Modeling of collaboration archetypes in digital market places

Dr. Paola Grosso
(on behalf of Lu Zhang)
What are multi-party collaboration relationships?

- Determined by the **trust** among participating parties
- Provide information about the rules of how data and compute are shared and used
- Defined from both DMP and application perspective
  - **DMP archetype** collaboration model
    - *Project DL4LD* defined multiple collaboration archetypes
    - A DMP may support a subset of archetypes
  - **Application collaboration model**
    - The application/user collaboration request comes from individual customer
    - **Hard and Soft Requests**
      - **Hard Request**: not negotiable and must be fulfilled
      - **Soft Request**: can be adjusted to better fit any existing DMP archetyp
Parties in the DMP may collaborate across a number of scopes: data, algorithm, result.

In each scope, a number, which we call collaboration level, describes the concrete approach of asset sharing between any source and target:

- E.g. Filetransfer or Remote filesystem mount
- This model is generic, more scopes and collaboration level could be extended

How to model multi-party collaboration relationships generically?
How to match application requests to DMP archetypes?

- Map any collaboration model as a point in discrete space – relative distance
  - Pre-processing block for more commensurate comparison
    - Reduce the influence of how we assign those participating parties
    - Aim to find an optimum fitness between two collaboration models

- The *closeness* of application request and the supported DMP archetypes can be identified
How to match application requests to DMP archetypes?

Filtering: Only consider the DMP candidates that fully satisfy the hard request of customer request!
Evaluation metrics of a DMP

**Motivation:**
- Provide a-priori information for DMP providers and potential customers
- Allow for an intelligent selection of DMPs

**Evaluation metrics**

- **Coverage**
  - how well the overall application requests can be satisfied by supported archetypes of a DMP

- **DMP Extensibility**
  - richness a DMP can achieve by decomposing and composing current archetypes

- **App Extensibility**
  - how elastic of an application request can achieve a perfect match with a given DMP

- **Precision**
  - how well the supported archetypes fits an application request

- **Flexibility**
  - how easily the application request from potential customer could be satisfied
Coverage

• A higher *coverage is* achieved by lowering customer satisfaction degree
  – Pre-define a tolerant distance \( D_A \)
  – Covered area of each archetype is modeled effectively as a sphere with radius \( D_A \)
  – Total covered area is of a DMP is the union of individual covered area

• Coverage of a DMP with under a fix \( D_A \) is calculated as

• An optimization algorithm for coverage calculation is designed for complexity reduction
How to use the proposed metrics for intelligent selection?

- Normally there are multiple DMP candidate and each DMP may support different sets of archetypes
- Evaluation metrics could be computed for each DMP with a specific application request
- An optimal DMP could be recommended to a potential customer for a given application

Algorithm 2: Metrics validation with a specific collaboration request

1. Input collaboration request → cr
2. Sort DMP candidates on coverage in descending order → DMP\textsubscript{rank}
3. for dmp\textsubscript{i} ∈ DMP\textsubscript{rank} do
4. if precision(dmp\textsubscript{i}, cr) = 1 then
5. dmp\textsubscript{i} → dmp\textsubscript{opt}
6. go to output
7. end if
8. end for
9. if flexibility(cr) > 0 then
10. if ∃E\textsubscript{A} ≥ 0 then
11. Select dmp\textsubscript{i} with maximum E\textsubscript{A}
12. dmp\textsubscript{i} → dmp\textsubscript{opt}
13. go to output
14. end if
15. end if
16. Extend DMP\textsubscript{rank} by primitive composition → DMP’\textsubscript{e}
17. for dmp\textsubscript{i} ∈ DMP’\textsubscript{e} do
18. if precision(dmp\textsubscript{i}, cr) = 1 then
19. dmp\textsubscript{i} → dmp\textsubscript{opt}
20. go to output
21. end if
22. end for
23. output:
24. Return dmp\textsubscript{opt}

Firstly, select the DMP candidate containing an exactly matched archetype

Then, select the DMP candidate containing an exactly matched archetype with minimum modification of soft requirements in the application request

Lastly, select the DMP candidate containing an exactly matched archetype by composing and decomposing current archetypes
Intelligent selection of DMPs

Hard Request: Air France and KLM trusts Dell in data scope

Soft Request: Air France prefer direct data transfer and KLM prefers direct mounting

<table>
<thead>
<tr>
<th>DMP</th>
<th>Supported Archetype Trust Models</th>
</tr>
</thead>
<tbody>
<tr>
<td>DMP₁</td>
<td>{1, 2, 3, 4, 7}</td>
</tr>
<tr>
<td>DMP₂</td>
<td>{1, 2, 3, 5, 7}</td>
</tr>
<tr>
<td>DMP₃</td>
<td>{1, 2, 3, 5, 6}</td>
</tr>
<tr>
<td>DMP₄</td>
<td>{1, 3, 4, 5, 7}</td>
</tr>
<tr>
<td>DMP₅</td>
<td>{2, 3, 4, 6, 7}</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>DMP₁</th>
<th>DMP₂</th>
<th>DMP₃</th>
<th>DMP₄</th>
<th>DMP₅</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coverage (1e(-12))</td>
<td>4.29</td>
<td>4.28</td>
<td>4.26</td>
<td>3.69</td>
<td>3.65</td>
</tr>
<tr>
<td>Precision</td>
<td>0.83</td>
<td>0.83</td>
<td>0.83</td>
<td>0.83</td>
<td>-0.67</td>
</tr>
<tr>
<td>Flexibility</td>
<td>0.06</td>
<td>0.06</td>
<td>0.06</td>
<td>0.06</td>
<td>0.06</td>
</tr>
<tr>
<td>Application extensibility</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>-∞</td>
</tr>
</tbody>
</table>

Archetype I in DMP₁ is the best matched candidate for the app request
DMP archetypes in DL4LD

Archetype I

Archetype II

Archetype III

Archetype IV

Archetype V

Archetype VI

Archetype VII