulatively these stages represent that material cycle, a subsystem is termed an ‘aggregate subsystem module’ (see Subsystem Specification and Disaggregation). Therefore, a subsystem boundary can be a subset of, the same as, or a superset of one or more system boundaries, or exist outside the system boundary (e.g., the *aggregate environment subsystem module* (*ENV.5*), Figure 3C), depending on how these boundaries are defined.

### Labels <heading level 3>

In Figure 3, flows are represented by arrows, whereas transformative and distributive processes are represented by dark grey squares and light grey circles, respectively. Process labels (located directly above processes in Figures 3B and 3C) are specified as *a.b.c.d.e*, where *a* represents the reference material defined by the system boundary (*a* = 1 for reference material $m_1$), *b* defines the aggregate subsystem module abbreviation, *c* is the subsystem code, *d* indicates the type of process, transformative (*T*) or distributive (*D*), and *e* is a process code that is unique to each process in each subsystem for reference material *a*. Flow labels (located adjacent to flow arrows in Figures 3B and 3C) are specified in the form *origin_destination*, where *origin* and *destination* specify the labels of the processes that a flow originates and terminates at, respectively (e.g., the flow from *1.PEM.1;1;1.D.2;2* to *1.F&M.2;1;2.T.1;1* is labeled *1.PEM.1;1;1.D.2;2_1.F&M.2;1;2.T.1;1*, where *PEM* refers to an 'aggregate production of engineering materials module'). A subsystem label is specified by the aggregate subsystem module abbreviation followed by a period and then the subsystem code, i.e., *b.c*. For example, the subsystem label for the *production subsystem* in Figure 3B is *PEM.1;1;1*.

### Codes <heading level 3>

A subsystem code (*c*) is specified according to the level of data disaggregation, with its character length excluding semi-colons specifying the disaggregation level. Process codes (*e*) indicate the positions of processes in subsystems (see Transforming Data into Matrix Format). These positions begin at matrix coordinates of 1;1 (*row*;*column*) in each subsystem (i.e., 1;1 indicates the top left corner cell in a subsystem). Semi-colons are used to separate numerical values in subsystem codes (*c*) and process codes (*e*) for clarity. For example, the subsystem represented by the abbreviation

*F&M* and subsystem code *2;1;2* in Figure 3B (i.e., the *F&M.2;1;2* subsystem) represents data for the second transformative process (*Manufacturing*) in the *F&M.2;1* subsystem (not shown in Figure 3). Therefore, it also exists within the aggregate *fabrication & manufacturing* subsystem module *F&M* on the third disaggregation level (character length excluding semi-colons(*2;1;2*) = 3).

**Names <heading level 3>**

Process names are displayed on processes, with ‘output’ used here to refer to transformative process outputs in general. Our vision is that process names will be unambiguously defined using an internationally standardized terminology in the future that is established and widely used by material stocks and flows data providers, which is also not specific to a particular material systems analysis technique, e.g., harmonized system (HS) codes; the development of this standardized classification system is beyond the scope of this work. Therefore, process names specified here are used to describe concepts and the initial implementation of UMIS only, which should be recognized as ‘place holders’ due to the absence of this standardized classification system.

**I/O Analysis and LCA in UMIS <heading level 3>**

Make and use tables are used in UMIS for consistency with I/O analysis. They are compiled in Figure 3A using flows within the system boundaries shown in Figures 3B and 3C only. This condition is imposed to simplify our illustration and so does not represent an intrinsic limitation of UMIS. The labels of flows used to construct the make and use tables are shown in purple (mining industry outputs), blue and red (mining outputs used in the construction and manufacturing industries, respectively), green (manufacturing industry outputs), and pink (construction industry outputs) text.

LCA inventory tables can be compiled using data structured by UMIS (Figure 3C). Here, processes (e.g., *mining type A*, *1.PEM.1;1;1;1.T.1;1*, *mining type B*, *1.PEM.1;1;1;1.T.3;3*, and *mining type C*, *1.PEM.1;1;1;1.T.5;5*) in the *mining* subsystem (*PEM.1;1;1;1*) are specified by disaggregating processes in the *production* subsystem *PEM.1;1;1* (in this case the *mining* and *production* subsystems are substitutable). Complete representation of the inventory data is achieved by specifying an aggregate *environment* subsystem module (*ENV*) and disaggregating all aggregate subsystem modules to the appropriate level such that all relevant flows to and from this aggregate *environment* subsystem module are explicit (it is necessary to disaggregate *PEM* in Figure 3B to explicitly show these flows in Figure 3C, shaded green boxes). The aggregate *environment* subsystem module is external to the system boundary in this example. The complete set of aggregate subsystem modules here, i.e., aggregate *production of engineering materials* (*PEM*), *fabrication & manufacturing* (*F&M*), and *environment* (*ENV*) subsystem modules, represents the combined anthropogenic and natural system boundary for a single reference material and reference timeframe.

## Transforming Data into a Matrix Format <heading level 2>

UMIS is visualized using matrix type UMIS diagrams. Visualizing MFA data in matrices is analogous to typical representations of I/O analysis and LCA data (e.g., physical I/O tables), and so facilitates convergence of these methods. Our effort here builds on existing matrix-based visualizations and computational analysis of material stocks and flows data (Pauliuk et al., 2015;

Nakamura and Nakajima, 2005; Eckelman and Daigo, 2008; Nakamura et al., 2011; Yamada et al., 2006). Material stocks and flows data visualized in matrix formats conform directly to the way in which these data are treated in computational models (as matrices). Therefore, material stocks and flows data structured in matrix format can be readily referenced in computational models and databases that require indexing of many data inputs, for which the natural indices are row and column coordinates.

### Processes and Flows <heading level 3>

Transformation of block flow type diagrams (Figures 3B-3C) into matrix format is achieved by specifying inputs to processes as columns and outputs from processes as rows (Figure 4A), with processes positioned along the matrix diagonal. The matrix for each subsystem is square because each transformative process has exactly one output that is assigned a distributive process. This set of processes, one transformative and one distributive process, represents the basic building block of UMIS.

UMIS diagrams are defined such that transformative (dark grey squares), distributive (light grey circles), and storage processes (small light grey rectangles), and flows (faded red diamonds), are illustrated using the standardized notation introduced in Figures 3 and 4. Flows originate and terminate at processes only. They follow a clockwise direction in UMIS diagrams; i.e., a flow originating at a process in the upper left of the matrix terminates at a process below and to the right of it, with its label located in the upper matrix triangle. The absence of a red diamond in a cell indicates no flow. An empty bottom right matrix quadrant is generated if flows crossing subsystem boundaries (i.e., cross boundary flows) are displayed in UMIS diagrams (Figure 4A). These matrix diagrams retain the same system boundary definitions as defined in block flow type diagrams (Figures 3B and 3C), i.e., defined in terms of a reference material, a reference timeframe, and a reference space.

Figure 4. (A) Key aspects of UMIS, illustrated using UMIS type diagrams for one of each transformative, distributive, and storage process, three flows, the virtual reservoir, and the metadata layer. (B) The virtual reservoir shown here can lie inside or outside the system boundary, but occurs inside of it here. The metadata layer contains additional information (e.g., uncertainty, system boundary properties) about processes, stocks, and/or flows positioned at the same matrix coordinates. Flows depicted by grey arrows in (A) and conceptual linkages depicted by black arrows in (B) are omitted in UMIS diagrams, and are only shown here to guide readers.

### Stock <heading level 3>

Conceptually, storage processes are connected to stocks residing in a ‘virtual reservoir’ that is implicitly described by UMIS diagrams (it is shown in Figure 4B to illustrate the concept). The reservoir is ‘virtual’ because in reality stocks reside within processes, whereas in UMIS they are conceptualized as residing in their own layer to facilitate better integration with flow-based material systems analysis methods such as I/O analysis. The virtual reservoir may lie outside (Graedel et al., 2005), inside (Müller et al., 2006), or both outside and inside the system boundary (Figure 4B), but typically lies outside the system boundary in MFA investigations with a reference timeframe of a single year, for which only stock accumulation and/or depletion is accounted.

### Metadata <heading level 3>

A ‘metadata layer’ is also implied in UMIS diagrams. This layer conceptually links data to additional information (i.e., ‘data about data’ or metadata e.g., reference space, reference timeframe, reference material, label, source, uncertainty, units, calculation details). Mass balance residuals exist in this metadata layer. Material stocks and flows data and their associated metadata are positioned at the same matrix coordinates (in terms of subsystem and process codes) in UMIS diagrams, meaning that these data are indexed within the UMIS structure by the same label. For example, metadata and (total, additions to, and removals from) stock associated with the transformative process in Figure 4A lie directly behind it, i.e., in the top left corner cell of each matrix, and are indexed in UMIS with the same process label. The inclusion of all metadata types in the metadata layer means that each data entry in UMIS can be explicitly associated with detailed supplementary information, including uncertainty, and tracked throughout material cycles.

## Subsystem Specification and Disaggregation <heading level 2>

The complete set of subsystems, aggregate subsystem modules, and the virtual reservoir represent the whole system (for all reference materials, reference spaces, and reference timeframes), containing the anthroposphere and the (natural) environment. Modularization of the whole system into subsystems adds key flexibility to UMIS because it enables linkages between material stocks and flows data at any level of disaggregation and provides a mechanism to eliminate double counting of data (revisited below). The subsystem concept is consistent with the way that data is structured in existing material cycle investigations, which often define aggregate production, fabrication, manufacturing, use, waste management, and environment processes (Talens Peiró et al., 2013). These aggregate process categories are thus natural choices for subsystems (and aggregate subsystem modules). Subsystems are also useful visualization tools, providing logical cutoffs to view parts of UMIS diagrams, and to confine updates to a single or partial set of subsystems rather than the whole system. These attributes are potentially important in complex computational analysis of highly disaggregated systems containing many processes, stocks, and flows.

However, UMIS does not preclude the specification of alternative subsystems (and aggregate subsystem modules) to the common aggregate processes or life cycle stages used in MFA investigations (Graedel et al., 2002). For example, an ‘engineering materials’ subsystem can be specified to describe the production of alloys and other engineering composites. In doing so, UMIS can recast the typical definition of the ‘production’ subsystem to precede an ‘engineering materials’ subsystem. Subsystem specification is thus completely left to user discretion.

**Consistent Subsystem Disaggregation <heading level 3>**

**Subsystem Specification, Stage One <heading level 4>**

The first stage of subsystem specification uses a three-step strategy in which the objectives are to:

1. Define a set of aggregate subsystem modules, each containing a single subsystem consisting of a transformative and storage process, with one outflow and an associated distributive and storage process. These aggregate subsystem modules individually represent stages in material cycles and together with the virtual reservoir comprise the reference material (*m*), reference timeframe (*t*), and reference space (*s*) component of the whole system. This step is shown in Figure 5A, where two aggregate subsystem modules are defined within the reference material $m_1$, reference space $s_1$, and reference timeframe $t_1$ system boundary (red dashed double dotted line). The aggregate subsystem modules are *ANT* (yellow shaded box) and *NAT* (blue shaded box), and their respective subsystems are *ANT.1* (*aggregate anthroposphere*) and *NAT.1* (*aggregate nature*). We note again that

subsystem specification (e.g., the specification of *ANT* and *NAT* here) is completely up to user discretion.

2. Select a single transformative and storage process, and one outflow and associated distributive and storage process. Define a subsystem by disaggregating these processes and flows to the next disaggregation level (one outflow and an associated distributive and storage process are again assigned to each disaggregated transformative and storage process). The newly defined subsystem is added to the UMIS diagram along the matrix diagonal within the same aggregate subsystem module, which is expanded as necessary. This step is shown in Figure 5B, where the *ANT.1;1* (*anthroposphere*) subsystem (green shaded box) is defined by disaggregating processes and flows in *ANT.1* (*aggregate anthroposphere*). *ANT.1;1* is specified in terms of *production and use* and *recycling and disposal* processes. These processes are added to the bottom right of *ANT.1* along the matrix diagonal within the aggregate subsystem module (*ANT*). Repeat this step until the relevant data for the aggregate subsystem module are fully defined.
3. Specify flows from each distributive process to every transformative process. This step is shown in Figure 5C.

Steps (1-3) guarantee that UMIS diagrams for any single reference space represent bipartite directed graphs. Processes and flows generated through steps (1-3) are given unique labels according to the aforementioned labeling rules. The first stage of subsystem specification defines the maximal set of processes and flows within a single reference material, reference space, and reference timeframe component of the whole system, for data disaggregated using a single consistent approach.

Figure 5. First stage of subsystem specification, which occurs in three steps. (A) Step 1, aggregate subsystem modules *ANT* and *NAT* are defined, which cumulatively represent the reference material $m_1$, reference space $s_1$, and reference timeframe $t_1$ component of the whole system. *ANT.1* and *NAT.2* subsystems are also defined. (B) Step 2, specification of the *ANT.1;1* subsystem to fully describe the available (consistently disaggregated) data for *ANT.1* and reference material $m_1$ in the reference space $s_1$ and reference timeframe $t_1$ component of the whole system. (C) Step 3, specification of all flows from distributive to transformative processes. (D) UMIS diagram produced with *production and use* (*ANT.1;1;1*) and *recycling and disposal* (*ANT.1;1;2*) subsystems, processes, and flows defined by disaggregating *ANT.1;1*. The virtual reservoir and metadata layer are omitted for clarity. Flows depicted by faded grey arrows in (C) and black arrows depicting subsystem disaggregation in (B) and (D) are omitted in UMIS diagrams, and are only shown here to guide readers. The black dashed lines represent subsystem boundaries, the red dashed double dotted lines represent system boundaries, and the solid black lines bordering UMIS diagrams represent whole system boundaries. A dynamic version of this figure is available as SI in Microsoft PowerPoint format.

### Divergent Subsystem Disaggregation <heading level 3>

We use tree-type data structure terminology in the following discussion. This terminology is particularly well suited to describing data in databases and elicitation diagrams, and thus also UMIS. A common and consistent approach to disaggregate material stocks and flows data is to define ‘child’ processes that describe more specific processes than their ‘parent’ processes. This approach is illustrated in Figure 5D, where the *aggregate anthroposphere* ‘root’ subsystem (*ANT.1*) is disaggregated into the *anthroposphere* ‘child’ subsystem (*ANT.1;1*). Here, *ANT.1* is also the parent of *ANT.1;1*. The *ANT.1;1* child subsystem is further disaggregated into *production and use* (*ANT.1;1;1*) and *recycling and disposal* (*ANT.1;1;2*) ‘grandchild’ subsystems. This disaggregation process can continue (e.g., from *production and use* (*ANT.1;1;1*) to *production* (*ANT.1;1;1;1*) and *use* (*ANT.1;1;1;2*)), until all the available data are described. However, different approaches can be used to disaggregate material stocks and flows data. For example, it is possible to disaggregate by material rather than by process specificity. In this case the *aggregate anthroposphere* root subsystem (*ANT.1*) could be disaggregated into an *anthroposphere* child subsystem (*ANT.1;1'*), and *metals* (*ANT.1;1';1*) and *non-metals* (*ANT.1;1';2*) grandchild subsystems.

#### Example: Four Car Types <heading level 4>

Here, divergent disaggregation approaches are illustrated using two types of data for *cars* in a *transport* system (Figure 6, 'nodes' are written in italics here). The *cars* data are disaggregated by size, either *big* or *small* (Figure 6A), or by color, either *red* or *blue* (Figure 6B). No other types of *cars* or *transport* exist in this example. These data are visualized as two 'material trees' within the same *transport* system. All four units of *transport* are *cars*. The four units of *cars* are constituted by either two *big* cars and two *small* cars, or one *red* car and three *blue* cars. However, no information is available on which *big* or *small* cars are *red* or *blue*, or vice versa; therefore, *cars* data can only be categorized by size (*big* or *small*, *cars*), or color (*red* or *blue*, *cars'*), and two material trees (with *cars* data disaggregated once in both) are needed to fully describe the *cars* data within the same *transport* system.

Two material trees for cars are specified as follows: the *transport* data in Figure 6A (four *cars*) is 'copied' as *transport* data into Figure 6B to specify the second material tree, i.e., the 'copied material tree'. Therefore, *transport* data in the *transport* 'fork node' and material tree (Figure 6A) is copied into the *transport* 'copied fork node' in the copied material tree (Figure 6B). A fork node is defined as a node at which copying occurs. The copied fork node is *transport* rather than *cars* or *cars'* because the data described by *transport* is the same (four *cars*) in either material tree, whereas *cars* and *cars'* describe different data (either *big* and *small* cars, or *red* and *blue* cars, respectively). Therefore, cars and cars' are colored differently in Figure 6. UMIS uses this method of copying nodes in material trees to universally structure material stocks and flows data at any level of disaggregation. Nodes in material trees are analogous to subsystems in subsystem sets in UMIS diagrams.

It is important to note here that if only one material tree is specified, then the *transport* system would not be able to simultaneously contain all four types of *cars* data. In this case, *big* and *small* cars would both need to be further disaggregated into *red* and *blue* cars to simultaneously describe *big*, *small*, *red*, and *blue* cars. However, the data needed to do this may not exist.

Figure 6. Divergent disaggregation of *cars* data into (A) *big* or *small* (*cars*), and (B) *red* or *blue* (*cars'*) types within the *transport* system. The *transport* data in (A), i.e., four *cars*, are 'copied' as *transport* data into (B) to describe both types of disaggregated *cars* data. Two cars are *big*, two cars are *small*, one car is *red*, and three cars are *blue*. Only data from a single material tree should be used by a modeler at any one time, either the (A) material tree or the (B) copied material tree, else the visualized system describes eight rather than four cars (i.e., to avoid double counting of data). Nodes in material trees are analogous to subsystems in subsystem sets in UMIS diagrams.

**Subsystem Specification, Stage Two <heading level 4>**

Divergent disaggregation approaches are reconciled in UMIS using the second stage of subsystem specification, which employs the following three-step strategy:

1. Define a 'fork subsystem' and then copy it by defining another subsystem with equivalent properties. The newly defined subsystem is termed a 'copied fork subsystem'. It is the 'root' subsystem in its 'copied subsystem set' (i.e., the first subsystem in the set). Fork and copied fork subsystems exist within the same aggregate subsystem module and are substitutable. This step is shown in Figure 7A, where the *aggregate anthroposphere* fork subsystem (*ANT.1*) is copied to yield the *aggregate anthroposphere* copied fork subsystem (*ANT.1*).

2. Disaggregate processes and flows in the copied fork subsystem following step 2 in the procedure for Subsystem Specification, Stage One except mark each newly defined subsystem code with an apostrophe. If that subsystem code already exists, mark each newly defined subsystem code with an additional apostrophe so that it has exactly one more than any existing subsystem code (e.g., if *ANT.1;1’* exists, the newly defined subsystem code is *ANT.1;1’’*). This step is shown in Figure 7B, where processes and flows in the copied fork subsystem *ANT.1* (*aggregate anthroposphere*, first data disaggregation level) are disaggregated and used to define an *ANT.1;1’* (*anthroposphere*) child subsystem (second data disaggregation level), and *ANT.1;1’;1* (*metals*) and *ANT.1;1’;2* (*non-metals*) grandchild subsystems (third data disaggregation level). These subsystems comprise the copied subsystem set and exist within the aggregate subsystem module *ANT*.
3. Add the newly defined copied subsystem set to the UMIS diagram along its matrix diagonal below the existing subsystem set and any existing copied subsystem sets, within its aggregate subsystem module. Specify flows from each distributive process to every transformative process. This step is shown in Figure 7C (note: processes and flows are omitted in Figure 7C to compact the plot).

Any subsystem can be specified as a fork subsystem and then be copied to define a copied fork subsystem using this procedure (e.g., the *ANT.1;1* subsystem in Figure 7D could be specified as a fork subsystem and then copied to define the copied fork subsystem *ANT.1;1*, which could then be disaggregated into *ANT.1;1;1’*, *ANT.1;1;2’* etc.). Fork subsystems are always specified such that the processes, stocks, and flows within their child subsystems are defined using more than one disaggregation approach.

Figure 7. Second stage of subsystem specification, which occurs in three steps. (A) Step 1, the fork subsystem *ANT.1* (*aggregate anthroposphere*) is copied to yield the copied fork subsystem *ANT.1* (*aggregate anthroposphere*). These subsystems are equivalent, substitutable, and occur within the same aggregate subsystem module (*ANT*). (B) Step 2, processes and flows in the copied fork subsystem *ANT.1* are disaggregated and *ANT.1;1'* (*anthroposphere*), *ANT.1;1';1* (*metals*), and *ANT.1;1';2* (*non-metals*) subsystems are defined to fully describe the available data for this copied subsystem set. (C) Step 3, the copied subsystem set (*ANT.1, ANT.1;1'*, *ANT.1;1';1*, and *ANT.1;1';2*) is added to the UMIS diagram and all flows from distributive to transformative processes are specified. This fully specifies the reference material $m_1$, reference space $s_1$, and reference timeframe $t_1$ component of the whole system. The virtual reservoir and metadata layer are omitted for clarity. Flows are omitted, and processes are omitted in *ANT.1* and *NAT.2* or otherwise replaced by grey shaded regions in (C) to simplify the diagram. Thick black arrows and lines depicting subsystem specification and disaggregation in (A-C), and shaded grey regions representing processes in (C), are omitted in UMIS diagrams, and are only shown here to guide readers. The black dashed lines represent subsystem boundaries, the red dashed double dotted lines represent system boundaries, and the solid black lines bordering UMIS diagrams represent whole system boundaries. A dynamic version of this figure is available as SI in Microsoft PowerPoint format.

## Avoiding Double Counting of Data <heading level 2>

Double counting of data occurs when differently disaggregated data are incorrectly summed, accounting for the same material mass twice. It is avoided in computational models utilizing UMIS by the modeler: (1) treating aggregate subsystem modules discretely; and then (2) specifying their constituent subsystems to fully represent data at (2a) a single disaggregation level only, and (2b) only using one fork or copied fork subsystem (including all their child, grandchild, etc. subsystems) at every instance where divergent disaggregation occurs (i.e., wherever subsystem forking occurs).

### Equivalent Representations of Different Data <heading level 3>

As shown in Figure 8, this treatment does not prohibit using data from different disaggregation levels. It also does not limit how UMIS structured data are archived in databases. Subsystems covering four levels of data disaggregation are shown in Figure 8:

1. *ANT.1* and *NAT.2*, which contain data on the first level;
2. *ANT.1;1* and *ANT.1;1'*, which contain data on the second level (produced by disaggregating data in *ANT.1*);
3. *ANT.1;1;1*, *ANT.1;1;2*, and *ANT.1;1;3* (produced by disaggregating *ANT.1;1*), and also *ANT.1;1';1*, *ANT.1;1';2*, and *ANT.1;1';3* (produced by disaggregating *ANT.1;1'*) contain data on the third level; and
4. *ANT.1;1;1;1* and *ANT.1;1;1;2* (produced by disaggregating *ANT.1;1;1*), and *ANT.1;1;3;1* and *ANT.1;1;3;2* (produced by disaggregating *ANT.1;1;3*), which contain data on the fourth level and are not disaggregated further.

In the example shown in Figure 8, the aggregate subsystem module *ANT* can only be fully represented by data on the first, second, or third disaggregation levels (condition 2a) because *ANT.1;1;2*, *ANT.1;1';1*, *ANT.1;1';2*, and *ANT.1;1';3* are not disaggregated further here. *ANT* is also specified by using only one fork subsystem (Figures 8B-8F) or copied fork subsystem (Figures 8G-8H) at the single instance where subsystem forking occurs, i.e., at *ANT.1* (condition 2b). Here, *ANT* and *NAT* are individual stages in a material cycle and together constitute the reference material $m_1$, reference space $s_1$, and reference timeframe $t_1$ component of the whole system (condition 1). The examples (Figures 8A-8I) show the flexibility of UMIS in defining a whole system in terms of aggregate subsystem modules, which can be comprised of differently disaggregated data depending on data availability or visualization priorities.

Figure 8. Equivalent representations (A-I) of the reference material $m_1$, reference space $s_1$, and reference timeframe $t_1$ component of the whole system, represented in terms of UMIS diagrams and excluding double counting of data. Processes are replaced by grey shaded regions or omitted in *ANT.1* and *NAT.2*, and flows are omitted. In (A and I), the aggregate subsystem modules *ANT* and *NAT*, and their relevant data are shown. In (B), *ANT* is represented using data on the first disaggregation level (*ANT.1*). *ANT* is represented using data on the second level of disaggregation only in (C) and (H), i.e., for the *ANT.1;1* and *ANT.1;1'* subsystems, respectively. In (D-G), *ANT* is represented by various combinations of data on the second, third, and fourth disaggregation levels. The virtual reservoir and metadata layer are omitted for clarity. The black dashed lines represent subsystem boundaries, the red dashed double dotted lines represent system boundaries, and the solid black lines bordering UMIS diagrams represent whole system boundaries.

### Selecting Data to Avoid Double Counting <heading level 3>

Allowing aggregate subsystem modules to be described by any relevant data that avoids double counting (regardless of the disaggregation level and type) facilitates the development of more reliable, flexible, and detailed whole system computational models and databases by giving the modeler extra choice. For a whole system with poor data availability at a more disaggregated level (e.g., for a child subsystem), but good data availability at a less disaggregated level (e.g., for its parent subsystem), this attribute of UMIS enables the modeler to choose to use the less disaggregated data for that particular subsystem without imposing any conditions outside of the (copied) subsystem set that contains these parent and child subsystems. Similarly, UMIS allows the modeler to choose between data represented by a subsystem set or differently disaggregated data represented by a copied subsystem set at each point of divergent disaggregation. It is noteworthy that differently disaggregated data are related through their common fork/copied fork subsystems; unknown data can be calculated by e.g., applying the mass conservation principle and Bayes' theorem of conditional probability (Lupton and Allwood).

Selecting data for a copied subsystem set (e.g., *ANT.1*, *ANT.1;1'*, *ANT.1;1';1*, and *ANT.1;1';2*) is done by ignoring all data associated with its complementary (copied) subsystem set(s) (e.g., *ANT.1*, *ANT.1;1*, *ANT.1;1;1*, and *ANT.1;1;2*), Figure 9A. The opposite scenario (i.e., selecting a subsystem set) is shown in Figure 9B. This flexible treatment of data in UMIS is key to its compatibility with MFA, I/O analysis, and LCA datasets, and also data for commodities containing various components, engineering materials, and substances that are reported by e.g., (inter)national statistical offices (United Nations Statistics Division, 2017; U.S. Geological Survey, 2011).

Figure 9. Selection of differently disaggregated data in UMIS to avoid double counting. Selection of data for the (A) copied subsystem set (*ANT.1*, *ANT.1;1'*, *ANT.1;1';1*, and *ANT.1;1';2*) and (B) subsystem set (*ANT.1*, *ANT.1;1*, *ANT.1;1;1* and *ANT.1;1;2*) are shown. Unselected subsystems (including their processes and flows) are covered by white blocks. Each UMIS diagram, (A) and (B), define the reference material $m_1$, reference space $s_1$, and reference timeframe $t_1$ component of the whole system, but do so using differently disaggregated data. Processes are omitted in *ANT.1* and *NAT.2* and replaced by grey shaded regions otherwise. Flows, the virtual reservoir, and the metadata layer are omitted for clarity. The black dashed lines represent subsystem boundaries, the red dashed double dotted lines represent system boundaries, and the solid black lines bordering UMIS diagrams represent whole system boundaries.

## Other Key Properties of UMIS <heading level 2>

### Cross Boundary Flows and Trade <heading level 3>

Cross boundary flows (*xs*) are defined in UMIS as flows between two reference spaces (*s*); a trade flow is a type of cross boundary flow that occurs between system boundaries that fully describe independent economic entities. They are implicitly represented in UMIS diagrams for a single reference space. This is because cross boundary flows always occur between two transformative or distributive processes with the same labels, which occur in subsystems with different reference spaces but otherwise the same attributes. Therefore, a single UMIS diagram defines the labels for every cross boundary flow associated with the subsystem(s) that it depicts. UMIS diagrams representing individual reference spaces can be combined to result in a multi-regional UMIS diagram that explicitly displays cross boundary flows (Figure 10). This treatment is analogous to the compilation of multi-regional I/O tables (Peters and Hertwich, 2006).

Figure 10. Conceptual visualization of cross boundary flows (*xs*) in a multi-regional UMIS diagram. The subsystem is fixed, the reference material and reference timeframe components of the whole system are fixed, and there are two reference spaces, $s_1$ and $s_2$. Cross boundary flows are shown as red diamonds in the blue shaded regions (faded grey arrows are shown here to guide readers only and are not normally displayed). The virtual reservoir and metadata layer are omitted for clarity.

**Intersecting Reference Materials <heading level 3>**

Simultaneous consideration of multiple material cycles adds substantial complexity to system-wide analyses of resources and materials, and is relatively infrequently reported (Nakajima et al., 2013). For example, copper-cobalt concentrate produced as a by-product from copper electrowinning (in the copper cycle) is typically recovered and then refined to cobalt metal (Donaldson and Beyersmann, 2000), although MFA diagrams for the cobalt cycle may only explicitly represent the latter recovery and refining steps (Harper et al., 2012). Therefore, information about the copper cycle, e.g., the concentration of cobalt in copper-cobalt concentrate and the amount of this material, can be used to determine material stocks and flows data in the cobalt cycle. In UMIS, materials that are not included in the defined reference material (which are thus outside the system boundary of interest) are termed ‘intersecting materials’. Information about intersecting materials that is used to determine material stocks and flows data in material cycles is represented in UMIS diagrams in the metadata layer.

**Temporal Metadata and Time Series Analysis <heading level 3>**

Similar to intersecting materials, material stocks and flows data at a particular reference timeframe can be determined using information from a different reference timeframe. For example, the global mass of stocked vehicles in the year 2000 can be used together with the additions and withdrawals of vehicles in the year 2001 to determine the vehicle stock in that year. This information is also present in the metadata layer in UMIS diagrams.

Material stocks and flows data along a time series is represented in UMIS by sequentially stacking 'snapshots' of UMIS diagrams at specific reference timeframes (Figure 11 shows four stacked snapshots of the whole system at reference timeframes of $t_1$ (least recent), $t_2$, $t_3$, and $t_4$ (most recent)), with older reference timeframes presented further in the background. These snapshots are implicitly linked by temporal metadata. The sequential structuring of time series data in terms of UMIS diagram snapshots (at reference timeframes, $t$), incorporating subsystem (and aggregate subsystem modules) and multi-regional (reference spaces, $s$, and cross boundary flows, $xs$) components, and (implicitly) virtual reservoirs and metadata layers at each reference timeframe, is the method by which the whole system is represented in UMIS across materials, space, and time. This time series representation facilitates the development of complex, computational, and dynamic models of material cycles.

Figure 11. UMIS diagram representation of the whole system, shown in terms of 'snapshots' at four reference timeframes ($t_1$ (least recent), $t_2$, $t_3$, and $t_4$ (most recent)), two reference spaces ($s_1$ and $s_2$), and a single reference material ($m_1$). Five aggregate subsystem modules (*PEM*, *F&M*, *USE*, *WMR*, *ENV*) are shown in yellow shaded boxes within each system boundary (represented by red alternating dashed double dotted lines). Cross boundary flows from reference spaces $s_1$ to $s_2$ ($xs_{1-2}$), and from reference spaces $s_2$ to $s_1$ ($xs_{2-1}$) are shown as blue shaded regions.

### **Querying UMIS Structured Data <heading level 3>**

#### By Name <heading level 4>

UMIS structures data so that the complete multiple reference material compositions of material stocks and flows can be queried across different material cycles. This is facilitated by assigning standardized names to each process (Figure 4), as discussed in the Names section. For example, the multi-reference material composition of stainless steel can be obtained by referencing all data related to processes named *stainless steel* across all (reference material specific) UMIS diagrams, i.e., for iron, chromium, nickel, etc. Flows adjacent to a distributive process and stock within a distributive process (in the virtual reservoir) are always of the same material type, which is specified by the distributive process name. Material stocked within a transformative process (in the virtual reservoir) is queried using its name or the material of its adjacent inflow (which in-turn is defined by the name of its adjacent distributive process).

#### By Label <heading level 4>

UMIS also enables hierarchical structuring of material stocks and flows data for commodities produced along material cycles (of any reference material composition), and within the whole system, to fully describe their component, engineering material, and substance constituents. This is achieved by: (1) specifying a general reference material, e.g., metallic elements, car-related materials, all materials, etc.; (2) using UMIS to structure and disaggregate material stocks and flows data such that all commodities related to the specified reference material are explicit (with the names of distributive processes defining these commodities); and then (3) disaggregating processes related to each commodity using the divergent disaggregation approach such that each of their components (sub-commodities), engineering materials, and elements are assigned distributive processes (the order in which commodities are disaggregated into their constituents is specified by the user). This is the method by which UMIS structures commodity-related data, e.g., monetary and mass trade statistics from the United Nations Comtrade Database (United Nations Statistics Division, 2017).

### Data in Non-Mass Units <heading level 3>

Data in mass and other units, e.g., monetary and energy, are similarly structured and visualized in UMIS, i.e., within the same integrated structure. All data types associated with a particular flow, stock, or process are represented by the same flow or process label, and distinguished by their units. It is this indexing feature (by process and flow label) and the flexible representation of differently disaggregated data (in aggregate subsystem modules) in UMIS that is exploited to simultaneously refer to MFA, I/O analysis, and LCA data in databases and computational models. Flows are similarly tracked in UMIS, I/O tables, and LCA process matrices, although UMIS additionally tracks stocks (in the virtual reservoir) and metadata (in the metadata layer). For example, data for “iron, gold, silver, and other metal ore mining” (2007 North American industry classification system (NAICS) code 2122A0) and “construction” (2007 NAICS code 23) may be structured in UMIS within aggregate subsystem modules such as *production of engineering materials* and *use*, respectively. An economic sector in an I/O table or in a make and use table may be constructed from (meta)data for a group of UMIS structured processes, stocks, and flows. Note that this may include processes representing e.g., a company in the services sector (employed people, computers, offices, etc., in an aggregate *use* subsystem module), to which quantitative monetary information are associated (in the metadata layer). An example application of UMIS to structure LCA data in mass and non-mass units is presented in the SI.

# Discussion <heading level 1>

In summary, UMIS can be readily and universally applied to transform diverse material stocks and flows data (e.g., mass, monetary, and energy) at any level of disaggregation into its standardized data structure without loss of information and avoiding double counting. Material cycles defined using UMIS will likely always contain data gaps. However, UMIS provides a methodology to unambiguously define and place material stocks and flows data into material cycles in their respective context(s). These missing data may therefore be estimated, e.g., using a Bayesian approach (Lupton and Allwood), and improved over time as additional data are generated and consolidated into the data structure.

UMIS comprehensively places material stocks and flows data into material systems contexts by uniquely labeling and visualizing subsystems, transformative, distributive, and storage processes, stocks, and also flows. This labeling system facilitates referencing of UMIS structured and visualized data, and their metadata e.g., uncertainty and system boundary properties, in complex computational code and databases. For example, UMIS can be used to holistically integrate material stocks and flows data describing vehicle value chains into a single systems context, such as: the (co-)production of vehicle-related elements in individual mine sites; element stocks in vehicles as functions of the country of sale, brand, and model; in-use phase greenhouse gas emissions; and international trading of down-cycled scrap metal. These data, and this single material system, could then be incorporated into a database and comprehensively visualized in a UMIS diagram. The UMIS diagram could then be used to develop a computational script to model this material system that has the flexibility to use these data at multiple levels of disaggregation at each (life) cycle stage whilst also avoid double counting. This script could be coded with the aim of producing material supply and demand scenarios for vehicles that are consistent with projected low-$CO_2$ emissions technology mixes (Fulton and Ward, 2011). Therefore, UMIS provides a flexible and comprehensive data structure that enables standardization, storage, and enhanced exchanging of material stocks and flows data. Such a data structure is a necessary step towards the complete and general standardization of material stocks and flows data. We believe that this development will eventually enable a step change improvement in the capabilities of material systems analysis, which will emerge as more (diverse) material stocks and flows data become available and get consolidated.

It is important to emphasize for clarity that UMIS is not a database, it is a data structure that can be used to place information about material systems into their respective context(s). These contextualized data can then be used to develop tools such as databases, elicitation diagrams, and computational models. A key motivation for developing UMIS comes from our work in integrating ~20 years of material cycle and criticality data generated within Yale's Center for Industrial Ecology into a single database. Here, UMIS is providing the data structure to comprehensively place these material stocks and flows data into their respective systems contexts. This database will be transferred to the United States Geological Survey upon completion, where it will be maintained in an openly accessible format, given wide access, and periodically updated and enhanced.

To illustrate the application and properties of UMIS, we have used UMIS to recast existing data published for the cobalt cycle, and material stocks and flows data represented by block flow type diagrams, system dynamics diagrams, Sankey diagrams, matrices, and also the EW-MFA classification system, to demonstrate how it can be applied to other existing (as well as yet to be published) data. These examples are presented in the SI.

We envisage that applications of UMIS to many diverse data sources will facilitate the development of whole system databases, similar to the database that we are currently developing at Yale's Center for Industrial Ecology. Our goal is for this database, and databases like it to which the community can add data, to become foundational tools to unify and accumulate material stocks and flows data. These data may then be extracted, analyzed, exchanged, and enhanced by diverse users, who can use UMIS-type elicitation diagrams to visualize these data and to perform complex computational data analyses. Key quantitative results from these analyses may then be flexibly visualized and shared in communication tools, such as Sankey diagrams (Lupton and Allwood, 2017). Therefore, UMIS can provide a key role in advancing the cumulative body of knowledge of material cycles in anthropogenic and natural systems.

# References <heading level 1>

Ayres, R. U. 1992. Toxic Heavy Metals: Materials Cycle Optimization. *Proceedings of the National Academy of Sciences of the United States of America* 89(3): 815-820.

Ayres, R. U. 1994. Industrial Metabolism: Theory and Policy. In *The Greening of Industrial Ecosystems*. Washington DC: National Academy Press.

Ayres, R. U., L. W. Ayres, J. A. Tarr, and R. C. Widgery. 1988. *An Historical Reconstruction of Major Pollutant Levels in the Hudson-Raritan Basin, 1880-1980, NOAA Technical Memorandum NOS OMA 43*. Rockville (MD), United States: National Oceanic and Atmospheric Administration (NOAA), United States Department of Commerce.

Brunner, P. H. and H. Rechberger. 2005. *Practical Handbook of Material Flow Analysis*. Boca Raton (FL), United States: Lewis Publishers, CRC Press.

Chen, W.-Q. and T. E. Graedel. 2012. Anthropogenic Cycles of the Elements: A Critical Review. *Environmental Science & Technology* 46(16): 8574-8586.

Chen, W.-Q., T. E. Graedel, P. Nuss, and H. Ohno. 2016. Building the Material Flow Networks of Aluminum in the 2007 U.S. Economy. *Environmental Science & Technology* 50(7): 3905-3912.

Cullen, J. M., J. M. Allwood, and M. D. Bambach. 2012. Mapping the Global Flow of Steel: From Steelmaking to End-Use Goods. *Environmental Science and Technology* 46(24): 13048-13055.

Davis, J., R. Geyer, J. Ley, J. He, R. Clift, A. Kwan, M. Sansom, and T. Jackson. 2007. Time-Dependent Material Flow Analysis of Iron and Steel in the UK: Part 2. Scrap Generation and Recycling. *Resources, Conservation and Recycling* 51(1): 118-140.

Donaldson, J. D. and D. Beyersmann. 2000. Cobalt and Cobalt Compounds. In *Ullmann's Encyclopedia of Industrial Chemistry*: John Wiley and Sons Inc.

Eckelman, M. J. and I. Daigo. 2008. Markov Chain Modeling of the Global Technological Lifetime of Copper. *Ecological Economics* 67(2): 265-273.

EUROSTAT. 2001. *Economy-Wide Material Flow Accounts and Derived Indicators - A Methodological Guide*. Luxembourg: European Communities.

EUROSTAT. 2008. *Eurostat Manual of Supply, Use and Input-Output Tables*. Luxembourg: European Commission.

Fischer-Kowalski, M., F. Krausmann, S. Giljum, S. Lutter, A. Mayer, S. Bringezu, Y. Moriguchi, H. Schütz, H. Schandl, and H. Weisz. 2011. Methodology and Indicators of Economy-Wide Material Flow Accounting. *Journal of Industrial Ecology* 15(6): 855-876.

Ford, A. 1999. *Modeling the Environment. An Introduction to System Dynamics Models of Environmental Systems*. Washington, D.C.: Island Press.

Frosch, R. A. and N. E. Gallopoulos. 1989. Strategies for Manufacturing. *Scientific American* 261: 144-152.

Fulton, L. and J. Ward, eds. 2011. *Technology Roadmap: Electric and Plug-In Hybrid Electric Vehicles*. Paris: International Energy Agency (IEA).

Graedel, T. E., M. Bertram, K. Fuse, R. B. Gordon, R. Lifset, H. Rechberger, and S. Spatari. 2002. The Contemporary European Copper Cycle: The Characterization of Technological Copper Cycles. *Ecological Economics* 42: 9-26.

Graedel, T. E., D. Van Beers, M. Bertram, K. Fuse, R. B. Gordon, A. Gritsinin, E. M. Harper, A. Kapur, R. J. Klee, R. Lifset, L. Memon, and S. Spatari. 2005. The Multilevel Cycle of Anthropogenic Zinc. *Journal of Industrial Ecology* 9(3): 67-90.

Graedel, T. E., D. van Beers, M. Bertram, K. Fuse, R. B. Gordon, A. Gritsinin, A. Kapur, R. J. Klee, R. J. Lifset, L. Memon, H. Rechberger, S. Spatari, and D. Vexler. 2004. Multilevel Cycle of Anthropogenic Copper. *Environmental Science & Technology* 38(4): 1242-1252.

Harper, E. M., G. Kavlak, and T. E. Graedel. 2012. Tracking the Metal of the Goblins: Cobalt's Cycle of Use. *Environmental Science & Technology* 46(2): 1079-1086.

Hawkins, T., C. Hendrickson, C. Higgins, H. S. Matthews, and S. Suh. 2007. A Mixed-Unit Input-Output Model for Environmental Life-Cycle Assessment and Material Flow Analysis. *Environmental Science & Technology* 41(3): 1024-1031.

Hellweg, S. and L. M. i Canals. 2014. Emerging Approaches, Challenges and Opportunities in Life Cycle Assessment. *Science* 344(6188): 1109.

Hendriks, C., R. Obernosterer, D. Muller, S. Kytzia, P. Baccini, and P. H. Brunner. 2000. Material Flow Analysis: A Tool to Support Environmental Policy Decision Making. Case-Studies on the City of Vienna and the Swiss Lowlands. *Local Environment* 5(3): 311.

Hoekman, P. and H. von Blottnitz. 2016. Cape Town's Metabolism: Insights from a Material Flow Analysis. *Journal of Industrial Ecology*: http://dx.doi.org/10.1111/jiec.12508.

Lenzen, M., A. Geschke, T. Wiedmann, J. Lane, N. Anderson, T. Baynes, J. Boland, P. Daniels, C. Dey, J. Fry, M. Hadjikakou, S. Kenway, A. Malik, D. Moran, J. Murray, S. Nettleton, L. Poruschi, C. Reynolds, H. Rowley, J. Ugon, D. Webb, and J. West. 2014. Compiling and Using Input–Output Frameworks Through Collaborative Virtual Laboratories. *Science of The Total Environment* 485–486: 241-251.

Lupton, R. C. and J. M. Allwood. Incremental Material Flow Analysis with Bayesian Inference. *Journal of Industrial Ecology*: accepted.

Lupton, R. C. and J. M. Allwood. 2017. Hybrid Sankey Diagrams: Visual Analysis of Multidimensional Data for Understanding Resource Use. *Resources, Conservation and Recycling* 124: 141-151.

Meylan, G. and B. K. Reck. 2017. The Anthropogenic Cycle of Zinc: Status Quo and Perspectives. *Resources, Conservation and Recycling* 123: 1-10.

Miller, R. E. and P. D. Blair. 2009. *Input-Output Analysis: Foundations and Extensions*. Cambridge, United Kingdom: Cambridge University Press.

Müller, D. B. 2006. Stock Dynamics for Forecasting Material Flows—Case Study for Housing in the Netherlands. *Ecological Economics* 59(1): 142-156.

Müller, D. B., T. Wang, B. Duval, and T. E. Graedel. 2006. Exploring the Engine of Anthropogenic Iron Cycles. *Proceedings of the National Academy of Sciences* 103(44): 16111-16116.

Nakajima, K., H. Ohno, Y. Kondo, K. Matsubae, O. Takeda, T. Miki, S. Nakamura, and T. Nagasaka. 2013. Simultaneous Material Flow Analysis of Nickel, Chromium, and Molybdenum Used in Alloy Steel by Means of Input–Output Analysis. *Environmental Science & Technology* 47(9): 4653-4660.

Nakamura, S. and K. Nakajima. 2005. Waste Input-Output Material Flow Analysis of Metals in the Japanese Economy. *Materials Transactions* 46(12): 2550-2553.

Nakamura, S., Y. Kondo, K. Matsubae, K. Nakajima, and T. Nagasaka. 2011. UPIOM: A New Tool of MFA and Its Application to the Flow of Iron and Steel Associated with Car Production. *Environmental Science & Technology* 45(3): 1114-1120.

Pauliuk, S., G. Majeau-Bettez, and D. B. Müller. 2015. A General System Structure and Accounting Framework for Socioeconomic Metabolism. *Journal of Industrial Ecology* 19(5): 728-741.

Pauliuk, S., R. L. Milford, D. B. Müller, and J. M. Allwood. 2013. The Steel Scrap Age. *Environmental Science & Technology* 47(7): 3448-3454.

Peters, G. P. and E. G. Hertwich. 2006. Structural Analysis of International Trade: Environmental Impacts of Norway. *Economic Systems Research* 18(2): 155-181.

Polis, G. A. and K. O. Winemiller. 1996. *Food Webs: Integration of Patterns & Dynamics*. New York: Chapman & Hall.

Rauch, J. N. and T. E. Graedel. 2007. Earth's Anthrobiogeochemical Copper Cycle. *Global Biogeochemical Cycles* 21(2): GB2010.

Rauch, J. N. and J. M. Pacyna. 2009. Earth's Global Ag, Al, Cr, Cu, Fe, Ni, Pb, and Zn Cycles. *Global Biogeochemical Cycles* 23(2): GB2001.

Schmidt, M. 2008. The Sankey Diagram in Energy and Material Flow Management. *Journal of Industrial Ecology* 12(2): 173-185.

Talens Peiró, L., G. Villalba Méndez, and R. U. Ayres. 2013. Lithium: Sources, Production, Uses, and Recovery Outlook. *JOM* 65(8): 986-996.

Tanimoto, A. H., X. Gabarrell Durany, G. Villalba, and A. C. Pires. 2010. Material Flow Accounting of the Copper Cycle in Brazil. *Resources, Conservation and Recycling* 55(1): 20-28.

U.S. Geological Survey. 2011. *Mineral Commodity Summaries 2011*: U.S. Geological Survey.

Uihlein, A., W.-R. Poganietz, and L. Schebek. 2006. Carbon Flows and Carbon Use in the German Anthroposphere: An Inventory. *Resources, Conservation and Recycling* 46(4): 410-429.

United Nations Statistics Division. 2017. UN Comtrade Database. www.comtrade.un.org/data/. Accessed 2 March 2017.

Wang, M., W. Chen, and X. Li. 2015. Substance Flow Analysis of Copper in Production Stage in the U.S. from 1974 to 2012. *Resources, Conservation and Recycling* 105A: 36-48.

Wang, T., D. B. Müller, and T. E. Graedel. 2007. Forging the Anthropogenic Iron Cycle. *Environmental Science and Technology* 41(14): 5120-5129.

Yamada, H., I. Daigo, Y. Matsuno, Y. Adachi, and Y. Kondo. 2006. Application of Markov Chain Model to Calculate the Average Number of Times of Use of a Material in Society. An Allocation Methodology for Open-Loop Recycling. Part 1: Methodology Development. *The International Journal of Life Cycle Assessment* 11(5): 354-360.

# Supporting Information <heading level 1>

Additional Supporting Information (SI) may be found in the online version of this article at the publisher's website: (1) Application of UMIS to recast cobalt cycle data reported by Harper et al. (2012) and three figures (Figures S1-S3) that illustrate this procedure (section S1.1); (2) a UMIS diagram for the cobalt cycle, a single reference space, a single reference timeframe, and all aggregate subsystem modules, presented as a comma separated value file (UMIS_diagram_cobalt.csv); the Python script used to generate this UMIS diagram, provided as (3) Python (UMIS_diagrams_1.0.py) and (4) IPython notebooks (UMIS_diagrams_1.0.ipynb), and also in (5) hypertext markup language (UMIS_diagrams_1.0.html); and (6) the input file for the Python script (transformative_processes_input_cobalt.csv). Example applications of UMIS to recast data published in a (7) block flow type diagram (section S1.2), a (8) system dynamics diagram (section S1.3), a (9) Sankey diagram (section S1.4), data structured using the (10) EW-MFA classification system (section S1.5), and data published for a (10) LCA system represented by a matrix and a block flow type diagram (section S1.6), their respective UMIS diagrams ((11) UMIS_diagram_bflow.csv, (12) UMIS_diagram_sdyn.csv, (13) UMIS_diagram_sankey.csv, (14) UMIS_diagram_ewmfa.csv, (15) UMIS_diagram_matrixlca.csv), and input files for the aforementioned Python script ((16) transformative_processes_input_bflow.csv, (17) transformative_processes_input_sdyn.csv, (18) transformative_processes_input_sankey.csv, (19) transformative_processes_input_ewmfa.csv, (20) transformative_processes_input_matrixlca.csv), are also provided as SI. We additionally provide dynamic versions of (21) Figure 5, (22) Figure 7, and (23) Figure S2 as SI in Microsoft PowerPoint format, a (24) pdf version of the UMIS diagram for the matrix-based LCA system (UMIS_diagram_matrixlca.pdf, note: flow labels are omitted in this diagram for simplicity), and also high resolution images of (25) Figure S2 and (26) Figure S3 as SI in pdf format. These examples demonstrate a variety of potential applications of UMIS and also exhibit some minor yet important features of UMIS not fully covered in the main text.

# Acknowledgements <heading level 1>

This work is supported by Grant #1636509 of the National Science Foundation. We are deeply thankful to Michael S. Baker, Nedal T. Nassar, Daniel B. Müller, Maren Lundhaug, Mark Uwe Simoni, Oliver Schwab, Benjamin Sprecher, David Font Vivanco, Ranran Wang, Edgar Hertwich, and Stefan Pauliuk for their insightful feedback. We also thank three anonymous reviewers whose comments helped to significantly improve the quality of this paper.

# Conflicts of Interests <heading level 1>

The authors declare no competing financial interests.

**List of Figure Captions <heading level 1>**

Figure 1. Relationships between material stocks and flows in anthropogenic and natural systems. Material stored in a particular reservoir undergoes processing, storage, distribution, and transformation, to again become stored in another (one or more) reservoir(s). Total mass is conserved but the location of the material changes. These relationships between reservoirs and processes provide a basis upon which a unified structure for material stocks and flows data can be built.

Figure 2. Exemplary block flow type diagram for the iron cycle, the year 2000, and the United States, adapted from (Müller et al., 2006). Mass quantities in Tg/year are displayed adjacent to each respective flow. Mass balance residuals are not shown (e.g., around the 'Blast Furnace' transformative process). Note that some distributive processes needed to avoid material flowing between two processes of the same type and thus to ensure consistency with the bipartite directed graph structure are omitted, e.g., between the 'Manuf.' and 'Scrap Process. & Waste Manag.' transformative. Production (dashed green box), engineering materials (dashed yellow box), fabrication & manufacturing (dashed purple box), use (dashed orange box), waste management (dashed red box), and environment (dashed blue box) subsystems are added to illustrate the subsystem concept (see Development of the Unified Materials Information System (UMIS)).

Figure 3. Relationships between (A) I/O analysis (make and use tables), (B) MFA (block flow type diagram), and (C) LCA (inventory) data. Transformative and distributive processes are shown as darker grey filled squares and lighter grey filled circles, respectively. Flows are displayed as arrows. Colored bold arrows (B-C) are flows that are entered into the make and use tables here (A). Subsystem, aggregate subsystem module, and system boundaries are shown as dashed, bold dashed, and alternating dashed double dotted lines, respectively. Process and flow labels are used to reference data between the respective methodologies; their formulation, and also labeling of subsystems, are described in the text. The environment subsystem is included in (C) to demonstrate the compilation of an inventory table, which is done by disaggregating the aggregate production of engineering materials subsystem module *(PEM.1*) (shaded green boxes in B and C) to account for all inflows to and outflows from the *aggregate environment subsystem module* (*ENV.5*) (black bold arrows).

Figure 4. (A) Key aspects of UMIS, illustrated using UMIS type diagrams for one of each transformative, distributive, and storage process, three flows, the virtual reservoir, and the metadata layer. (B) The virtual reservoir shown here can lie inside or outside the system boundary, but occurs inside of it here. The metadata layer contains additional information (e.g., uncertainty, system boundary properties) about processes, stocks, and/or flows positioned at the same matrix coordinates. Flows depicted by grey arrows in (A) and conceptual linkages depicted by black arrows in (B) are omitted in UMIS diagrams, and are only shown here to guide readers.

Figure 5. First stage of subsystem specification, which occurs in three steps. (A) Step 1, aggregate subsystem modules *ANT* and *NAT* are defined, which cumulatively represent the reference material $m_1$, reference space $s_1$, and reference timeframe $t_1$ component of the whole system. *ANT.1* and *NAT.2* subsystems are also defined. (B) Step 2, specification of the *ANT.1;1* subsystem to fully describe the available (consistently disaggregated) data for *ANT.1* and reference material $m_1$ in the reference space $s_1$ and reference timeframe $t_1$ component of the whole system. (C) Step 3, specification of all flows from distributive to transformative processes. (D) UMIS diagram produced with *production and use* (*ANT.1;1;1*) and *recycling and disposal* (*ANT.1;1;2*) subsystems, processes, and flows defined by disaggregating *ANT.1;1*. The virtual reservoir and metadata layer are omitted for clarity. Flows depicted by faded grey arrows in (C) and black arrows depicting subsystem disaggregation in (B) and (D) are omitted in UMIS diagrams, and are only shown here to guide readers. The black dashed lines represent subsystem boundaries, the red dashed double dotted lines represent system boundaries, and the solid black lines bordering UMIS diagrams represent whole system boundaries. A dynamic version of this figure is available as SI in Microsoft PowerPoint format.

Figure 6. Divergent disaggregation of *cars* data into (A) *big* or *small* (*cars*), and (B) *red* or *blue* (*cars'*) types within the *transport* system. The *transport* data in (A), i.e., four *cars*, are 'copied' as *transport* data into (B) to describe both types of disaggregated *cars* data. Two cars are *big*, two cars are *small*, one car is *red*, and three cars are *blue*. Only data from a single material tree should be used by a modeler at any one time, either the (A) material tree or the (B) copied material tree, else the visualized system describes eight rather than four cars (i.e., to avoid double counting of data). Nodes in material trees are analogous to subsystems in subsystem sets in UMIS diagrams.

Figure 7. Second stage of subsystem specification, which occurs in three steps. (A) Step 1, the fork subsystem *ANT.1* (*aggregate anthroposphere*) is copied to yield the copied fork subsystem *ANT.1* (*aggregate anthroposphere*). These subsystems are equivalent, substitutable, and occur within the same aggregate subsystem module (*ANT*). (B) Step 2, processes and flows in the copied fork subsystem *ANT.1* are disaggregated and *ANT.1;1'* (*anthroposphere*), *ANT.1;1';1* (*metals*), and *ANT.1;1';2* (*non-metals*) subsystems are defined to fully describe the available data for this copied subsystem set. (C) Step 3, the copied subsystem set (*ANT.1, ANT.1;1'*, *ANT.1;1';1*, and *ANT.1;1';2*) is added to the UMIS diagram and all flows from distributive to transformative processes are specified. This fully specifies the reference material $m_1$, reference space $s_1$, and reference timeframe $t_1$ component of the whole system. The virtual reservoir and metadata layer are omitted for clarity. Flows are omitted, and processes are omitted in *ANT.1* and *NAT.2* or otherwise replaced by grey shaded regions in (C) to simplify the diagram. Thick black arrows and lines depicting subsystem specification and disaggregation in (A-C), and shaded grey regions representing processes in (C), are omitted in UMIS diagrams, and are only shown here to

guide readers. The black dashed lines represent subsystem boundaries, the red dashed double dotted lines represent system boundaries, and the solid black lines bordering UMIS diagrams represent whole system boundaries. A dynamic version of this figure is available as SI in Microsoft PowerPoint format.

Figure 8. Equivalent representations (A-I) of the reference material $m_1$, reference space $s_1$, and reference timeframe $t_1$ component of the whole system, represented in terms of UMIS diagrams and excluding double counting of data. Processes are replaced by grey shaded regions or omitted in *ANT.1* and *NAT.2*, and flows are omitted. In (A and I), the aggregate subsystem modules *ANT* and *NAT*, and their relevant data are shown. In (B), *ANT* is represented using data on the first disaggregation level (*ANT.1*). *ANT* is represented using data on the second level of disaggregation only in (C) and (H), i.e., for the *ANT.1;1* and *ANT.1;1'* subsystems, respectively. In (D-G), *ANT* is represented by various combinations of data on the second, third, and fourth disaggregation levels. The virtual reservoir and metadata layer are omitted for clarity. The black dashed lines represent subsystem boundaries, the red dashed double dotted lines represent system boundaries, and the solid black lines bordering UMIS diagrams represent whole system boundaries.

Figure 9. Selection of differently disaggregated data in UMIS to avoid double counting. Selection of data for the (A) copied subsystem set (*ANT.1*, *ANT.1;1'*, *ANT.1;1';1*, and *ANT.1;1';2*) and (B) subsystem set (*ANT.1*, *ANT.1;1*, *ANT.1;1;1* and *ANT.1;1;2*) are shown. Unselected subsystems (including their processes and flows) are covered by white blocks. Each UMIS diagram, (A) and (B), define the reference material $m_1$, reference space $s_1$, and reference timeframe $t_1$ component of the whole system, but do so using differently disaggregated data. Processes are omitted in *ANT.1* and *NAT.2* and replaced by grey shaded regions otherwise. Flows, the virtual reservoir, and the metadata layer are omitted for clarity. The black dashed lines represent subsystem boundaries, the red dashed double dotted lines represent system boundaries, and the solid black lines bordering UMIS diagrams represent whole system boundaries.

Figure 10. Conceptual visualization of cross boundary flows (*xs*) in a multi-regional UMIS diagram. The subsystem is fixed, the reference material and reference timeframe components of the whole system are fixed, and there are two reference spaces, $s_1$ and $s_2$. Cross boundary flows are shown as red diamonds in the blue shaded regions (faded grey arrows are shown here to guide readers only and are not normally displayed). The virtual reservoir and metadata layer are omitted for clarity.

Figure 11. UMIS diagram representation of the whole system, shown in terms of 'snapshots' at four reference timeframes ($t_1$ (least recent), $t_2$, $t_3$, and $t_4$ (most recent)), two reference spaces ($s_1$

and $s_2$), and a single reference material ($m_1$). Five aggregate subsystem modules (*PEM*, *F&M*, *USE*, *WMR*, *ENV*) are shown in yellow shaded boxes within each system boundary (represented by red alternating dashed double dotted lines). Cross boundary flows from reference spaces $s_1$ to $s_2$ ($xs_{1\text{-}2}$), and from reference spaces $s_2$ to $s_1$ ($xs_{2\text{-}1}$) are shown as blue shaded regions.