



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

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 2019, 'Unified Materials Information System (UMIS): An Integrated Material Stocks and Flows Data Structure', *Journal of Industrial Ecology*, vol. 23, no. 1, pp. 222-240. <https://doi.org/10.1111/jiec.12730>

Digital Object Identifier (DOI):

[10.1111/jiec.12730](https://doi.org/10.1111/jiec.12730)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Peer reviewed version

Published In:

Journal of Industrial Ecology

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



1 **Unified Materials Information System (UMIS): An Integrated Material Stocks and Flows**

2 **Data Structure**

3
4 Rupert J. Myers ^{1,a,*}, Tomer Fishman ^{1,b}, Barbara K. Reck ^{1,c}, T. E. Graedel ^{1,d}

5
6 ¹ Yale School of Forestry & Environmental Studies, Yale University, 195 Prospect St, New
7 Haven 06511 Connecticut, United States

8
9 *Corresponding author. Email: ^a rupert.myers@gmail.com; ^b tomer.fishman@yale.edu; ^c
10 barbara.reck@yale.edu; ^d thomas.graedel@yale.edu.

11
12 **Abstract <heading level 1>**

13 Modern society depends on the use of many diverse materials. Effectively managing these
14 materials is becoming increasingly important and complex, from the analysis of supply chains, to
15 quantifying their environmental impacts, to understanding future resource availability. Material
16 stocks and flows data enable such analyses but currently exist mainly as discrete packages, with
17 highly varied type, scope, and structure. These factors constitute a powerful barrier to holistic
18 integration and thus universal analysis of existing and yet to be published material stocks and flows
19 data. We present the Unified Materials Information System (UMIS) to overcome this barrier by
20 enabling material stocks and flows data to be comprehensively integrated across space, time,
21 materials, and data type independent of their disaggregation, without loss of information, and
22 avoiding double counting. UMIS can therefore be applied to structure diverse material stocks and
23 flows data and their metadata across material systems analysis methods such as material flow

24 analysis (MFA), input-output (I/O) analysis, and life cycle assessment (LCA). UMIS uniquely
25 labels and visualizes processes and flows in UMIS diagrams; therefore, material stocks and flows
26 data visualized in UMIS diagrams can be individually referenced in databases and computational
27 models. Applications of UMIS to restructure existing material stocks and flows data represented
28 by block flow diagrams, system dynamics diagrams, Sankey diagrams, matrices, and derived using
29 the ‘economy-wide’ MFA classification system are presented to exemplify use. UMIS advances
30 the capabilities with which complex quantitative material systems analysis, archiving, and
31 computation of material stocks and flows data can be performed.

32

33 **Introduction <heading level 1>**

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

43

44 Material systems analysis fundamentally involves the analysis of the type and quantity of existing
45 materials, how and to what extent they get transformed in and distributed among (enter and leave)
46 processes such as production, use, and recycling in anthropogenic systems, and their associated

47 impacts on economic and natural systems (i.e., environmental impacts). The natural system is
48 constituted by natural processes such as nutrient cycling among organisms in marine ecosystems
49 excluding humans, which is depicted in food webs (Polis and Winemiller, 1996), whereas the
50 anthropogenic system is constituted by anthropogenic processes such as manufacturing,
51 construction, transportation etc. (Ayres, 1994) typically along industrial supply and value chains.
52 Therefore, a process such as fishing represents a linkage, possibly the transformation (e.g., from
53 alive to dead fish), distribution (e.g., from the ocean to boat), and/or storage (e.g., withdrawal from
54 the ocean and deposition into a bucket), of material between anthropogenic and natural systems
55 (Figure 1). It is notable that material stocks and flows data are treated similarly in the analysis of
56 natural (e.g., food webs) and anthropogenic (e.g., supply and value chains) systems, and that
57 material processing changes the location but not the cumulative mass of material in the combined
58 anthropogenic and natural system (excluding nuclear reactions). These data can thus be reconciled
59 into a single unified structure. Consideration of both natural and anthropogenic processes is
60 essential to the holistic analysis of material systems.

61

62

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

69

70 Material stocks and flows data have been individually compiled and published for decades in
71 diverse and seemingly inconsistent formats that typically serve small sections of the material
72 systems analysis research community. Although these data have proliferated in recent years, it is
73 challenging to synthesize, build on, and enhance them due to their diverse and inconsistent

74 formatting. For example, the combined use of material stocks and flows data in monetary and mass
75 units can provide a greatly enhanced description of anthropogenic systems relative to what can be
76 accomplished using only one of these data types, and there is an abundance of both types of data
77 (Chen and Graedel, 2012; Lenzen et al., 2014), however these data are relatively infrequently used
78 together in holistic material cycle investigations (Nakajima et al., 2013; Chen et al., 2016). This
79 effort of combining multiple data types is hampered by the absence of a single flexible, universally
80 applicable, standardized, and generic machine readable data structure that can be applied without
81 loss of information. Reconciliation of material stocks and flows data into such a structure has not
82 yet been achieved but would provide a foundation to develop substantially more functional,
83 holistic, and higher complexity databases and quantitative computational models of anthropogenic
84 and natural systems. It would therefore improve data availability, increase the reproducibility of
85 research results, eliminate repetition of work, integrate research efforts to advance our
86 understanding of material systems issues such as the sustainability and resilience of industrial
87 supply chains, and increase the effectiveness of the material systems analysis research community.

88

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

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

97 materials, e.g., production and biomass, respectively) or rather high (Meylan and Reck,
98 2017) (e.g., ‘copper; strip, of a thickness exceeding 0.15 mm, of copper-zinc base alloys
99 (brass), in coils’). MFA data often describe partial or complete material cycles (Graedel et
100 al., 2004), but also frequently describe more aggregate data and indicators such as domestic
101 extraction in ‘economy wide’ MFA (EW-MFA); such data can exist on the firm level and
102 sub-national (e.g., river basins and cities), country, international, and global scales
103 (EUROSTAT, 2001; Fischer-Kowalski et al., 2011).

104 2. Life cycle assessment (LCA), which has as its objective to “[compile] and [evaluate] the
105 inputs, outputs, and potential environmental impacts of a product system throughout its life
106 cycle” (Hellweg and i Canals, 2014). LCA data are normally relatively highly
107 disaggregated and refer to multiple materials, owing to the need to describe the full
108 ensemble of environmental inputs and outputs relevant to a product system, yet often use
109 generic or non-process specific data.

110 3. Input-output (I/O) analysis, which differs from LCA and MFA in that it tracks monetary
111 flows through the economy in matrices that are “generally constructed from observed
112 economic data for a specific geographic region” (Miller and Blair, 2009), to e.g., allocate
113 environmental impacts to products and services. More aggregated descriptions of the
114 economy are typically investigated using I/O analysis rather than LCA, consistent with
115 economic data published by e.g., national statistical offices. I/O analysis and LCA data
116 have been harmonized in multi-regional I/O tables (Lenzen et al., 2014) and I/O-LCA
117 models (Hawkins et al., 2007) by reconciling differences in data (dis)aggregation.
118 However, I/O analysis and MFA data, despite sharing some key concepts (e.g., accounting
119 of material flows), are often disaggregated differently. The former normally describe

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

123
124 Pauliuk et al. (2015) recently showed that material stocks and flows data can be unified across
125 MFA, I/O analysis, and LCA by employing the make and use table approach used to compile I/O
126 tables (EUROSTAT, 2008). Consistency with this approach can be achieved by transforming
127 material stocks and flows data into the bipartite directed graph structure (i.e., a graph representing
128 a system containing two types of processes and only flows between processes of different type).
129 In practice, the bipartite directed graph structure can be attained by ensuring that transformative
130 processes are always followed by one or more flows that each terminate at distributive processes,
131 and vice versa. This representation is realistic because transformed materials are typically
132 distributed to locations different from where they were produced. We build on these insights and
133 address the challenge of unifying material stocks and flows data across MFA, I/O analysis, and
134 LCA methods by:

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

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

144

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

147

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

149 The basic attributes of MFA are that the mass conservation principle is respected and that the
150 investigated system is represented by processes, stocks, and flows. The investigated system is
151 specified using a ‘system boundary’ defined in terms of space (reference space), time (reference
152 timeframe), and one or more materials (reference material) (Brunner and Rechberger, 2005). The
153 reference timeframe can be a time period, e.g., a year, or a specific point in time, e.g., the end of a
154 year. Exemplary block flow type diagrams (Figure 2) depict this information by differentiating
155 among transformative, distributive, and storage processes. They also differentiate among flows
156 that are internal to (hereafter termed ‘flows’) and cross the system boundaries (hereafter termed
157 ‘cross boundary flows’, or ‘trade flows’ if the reference spaces represent independent economic
158 entities e.g., countries in Figure 2) (Pauliuk et al., 2015; Müller et al., 2006). Transformative,
159 distributive, and storage processes transform process inputs to outputs, distribute process outputs
160 to inputs, and produce or release stocks, respectively. It is typical to assign processes to each major
161 stage in anthropogenic material cycles (these are often production, fabrication & manufacturing,
162 use, and waste management) (Graedel et al., 2002). MFA diagrams sometimes display uncertainty
163 (Rauch and Pacyna, 2009) and also differences that result from applications of the mass
164 conservation principle (i.e., a ‘mass balance’) when compared to the observed data (leading to

165 ‘mass balance residuals’) (Graedel et al., 2004). However, MFA diagrams that incompletely
166 distinguish among the aforementioned types of processes and flows dominate (the distinction is
167 often either implied or unnecessary if the system boundary coincides with a single transformative
168 process) (Hendriks et al., 2000; Tanimoto et al., 2010; Uihlein et al., 2006; Davis et al., 2007;
169 Müller, 2006). Material stocks and flows data are also visualized using other types of diagrams,
170 e.g., Sankey (Schmidt, 2008) and system dynamics (Ford, 1999) diagrams, which share some of
171 these attributes.

172

173

174 Figure 2. Exemplary block flow type diagram for the iron cycle, the year 2000, and the United
175 States, adapted from (Müller et al., 2006). Mass quantities in Tg/year are displayed adjacent to
176 each respective flow. Mass balance residuals are not shown (e.g., around the ‘Blast Furnace’
177 transformative process). Note that some distributive processes needed to avoid material flowing
178 between two processes of the same type and thus to ensure consistency with the bipartite directed
179 graph structure are omitted, e.g., between the ‘Manuf.’ and ‘Scrap Process. & Waste Manag.’
180 transformative. Production (dashed green box), engineering materials (dashed yellow box),
181 fabrication & manufacturing (dashed purple box), use (dashed orange box), waste management
182 (dashed red box), and environment (dashed blue box) subsystems are added to illustrate the
183 subsystem concept (see Development of the Unified Materials Information System (UMIS)).
184

185 However, most MFA diagrams are used to communicate key messages and quantitative results
186 rather than to place and show data in complete detail and in their exact context within material
187 systems. Therefore, the formatting of these MFA ‘communication diagrams’ changes greatly
188 depending on the number of processes displayed, data availability, and investigation scope (Lupton
189 and Allwood, 2017). Consequently, most are significantly mismatched with one another in style
190 and detail even when describing similar systems (Wang et al., 2007; Pauliuk et al., 2013; Cullen
191 et al., 2012; Müller et al., 2006). MFA communication diagrams also typically do not normally
192 use explicit, standardized labeling systems to annotate processes and flows. These attributes hinder
193 their utility to illustrate the kind of highly structured and detailed (meta)data that are used in

194 databases and computational models of complex material systems (e.g., the exact positions of
195 material stocks and flows data in highly and differently disaggregated material cycles). Explicitly
196 and comprehensively indexing material stocks and flows data visualizations is beneficial in
197 computational modeling of complex material systems because it allows visualized information to
198 be precisely referenced. Therefore, the increasing complexity of data analysis and availability of
199 data in industrial ecology is creating a growing need to develop ‘elicitation diagrams’ that can
200 visualize fully detailed material stocks and flows data in their exact systems context within a
201 standardized and labeled structure.

202

203 The goal of this paper is thus to develop a Unified Materials Information System (UMIS) to
204 structure, label, and visualize diverse material stocks and flows data and their metadata (e.g.,
205 uncertainty, system boundary properties) into a single standardized format. UMIS could then
206 consolidate datasets across the major material systems analysis methods, e.g., MFA, I/O analysis,
207 and LCA. Here, the ‘whole system’ describes the entire system in its most general sense, including
208 the anthroposphere and nature, for all reference spaces, reference timeframes, and reference
209 materials. UMIS is visualized in terms of matrix type ‘UMIS diagrams’ showing material inputs,
210 outputs, and processing. The UMIS diagram for each reference material is unique because the
211 processes, stocks, and flows that comprise each material cycle are unique. For example, UMIS
212 diagrams for iron in the United States in the year 2000 and for iron in Australia in the year 2017
213 are equivalent, but both are different from the UMIS diagram for copper in the United States in
214 the year 2017. This representation means that any irrelevant (e.g., obsolete) processes and flows
215 for a reference material in a particular reference timeframe or reference space remain in UMIS
216 diagrams and are associated with zero material mass. The effort focuses on materials and mass,

217 two fundamental foci of material systems analysis research. Such an approach is naturally aligned
218 with MFA methodology although we show that it can be readily applied to other data types (e.g.,
219 monetary and energy) and methods (e.g., I/O analysis and LCA). This paper also aims to develop
220 UMIS so that data visualized in UMIS diagrams can be readily referenced in databases and
221 computational models. Another aim of this paper is to demonstrate how UMIS is used to transform
222 and visualize material stocks and flows data into its standardized structure (these demonstrations
223 are presented as Supporting Information, SI).

224

225 **Development of UMIS <heading level 1>**

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

232

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

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

240

241

242 Figure 3. Relationships between (A) I/O analysis (make and use tables), (B) MFA (block flow
243 type diagram), and (C) LCA (inventory) data. Transformative and distributive processes are
244 shown as darker grey filled squares and lighter grey filled circles, respectively. Flows are
245 displayed as arrows. Colored bold arrows (B-C) are flows that are entered into the make and use
246 tables here (A). Subsystem, aggregate subsystem module, and system boundaries are shown as
247 dashed, bold dashed, and alternating dashed double dotted lines, respectively. Process and flow
248 labels are used to reference data between the respective methodologies; their formulation, and
249 also labeling of subsystems, are described in the text. The environment subsystem is included in
250 (C) to demonstrate the compilation of an inventory table, which is done by disaggregating the
251 aggregate production of engineering materials subsystem module (*PEM.1*) (shaded green boxes
252 in B and C) to account for all inflows to and outflows from the *aggregate environment subsystem*
253 *module (ENV.5)* (black bold arrows).
254

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

258

259 **A Prescriptive Condition <heading level 3>**

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

268

269 **Subsystems <heading level 3>**

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

281
282 Subsystems can be infinitely disaggregated to describe more specific material stocks and flows
283 data. Subsystem disaggregation is shown in Figures 3B and 3C, where the *production subsystem*
284 (*PEM.1;1;1*) (Figure 3B) is disaggregated into a *mining subsystem* (*PEM.1;1;1;1*) (Figure 3C).
285 The most aggregated subsystem represents the whole system. If a subsystem is defined to represent
286 a stage in a material cycle (e.g., the ‘fabrication & manufacturing’ stage in Figure 1) and where
287 cumulatively these stages represent that material cycle, a subsystem is termed an ‘aggregate
288 subsystem module’ (see Subsystem Specification and Disaggregation). Therefore, a subsystem
289 boundary can be a subset of, the same as, or a superset of one or more system boundaries, or exist
290 outside the system boundary (e.g., the *aggregate environment subsystem module* (*ENV.5*), Figure
291 3C), depending on how these boundaries are defined.

292

293 **Labels <heading level 3>**

294 In Figure 3, flows are represented by arrows, whereas transformative and distributive processes
295 are represented by dark grey squares and light grey circles, respectively. Process labels (located
296 directly above processes in Figures 3B and 3C) are specified as *a.b.c.d.e*, where *a* represents the
297 reference material defined by the system boundary (*a* = 1 for reference material *m₁*), *b* defines the
298 aggregate subsystem module abbreviation, *c* is the subsystem code, *d* indicates the type of process,
299 transformative (*T*) or distributive (*D*), and *e* is a process code that is unique to each process in each
300 subsystem for reference material *a*. Flow labels (located adjacent to flow arrows in Figures 3B and
301 3C) are specified in the form *origin_destination*, where *origin* and *destination* specify the labels
302 of the processes that a flow originates and terminates at, respectively (e.g., the flow from
303 *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*,
304 where *PEM* refers to an ‘aggregate production of engineering materials module’). A subsystem
305 label is specified by the aggregate subsystem module abbreviation followed by a period and then
306 the subsystem code, i.e., *b.c*. For example, the subsystem label for the *production subsystem* in
307 Figure 3B is *PEM.1;1;1*.

308

309 **Codes <heading level 3>**

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

316 *F&M* and subsystem code 2;1;2 in Figure 3B (i.e., the *F&M.2;1;2* subsystem) represents data for
317 the second transformative process (*Manufacturing*) in the *F&M.2;1* subsystem (not shown in
318 Figure 3). Therefore, it also exists within the aggregate *fabrication & manufacturing* subsystem
319 module *F&M* on the third disaggregation level (character length excluding semi-colons(2;1;2) =
320 3).

321

322 **Names <heading level 3>**

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

331

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

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

338 manufacturing industries, respectively), green (manufacturing industry outputs), and pink
339 (construction industry outputs) text.

340

341 LCA inventory tables can be compiled using data structured by UMIS (Figure 3C). Here, processes
342 (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*
343 *C*, *1.PEM.1;1;1;1.T.5;5*) in the *mining* subsystem (*PEM.1;1;1;1*) are specified by disaggregating
344 processes in the *production* subsystem *PEM.1;1;1* (in this case the *mining* and *production*
345 subsystems are substitutable). Complete representation of the inventory data is achieved by
346 specifying an aggregate *environment* subsystem module (*ENV*) and disaggregating all aggregate
347 subsystem modules to the appropriate level such that all relevant flows to and from this aggregate
348 *environment* subsystem module are explicit (it is necessary to disaggregate *PEM* in Figure 3B to
349 explicitly show these flows in Figure 3C, shaded green boxes). The aggregate *environment*
350 subsystem module is external to the system boundary in this example. The complete set of
351 aggregate subsystem modules here, i.e., aggregate *production of engineering materials (PEM)*,
352 *fabrication & manufacturing (F&M)*, and *environment (ENV)* subsystem modules, represents the
353 combined anthropogenic and natural system boundary for a single reference material and reference
354 timeframe.

355

356 **Transforming Data into a Matrix Format <heading level 2>**

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

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

367

368 **Processes and Flows <heading level 3>**

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

375

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

384 diagrams retain the same system boundary definitions as defined in block flow type diagrams
385 (Figures 3B and 3C), i.e., defined in terms of a reference material, a reference timeframe, and a
386 reference space.

387

388

389 Figure 4. (A) Key aspects of UMIS, illustrated using UMIS type diagrams for one of each
390 transformative, distributive, and storage process, three flows, the virtual reservoir, and the
391 metadata layer. (B) The virtual reservoir shown here can lie inside or outside the system
392 boundary, but occurs inside of it here. The metadata layer contains additional information (e.g.,
393 uncertainty, system boundary properties) about processes, stocks, and/or flows positioned at the
394 same matrix coordinates. Flows depicted by grey arrows in (A) and conceptual linkages depicted
395 by black arrows in (B) are omitted in UMIS diagrams, and are only shown here to guide readers.
396

397 **Stock <heading level 3>**

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

406

407 **Metadata <heading level 3>**

408 A ‘metadata layer’ is also implied in UMIS diagrams. This layer conceptually links data to
409 additional information (i.e., ‘data about data’ or metadata e.g., reference space, reference
410 timeframe, reference material, label, source, uncertainty, units, calculation details). Mass balance

411 residuals exist in this metadata layer. Material stocks and flows data and their associated metadata
412 are positioned at the same matrix coordinates (in terms of subsystem and process codes) in UMIS
413 diagrams, meaning that these data are indexed within the UMIS structure by the same label. For
414 example, metadata and (total, additions to, and removals from) stock associated with the
415 transformative process in Figure 4A lie directly behind it, i.e., in the top left corner cell of each
416 matrix, and are indexed in UMIS with the same process label. The inclusion of all metadata types
417 in the metadata layer means that each data entry in UMIS can be explicitly associated with detailed
418 supplementary information, including uncertainty, and tracked throughout material cycles.

419

420 **Subsystem Specification and Disaggregation <heading level 2>**

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

433 computational analysis of highly disaggregated systems containing many processes, stocks, and
434 flows.

435
436 However, UMIS does not preclude the specification of alternative subsystems (and aggregate
437 subsystem modules) to the common aggregate processes or life cycle stages used in MFA
438 investigations (Graedel et al., 2002). For example, an ‘engineering materials’ subsystem can be
439 specified to describe the production of alloys and other engineering composites. In doing so, UMIS
440 can recast the typical definition of the ‘production’ subsystem to precede an ‘engineering
441 materials’ subsystem. Subsystem specification is thus completely left to user discretion.

442

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

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

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

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

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

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

470 3. Specify flows from each distributive process to every transformative process. This step is
471 shown in Figure 5C.

472
473 Steps (1-3) guarantee that UMIS diagrams for any single reference space represent bipartite
474 directed graphs. Processes and flows generated through steps (1-3) are given unique labels
475 according to the aforementioned labeling rules. The first stage of subsystem specification defines
476 the maximal set of processes and flows within a single reference material, reference space, and
477 reference timeframe component of the whole system, for data disaggregated using a single
478 consistent approach.

479

480

481 Figure 5. First stage of subsystem specification, which occurs in three steps. (A) Step 1,
482 aggregate subsystem modules *ANT* and *NAT* are defined, which cumulatively represent the
483 reference material m_I , reference space s_I , and reference timeframe t_I component of the whole
484 system. *ANT.1* and *NAT.2* subsystems are also defined. (B) Step 2, specification of the *ANT.1;1*
485 subsystem to fully describe the available (consistently disaggregated) data for *ANT.1* and
486 reference material m_I in the reference space s_I and reference timeframe t_I component of the
487 whole system. (C) Step 3, specification of all flows from distributive to transformative processes.
488 (D) UMIS diagram produced with *production and use (ANT.1;1;1)* and *recycling and disposal*
489 (*ANT.1;1;2*) subsystems, processes, and flows defined by disaggregating *ANT.1;1*. The virtual
490 reservoir and metadata layer are omitted for clarity. Flows depicted by faded grey arrows in (C)
491 and black arrows depicting subsystem disaggregation in (B) and (D) are omitted in UMIS
492 diagrams, and are only shown here to guide readers. The black dashed lines represent subsystem
493 boundaries, the red dashed double dotted lines represent system boundaries, and the solid black
494 lines bordering UMIS diagrams represent whole system boundaries. A dynamic version of this
495 figure is available as SI in Microsoft PowerPoint format.
496

497 **Divergent Subsystem Disaggregation <heading level 3>**

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

511 subsystem (*ANT.1;1'*), and *metals* (*ANT.1;1';1*) and *non-metals* (*ANT.1;1';2*) grandchild
512 subsystems.

513

514 **Example: Four Car Types <heading level 4>**

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

525

526 Two material trees for cars are specified as follows: the *transport* data in Figure 6A (four *cars*) is
527 'copied' as *transport* data into Figure 6B to specify the second material tree, i.e., the 'copied
528 material tree'. Therefore, *transport* data in the *transport* 'fork node' and material tree (Figure 6A)
529 is copied into the *transport* 'copied fork node' in the copied material tree (Figure 6B). A fork node
530 is defined as a node at which copying occurs. The copied fork node is *transport* rather than *cars*
531 or *cars'* because the data described by *transport* is the same (four *cars*) in either material tree,
532 whereas *cars* and *cars'* describe different data (either *big* and *small* cars, or *red* and *blue* cars,
533 respectively). Therefore, *cars* and *cars'* are colored differently in Figure 6. UMIS uses this method

534 of copying nodes in material trees to universally structure material stocks and flows data at any
535 level of disaggregation. Nodes in material trees are analogous to subsystems in subsystem sets in
536 UMIS diagrams.

537

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

542

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

551

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

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

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

562 2. Disaggregate processes and flows in the copied fork subsystem following step 2 in the
563 procedure for Subsystem Specification, Stage One except mark each newly defined
564 subsystem code with an apostrophe. If that subsystem code already exists, mark each newly
565 defined subsystem code with an additional apostrophe so that it has exactly one more than
566 any existing subsystem code (e.g., if *ANT.1;1'* exists, the newly defined subsystem code is
567 *ANT.1;1''*). This step is shown in Figure 7B, where processes and flows in the copied fork
568 subsystem *ANT.1* (*aggregate anthroposphere*, first data disaggregation level) are
569 disaggregated and used to define an *ANT.1;1'* (*anthroposphere*) child subsystem (second
570 data disaggregation level), and *ANT.1;1';1* (*metals*) and *ANT.1;1';2* (*non-metals*)
571 grandchild subsystems (third data disaggregation level). These subsystems comprise the
572 copied subsystem set and exist within the aggregate subsystem module *ANT*.

573 3. Add the newly defined copied subsystem set to the UMIS diagram along its matrix diagonal
574 below the existing subsystem set and any existing copied subsystem sets, within its
575 aggregate subsystem module. Specify flows from each distributive process to every
576 transformative process. This step is shown in Figure 7C (note: processes and flows are
577 omitted in Figure 7C to compact the plot).

578
579 Any subsystem can be specified as a fork subsystem and then be copied to define a copied fork
580 subsystem using this procedure (e.g., the *ANT.1;1* subsystem in Figure 7D could be specified as a
581 fork subsystem and then copied to define the copied fork subsystem *ANT.1;1*, which could then be
582 disaggregated into *ANT.1;1;1'*, *ANT.1;1;2'* etc.). Fork subsystems are always specified such that
583 the processes, stocks, and flows within their child subsystems are defined using more than one
584 disaggregation approach.

585

586

587 Figure 7. Second stage of subsystem specification, which occurs in three steps. (A) Step 1, the
588 fork subsystem *ANT.1 (aggregate anthroposphere)* is copied to yield the copied fork subsystem
589 *ANT.1 (aggregate anthroposphere)*. These subsystems are equivalent, substitutable, and occur
590 within the same aggregate subsystem module (*ANT*). (B) Step 2, processes and flows in the
591 copied fork subsystem *ANT.1* are disaggregated and *ANT.1;1'* (*anthroposphere*), *ANT.1;1';1*
592 (*metals*), and *ANT.1;1';2 (non-metals)* subsystems are defined to fully describe the available data
593 for this copied subsystem set. (C) Step 3, the copied subsystem set (*ANT.1*, *ANT.1;1'*,
594 *ANT.1;1';1*, and *ANT.1;1';2*) is added to the UMIS diagram and all flows from distributive to
595 transformative processes are specified. This fully specifies the reference material m_l , reference
596 space s_l , and reference timeframe t_l component of the whole system. The virtual reservoir and
597 metadata layer are omitted for clarity. Flows are omitted, and processes are omitted in *ANT.1* and
598 *NAT.2* or otherwise replaced by grey shaded regions in (C) to simplify the diagram. Thick black
599 arrows and lines depicting subsystem specification and disaggregation in (A-C), and shaded grey
600 regions representing processes in (C), are omitted in UMIS diagrams, and are only shown here to
601 guide readers. The black dashed lines represent subsystem boundaries, the red dashed double
602 dotted lines represent system boundaries, and the solid black lines bordering UMIS diagrams
603 represent whole system boundaries. A dynamic version of this figure is available as SI in
604 Microsoft PowerPoint format.
605

606 **Avoiding Double Counting of Data <heading level 2>**

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

614

615 **Equivalent Representations of Different Data <heading level 3>**

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

- 619 1. *ANT.1* and *NAT.2*, which contain data on the first level;
- 620 2. *ANT.1;1* and *ANT.1;1'*, which contain data on the second level (produced by
621 disaggregating data in *ANT.1*);
- 622 3. *ANT.1;1;1*, *ANT.1;1;2*, and *ANT.1;1;3* (produced by disaggregating *ANT.1;1*), and also
623 *ANT.1;1';1*, *ANT.1;1';2*, and *ANT.1;1';3* (produced by disaggregating *ANT.1;1'*) contain
624 data on the third level; and
- 625 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*
626 and *ANT.1;1;3;2* (produced by disaggregating *ANT.1;1;3*), which contain data on the
627 fourth level and are not disaggregated further.

628

629 In the example shown in Figure 8, the aggregate subsystem module *ANT* can only be fully
630 represented by data on the first, second, or third disaggregation levels (condition 2a) because
631 *ANT.1;1;2*, *ANT.1;1';1*, *ANT.1;1';2*, and *ANT.1;1';3* are not disaggregated further here. *ANT* is
632 also specified by using only one fork subsystem (Figures 8B-8F) or copied fork subsystem (Figures
633 8G-8H) at the single instance where subsystem forking occurs, i.e., at *ANT.1* (condition 2b). Here,
634 *ANT* and *NAT* are individual stages in a material cycle and together constitute the reference
635 material m_I , reference space s_I , and reference timeframe t_I component of the whole system
636 (condition 1). The examples (Figures 8A-8I) show the flexibility of UMIS in defining a whole
637 system in terms of aggregate subsystem modules, which can be comprised of differently
638 disaggregated data depending on data availability or visualization priorities.

639

640

641 Figure 8. Equivalent representations (A-I) of the reference material m_I , reference space s_I , and
642 reference timeframe t_I component of the whole system, represented in terms of UMIS diagrams
643 and excluding double counting of data. Processes are replaced by grey shaded regions or omitted
644 in *ANT.1* and *NAT.2*, and flows are omitted. In (A and I), the aggregate subsystem modules *ANT*
645 and *NAT*, and their relevant data are shown. In (B), *ANT* is represented using data on the first
646 disaggregation level (*ANT.1*). *ANT* is represented using data on the second level of
647 disaggregation only in (C) and (H), i.e., for the *ANT.1;1* and *ANT.1;1'* subsystems, respectively.

648 In (D-G), *ANT* is represented by various combinations of data on the second, third, and fourth
649 disaggregation levels. The virtual reservoir and metadata layer are omitted for clarity. The black
650 dashed lines represent subsystem boundaries, the red dashed double dotted lines represent system
651 boundaries, and the solid black lines bordering UMIS diagrams represent whole system
652 boundaries.
653

654 **Selecting Data to Avoid Double Counting <heading level 3>**

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

668

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

676

677

678 Figure 9. Selection of differently disaggregated data in UMIS to avoid double counting.

679

680 Selection of data for the (A) copied subsystem set (*ANT.1*, *ANT.1;1'*, *ANT.1;1';1*, and
681 *ANT.1;1';2*) and (B) subsystem set (*ANT.1*, *ANT.1;1*, *ANT.1;1;1* and *ANT.1;1;2*) are shown.

682

683 Unselected subsystems (including their processes and flows) are covered by white blocks. Each
684 UMIS diagram, (A) and (B), define the reference material m_l , reference space s_l , and reference
685 timeframe t_l component of the whole system, but do so using differently disaggregated data.

686

687 Processes are omitted in *ANT.1* and *NAT.2* and replaced by grey shaded regions otherwise.

688

689 Flows, the virtual reservoir, and the metadata layer are omitted for clarity. The black dashed lines
690 represent subsystem boundaries, the red dashed double dotted lines represent system boundaries,
691 and the solid black lines bordering UMIS diagrams represent whole system boundaries.

692

689 Other Key Properties of UMIS <heading level 2>

690 Cross Boundary Flows and Trade <heading level 3>

691 Cross boundary flows (x_s) are defined in UMIS as flows between two reference spaces (s); a trade
692 flow is a type of cross boundary flow that occurs between system boundaries that fully describe
693 independent economic entities. They are implicitly represented in UMIS diagrams for a single
694 reference space. This is because cross boundary flows always occur between two transformative
695 or distributive processes with the same labels, which occur in subsystems with different reference
696 spaces but otherwise the same attributes. Therefore, a single UMIS diagram defines the labels for
697 every cross boundary flow associated with the subsystem(s) that it depicts. UMIS diagrams

698 representing individual reference spaces can be combined to result in a multi-regional UMIS
699 diagram that explicitly displays cross boundary flows (Figure 10). This treatment is analogous to
700 the compilation of multi-regional I/O tables (Peters and Hertwich, 2006).

701

702

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

709

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

711 Simultaneous consideration of multiple material cycles adds substantial complexity to system-
712 wide analyses of resources and materials, and is relatively infrequently reported (Nakajima et al.,
713 2013). For example, copper-cobalt concentrate produced as a by-product from copper
714 electrowinning (in the copper cycle) is typically recovered and then refined to cobalt metal
715 (Donaldson and Beyersmann, 2000), although MFA diagrams for the cobalt cycle may only
716 explicitly represent the latter recovery and refining steps (Harper et al., 2012). Therefore,
717 information about the copper cycle, e.g., the concentration of cobalt in copper-cobalt concentrate
718 and the amount of this material, can be used to determine material stocks and flows data in the
719 cobalt cycle. In UMIS, materials that are not included in the defined reference material (which are
720 thus outside the system boundary of interest) are termed ‘intersecting materials’. Information about
721 intersecting materials that is used to determine material stocks and flows data in material cycles is
722 represented in UMIS diagrams in the metadata layer.

723

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

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

730

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

742

743

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

750

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

752 **By Name <heading level 4>**

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

763

764 **By Label <heading level 4>**

765 UMIS also enables hierarchical structuring of material stocks and flows data for commodities
766 produced along material cycles (of any reference material composition), and within the whole
767 system, to fully describe their component, engineering material, and substance constituents. This
768 is achieved by: (1) specifying a general reference material, e.g., metallic elements, car-related
769 materials, all materials, etc.; (2) using UMIS to structure and disaggregate material stocks and
770 flows data such that all commodities related to the specified reference material are explicit (with
771 the names of distributive processes defining these commodities); and then (3) disaggregating
772 processes related to each commodity using the divergent disaggregation approach such that each
773 of their components (sub-commodities), engineering materials, and elements are assigned
774 distributive processes (the order in which commodities are disaggregated into their constituents is

775 specified by the user). This is the method by which UMIS structures commodity-related data, e.g.,
776 monetary and mass trade statistics from the United Nations Comtrade Database (United Nations
777 Statistics Division, 2017).

778

779 **Data in Non-Mass Units <heading level 3>**

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

797

798 **Discussion <heading level 1>**

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

807

808 UMIS comprehensively places material stocks and flows data into material systems contexts by
809 uniquely labeling and visualizing subsystems, transformative, distributive, and storage processes,
810 stocks, and also flows. This labeling system facilitates referencing of UMIS structured and
811 visualized data, and their metadata e.g., uncertainty and system boundary properties, in complex
812 computational code and databases. For example, UMIS can be used to holistically integrate
813 material stocks and flows data describing vehicle value chains into a single systems context, such
814 as: the (co-)production of vehicle-related elements in individual mine sites; element stocks in
815 vehicles as functions of the country of sale, brand, and model; in-use phase greenhouse gas
816 emissions; and international trading of down-cycled scrap metal. These data, and this single
817 material system, could then be incorporated into a database and comprehensively visualized in a
818 UMIS diagram. The UMIS diagram could then be used to develop a computational script to model
819 this material system that has the flexibility to use these data at multiple levels of disaggregation at
820 each (life) cycle stage whilst also avoid double counting. This script could be coded with the aim

821 of producing material supply and demand scenarios for vehicles that are consistent with projected
822 low-CO₂ emissions technology mixes (Fulton and Ward, 2011). Therefore, UMIS provides a
823 flexible and comprehensive data structure that enables standardization, storage, and enhanced
824 exchanging of material stocks and flows data. Such a data structure is a necessary step towards the
825 complete and general standardization of material stocks and flows data. We believe that this
826 development will eventually enable a step change improvement in the capabilities of material
827 systems analysis, which will emerge as more (diverse) material stocks and flows data become
828 available and get consolidated.

829

830 It is important to emphasize for clarity that UMIS is not a database, it is a data structure that can
831 be used to place information about material systems into their respective context(s). These
832 contextualized data can then be used to develop tools such as databases, elicitation diagrams, and
833 computational models. A key motivation for developing UMIS comes from our work in integrating
834 ~20 years of material cycle and criticality data generated within Yale's Center for Industrial
835 Ecology into a single database. Here, UMIS is providing the data structure to comprehensively
836 place these material stocks and flows data into their respective systems contexts. This database
837 will be transferred to the United States Geological Survey upon completion, where it will be
838 maintained in an openly accessible format, given wide access, and periodically updated and
839 enhanced.

840

841 To illustrate the application and properties of UMIS, we have used UMIS to recast existing data
842 published for the cobalt cycle, and material stocks and flows data represented by block flow type
843 diagrams, system dynamics diagrams, Sankey diagrams, matrices, and also the EW-MFA

844 classification system, to demonstrate how it can be applied to other existing (as well as yet to be
845 published) data. These examples are presented in the SI.

846

847 We envisage that applications of UMIS to many diverse data sources will facilitate the
848 development of whole system databases, similar to the database that we are currently developing
849 at Yale's Center for Industrial Ecology. Our goal is for this database, and databases like it to which
850 the community can add data, to become foundational tools to unify and accumulate material stocks
851 and flows data. These data may then be extracted, analyzed, exchanged, and enhanced by diverse
852 users, who can use UMIS-type elicitation diagrams to visualize these data and to perform complex
853 computational data analyses. Key quantitative results from these analyses may then be flexibly
854 visualized and shared in communication tools, such as Sankey diagrams (Lupton and Allwood,
855 2017). Therefore, UMIS can provide a key role in advancing the cumulative body of knowledge
856 of material cycles in anthropogenic and natural systems.

857

858 **References <heading level 1>**

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

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

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

- 867 Brunner, P. H. and H. Rechberger. 2005. *Practical Handbook of Material Flow Analysis*. Boca
868 Raton (FL), United States: Lewis Publishers, CRC Press.
- 869 Chen, W.-Q. and T. E. Graedel. 2012. Anthropogenic Cycles of the Elements: A Critical Review.
870 *Environmental Science & Technology* 46(16): 8574-8586.
- 871 Chen, W.-Q., T. E. Graedel, P. Nuss, and H. Ohno. 2016. Building the Material Flow Networks of
872 Aluminum in the 2007 U.S. Economy. *Environmental Science & Technology* 50(7): 3905-3912.
- 873 Cullen, J. M., J. M. Allwood, and M. D. Bambach. 2012. Mapping the Global Flow of Steel: From
874 Steelmaking to End-Use Goods. *Environmental Science and Technology* 46(24): 13048-13055.
- 875 Davis, J., R. Geyer, J. Ley, J. He, R. Clift, A. Kwan, M. Sansom, and T. Jackson. 2007. Time-
876 Dependent Material Flow Analysis of Iron and Steel in the UK: Part 2. Scrap Generation and
877 Recycling. *Resources, Conservation and Recycling* 51(1): 118-140.
- 878 Donaldson, J. D. and D. Beyersmann. 2000. Cobalt and Cobalt Compounds. In *Ullmann's*
879 *Encyclopedia of Industrial Chemistry*: John Wiley and Sons Inc.
- 880 Eckelman, M. J. and I. Daigo. 2008. Markov Chain Modeling of the Global Technological
881 Lifetime of Copper. *Ecological Economics* 67(2): 265-273.
- 882 EUROSTAT. 2001. *Economy-Wide Material Flow Accounts and Derived Indicators - A*
883 *Methodological Guide*. Luxembourg: European Communities.
- 884 EUROSTAT. 2008. *Eurostat Manual of Supply, Use and Input-Output Tables*. Luxembourg:
885 European Commission.
- 886 Fischer-Kowalski, M., F. Krausmann, S. Giljum, S. Lutter, A. Mayer, S. Bringezu, Y. Moriguchi,
887 H. Schütz, H. Schandl, and H. Weisz. 2011. Methodology and Indicators of Economy-Wide
888 Material Flow Accounting. *Journal of Industrial Ecology* 15(6): 855-876.
- 889 Ford, A. 1999. *Modeling the Environment. An Introduction to System Dynamics Models of*
890 *Environmental Systems*. Washington, D.C.: Island Press.

- 891 Frosch, R. A. and N. E. Gallopoulos. 1989. Strategies for Manufacturing. *Scientific American* 261:
892 144-152.
- 893 Fulton, L. and J. Ward, eds. 2011. *Technology Roadmap: Electric and Plug-In Hybrid Electric*
894 *Vehicles*. Paris: International Energy Agency (IEA).
- 895 Graedel, T. E., M. Bertram, K. Fuse, R. B. Gordon, R. Lifset, H. Rechberger, and S. Spatari. 2002.
896 The Contemporary European Copper Cycle: The Characterization of Technological Copper
897 Cycles. *Ecological Economics* 42: 9-26.
- 898 Graedel, T. E., D. Van Beers, M. Bertram, K. Fuse, R. B. Gordon, A. Gritsinin, E. M. Harper, A.
899 Kapur, R. J. Klee, R. Lifset, L. Memon, and S. Spatari. 2005. The Multilevel Cycle of
900 Anthropogenic Zinc. *Journal of Industrial Ecology* 9(3): 67-90.
- 901 Graedel, T. E., D. van Beers, M. Bertram, K. Fuse, R. B. Gordon, A. Gritsinin, A. Kapur, R. J.
902 Klee, R. J. Lifset, L. Memon, H. Rechberger, S. Spatari, and D. Vexler. 2004. Multilevel Cycle of
903 Anthropogenic Copper. *Environmental Science & Technology* 38(4): 1242-1252.
- 904 Harper, E. M., G. Kavlak, and T. E. Graedel. 2012. Tracking the Metal of the Goblins: Cobalt's
905 Cycle of Use. *Environmental Science & Technology* 46(2): 1079-1086.
- 906 Hawkins, T., C. Hendrickson, C. Higgins, H. S. Matthews, and S. Suh. 2007. A Mixed-Unit Input-
907 Output Model for Environmental Life-Cycle Assessment and Material Flow Analysis.
908 *Environmental Science & Technology* 41(3): 1024-1031.
- 909 Hellweg, S. and L. M. i Canals. 2014. Emerging Approaches, Challenges and Opportunities in
910 Life Cycle Assessment. *Science* 344(6188): 1109.
- 911 Hendriks, C., R. Obernosterer, D. Muller, S. Kytzia, P. Baccini, and P. H. Brunner. 2000. Material
912 Flow Analysis: A Tool to Support Environmental Policy Decision Making. Case-Studies on the
913 City of Vienna and the Swiss Lowlands. *Local Environment* 5(3): 311.
- 914 Hoekman, P. and H. von Blottnitz. 2016. Cape Town's Metabolism: Insights from a Material Flow
915 Analysis. *Journal of Industrial Ecology*: <http://dx.doi.org/10.1111/jiec.12508>.

- 916 Lenzen, M., A. Geschke, T. Wiedmann, J. Lane, N. Anderson, T. Baynes, J. Boland, P. Daniels,
917 C. Dey, J. Fry, M. Hadjikakou, S. Kenway, A. Malik, D. Moran, J. Murray, S. Nettleton, L.
918 Poruschi, C. Reynolds, H. Rowley, J. Ugon, D. Webb, and J. West. 2014. Compiling and Using
919 Input–Output Frameworks Through Collaborative Virtual Laboratories. *Science of The Total*
920 *Environment* 485–486: 241-251.
- 921 Lupton, R. C. and J. M. Allwood. Incremental Material Flow Analysis with Bayesian Inference.
922 *Journal of Industrial Ecology*: accepted.
- 923 Lupton, R. C. and J. M. Allwood. 2017. Hybrid Sankey Diagrams: Visual Analysis of
924 Multidimensional Data for Understanding Resource Use. *Resources, Conservation and Recycling*
925 124: 141-151.
- 926 Meylan, G. and B. K. Reck. 2017. The Anthropogenic Cycle of Zinc: Status Quo and Perspectives.
927 *Resources, Conservation and Recycling* 123: 1-10.
- 928 Miller, R. E. and P. D. Blair. 2009. *Input-Output Analysis: Foundations and Extensions*.
929 Cambridge, United Kingdom: Cambridge University Press.
- 930 Müller, D. B. 2006. Stock Dynamics for Forecasting Material Flows—Case Study for Housing in
931 the Netherlands. *Ecological Economics* 59(1): 142-156.
- 932 Müller, D. B., T. Wang, B. Duval, and T. E. Graedel. 2006. Exploring the Engine of Anthropogenic
933 Iron Cycles. *Proceedings of the National Academy of Sciences* 103(44): 16111-16116.
- 934 Nakajima, K., H. Ohno, Y. Kondo, K. Matsubae, O. Takeda, T. Miki, S. Nakamura, and T.
935 Nagasaka. 2013. Simultaneous Material Flow Analysis of Nickel, Chromium, and Molybdenum
936 Used in Alloy Steel by Means of Input–Output Analysis. *Environmental Science & Technology*
937 47(9): 4653-4660.
- 938 Nakamura, S. and K. Nakajima. 2005. Waste Input-Output Material Flow Analysis of Metals in
939 the Japanese Economy. *Materials Transactions* 46(12): 2550-2553.

- 940 Nakamura, S., Y. Kondo, K. Matsubae, K. Nakajima, and T. Nagasaka. 2011. UPIOM: A New
941 Tool of MFA and Its Application to the Flow of Iron and Steel Associated with Car Production.
942 *Environmental Science & Technology* 45(3): 1114-1120.
- 943 Pauliuk, S., G. Majeau-Bettez, and D. B. Müller. 2015. A General System Structure and
944 Accounting Framework for Socioeconomic Metabolism. *Journal of Industrial Ecology* 19(5): 728-
945 741.
- 946 Pauliuk, S., R. L. Milford, D. B. Müller, and J. M. Allwood. 2013. The Steel Scrap Age.
947 *Environmental Science & Technology* 47(7): 3448-3454.
- 948 Peters, G. P. and E. G. Hertwich. 2006. Structural Analysis of International Trade: Environmental
949 Impacts of Norway. *Economic Systems Research* 18(2): 155-181.
- 950 Polis, G. A. and K. O. Winemiller. 1996. *Food Webs: Integration of Patterns & Dynamics*. New
951 York: Chapman & Hall.
- 952 Rauch, J. N. and T. E. Graedel. 2007. Earth's Anthrobiogeochemical Copper Cycle. *Global*
953 *Biogeochemical Cycles* 21(2): GB2010.
- 954 Rauch, J. N. and J. M. Pacyna. 2009. Earth's Global Ag, Al, Cr, Cu, Fe, Ni, Pb, and Zn Cycles.
955 *Global Biogeochemical Cycles* 23(2): GB2001.
- 956 Schmidt, M. 2008. The Sankey Diagram in Energy and Material Flow Management. *Journal of*
957 *Industrial Ecology* 12(2): 173-185.
- 958 Talens Peiró, L., G. Villalba Méndez, and R. U. Ayres. 2013. Lithium: Sources, Production, Uses,
959 and Recovery Outlook. *JOM* 65(8): 986-996.
- 960 Tanimoto, A. H., X. Gabarrell Durany, G. Villalba, and A. C. Pires. 2010. Material Flow
961 Accounting of the Copper Cycle in Brazil. *Resources, Conservation and Recycling* 55(1): 20-28.
- 962 U.S. Geological Survey. 2011. *Mineral Commodity Summaries 2011*: U.S. Geological Survey.
- 963 Uihlein, A., W.-R. Poganietz, and L. Schebek. 2006. Carbon Flows and Carbon Use in the German
964 Anthroposphere: An Inventory. *Resources, Conservation and Recycling* 46(4): 410-429.

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

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

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

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

975

976 **Supporting Information <heading level 1>**

977 Additional Supporting Information (SI) may be found in the online version of this article at the
978 publisher's website: (1) Application of UMIS to recast cobalt cycle data reported by Harper et al.
979 (2012) and three figures (Figures S1-S3) that illustrate this procedure (section S1.1); (2) a UMIS
980 diagram for the cobalt cycle, a single reference space, a single reference timeframe, and all
981 aggregate subsystem modules, presented as a comma separated value file
982 (UMIS_diagram_cobalt.csv); the Python script used to generate this UMIS diagram, provided as
983 (3) Python (UMIS_diagrams_1.0.py) and (4) IPython notebooks (UMIS_diagrams_1.0.ipynb),
984 and also in (5) hypertext markup language (UMIS_diagrams_1.0.html); and (6) the input file for
985 the Python script (transformative_processes_input_cobalt.csv). Example applications of UMIS to
986 recast data published in a (7) block flow type diagram (section S1.2), a (8) system dynamics
987 diagram (section S1.3), a (9) Sankey diagram (section S1.4), data structured using the (10) EW-
988 MFA classification system (section S1.5), and data published for a (10) LCA system represented

989 by a matrix and a block flow type diagram (section S1.6), their respective UMIS diagrams ((11)
990 UMIS_diagram_bflow.csv, (12) UMIS_diagram_sdyn.csv, (13) UMIS_diagram_sankey.csv, (14)
991 UMIS_diagram_ewmfa.csv, (15) UMIS_diagram_matrixlca.csv), and input files for the
992 aforementioned Python script ((16) transformative_processes_input_bflow.csv, (17)
993 transformative_processes_input_sdyn.csv, (18) transformative_processes_input_sankey.csv, (19)
994 transformative_processes_input_ewmfa.csv, (20) transformative_processes_input_matrixlca.csv),
995 are also provided as SI. We additionally provide dynamic versions of (21) Figure 5, (22) Figure 7,
996 and (23) Figure S2 as SI in Microsoft PowerPoint format, a (24) pdf version of the UMIS diagram
997 for the matrix-based LCA system (UMIS_diagram_matrixlca.pdf, note: flow labels are omitted in
998 this diagram for simplicity), and also high resolution images of (25) Figure S2 and (26) Figure S3
999 as SI in pdf format. These examples demonstrate a variety of potential applications of UMIS and
1000 also exhibit some minor yet important features of UMIS not fully covered in the main text.

1001

1002 **Acknowledgements <heading level 1>**

1003 This work is supported by Grant #1636509 of the National Science Foundation. We are deeply
1004 thankful to Michael S. Baker, Nedal T. Nassar, Daniel B. Müller, Maren Lundhaug, Mark Uwe
1005 Simoni, Oliver Schwab, Benjamin Sprecher, David Font Vivanco, Ranran Wang, Edgar Hertwich,
1006 and Stefan Pauliuk for their insightful feedback. We also thank three anonymous reviewers whose
1007 comments helped to significantly improve the quality of this paper.

1008

1009 **Conflicts of Interests <heading level 1>**

1010 The authors declare no competing financial interests.

1011

1012 **List of Figure Captions <heading level 1>**

1013 Figure 1. Relationships between material stocks and flows in anthropogenic and natural systems.
1014 Material stored in a particular reservoir undergoes processing, storage, distribution, and
1015 transformation, to again become stored in another (one or more) reservoir(s). Total mass is
1016 conserved but the location of the material changes. These relationships between reservoirs and
1017 processes provide a basis upon which a unified structure for material stocks and flows data can
1018 be built.
1019

1020
1021 Figure 2. Exemplary block flow type diagram for the iron cycle, the year 2000, and the United
1022 States, adapted from (Müller et al., 2006). Mass quantities in Tg/year are displayed adjacent to
1023 each respective flow. Mass balance residuals are not shown (e.g., around the ‘Blast Furnace’
1024 transformative process). Note that some distributive processes needed to avoid material flowing
1025 between two processes of the same type and thus to ensure consistency with the bipartite directed
1026 graph structure are omitted, e.g., between the ‘Manuf.’ and ‘Scrap Process. & Waste Manag.’
1027 transformative. Production (dashed green box), engineering materials (dashed yellow box),
1028 fabrication & manufacturing (dashed purple box), use (dashed orange box), waste management
1029 (dashed red box), and environment (dashed blue box) subsystems are added to illustrate the
1030 subsystem concept (see Development of the Unified Materials Information System (UMIS)).
1031

1032
1033 Figure 3. Relationships between (A) I/O analysis (make and use tables), (B) MFA (block flow
1034 type diagram), and (C) LCA (inventory) data. Transformative and distributive processes are
1035 shown as darker grey filled squares and lighter grey filled circles, respectively. Flows are
1036 displayed as arrows. Colored bold arrows (B-C) are flows that are entered into the make and use
1037 tables here (A). Subsystem, aggregate subsystem module, and system boundaries are shown as
1038 dashed, bold dashed, and alternating dashed double dotted lines, respectively. Process and flow
1039 labels are used to reference data between the respective methodologies; their formulation, and
1040 also labeling of subsystems, are described in the text. The environment subsystem is included in
1041 (C) to demonstrate the compilation of an inventory table, which is done by disaggregating the
1042 aggregate production of engineering materials subsystem module (*PEM.1*) (shaded green boxes
1043 in B and C) to account for all inflows to and outflows from the *aggregate environment subsystem*
1044 *module (ENV.5)* (black bold arrows).
1045

1046
1047 Figure 4. (A) Key aspects of UMIS, illustrated using UMIS type diagrams for one of each
1048 transformative, distributive, and storage process, three flows, the virtual reservoir, and the
1049 metadata layer. (B) The virtual reservoir shown here can lie inside or outside the system
1050 boundary, but occurs inside of it here. The metadata layer contains additional information (e.g.,
1051 uncertainty, system boundary properties) about processes, stocks, and/or flows positioned at the
1052 same matrix coordinates. Flows depicted by grey arrows in (A) and conceptual linkages depicted
1053 by black arrows in (B) are omitted in UMIS diagrams, and are only shown here to guide readers.

1054

1055

1056 Figure 5. First stage of subsystem specification, which occurs in three steps. (A) Step 1,
1057 aggregate subsystem modules *ANT* and *NAT* are defined, which cumulatively represent the
1058 reference material m_I , reference space s_I , and reference timeframe t_I component of the whole
1059 system. *ANT.1* and *NAT.2* subsystems are also defined. (B) Step 2, specification of the *ANT.1;1*
1060 subsystem to fully describe the available (consistently disaggregated) data for *ANT.1* and
1061 reference material m_I in the reference space s_I and reference timeframe t_I component of the
1062 whole system. (C) Step 3, specification of all flows from distributive to transformative processes.
1063 (D) UMIS diagram produced with *production and use (ANT.1;1;1)* and *recycling and disposal*
1064 (*ANT.1;1;2*) subsystems, processes, and flows defined by disaggregating *ANT.1;1*. The virtual
1065 reservoir and metadata layer are omitted for clarity. Flows depicted by faded grey arrows in (C)
1066 and black arrows depicting subsystem disaggregation in (B) and (D) are omitted in UMIS
1067 diagrams, and are only shown here to guide readers. The black dashed lines represent subsystem
1068 boundaries, the red dashed double dotted lines represent system boundaries, and the solid black
1069 lines bordering UMIS diagrams represent whole system boundaries. A dynamic version of this
1070 figure is available as SI in Microsoft PowerPoint format.
1071

1072

1073 Figure 6. Divergent disaggregation of *cars* data into (A) *big* or *small (cars)*, and (B) *red* or *blue*
1074 (*cars* ') types within the *transport* system. The *transport* data in (A), i.e., four *cars*, are 'copied'
1075 as *transport* data into (B) to describe both types of disaggregated *cars* data. Two cars are *big*,
1076 two cars are *small*, one car is *red*, and three cars are *blue*. Only data from a single material tree
1077 should be used by a modeler at any one time, either the (A) material tree or the (B) copied
1078 material tree, else the visualized system describes eight rather than four cars (i.e., to avoid double
1079 counting of data). Nodes in material trees are analogous to subsystems in subsystem sets in
1080 UMIS diagrams.
1081

1082

1083 Figure 7. Second stage of subsystem specification, which occurs in three steps. (A) Step 1, the
1084 fork subsystem *ANT.1 (aggregate anthroposphere)* is copied to yield the copied fork subsystem
1085 *ANT.1 (aggregate anthroposphere)*. These subsystems are equivalent, substitutable, and occur
1086 within the same aggregate subsystem module (*ANT*). (B) Step 2, processes and flows in the
1087 copied fork subsystem *ANT.1* are disaggregated and *ANT.1;1' (anthroposphere)*, *ANT.1;1';1*
1088 (*metals*), and *ANT.1;1';2 (non-metals)* subsystems are defined to fully describe the available data
1089 for this copied subsystem set. (C) Step 3, the copied subsystem set (*ANT.1, ANT.1;1',*
1090 *ANT.1;1';1, and ANT.1;1';2*) is added to the UMIS diagram and all flows from distributive to
1091 transformative processes are specified. This fully specifies the reference material m_I , reference
1092 space s_I , and reference timeframe t_I component of the whole system. The virtual reservoir and
1093 metadata layer are omitted for clarity. Flows are omitted, and processes are omitted in *ANT.1* and
1094 *NAT.2* or otherwise replaced by grey shaded regions in (C) to simplify the diagram. Thick black
1095 arrows and lines depicting subsystem specification and disaggregation in (A-C), and shaded grey
1096 regions representing processes in (C), are omitted in UMIS diagrams, and are only shown here to

1097 guide readers. The black dashed lines represent subsystem boundaries, the red dashed double
1098 dotted lines represent system boundaries, and the solid black lines bordering UMIS diagrams
1099 represent whole system boundaries. A dynamic version of this figure is available as SI in
1100 Microsoft PowerPoint format.
1101

1102
1103 Figure 8. Equivalent representations (A-I) of the reference material m_I , reference space s_I , and
1104 reference timeframe t_I component of the whole system, represented in terms of UMIS diagrams
1105 and excluding double counting of data. Processes are replaced by grey shaded regions or omitted
1106 in *ANT.1* and *NAT.2*, and flows are omitted. In (A and I), the aggregate subsystem modules *ANT*
1107 and *NAT*, and their relevant data are shown. In (B), *ANT* is represented using data on the first
1108 disaggregation level (*ANT.1*). *ANT* is represented using data on the second level of
1109 disaggregation only in (C) and (H), i.e., for the *ANT.1;1* and *ANT.1;1'* subsystems, respectively.
1110 In (D-G), *ANT* is represented by various combinations of data on the second, third, and fourth
1111 disaggregation levels. The virtual reservoir and metadata layer are omitted for clarity. The black
1112 dashed lines represent subsystem boundaries, the red dashed double dotted lines represent system
1113 boundaries, and the solid black lines bordering UMIS diagrams represent whole system
1114 boundaries.
1115

1116
1117 Figure 9. Selection of differently disaggregated data in UMIS to avoid double counting.
1118 Selection of data for the (A) copied subsystem set (*ANT.1*, *ANT.1;1'*, *ANT.1;1';1*, and
1119 *ANT.1;1';2*) and (B) subsystem set (*ANT.1*, *ANT.1;1*, *ANT.1;1;1* and *ANT.1;1;2*) are shown.
1120 Unselected subsystems (including their processes and flows) are covered by white blocks. Each
1121 UMIS diagram, (A) and (B), define the reference material m_I , reference space s_I , and reference
1122 timeframe t_I component of the whole system, but do so using differently disaggregated data.
1123 Processes are omitted in *ANT.1* and *NAT.2* and replaced by grey shaded regions otherwise.
1124 Flows, the virtual reservoir, and the metadata layer are omitted for clarity. The black dashed lines
1125 represent subsystem boundaries, the red dashed double dotted lines represent system boundaries,
1126 and the solid black lines bordering UMIS diagrams represent whole system boundaries.
1127

1128
1129 Figure 10. Conceptual visualization of cross boundary flows (xs) in a multi-regional UMIS
1130 diagram. The subsystem is fixed, the reference material and reference timeframe components of
1131 the whole system are fixed, and there are two reference spaces, s_1 and s_2 . Cross boundary flows
1132 are shown as red diamonds in the blue shaded regions (faded grey arrows are shown here to
1133 guide readers only and are not normally displayed). The virtual reservoir and metadata layer are
1134 omitted for clarity.
1135

1136
1137 Figure 11. UMIS diagram representation of the whole system, shown in terms of ‘snapshots’ at
1138 four reference timeframes (t_1 (least recent), t_2 , t_3 , and t_4 (most recent)), two reference spaces (s_1

1139 and s_2), and a single reference material (m_1). Five aggregate subsystem modules (PEM , $F\&M$,
1140 USE , WMR , ENV) are shown in yellow shaded boxes within each system boundary (represented
1141 by red alternating dashed double dotted lines). Cross boundary flows from reference spaces s_1 to
1142 s_2 (x_{s_1-2}), and from reference spaces s_2 to s_1 (x_{s_2-1}) are shown as blue shaded regions.
1143