# Ontology engineering approach to support process of model and data integration

Linsey Koo, Edlira Kalemi, Franjo Cecelja

1

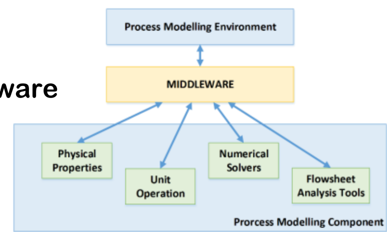
## Background

- Increasing numbers of modelling methods using heterogeneous tools
  - Models remain **implicit** to whom built them
  - Limits the potential of **reusability**
  - Time consuming & **redundant work**
- Lack of complete **libraries** of bio-chemical processes
- To retain the valuable models and data in biorefining
  - Systematic approach to identifying, capturing, retrieving, sharing and effectively reusing these models and data
- Need to build new models or integrate existing ones

2

# Existing Framework: CAPE-OPEN

- Concept
  - Standardisation of interfaces to enable interoperability between simulator software components from different sources
  - Integration of models and tools to take advantage of characteristics that vary between simulation environments
- **Middleware Architecture**
- Widely **supported** by existing software




- Rely on user intervention
- Limits possibility of reusing existing models

18AIChE

UNIVERSITY OF SURREY



Message from **CO-LaN** - the CAPE-OPEN Laboratories Network:
'… speaking <u>**negatively**</u> on CAPE-OPEN …'

CO-LaN

We **benefit** from CAPE-OPEN achievements **greatly** and only propose to build upon it:

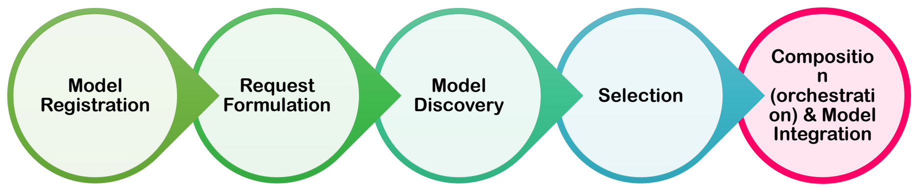
- Introducing higher level of **flexibility** in model/data integration by introducing knowledge modelling - ontology (repository)
- To assist user decision by introducing **partial matching** and relevance **ranking**
- Benefitting from well-proven (service) **integration** technologies such as SWS

18AIChE

UNIVERSITY OF SURREY

# Ontology Engineering Approach

- Build on previous work – **CAPE-OPEN Framework**
- **Semantic** model/data integration




- Web Based **Repository**
- Parsing Ontology

- User specification
- Input specification

- Input-Output Matching
- **Partial Matching**

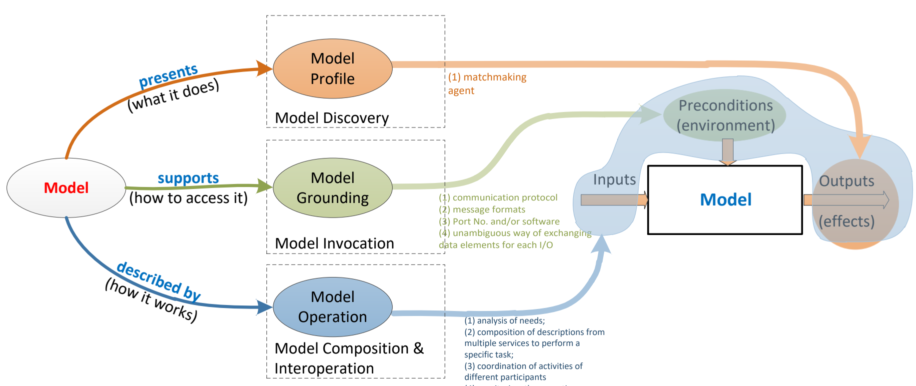
- Match **ranking**
- Best match to the request

- Ontology Web Language for Services Framework



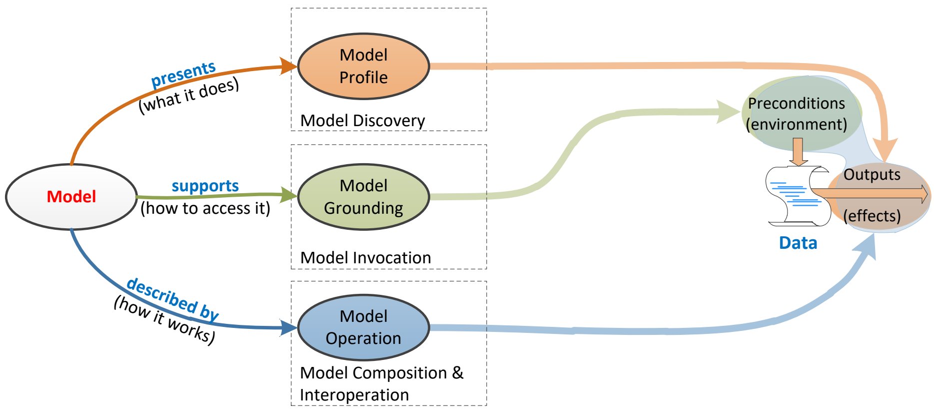
# Model Representation/Description

**Generic** model representation with **semantic** description using ontologies: the **semantic web service (SWS) description** in **OWL-S** framework



# Data Representation/Description

Data uses **similar** principle to model representation and SWS description:




18AIChE

UNIVERSITY OF SURREY

7

# InterCAPEmodel Ontology

- **Knowledge Representation in the domain of biorefining**

  - **Expand knowledge of process models & data**
  - **Provide a classification & characterisation of models & data**
  - **Derive implicit information through the analysis of explicit knowledge**
- **Classification of Models**
  - **Functionality, (biorefining) Platform, Characteristics, Input & Output**
- **Inputs & Outputs of Models**
  - **No. of Inputs & Outputs**
  - **Type of Inputs & Outputs (i.e. material, energy, etc.)**
  - **Parameters of Inputs & Outputs**

18AIChE

UNIVERSITY OF SURREY

8

# InterCAPEmodel Ontology

- **Classification** of Models

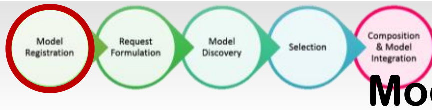
| ModelByFunctionality | ModelByBiorefiningPlatform | ModelByCharacteristics | ModelByInputType | ModelByOutputType |
|---|---|---|---|---|
| FunctionalityForEquipmentLevel | SugarPlatform | ModellingScope | MaterialInput | MaterialOutput |
| Reaction | C5SugarPlatform | ModellingAndSimulation | FeedstockByType | ProductType |
| BiochemicalReaction | C6SugarPlatform | ProcessSynthesisAndDesign | VirginResource | BiochemicalProduct |
| ThermochemicalReaction | Bio-OilPlatform | PlanningAndScheduling | WasteResource | Biofuel |
| ChemicalReaction | BiogasPlatform | ProcessMonitoringAndControl | FeedstockBySource | Biomaterial |
| HeatExchange | SyngasPlatform | IntegratedApproach | EnergyCrop | ProductByIndustrySector |
| Heating | HydrogenlPlatform | ComplexityOfModel | PrimaryResidue | CommunicationSector |
| Cooling | OrganicJuicePlatform | Rigorous | Wastes | EnvironmentSector |
| PressureChanger | PyrolyticLiquidPlatform | Shortcut | ChemicalComponent | HealthAndHygieneSector |
| IncreaseInPressure | LigninPlatform | Conceptual | EnergyInput | HousingSector |
| DecreaseInPressure | ElectricityAndHeatPlatform | NatureOfModel | Steam | IndustrialSector |
| Mixing | | Mechanistic | Heat | RecreationSector |
| Splitting | | Empirical | Electricity | SafeFoodSupplySector |
| Separation | | EquationFormOfModel | | TextileSector |
| HomogeneousSeparation | | Dynamic | | TransportationSector |
| HeterogeneousSeparation | | SteadyState | | ChemicalComponent |
| FunctionalityForProcessLevel | | ScaleOfModel | | EnergyOutput |
| PretreatmentProcess | | IndividualOperatingUnit | | Steam |
| SizeReduction | | FunctionalProcess | | Heat |
| Densification | | ProcessPlant | | Electricity |
| Physico-chemicalProcess | | SupplyChain | | |
| ChemicalProcess | | ModellingType | | |
| BiologicalProcess | | SequentialModularApproach | | |
| Densification | | EquationOrientedApproach | | |
| ConversionProcess | | StatisticalModelling | | |
| BiochemicalConversion | | BlockDiagramOriented (ForControl) | | |
| ThermochemicalConversion | | ComputationalFluidDynamics | | |



# InterCAPEmodel Ontology

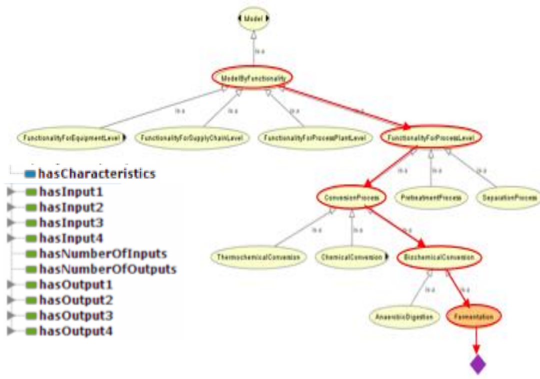
- **Inputs** & **Outputs** of Model (as data/object properties and as ontology entities)

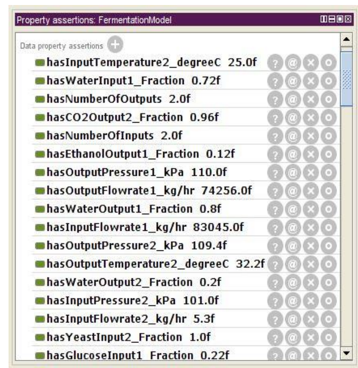
| Input | Output | Description |
|---|---|---|
| HasInput | hasOutput | Define direction of flow |
| hasNumberofInputs | hasNumberofOutputs | Define number of ports required for the model by type of inputs and outputs |
| hasNumberofMaterialInputs | hasNumberofMaterialOutputs | |
| hasNumberofEnergyInputs | hasNumberofEnergyOutputs | |
| hasMaterialInputs | hasMaterialOutputs | Define value of material composition for each stream |
| hasMaterialInput1 | hasMaterialOutput1 | |
| hasMaterialInput2 | hasMaterialOutput2 | |
| hasMaterialInput3 | hasMaterialOutput3 | |
| : | : | |
| hasEnergyInputs | hasEnergyOutputs | Define value of energy composition for each stream |
| hasEnergyInput1 | hasEnergyOutput1 | |
| hasEnergyInput2 | hasEnergyOutput2 | |
| hasEnergyInput3 | hasEnergyOutput3 | |
| : | : | |
| hasInputParameters | hasOutputParameters | Define parameters of input/output and set values for each parameter in SI units |
| hasInputFlowrate | hasOutputFlowrate | |
| hasMassFlowrate | hasMassFlowrate | |
| hasMolarFlowrate | hasMolarFlowrate | |
| hasVolumetricFlowrate | hasVolumetricFlowrate | |
| hasPhaseFraction | hasPhaseFraction | |
| hasTemperature | hasTemperature | |
| hasPressure | hasPressure | |
| : | : | |




# Model/Data Registration

- **Registration** path is created by **parsing** ontology – subsumption and object properties, **inferred ontology**




- Incorporate **explicit knowledge** of each model/data in public repository(ies)
- Increase **share & reuse** of existing models/datasets

18AIChE

UNIVERSITY OF SURREY

11

# Model Reg. by Ontology Parsing

**Fatty Acid Methyl Esters (FAME) Purification Model**

Aspen Plus Software
Restriction (Material): Presence of Glycerol

**Material Input 1**

Compositions
- FAME
- Methanol

Parameters
- Concentration, x
- Flowrate, m
- Temperature, t
- Pressure, p

**Model Description**
- Functionality: Separation
- Equipment Type: Distillation column with 5 equilibrium stages
- Scale: Functional Process Unit
- Model Method: NRTL
- Model Type: Simulation
- Purity of product: 99.8%
- Number of Input Data:1
- Number of Output Data: 2

**Material Output 1**

Composition
- FAME

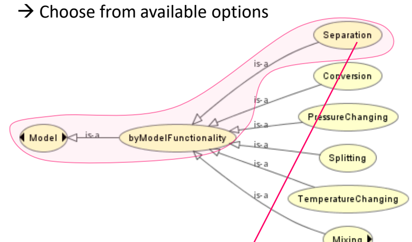
Parameters
- x, m, t, p

**Material Output 2**

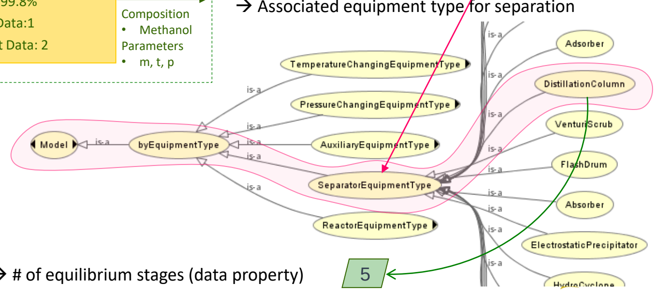
Composition
- Methanol

Parameters
- m, t, p

Q) Functionality of the model?
→ Choose from available options




→ Associated equipment type for separation

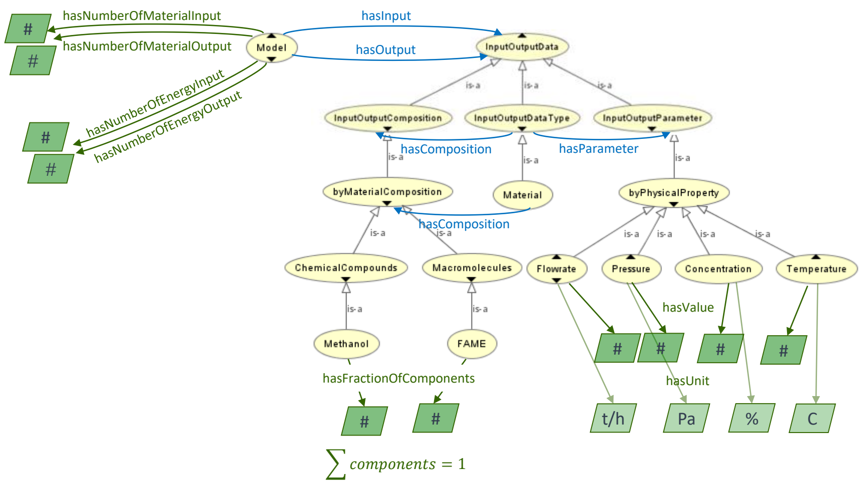



→ # of equilibrium stages (data property)

18AIChE

12

# Registration of Model's Input & Output

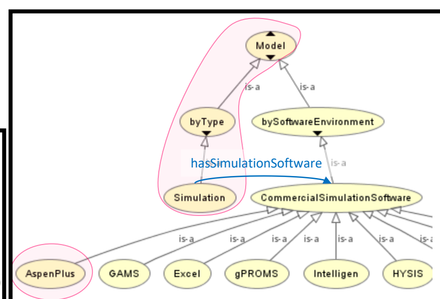



Trokanas, N., F. Cecelja and T. Raafat (2014). ***Semantic Input/Output Matching for Waste Processing in Industrial Symbiosis***, *Computers & Chemical Engineering*, 66, pp 259 - 268..

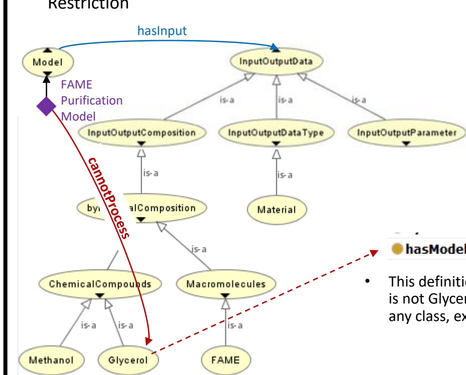
13

# Precondition & Restriction

Precondition → modelling software → AspenPlus



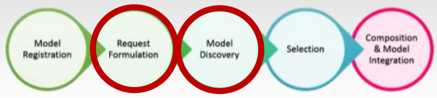

Restriction




**hasModelInputs some (not (Glycerol))**

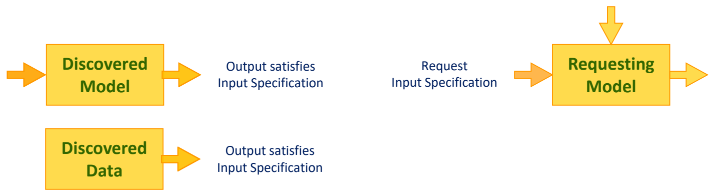
- This definition states FAME Purification model is a model with an input that is not Glycerol. Thus any instance of this class must have any inputs that of any class, except Glycerol.

14




# Model Discovery

- Process of **finding** a **candidate** model/dataset which is to be integrated with the requesting model

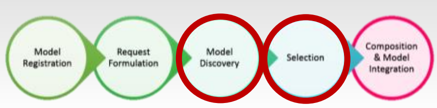



- Fully or partially satisfy the requested **functionality** and the output **properties**
  - **Functionality:** distance measurement matching & similarity measure
  - **Input/Output property:** data property matching & similarity measure
- Requested parameter may come from **different models** or data, or more than one model or data

18AIChE

UNIVERSITY OF SURREY

15




# Model Selection

- Matching process

**Elimination**
- Process of **reducing redundant matching** to avoid performance deficiency
- **Key components** required for the input of the requesting model is considered as elimination criterion

Still refining elimination process

**Semantic Matching**
- Process of quantifying the **semantic relevance** between the requesting model and models residing in the repository
  - Distance measurement between respective concepts representing tacit knowledge which is measured along the hierarchical and object property relationship in the ontology
  - Property similarity that calculates values of properties that characterise explicit knowledge in the form of vectors

**Performance Ranking**
- Based on the **level of match**, the discovered models that fully or partially satisfy matching conditions are ranked

F. Cecelja, N. Trokanas, T. Raafat *et al.*, **Optimising Environmental Performance of Symbiotic Networks Using Semantics**, *Journal of Computers & Chemical Engineering*, vol. submitted, 2014

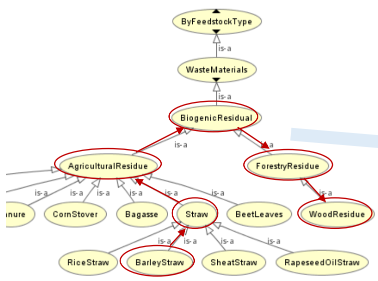
18AIChE

UNIVERSITY OF SURREY

16

## Input/Output Matching

- Graphical Method
  - Functionality & Input/Output Type
  - Process Synthesis Logic




- Property Matching
  - Explicit Knowledge
    - Cosine Similarity

$$h_k^{V,C} = \frac{p_r \cdot p_i}{\|p_r\|\|p_i\|} = \frac{\sum_{i=1}^{n} p_{r,i} \times p_{i,i}}{\sqrt{\sum_{i=1}^{n}(p_{r,i})^2} \times \sqrt{\sum_{i=1}^{n}(p_{i,i})^2}}$$

    - Euclidean Similarity

$$h_k^{V,E} = 1 - \frac{\sqrt{\sum_{i=1}^{n}(p_{r,i} - p_{i,i})^2}}{max\sqrt{\sum_{i=1}^{n}(p_{r,i} - p_{i,i})^2}}$$

    - Aggregated Similarity

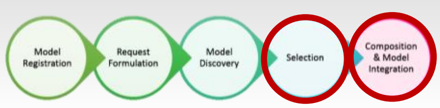
$$h_k^{V} = \frac{h_k^{V,C} + h_k^{V,E}}{2}$$

➢ Partial Matching Allowed

Trokanas, N., F. Cecelja and T. Raafat (2014). ***Semantic Input/Output Matching for Waste Processing in Industrial Symbiosis***, *Computers & Chemical Engineering*, 66, pp 259 - 268..

UNIVERSITY OF SURREY

17




## Model Integration

- **Candidate models and data are ranked based on semantic relevance**
- **Best matches that satisfy the requestor's functionality and output property are proposed – the user makes decision**

- **Enable a (semantic) flexible and user customised model integration fully coordinated by SWS**

18AIChE

UNIVERSITY OF SURREY

18

## Demonstration

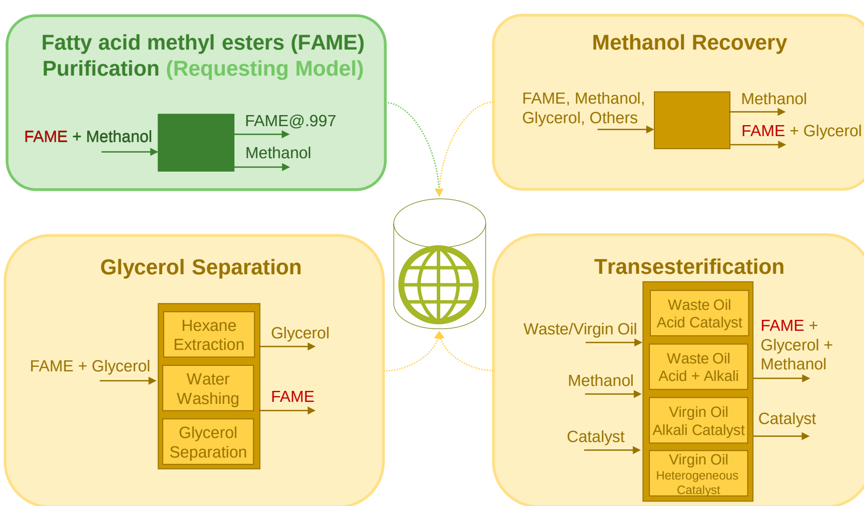
| Process Model | # Input | Required input components | FAME % | # Output | Main Output components | FAME % |
|---|---|---|---|---|---|---|
| FAME Purification | 1 | S1. FAME, Oil | .852 - .969 | 2 | S1. FAME<br>S2. Oil | .997 - .998 |
| Hexane Extraction | 2 | S1. FAME, Glycerol<br>S2. Hexane | .830 - .852 | 2 | S1. FAME<br>S2. Glycerol | .852 – .969 |
| Water Washing | 2 | S1. FAME, Glycerol<br>S2. Water | | | | |
| Glycerol Separation | 1 | S1. FAME, Glycerol | | | | |
| Methanol Recovery | 1 | S1. Methanol, FAME, Glycerol | .719 - .830 | 2 | S1. Methanol (Recycle)<br>S2. FAME, Glycerol | .719 – .830 |
| Transesterification 1<br>Transesterification 2 | 3 | S1. Waste Oil<br>S2. Methanol<br>S3. Catalyst:<br>T1 - $H_2SO_4$<br>T2 - $H_2SO_4$, NaOH | .000 | 2 | S1. FAME, Glycerol, Methanol<br>S2. Catalyst (Recycle) | .719<br><br><br>.830 |
| Transesterification 3<br>Transesterification 4 | 3 | S1. Virgin Oil<br>S2. Methanol<br>S3. Catalyst:<br>T1 - NaOH<br>T2 - $Ca_3La_1$ | | | | <br>.779 |

18AIChE

UNIVERSITY OF SURREY

19

## Demonstration

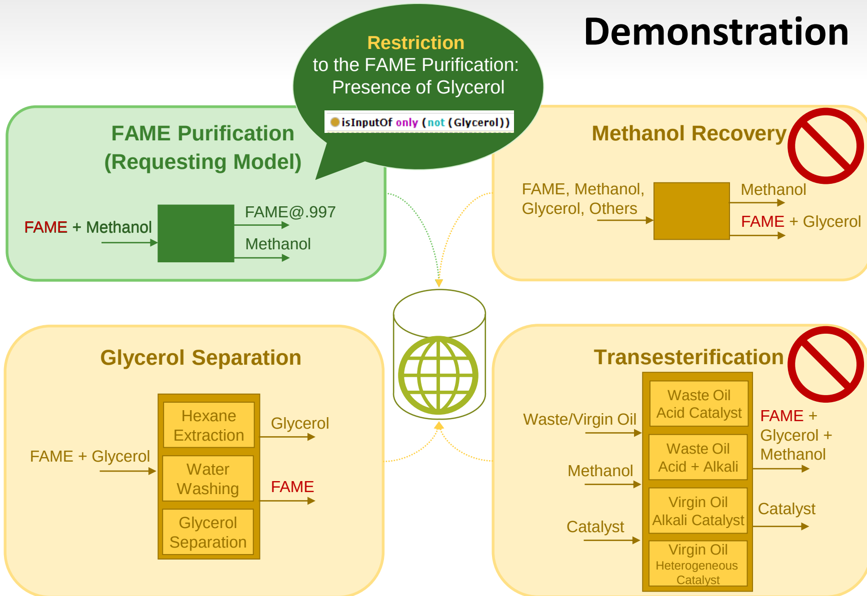



18AIChE

UNIVERSITY OF SURREY

20

# Demonstration

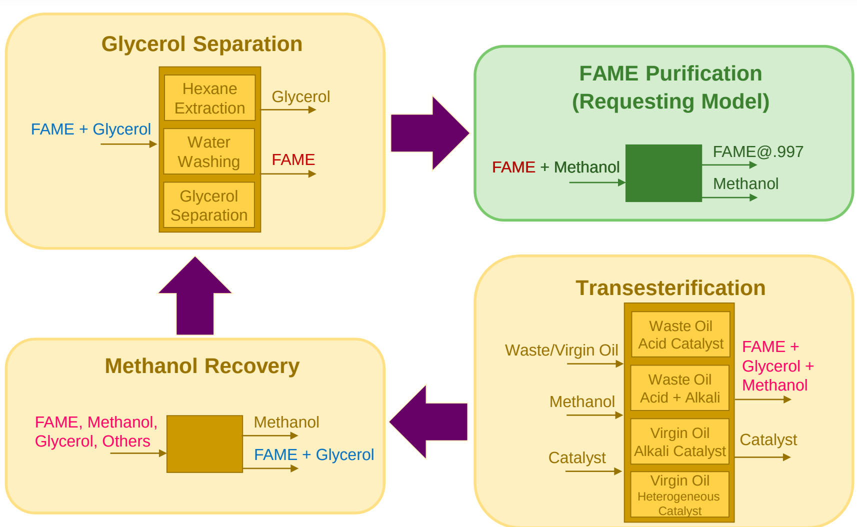



18AIChE

UNIVERSITY OF SURREY

21

# Demonstration: Backward Matching



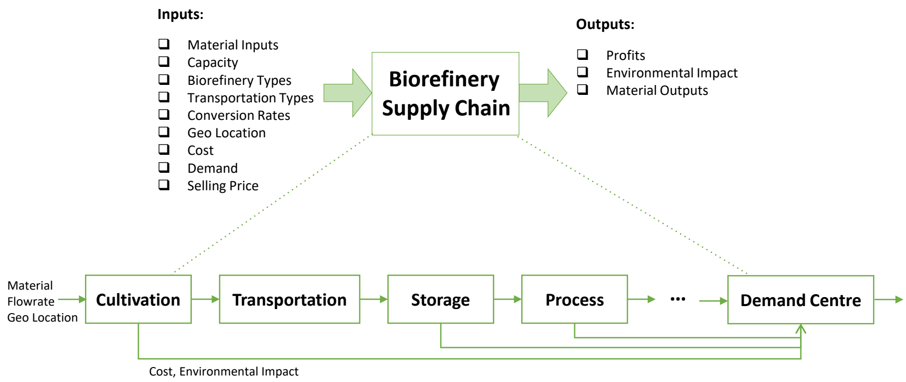

18AIChE

UNIVERSITY OF SURREY

22

# Demonstration

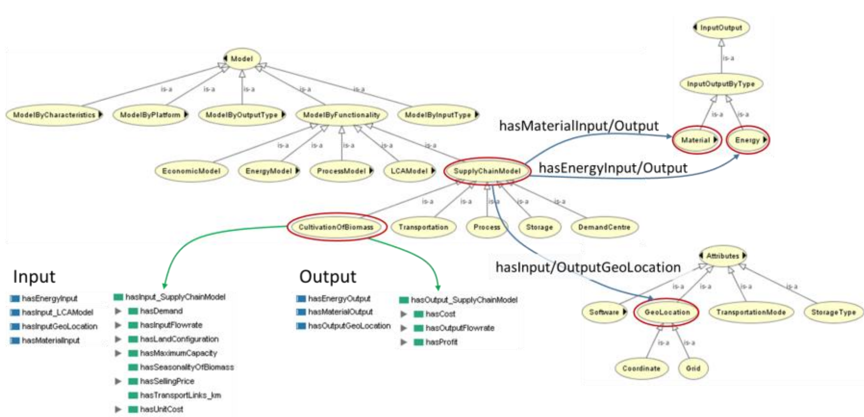
- **Biorefinery supply chain model**




23

# Demonstration

- **Representation of supply chain network using taxonomy, attribute, and relation in InterCAPE Ontology**




24

# Demonstration

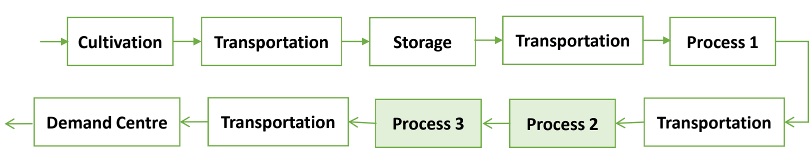
- **Key parameters** of Input/Output for supply chain

| | Input | Output | Environment |
|---|---|---|---|
| Cultivation | ○ Material Input<br>○ Yield of Biomass<br>○ Seasonality of Biomass<br>○ Input Flowrate<br>○ Land Configuration<br>○ Geo. Location<br>○ Unit Cost of Cultivation | ○ Material Output<br>○ Output Flowrate<br>○ Geo. Location<br>○ Cost of Cultivation | ○ Software |
| Transportation | ○ Material Input<br>○ Max. Capacity<br>○ Input Flowrate<br>○ Geo. Location<br>○ Unit Cost of Transport | ○ Material Output<br>○ Output Flowrate<br>○ Geo. Location<br>○ Cost of Transport | ○ Software |
| Storage | ○ Material Input<br>○ Max. Capacity<br>○ Input Flowrate<br>○ Geo. Location<br>○ Unit Cost of Storage | ○ Material Output<br>○ Output Flowrate<br>○ Geo. Location<br>○ Cost of Storage | ○ Software |
| Process | ○ Material Input<br>○ Max. Capacity<br>○ Yield of Product<br>○ Input Flowrate<br>○ Geo. Location<br>○ Unit Cost (CAPEX & OPEX) of Process | ○ Material Output<br>○ Output Flowrate<br>○ Geo. Location<br>○ Cost of Process | ○ Software |
| Demand Centre | ○ Material Input<br>○ Min. Demand<br>○ Selling Price<br>○ Input Flowrate<br>○ Geo. Location | ○ Profit | ○ Software |

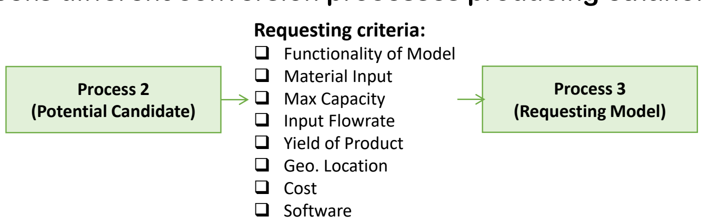


# Case Study

- Demonstrate performance of ontological approach to coordinate interoperability using lignocellulosic biorefining supply chain model that maximises total profit




- Process 3 is a requesting model (separation process) that seeks different conversion processes producing ethanol



# Case Study

- Requirement of Requesting Model

| | Model Functionality for Process | Complexity | Flowrate | Yield |
|---|---|---|---|---|
| Process 3 | Co Fermentation | Shortcut | 50,000 kg/hr | 0.075 |

- List of Models in Repository

| | Model Scale | Model Functionality | Model Functionality | Ethanol | Complexity | Flowrate | Yield | Software | Reference |
|---|---|---|---|---|---|---|---|---|---|
| MODEL 1 | Process | Conversion | C6 Fermentation | Yes | Shortcut | 117,233 | 0.116 | gProms | (Siougkrou et al. 2016) |
| MODEL 2 | Process | Conversion | SSF* | Yes | Shortcut | 449,353 | 0.055 | AspenPlus | (Humbird et al. 2011) |
| MODEL 3 | Process | Conversion | Transesterification | No | Detailed | 1,004 | 0.000 | AspenPlus | (Zhang et al. 2003) |
| MODEL 4 | Process | Conversion | Gasification | Yes | Conceptual | 3,967 | 0.066 | Data | (Wei et al. 2009) |
| MODEL 5 | Process | Conversion | C6 Fermentation | Yes | Detailed | 74,256 | 0.121 | AspenPlus | (AspenPlus 2007) |
| MODEL 6 | Process | Conversion | Gasification | Yes | Conceptual | 1,653 | 0.114 | Data | (Wei et al. 2009) |
| MODEL 7 | Process | Conversion | SSF* | Yes | Conceptual | 10,722 | 0.016 | Data | (Wei et al. 2009) |
| MODEL 8 | Process | Conversion | C6 Fermentation | Yes | Conceptual | 47,191 | 0.075 | AspenPlus | (Siougkrou et al. 2016) |
| MODEL 9 | Process | Conversion | Indirect Gasification | No | Detailed | 6,507 | 0.000 | AspenPlus | (Spath et al. 2005) |

18AIChE

UNIVERSITY OF SURREY

27

# Case Study

- Elimination:
  - No Ethanol & No Cost available

| | Model Scale | Model Functionality | Model Functionality | Ethanol | Complexity | Flowrate | Yield | Software | Reference |
|---|---|---|---|---|---|---|---|---|---|
| MODEL 1 | Process | Conversion | C6 Fermentation | Yes | Shortcut | 117,233 | 0.116 | gProms | (Siougkrou et al. 2016) |
| MODEL 2 | Process | Conversion | SSF* | Yes | Shortcut | 449,353 | 0.055 | AspenPlus | (Humbird et al. 2011) |
| MODEL 3 | Process | Conversion | Transesterification | No | Detailed | 1,004 | 0.000 | AspenPlus | (Zhang et al. 2003) |
| MODEL 4 | Process | Conversion | Gasification | Yes | Conceptual | 3,967 | 0.066 | Data | (Wei et al. 2009) |
| MODEL 5 | Process | Conversion | C6 Fermentation | Yes | Detailed | 74,256 | 0.121 | AspenPlus | (AspenPlus 2007) |
| MODEL 6 | Process | Conversion | Gasification | Yes | Conceptual | 1,653 | 0.114 | Data | (Wei et al. 2009) |
| MODEL 7 | Process | Conversion | SSF* | Yes | Conceptual | 10,722 | 0.016 | Data | (Wei et al. 2009) |
| MODEL 8 | Process | Conversion | C6 Fermentation | Yes | Conceptual | 47,191 | 0.075 | AspenPlus | (Siougkrou et al. 2016) |
| MODEL 9 | Process | Conversion | Indirect Gasification | No | Detailed | 6,507 | 0.000 | AspenPlus | (Spath et al. 2005) |

- Matching & Performance Ranking Result

| | Semantic Similarity | Cosine Similarity | Euclidean Similarity | Property Similarity | Aggregated Similarity |
|---|---|---|---|---|---|
| MODEL 1 | 0.833 | 1.000 | 0.832 | 0.916 | 0.875 |
| MODEL 2 | 0.750 | 1.000 | 0.000 | 0.500 | 0.625 |
| MODEL 5 | 0.667 | 1.000 | 0.939 | 0.970 | 0.818 |
| MODEL 8 | 0.667 | 1.000 | 0.993 | 0.996 | 0.832 |

18AIChE

UNIVERSITY OF SURREY

28

# Acknowledgement