Representing Natural Kinds by Spatial Inclusion and Containment

Stefan Schulz

UNIVERSITÄTS FREIBURG KLINIKUM

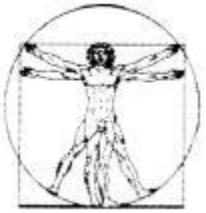
Medical Informatics, Freiburg University, Germany
http://www.imbi.uni-freiburg.de/medinf

Udo Hahn

Language and Information Engineering (JULIE) Lab,
Jena University, Germany

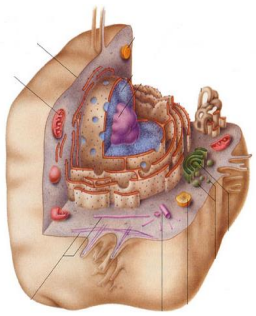
Ontologies of Biological Structure

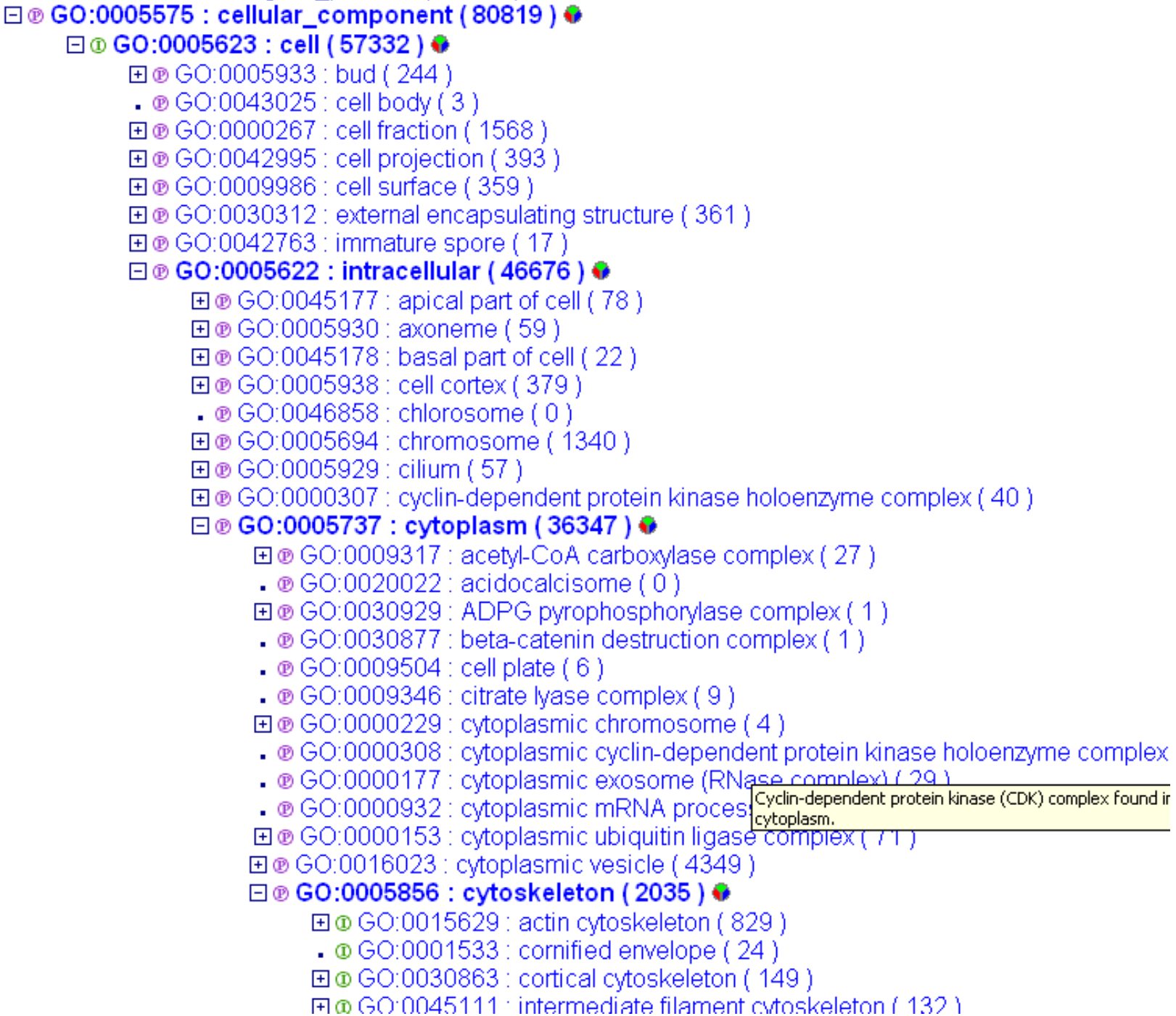
- Representation of Physical Parts of Organisms
- Large terminological repositories in biology exist and grow ...
  - Foundational Model of Anatomy: Human
    sig.biostr.washington.edu/projects/fm/



  - Open Biological Ontologies (OBO): Mouse, Fly, Fish, Worm, Fungi, …
    obo.sourceforge.net

  - Gene Ontology (GO) cellular component: Species independent
    www.geneontology.org/




GO:0005575 : cellular_component (80819)
GO:0005623 : cell (57332)
GO:0005933 : bud (244)
GO:0043025 : cell body (3)
GO:0000267 : cell fraction (1568)
GO:0042995 : cell projection (393)
GO:0009986 : cell surface (359)
GO:0030312 : external encapsulating structure (361)
GO:0042763 : immature spore (17)
GO:0005622 : intracellular (46676)
GO:0045177 : apical part of cell (78)
GO:0005930 : axoneme (59)
GO:0045178 : basal part of cell (22)
GO:0005938 : cell cortex (379)
GO:0046858 : chlorosome (0)
GO:0005694 : chromosome (1340)
GO:0005929 : cilium (57)
GO:0000307 : cyclin-dependent protein kinase holoenzyme complex (40)
GO:0005737 : cytoplasm (36347)
GO:0009317 : acetyl-CoA carboxylase complex (27)
GO:0020022 : acidocalcisome (0)
GO:0030929 : ADPG pyrophosphorylase complex (1)
GO:0030877 : beta-catenin destruction complex (1)
GO:0009504 : cell plate (6)
GO:0009346 : citrate lyase complex (9)
GO:0000229 : cytoplasmic chromosome (4)
GO:0000308 : cytoplasmic cyclin-dependent protein kinase holoenzyme complex
GO:0000177 : cytoplasmic exosome (RNase complex) (29)
GO:0000932 : cytoplasmic mRNA proces
Cyclin-dependent protein kinase (CDK) complex found ir
cytoplasm.
GO:0000153 : cytoplasmic ubiquitin ligase complex (71)
GO:0016023 : cytoplasmic vesicle (4349)
GO:0005856 : cytoskeleton (2035)
GO:0015629 : actin cytoskeleton (829)
GO:0001533 : cornified envelope (24)
GO:0030863 : cortical cytoskeleton (149)
GO:0045111 : intermediate filament cytoskeleton (132)

Canonic Representation of Mereotopological Structure

- *rel* $(Class_1, Class_2)$, *e.g.*

  part-of (*CellNucleus, Cell*)

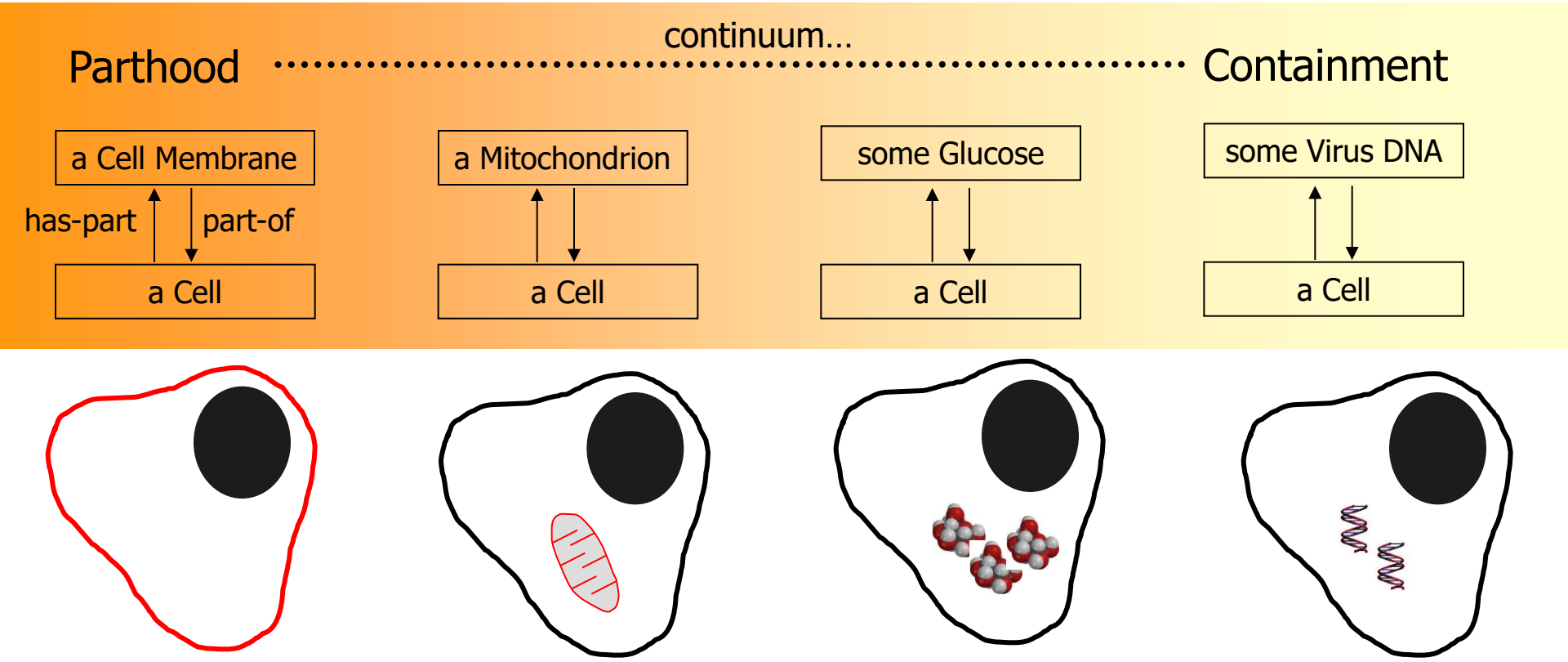
Canonic Representation of Mereotopological Structure

- *rel* $(Class_1, Class_2)$, *e.g.*
  part-of (CellNucleus, Cell)
- Open questions:
  - What is the meaning of mereotopological relations in Biology
  - Class level reading of mereotopological relations, how to interpret ?

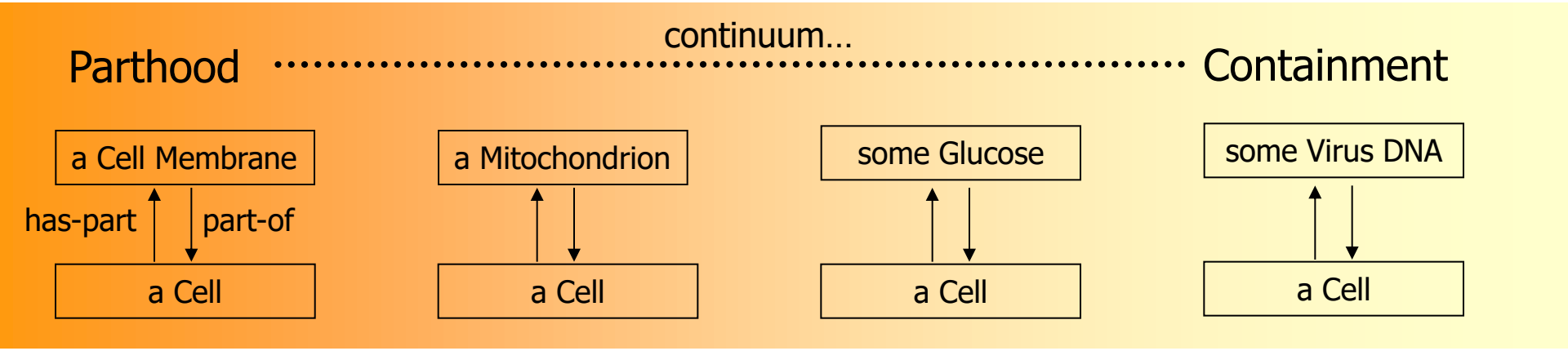
Canonic Representation of Mereotopological Structure

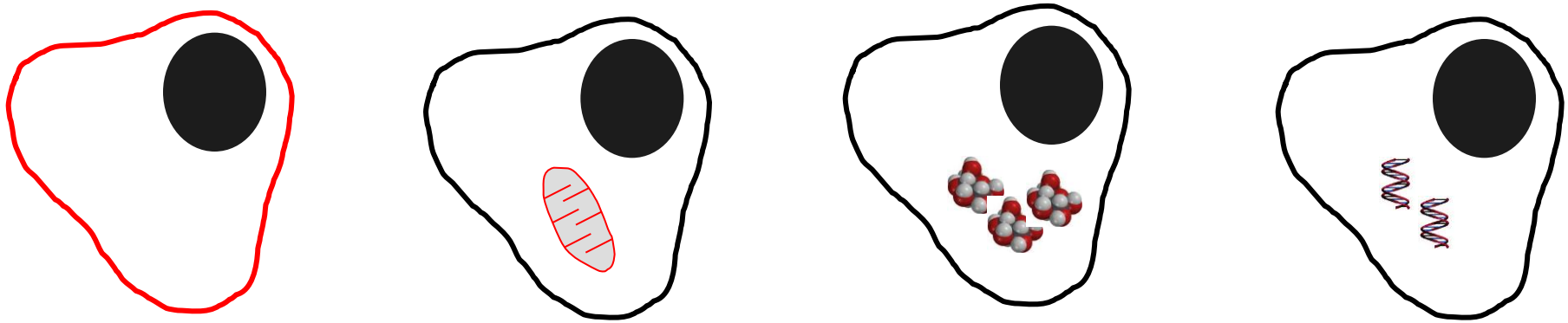
- *rel* $(Class_1, Class_2)$, *e.g.*
  part-of (CellNucleus, Cell)
- Open questions:
  - What is the meaning of mereotopological relations in Biology
  - Class level reading of mereotopological relations , how to interpret ?

Fuzziness of Mereotopological Relations in Biology



Fuzziness of Mereotopological Relations in Biology





includes / containee-of

has-part / part-of

location-of / has-location

Canonic Representation of Mereotopological Structure

- *rel* $(Class_1, Class_2)$, *e.g.*
  part-of (CellNucleus, Cell)
- Open questions:
  - What is the meaning of mereotopological relations in Biology
  - Class level reading of mereotopological relations , how to interpret ?

Class level reading of mereotopological relations

$R(C_1, C_2)$ expression, e.g. $Includes(Cell, CellNucleus)$

Alternative, conflicting interpretations:

- expresses ontological dependencies
  - "*each* cell includes a cell nucleus" and / or
  - "*each* cell nucleus is included in a cell"
- permits possible relations
  - "a cell *may* include a cell nucleus"
- reject any assertion not sanctioned by a *R(C1 ,C2)* expression:
  - "there is a cell nucleus which included in a protein molecule"

Instance and Class level reading of mereotopological relations (I)

Instance level:

containee-of, *includes:* transitive, reflexive, antisymmetric

$containee\text{-}of(x,y) \leftrightarrow includes(y,x)$

Class level:

- class A is a **specific containee** of class B:

$$SC(A,B) =_{def} \forall x: instance\text{-}of(x, A) \rightarrow \exists y: instance\text{-}of(y, B) \wedge containee\text{-}of(y,x)$$

- class B is a **specific includer** of class A:

$$SI(B,A) =_{def} \forall y: instance\text{-}of(y, B) \rightarrow \exists x: instance\text{-}of(x, A) \wedge includes(y,x)$$

Instance and Class level reading of mereotopological relations (II)

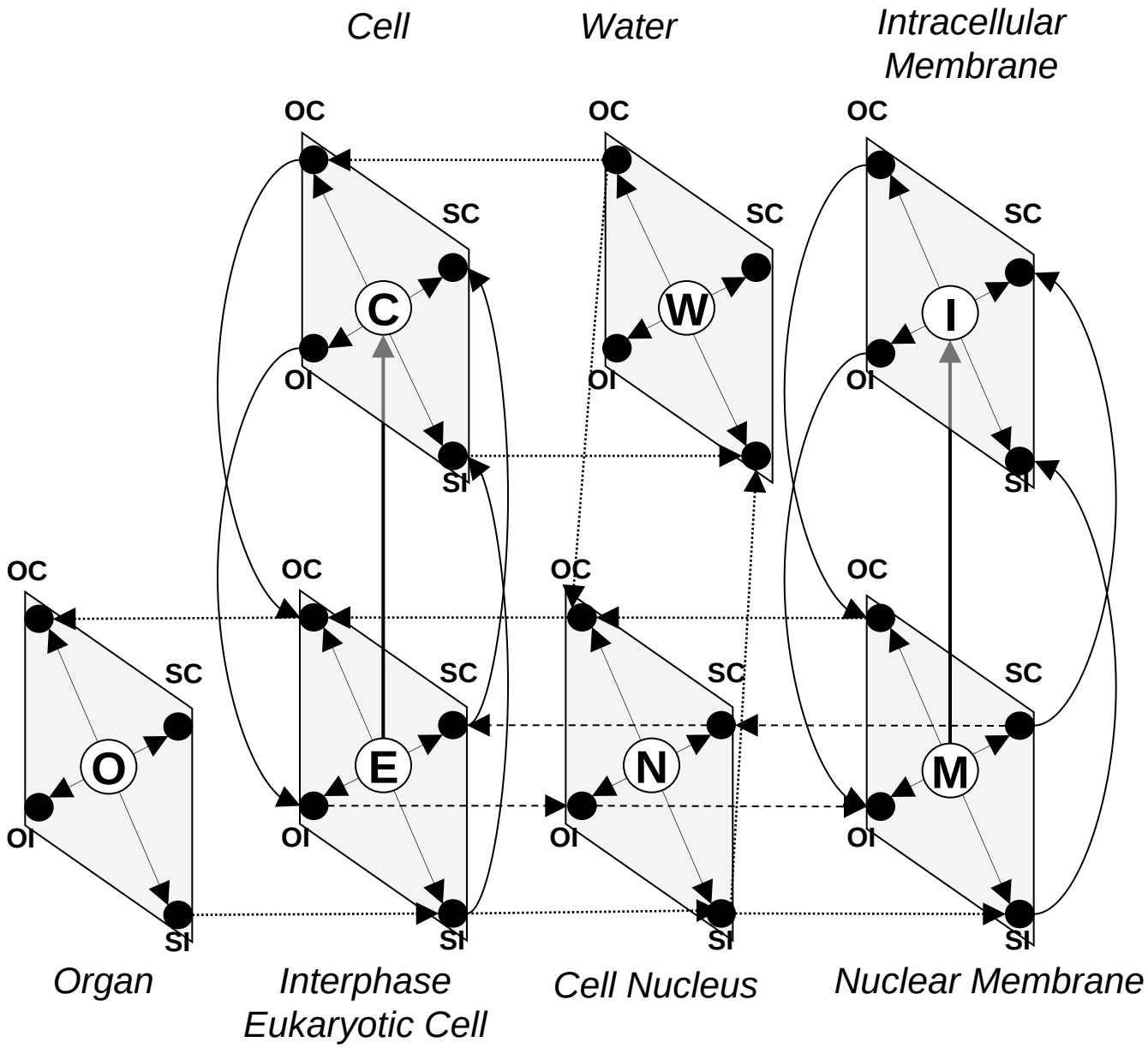
- $SC(A,B)$ is not the inverse of $SI(B,A)$
- class A is an **obligatory containee** of class B
  $OC(A,B) \leftrightarrow_{def} SI(B,A)$
- class B is an **obligatory includer** of class A
  $OI(B,A) \leftrightarrow_{def} SC(A,B)$
- SC, SI, OC, OI: transitive, reflexive, antisymmetric
- SC and SI propagate via $Is\text{-}A$:
  if A is a specific containee (includer) of B, every subclass of A is a specific containee (includer) of B, too
- OC and OI do not propagate via $Is\text{-}A$:
  if A is an obligatory containee (includer) of B, not any subclass of A is an obligatory containee (includer) of B

Arg1 \ Arg2	Cell	Neuron	Axon	Cell Membrane	Organ	Cell Nucleus	Heart	Water	Blood	Insulin	Electron
Cell		**Is-A^{-1}**	**OI**	**SI** **OI**	**OC**	**OI**	**OC**	**SI**	**OC**		**SI**
Neuron	**Is-A**		**OI**	**SI**		**SI**		**SI**			**SI**
Axon	**SC**	**SC**						**SI**			**SI**
Cell Membrane	**SC** **OC**	**OC**			**OC**		**OC**		**OC**		**SI**
Organ	**SI**			**SI**		**SI**	**Is-A^{-1}**	**SI**	**SI**	**SI**	**SI**
Cell Nucleus	**SC**	**OC**			**OC**		**OC**	**SI**	**OC**		**SI**
Heart	**SI**			**SI**	**IS-A**	**SI**		**SI**	**SI**	**SI**	**SI**
Water	**OC**	**OC**	**OC**		**OC**	**OC**	**OC**		**OC**		**SI**
Blood	**SI**			**SI**	**OC**	**SI**	**OC**	**SI**		**SI**	**SI**
Insulin					**OC**		**OC**		**OC**		**SI**
Electron	**OC**	**OC**	**OC**	**OC**	**OC**	**OC**	**OC**	**OC**	**OC**	**OC**	

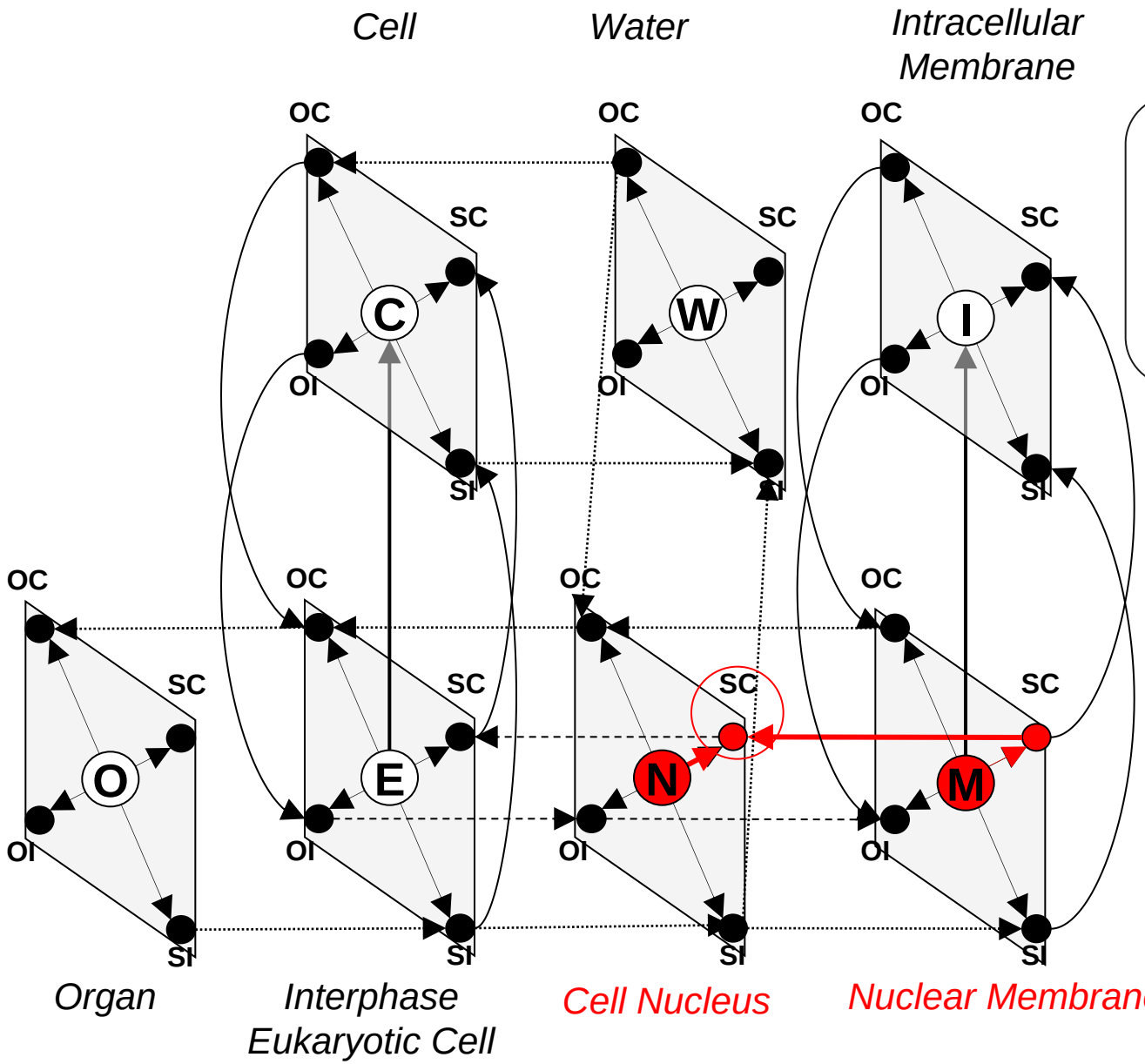
Extended Taxonomies

- Express mereotopological hierarchies as taxonomies
- Purpose: Better performance in large knowledge bases
- Introduction of reificator nodes:
  real classes:
  $Is\text{-}A(A,B_{SC}) =_{def} SC(A,B)$
  $Is\text{-}A(A,B_{SI}) =_{def} SI(A,B)$
  "pseudo classes" (do not capture properties of all instances !):
  $Is\text{-}A(A,B_{OC}) =_{def} OC(A,B)$
  $Is\text{-}A(A,B_{OI}) =_{def} OI(A,B)$
  } *A must be terminal nodes*

Extended Taxonomies

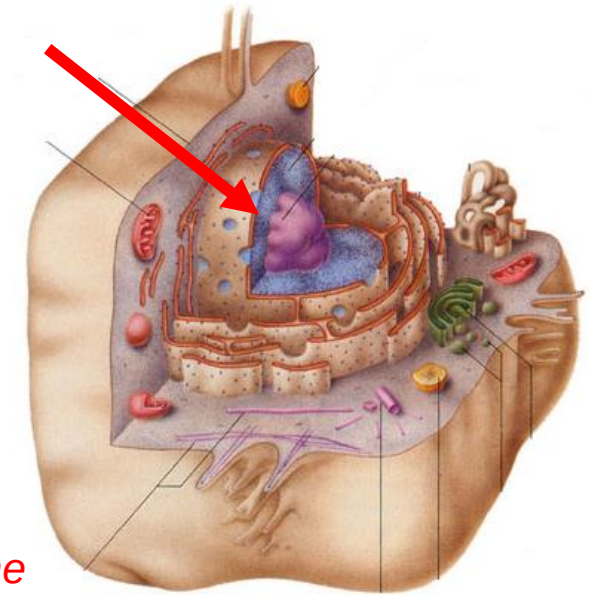


Extended Taxonomies

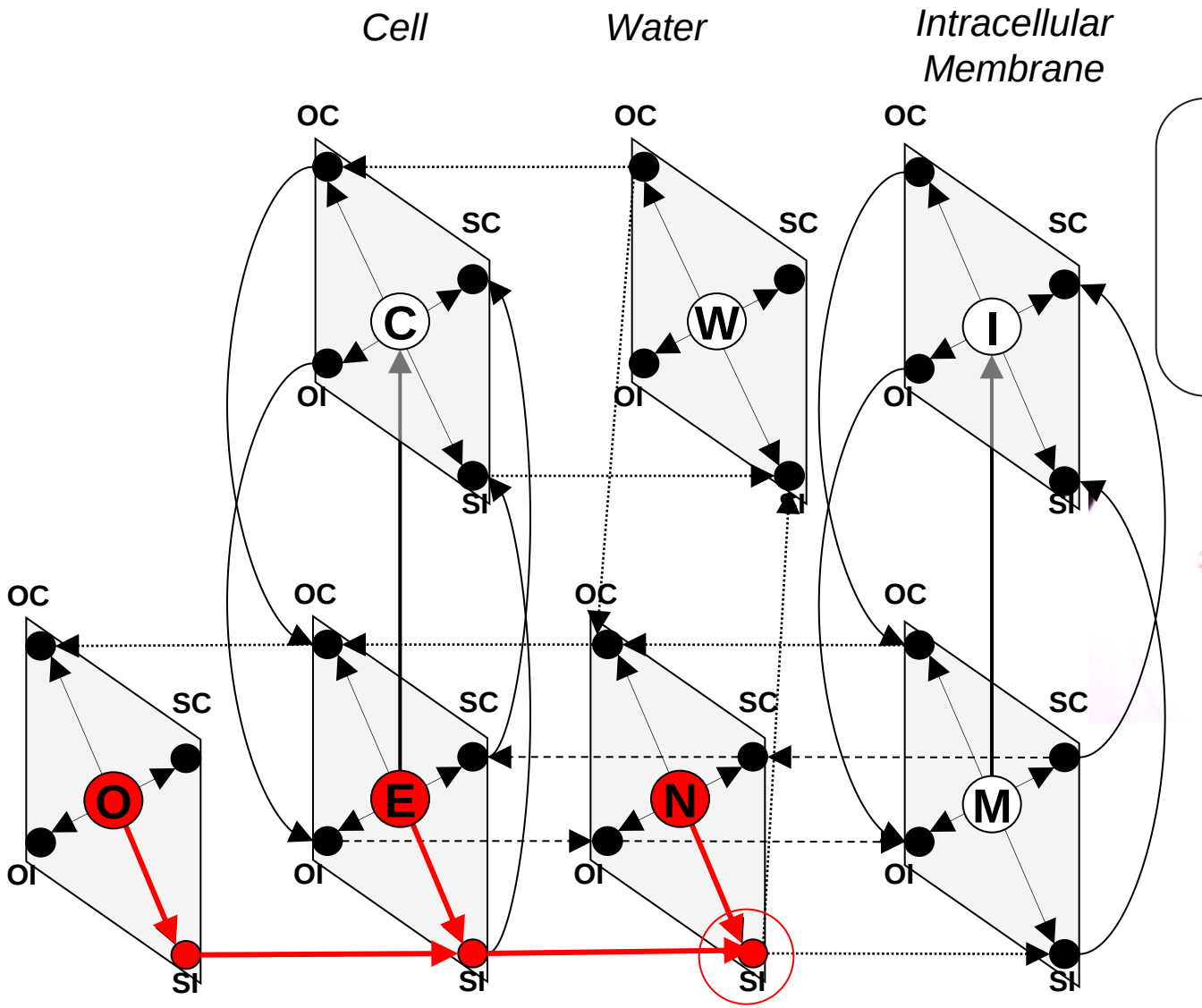


Specific Containees of Cell Nucleus

Cell Nucleus
Nuclear Membrane

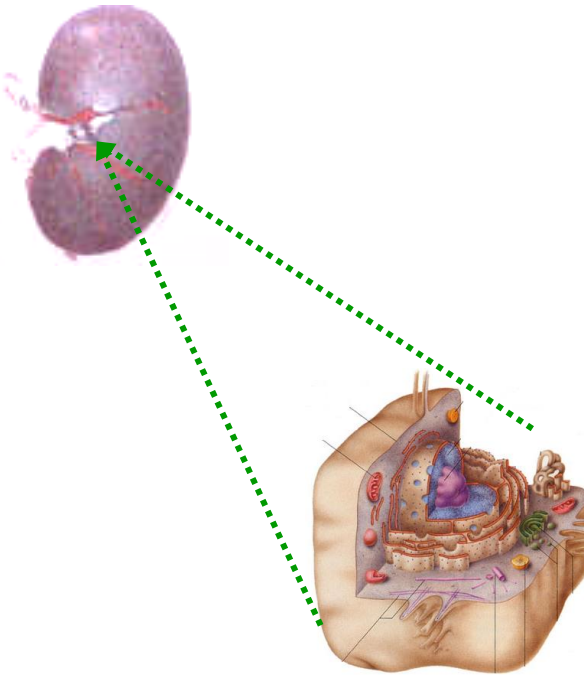


Extended Taxonomies

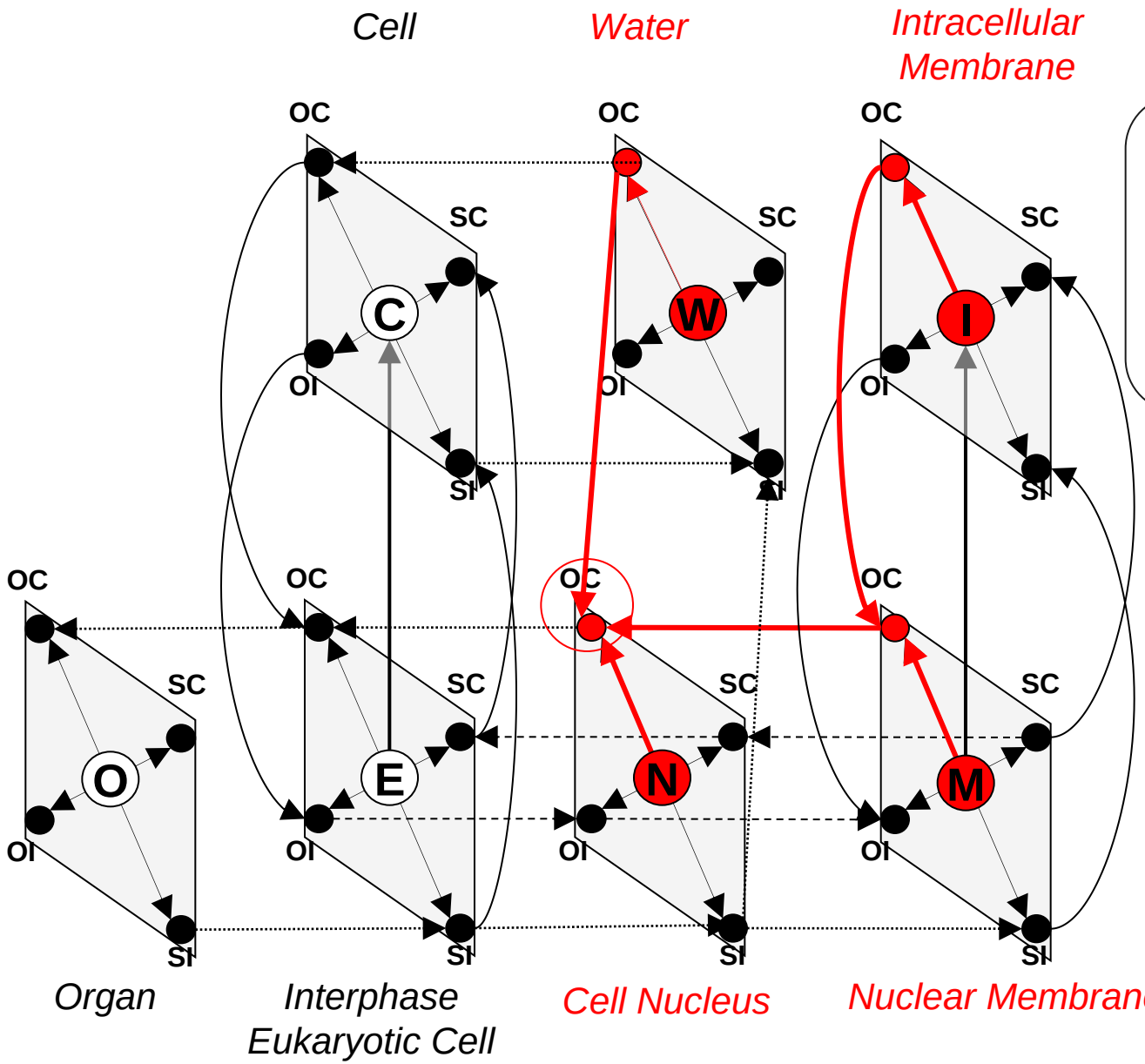


Specific Includers
of Cell Nucleus

Interphase Eukaryotic Cell, Tissue, Organ, Blood, Organism (...)

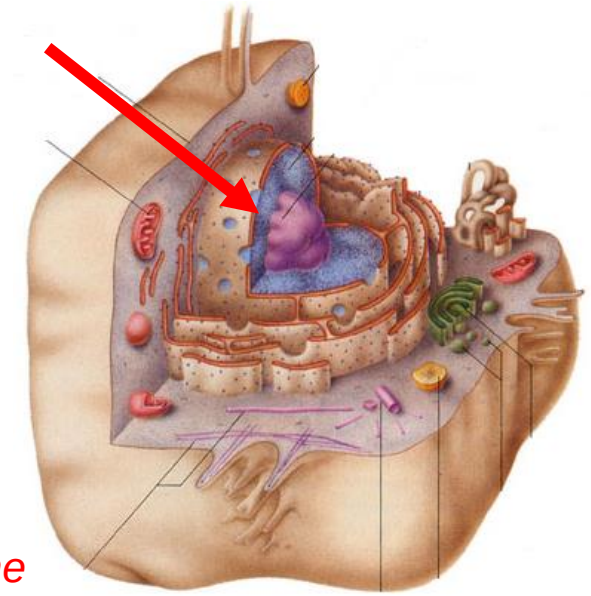


Extended Taxonomies

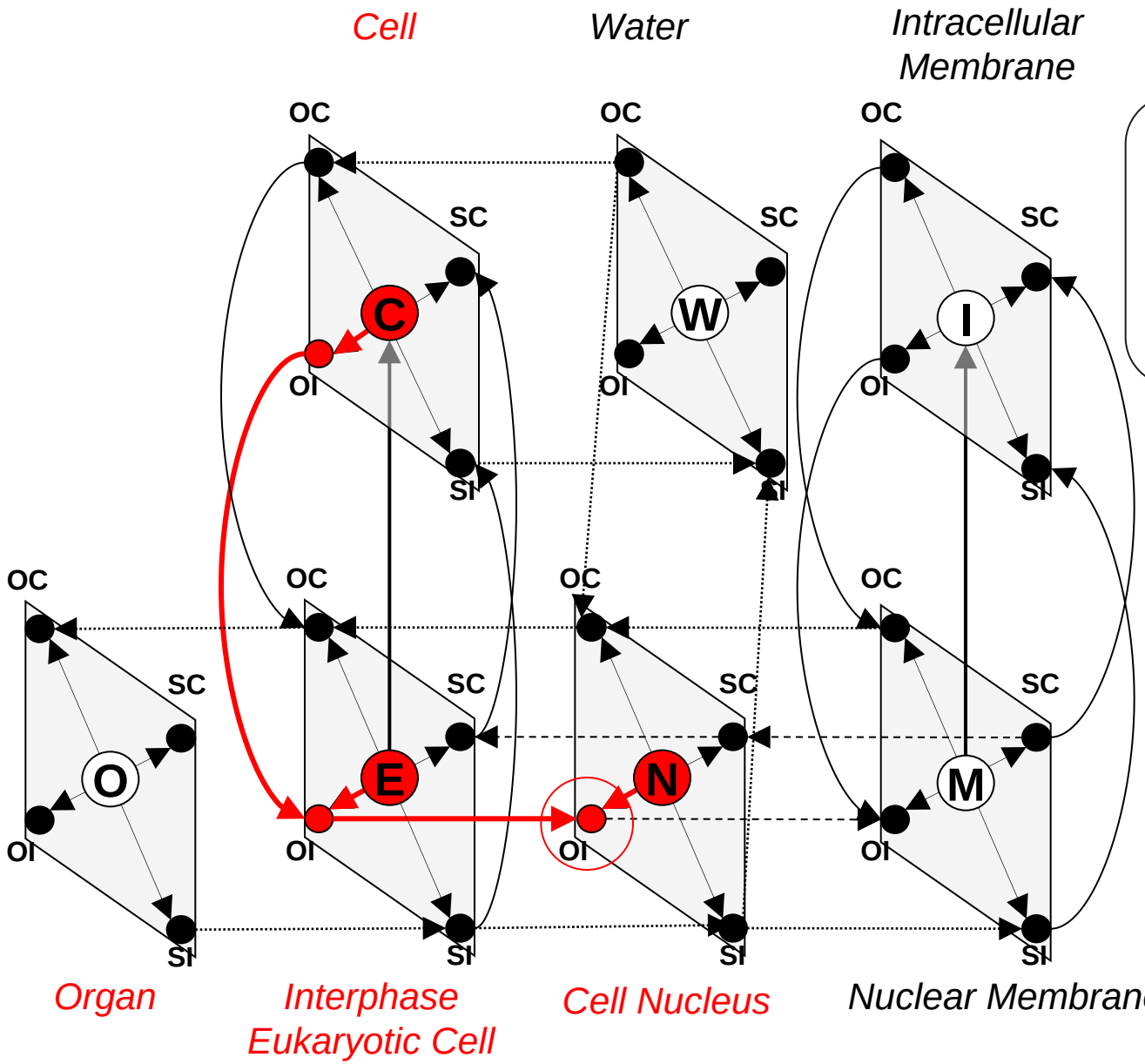


Obligatory Containees of Cell Nucleus

Cell Nucleus
Nuclear Membrane
Intracellular Membrane
Water

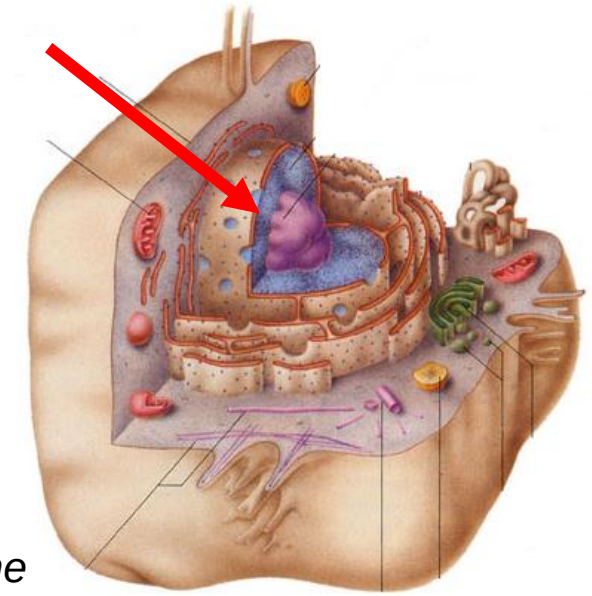


Extended Taxonomies



Obligatory Includers of Cell Nucleus

Interphase Eukaryotic Cell, Cell



Conclusion

- Capturing mereotopological basics in biomedical ontologies: Two recommendations:
  - Create consensus by conflating part-whole and locative relations to one base relation (containee-of / includes)
  - Eliminate ambiguity by explicitly introducing class level relations with a precise semantics
- Using extended taxonomies: Improve reasoning in large knowledge bases
- To do: express other mereotopological relations (overlap, disconnectedness) in terms of class-level predicates

www.imbi.uni-freiburg.de/medinf