Complexity and Diversity Management of Application Landscapes

29.06.2016, Prof. Dr. Florian Matthes

Software Engineering für betriebliche Informationssysteme (sebis)
Fakultät für Informatik
Technische Universität München

wwwmatthes.in.tum.de
1. Complexity and Diversity Management of Application Landscapes
Motivation

Business challenges
- Increasing uncertainty in the business environment → organizational agility
- Balance of exploration and exploitation → ambidexterity

IT challenges
- Uncontrolled growth of application landscapes → business silos
- Shrinking IT budgets → cost cutting in IT operations through standardization

Challenge: Manage diversity to decrease costs but enable innovation.
**Motivation**

**Exploration**

- Competitive Advantage
- New Digital Options

**Exploitation**

- Cost savings
- Economies of Scale

**Ambidexterity**

**AL Diversity (Diversification)**

**AL Diversity (Standardization)**

**Challenge:** How to manage application landscape (AL) diversity to decrease costs but enable innovation and co-evolution with the environment?

Theoretical foundation: The resource-based view

- Value-creating strategy
- Maintain a combination of resources that cannot be possessed or built up in a similar manner by competitors
- Information-based organizational processes
- Firm-specific
- Improve the productivity of resources
- Valuable
- Rare
- In-imitable
- Non-substitutable

Theoretical foundation: The value of digital options

Resource-based view of the firm (Ferrier et al. 1999)

→ Firms that possess a more complex base of resources and capabilities will be in an advantageous position to launch competitive actions.

Capability-Building and Entrepreneurial Action (Sambamurthy et al. 2003)

<table>
<thead>
<tr>
<th>Capability-Building Process</th>
<th>Entrepreneurial Action Process</th>
</tr>
</thead>
<tbody>
<tr>
<td>IT Competence</td>
<td></td>
</tr>
<tr>
<td>• Investment scale</td>
<td></td>
</tr>
<tr>
<td>• IT capabilities</td>
<td></td>
</tr>
<tr>
<td>Digital Options</td>
<td>Agility</td>
</tr>
<tr>
<td>• Process reach</td>
<td>• Customer agility</td>
</tr>
<tr>
<td>• Process richness</td>
<td>• Partnering agility</td>
</tr>
<tr>
<td>• Knowledge reach</td>
<td>• Operational agility</td>
</tr>
<tr>
<td>• Knowledge richness</td>
<td></td>
</tr>
</tbody>
</table>

Can be described by number and their diversity

A set of IT-enabled capabilities in the form of digitized enterprise work processes and knowledge systems.


### Epistemic goal

**Develop a deeper understanding of the concept diversity in the context of application landscapes.**

1. **Conceptualization**: What aspects contribute to the diversity of AL?
2. **Perception in practice**: What benefits and drawbacks relate to AL diversity?

### Design goal

**Design metrics for supporting decisions regarding the diversity of application landscapes.**

3. **Quantification**: What metrics quantify different aspects of diversity and which data is needed for their calculation?
4. **Software support**: How can diversity quantification be integrated in an EAM tool?
5. **Steering mechanisms**: How to influence AL diversity?

---

Research process (design science)


Complexity of Application Landscapes

Focus group results

AL Complexity

Skill dependency

Error rate

Agility

Shadow IT

Local Decisions

Business Complexity

Technological progress

Legal requirements

Tactical solutions

Cost

Skill dependency

Error rate

Agility

Shadow IT

Local Decisions

Business Complexity

Technological progress

Legal requirements

Tactical solutions

Diversity of Application Landscapes

AL Diversity

- Error rate: 84%
- Skill dependency: 96%
- Local Decisions: 88%
- Business Complexity: 96%
- Legal requirements: 79%
- Tactical solutions: 94%
- Technological progress: 96%
- Agilit: 60%
- Shadow IT: 78%
- Cost: +
- Local Decisions: +
- Business Complexity: +
- Legal requirements: +
- Tactical solutions: +
- Technological progress: +
- Skill dependency: +
- Error rate: +
- Shadow IT: -

n = 43
Why companies decrease their application landscape diversity

- Cost savings: 86% (only 12% measure cost savings)
- Preservation of manageability: 67%
- Implementation of security guidelines: 47%

Preliminary results for diversity effects

- AL diversity increases costs for IT operations: 97%
- AL diversity decreases time-to-market of new products: 94%
- AL diversity decreases technology dependence: 83%
- AL diversity is required for innovation: 92%
- AL diversity enables better business solutions: 74%

How companies influence application landscape diversity

- Defining standard products for new developments: 86%
- Definition of technical platforms: 80%
- Concessions for non-standard products: 71%
- Actively increase diversity: 44% (Multi-product strategies and subcompanies)
Research Design

Identify and describe different notions of complexity

- Literature review on EA complexity metrics
- Focus group with experienced practitioners
- Metric calculation with real-world data
- Qualitative evaluation with experts

3 approaches 6 patterns 4 use cases benefits & drawbacks

Schneider, Alexander W.; Reschenhofer, Thomas; Schuetz, Alexander; Matthes, Florian (2015): Empirical Results for Application Landscape Complexity. In: 48th Hawaii International Conference on System Sciences (HICSS). Kauai, HI, USA.
Complexity notions differ in terms of:

- the role of the observer
- the role of time
- the role of measures
- the role of statistics

Data collection

- 6 companies (Financial services and Automotive from Germany and Switzerland)
- More than 20 metrics found

Identified metric patterns

<table>
<thead>
<tr>
<th># Business Functions</th>
<th># Applications</th>
<th># Information Flows</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td># Infrastructure Components</td>
<td>Functional Redundancy</td>
<td>Standard Conformity</td>
</tr>
<tr>
<td>4</td>
<td>6</td>
<td>4</td>
</tr>
</tbody>
</table>

Usually calculated on both the application and the domain level

Schneider, Alexander W.; Reschenhofer, Thomas; Schuetz, Alexander; Matthes, Florian (2015): Empirical Results for Application Landscape Complexity. In: 48th Hawaii International Conference on System Sciences (HICSS). Kauai, HI, USA.
A generic information model for complexity measurements

Information model

Model derivation

- Summary of all data necessary to calculate the 3 types of metrics
- Regards also actually available data in industry

Schneider, Alexander W.; Reschenhofer, Thomas; Schuetz, Alexander; Matthes, Florian (2015): Empirical Results for Application Landscape Complexity. In: 48th Hawaii International Conference on System Sciences (HICSS). Kauai, HI, USA.
Data quality and collection effort

Case 1
- Financials (Switzerland)
- >10,000 employees

Case 2
- Financials (Germany)
- > 8,000 employees

Case 3
- Financials (Germany)
- > 52,000 employees

Case 4
- Financials (Germany)
- > 14,000 employees

<table>
<thead>
<tr>
<th>Concept</th>
<th>C1</th>
<th>C2</th>
<th>C3</th>
<th>C4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Business application</td>
<td>432</td>
<td>247</td>
<td>1898</td>
<td>234</td>
</tr>
<tr>
<td>- customization level</td>
<td>100 %</td>
<td>89 %</td>
<td>100 %</td>
<td>100 %</td>
</tr>
<tr>
<td>- function points</td>
<td>–</td>
<td>99 %</td>
<td>24 %</td>
<td>98 %</td>
</tr>
<tr>
<td>- programming languages</td>
<td>–</td>
<td>45 %</td>
<td>–</td>
<td>72 %</td>
</tr>
<tr>
<td>- implements (Use Case)</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>- supports (Business function)</td>
<td>–</td>
<td>97 %</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>- belongs to (Functional domain)</td>
<td>100 %</td>
<td>99 %</td>
<td>87 %</td>
<td>100 %</td>
</tr>
<tr>
<td>- uses (Operating system)</td>
<td>–</td>
<td>67 %</td>
<td>90 %</td>
<td>83 %</td>
</tr>
<tr>
<td>- uses (Data base system)</td>
<td>–</td>
<td>48 %</td>
<td>60 %</td>
<td>67 %</td>
</tr>
<tr>
<td>Business function</td>
<td>–</td>
<td>580</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Functional domain</td>
<td>13</td>
<td>16</td>
<td>33</td>
<td>41</td>
</tr>
<tr>
<td>Infrastructure component (Operating system)</td>
<td>–</td>
<td>9</td>
<td>15</td>
<td>12</td>
</tr>
<tr>
<td>- type</td>
<td>–</td>
<td>–</td>
<td>100 %</td>
<td>100 %</td>
</tr>
<tr>
<td>Infrastructure component (Data base system)</td>
<td>–</td>
<td>10</td>
<td>8</td>
<td>16</td>
</tr>
<tr>
<td>- type</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Information flow</td>
<td>539</td>
<td>827</td>
<td>8252</td>
<td>1214</td>
</tr>
<tr>
<td>- protocol</td>
<td>100 %</td>
<td>–</td>
<td>100 %</td>
<td>100 %</td>
</tr>
<tr>
<td>- from (Business application)</td>
<td>100 %</td>
<td>100 %</td>
<td>100 %</td>
<td>100 %</td>
</tr>
<tr>
<td>- to (Business application)</td>
<td>100 %</td>
<td>100 %</td>
<td>100 %</td>
<td>100 %</td>
</tr>
</tbody>
</table>

Schneider, Alexander W.; Reschenhofer, Thomas; Schuetz, Alexander; Matthes, Florian (2015): Empirical Results for Application Landscape Complexity. In: 48th Hawaii International Conference on System Sciences (HICSS). Kauai, HI, USA.
### Exemplary Results: Classification of Applications

<table>
<thead>
<tr>
<th>Class</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Core</td>
<td>38.24%</td>
<td>55.47%</td>
<td>44.89%</td>
<td>36.75%</td>
</tr>
<tr>
<td>Control</td>
<td>14.12%</td>
<td>7.29%</td>
<td>7.11%</td>
<td>17.09%</td>
</tr>
<tr>
<td>Shared</td>
<td>8.24%</td>
<td>10.53%</td>
<td>13.33%</td>
<td>2.99%</td>
</tr>
<tr>
<td>Peripherie</td>
<td>39.41%</td>
<td>26.72%</td>
<td>34.67%</td>
<td>34.19%</td>
</tr>
<tr>
<td>Propagation cost</td>
<td>24.39%</td>
<td>41.44%</td>
<td>30.28%</td>
<td>21.40%</td>
</tr>
</tbody>
</table>

Changes of **Core** or **Shared** Applications might have unexpected impacts.

Schneider, Alexander W.; Reschenhofer, Thomas; Schuetz, Alexander; Matthes, Florian (2015): Empirical Results for Application Landscape Complexity. In: 48th Hawaii International Conference on System Sciences (HICSS). Kauai, HI, USA.
Results: Heterogeneity Metrics

Calculated metrics (selection)
- Information flow protocol (Application & Domain)
- Customization level (Domain)
- Operating system (Domain)
- Database system (Domain/Landscape) \[C_1: -; C_2: 5.2; C_3: 4.7; C_4: 3.1\]
- Programming language (Domain)
- Coupled domains (Domain)

Identified cluster (correlation)
- Heterogeneity of Customization level, Database system & Operating system

No single metric correlates with the classification metric.

Schneider, Alexander W.; Reschenhofer, Thomas; Schuetz, Alexander; Matthes, Florian (2015): Empirical Results for Application Landscape Complexity. In: 48th Hawaii International Conference on System Sciences (HICSS). Kauai, HI, USA.
### Results: Industry Metrics (selection)

#### Application level

<table>
<thead>
<tr>
<th></th>
<th>C1</th>
<th>C2</th>
<th>C3</th>
<th>C4</th>
</tr>
</thead>
<tbody>
<tr>
<td>#InformationFlows</td>
<td>[0;108], 2.5</td>
<td>[0;70], 6.7</td>
<td>[0;319], 8.7</td>
<td>[0;122], 10.4</td>
</tr>
<tr>
<td>#Infrastr. Components</td>
<td>-</td>
<td>[0;20], 4.7</td>
<td>[0;22], 0.4</td>
<td>[0;13], 3.9</td>
</tr>
<tr>
<td>#FunctionPoints</td>
<td>-</td>
<td>[1;34], 6.7</td>
<td>[1;56], 7.3</td>
<td>[10;400], 178.9</td>
</tr>
</tbody>
</table>

#### Domain level

<table>
<thead>
<tr>
<th></th>
<th>C1</th>
<th>C2</th>
<th>C3</th>
<th>C4</th>
</tr>
</thead>
<tbody>
<tr>
<td>#Applications</td>
<td>[6;28], 14</td>
<td>[1;53], 16</td>
<td>[1;194], 50</td>
<td>[6;88], 24.8</td>
</tr>
<tr>
<td>#InformationFlows</td>
<td>[10;195], 83</td>
<td>[1;390], 110</td>
<td>[0;1316], 491</td>
<td>[23;580], 281</td>
</tr>
<tr>
<td>#StandardConformity</td>
<td>[1.4;3.9], 3</td>
<td>[1.4;5], 2.7</td>
<td>[1;3.5], 2.5</td>
<td>[2;3.3], 2.5</td>
</tr>
</tbody>
</table>

Schneider, Alexander W.; Reschenhofer, Thomas; Schuetz, Alexander; Matthes, Florian (2015): Empirical Results for Application Landscape Complexity. In: 48th Hawaii International Conference on System Sciences (HICSS). Kauai, HI, USA.
Interesting metric: Coupled domain heterogeneity

\[
\forall d_1, d_2 \in D \, d_1 \neq d_2 : [\text{coupled}(d_1, d_2) \iff \exists a_1 \in A_{d_1}, \exists a_2 \in A_{d_2} : (a_1, a_2) \in I ]
\]

A...Set of business applications
D...Set of domains
\(A_d\)...Applications of domain \(d\)
\(I \subseteq A \times A\)...Set of information flows, whereas \((a_1, a_2)\) defines an information flow from application \(a_1\) to application \(a_2\)
Contribution 1: Conceptual diversity framework

Diversity exemplified with Databases

Variation
What are differences between individuals?

Contribution 1: Conceptual diversity framework

Diversity exemplified with Databases

**Variety**
*How many different concepts do we have?*

**Balance**
*How are individuals distributed?*

**Disparity**
*To which degree do our concepts differ?*

**Variation**
*What are differences between individuals?*

---

Contribution 2: A novel metric to quantify balance

Novel balance metric based on ordinal Skewness

- Entropy is a measure of concentration → context-free
- Skewness is applicable to ordinal data → context-sensitive

\[ SO_X = -(k - 1) + 2 \sum_{i=1}^{k-1} P(X \leq x_i) \]

\( X : \) DB versions
\( k = |X| \)

Example: Versions of database systems

\[ SO_X = - (3-1) + 2 \left( \frac{1}{10} + \frac{3}{10} \right) = -1.2 \]
Normalization:
\[ SO_X^* = -1.2 / 3 = -0.4 \]

\[ SO_X = - (4-1) + 2 \left( \frac{2}{10} + \frac{6}{10} + \frac{8}{10} \right) = 0.2 \]
Normalization:
\[ SO_X^* = 0.2 / 4 = 0.05 \]

Novel disparity metric based on Weitzman (1992)

- Distance of concepts is defined by a pair-wise distance function
- Provides decision support for concept extinction

\[
V(S) = \max_{x \in S} \{ V(S \setminus x) + d(x, S \setminus x) \} \quad S: \text{Databases}
\]

Exemplary distance function

<table>
<thead>
<tr>
<th></th>
<th>Oracle</th>
<th>SQL Server</th>
<th>MongoDB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paradigm</td>
<td>SQL</td>
<td>SQL</td>
<td>NoSQL</td>
</tr>
<tr>
<td>License cost</td>
<td>High</td>
<td>Medium</td>
<td>Low</td>
</tr>
</tbody>
</table>

Exemplary Calculation

1. Let S equal the empty set and \( V(S) = 0 \)
2. Chose a concept \( y \) of least distance to a member of \( S \) and add \( y \) to \( S \) (repeat)

\[
V(S) = 3
\]
Selecting the case study method

- Research is at an early, formative stage (Roethlisberger, 1977)
- Research targets a practice-based problem where the experiences of the actors is important and the context of action is critical (Benbasat et al., 1987)

Evaluated artifacts

<table>
<thead>
<tr>
<th>Case</th>
<th>Disparity</th>
<th>Balance</th>
<th>SW Prototype</th>
<th>Industry</th>
<th>FTEs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case 1</td>
<td>●</td>
<td></td>
<td></td>
<td>Banking</td>
<td>&gt;100,000</td>
</tr>
<tr>
<td>Case 2</td>
<td></td>
<td>●</td>
<td>●</td>
<td>Manufacturing</td>
<td>&gt;20,000</td>
</tr>
<tr>
<td>Case 3</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>Insurance</td>
<td>&gt;21,000</td>
</tr>
<tr>
<td>Case 4</td>
<td>●</td>
<td></td>
<td></td>
<td>Automotive</td>
<td>&gt;230,000</td>
</tr>
</tbody>
</table>

Case study conduction

- Duration: Between 1 and 3 months
- Main partners: Enterprise architects, in 3 cases management involvement
- Access to company EA data
**Metric evaluation results**

**Case 3: Balance metric**
- **Context**: Identification of adolescence in Operating Systems
- **Six different products with up to eleven different releases**

<table>
<thead>
<tr>
<th>OS</th>
<th>v1</th>
<th>v2</th>
<th>v3</th>
<th>v4</th>
<th>v5</th>
<th>v6</th>
<th>v7</th>
<th>v8</th>
<th>v9</th>
<th>v10</th>
<th>v11</th>
<th>Balance</th>
</tr>
</thead>
<tbody>
<tr>
<td>OS1</td>
<td>20</td>
<td>235</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.843</td>
</tr>
<tr>
<td>OS2</td>
<td>243</td>
<td>3</td>
<td>77</td>
<td>2</td>
<td>397</td>
<td>5</td>
<td>6</td>
<td>5</td>
<td>1167</td>
<td>5</td>
<td>32</td>
<td>0.019</td>
</tr>
<tr>
<td>OS3</td>
<td>6</td>
<td>14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.400</td>
</tr>
<tr>
<td>OS4</td>
<td>2</td>
<td>1578</td>
<td>1496</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.010</td>
</tr>
</tbody>
</table>

**Result**: Architects and manager were able to easily identify OS with adolescence.

**Case 4: Disparity metric**
- **Context**: Diversification of Web Servers

- **Distance functions**: Source Code Access, JEE Support, Footprint
- **Result**: More structured and repeatable (incremental) approach than ad-hoc
Variety: Number of Different Types

Diversity increases with the number of different types.

Example

- Oracle
- MS SQL Server
- MySQL
- Oracle
- MS SQL Server
- MySQL
- DB2
- IMS
- Sybase
- MongoDB

Typical metric
- Number of types
- Variety(A) = 3, Variety(B) = 7


Schneider, Alexander W.; Reschenhofer, Thomas; Schuetz, Alexander; Matthes, Florian (2015): Empirical Results for Application Landscape Complexity. In: 48th Hawaii International Conference on System Sciences (HICSS). Kauai, HI, USA.
Balance: Instance Distribution among Types

Diversity increases the more balanced the distribution is.

Example

**Typical metric**

- Shannon Entropy

\[ - \sum_{i=1}^{n} p_i \ln(p_i) \]

Disparity: Difference among Concepts

Diversity increases when the distance among types increases.

Example

- Alligator
- Crocodile
- Caiman
- Alligator
- Elephant
- Piranha

Typical metric

- Weitzman metric

\[
\max_{x \in S} \{ V(S \setminus x) + d(x, S \setminus x) \}
\]

A Novel Metric to Quantify Disparity

Limitation of current approaches

- Disparity has not been considered yet in EA

Novel metric based on Weitzman (1992)

- Distance among concepts is defined by a pair-wise distance function
- Calculated by incrementally increasing it by the distance value to the closest concept
- Provides decision support for concept extinction while maximizing diversity

Example
How to Calculate Disparity?

Step 1: Element assessment
- Identify aspects contributing to disparity
- Assess each element regarding these aspects

Example: Database disparity
- Paradigm (SQL vs. NoSQL)
- Use case (OLTP vs. OLAP)
- Cost (cheap vs. expensive)

<table>
<thead>
<tr>
<th>Element</th>
<th>SQL</th>
<th>NoSQL</th>
<th>OLTP</th>
<th>OLAP</th>
<th>Cheap</th>
<th>Expensive</th>
</tr>
</thead>
<tbody>
<tr>
<td>SQL Server</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>DB2</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>Oracle</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>IMS</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>Sybase</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
</tr>
</tbody>
</table>
How to Calculate Disparity?

Step 2: Distance calculation

- Here Euclidean distance: yes = 1 and no = 0
- \( d(x,y) = \sum [(x_i - y_i)^2]^{1/2} \)

Example: Database disparity

<table>
<thead>
<tr>
<th>Element</th>
<th>SQL Server</th>
<th>DB2</th>
<th>Oracle</th>
<th>IMS</th>
<th>Sybase</th>
</tr>
</thead>
<tbody>
<tr>
<td>SQL Server</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DB2</td>
<td>1,41</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oracle</td>
<td>1,41</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IMS</td>
<td>2</td>
<td>1,41</td>
<td>1,41</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Sybase</td>
<td>2,45</td>
<td>2</td>
<td>2</td>
<td>1,41</td>
<td>0</td>
</tr>
</tbody>
</table>

Disparity: 1,41
**How to Calculate Disparity?**

**Step 2: Distance calculation**
- Here Euclidean distance: yes = 1 and no = 0
- \[d(x,y) = \sum [(x_i - y_i)^2]^{1/2}\]

**Example: Database disparity**

<table>
<thead>
<tr>
<th>Element</th>
<th>SQL Server</th>
<th>DB2</th>
<th>Oracle</th>
<th>IMS</th>
<th>Sybase</th>
</tr>
</thead>
<tbody>
<tr>
<td>SQL Server</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DB2</td>
<td>1.41</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oracle</td>
<td>1.41</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IMS</td>
<td>2</td>
<td>1.41</td>
<td>1.41</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Sybase</td>
<td>2.45</td>
<td>2</td>
<td>2</td>
<td>1.41</td>
<td>0</td>
</tr>
</tbody>
</table>

**Disparity:** 1.41
How to Calculate Disparity?

Step 2: Distance calculation

- Here Euclidean distance: yes = 1 and no = 0
- \( d(x,y) = \sum [(x_i - y_i)^2]^{1/2} \)

Example: Database disparity

<table>
<thead>
<tr>
<th>Element</th>
<th>SQL Server</th>
<th>DB2</th>
<th>Oracle</th>
<th>IMS</th>
<th>Sybase</th>
</tr>
</thead>
<tbody>
<tr>
<td>SQL Server</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DB2</td>
<td>1.41</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oracle</td>
<td>1.41</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IMS</td>
<td>2</td>
<td>1.41</td>
<td>1.41</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Sybase</td>
<td>2.45</td>
<td>2</td>
<td>2</td>
<td>1.41</td>
<td>0</td>
</tr>
</tbody>
</table>
How to Calculate Disparity?

Step 2: Distance calculation

- Here Euclidean distance: yes = 1 and no = 0
- \[ d(x,y) = \sum[(x_i - y_i)^2]^{1/2} \]

Example: Database disparity

<table>
<thead>
<tr>
<th>Element</th>
<th>SQL Server</th>
<th>DB2</th>
<th>Oracle</th>
<th>IMS</th>
<th>Sybase</th>
</tr>
</thead>
<tbody>
<tr>
<td>SQL Server</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DB2</td>
<td>1,41</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oracle</td>
<td>1,41</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IMS</td>
<td>2</td>
<td>1,41</td>
<td>1,41</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Sybase</td>
<td>2,45</td>
<td>2</td>
<td>2</td>
<td>1,41</td>
<td>0</td>
</tr>
</tbody>
</table>

Disparity: 3,23
### CALM³ Diversity Assessment for Databases

<table>
<thead>
<tr>
<th>Diversity aspect</th>
<th>Case 1</th>
<th>Case 2</th>
<th>Case 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variety</td>
<td>8</td>
<td>10</td>
<td>8</td>
</tr>
<tr>
<td>Balance</td>
<td>1.653</td>
<td>1.568</td>
<td>1.122</td>
</tr>
<tr>
<td>Balance (normalized)</td>
<td>5.227</td>
<td>4.795</td>
<td>3.074</td>
</tr>
<tr>
<td>Disparity</td>
<td>8.485</td>
<td>8.485</td>
<td>11.314</td>
</tr>
</tbody>
</table>

Criteria used to calculate distance among types:

- Paradigm (SQL vs. NoSQL)
- Use case (OLTP vs. OLAP)
- Cost (cheap vs. expensive)
## CALM³ Diversity Assessment for Databases

### Table: Domain (DBMS instances) Diversity Assessment

<table>
<thead>
<tr>
<th>Domain (DBMS instances)</th>
<th>Variety</th>
<th>Balance</th>
<th>Disparity</th>
<th>Disparity (weighted)</th>
<th>Types</th>
</tr>
</thead>
<tbody>
<tr>
<td>Compliance Services (4)</td>
<td>2</td>
<td>0.693</td>
<td>2.83</td>
<td>2.83</td>
<td>MS SQL, Oracle</td>
</tr>
<tr>
<td>Operational Risk (2)</td>
<td>2</td>
<td>0.693</td>
<td>2.83</td>
<td>4</td>
<td>Sybase, Oracle</td>
</tr>
<tr>
<td>Product Conditions (4)</td>
<td>2</td>
<td>0.693</td>
<td>2.83</td>
<td>2.83</td>
<td>DB2, IMS</td>
</tr>
<tr>
<td>Legal (2)</td>
<td>2</td>
<td>0.693</td>
<td>4</td>
<td>4.90</td>
<td>MS SQL, Sybase</td>
</tr>
<tr>
<td>Enterprise App. Integration (2)</td>
<td>2</td>
<td>0.693</td>
<td>0</td>
<td>0</td>
<td>DB2, Oracle</td>
</tr>
<tr>
<td>Technical Services (2)</td>
<td>2</td>
<td>0.693</td>
<td>2.83</td>
<td>2.83</td>
<td>Oracle, IMS</td>
</tr>
</tbody>
</table>

### In-depth analysis of case 2

- Variety and balance are equal for all of these domains
- Concrete types of used database systems differ significantly
- Disparity highlights potential standardization areas
Tricia integrated solution

- select data
- MxL query
- visualization
- MxL Execution (Evaluation of expressions)
- load & transform
- parse
- exec
- R
- temp.csv
- Store
- Computing Environments

Loosely-coupled solution

- select data
- SQL query
- EAM Tool
- load & transform
- exec
- R
- temp.csv
- Store
- Computing Environments

Tasks:
- automated
- manual

Tasks:
A General Distance Function for DBMS

Used criteria:
- Paradigm (SQL vs. NoSQL)
- Use case (OLTP vs. OLAP)
- Cost (cheap vs. expensive)
Evaluation

Case study candidates

Questions to be answered

- Do identified benefits and drawbacks support decisions regarding desired diversity?
- Are all relevant aspects of diversity covered by the conceptual framework?
- Do diversity metrics yield new insights thus supporting diversity decisions?
- Is it possible to develop appropriate distance functions?
- Is the novel disparity metric necessary to complement existing metrics?
- Does the proposed software support meet practitioners requirements?
Diversity dependencies across EA layers

Cross-layer induced diversity

- In case of dependencies between elements of different architectural layers diversity can be considered *cross-layer induced*.
- Example: Applications A and B use custom functionality of different databases.

Self-induced diversity

- In case of consistent interfaces and zero disparity the diversity can be considered *self-induced*.
- Example: Applications C and D both require a relational database.
Summary

Diversity reasons and consequences are manifold

- Business requirements
- Legal requirements
- Technological progress
- ... 

- Costs
- Increasing time-to-market
- Potential for innovation
- ... 

Diversity contributes significantly to complexity

- Variety covers the amount of types
- Balance covers the distribution of instances among types
- Disparity covers the degree of differences between types

Diversity assessment/measurement is essential for its management
AL diversity: enable organizational agility and keep costs under control

Diversity consists of four dimensions:

- Two metrics for their quantification have been developed and implemented
- Applicability validated in four case studies
- Examination of the relationship between AL diversity and AL complexity
- Identification of steering mechanisms for enterprise architects

Future Research

- Observe metric behavior in more detail
- Investigate relationships between diversity and other IT related aspects
- Develop a governance model for application landscape diversity management
Summary & Future work

Conclusion

- Up to now, literature provides three different approaches to measure EA complexity
- Practitioners developed additional useful metrics
- Necessary information for calculation is usually available
- Complexity metric values differ significantly in practice
- Model and domain design limits comparability
- Each type of metric has its own benefits and drawbacks

Open questions

- What are desirable target values for complexity metrics?
- How does complexity evolve over time?
- Can we identify the impact of a certain business architecture on the application landscape complexity?
Diversity is one aspect of complexity that seems to be particularly relevant for application landscape management. Diversity should be studied independently from complexity. Diversity can be further broken down into variety, balance, and disparity:

- **Variety** covers the number of different types.
- **Balance** covers the distribution of instances among types.
- **Disparity** covers the degree of differences between types.

Calculation of disparity requires additional information about distances. Required effort: Proportional to the number of types squared.
Thank you for your attention. Questions?