Heterogeneity in IT Landscapes and Monopoly Power of Firms: A Model to Quantify Heterogeneity

Completed Research Paper

Thomas Widjaja
Technische Universität Darmstadt
Chair of Information Systems
Hochschulstraße 1, 64289 Darmstadt
Germany
widjaja@is.tu-darmstadt.de

Jasmin Kaiser
Technische Universität Darmstadt
Chair of Information Systems
Hochschulstraße 1, 64289 Darmstadt
Germany
kaiser@is.tu-darmstadt.de

Dennis Tepel
Technische Universität Darmstadt
Chair of Information Systems
Hochschulstraße 1, 64289 Darmstadt
Germany
tepel@is.tu-darmstadt.de

Peter Buxmann
Technische Universität Darmstadt
Chair of Information Systems
Hochschulstraße 1, 64289 Darmstadt
Germany
buxmann@is.tu-darmstadt.de

Abstract

The term "heterogeneity" is widely used to describe complex IT systems and can refer to various characteristics, such as vendor, technology, or semantic diversity of the systems components. However, no commonly accepted definition or quantification of this "diversity" can be found in IS literature. In this article we transfer well-studied measures from other disciplines (especially economics and the anti-monopoly legislation) to heterogeneity in IT landscapes. The two main contributions of this article are A) the definition of heterogeneity in an IT landscape as a statistical property – which therefore can be measured by statistical indexes and B) a generic approach to quantify heterogeneity in IT landscapes. The applicability of the conceptualization and the approach to quantify heterogeneity is demonstrated in two real IT organizations.

Keywords: Enterprise architecture, enterprise software/systems, heterogeneity, information system complexity, IS planning

Introduction

Over time most enterprises' IT landscapes[1] evolved figuratively to "patchworks" of soft- and hardware solutions provided by various vendors which are in most cases also supplemented by self-developed components. If those heterogeneous IT landscapes stay untouched, they are "at best" difficult to manage and due to incompatibilities costly to maintain. Heterogeneity becomes especially relevant in cases that require flexibility of the information systems architecture, for example, in the context of post-merger IT integration, business process reengineering and development of innovative IT solutions.

For those reasons, it is often a goal of CIOs to manage the "heterogeneity" of IT (Boh and Yellin 2006; Ross et al. 2006; Tamm et al. 2011). This is, in particular, reflected in various IS literature streams such as IT standardization (e.g., Boh and Yellin 2006; Weitzel et al. 2006), Vendor Management (e.g., Cousins and Spekman 2003; Lacity and Willcocks 1998; Rottman and Lacity 2006) and Enterprise Architecture Management (e.g., Richardson et al. 1990; Ross et al. 2006; Tamm et al. 2011) which – at least in passing – address the question regarding the optimal degree of heterogeneity. In those streams of literature/fields of research, the term "heterogeneity" can relate to a variety of IT elements, e.g., "heterogeneity of hardware suppliers", "heterogeneity of software vendors", "heterogeneity of used technologies", and "semantic heterogeneity" in databases. So far, however, there has been little discussion about the definition of the term and in particular no concrete measure for heterogeneity has been proposed in the IS literature.

The two main contributions of the article are A) the definition of heterogeneity in an IT landscape as a statistical property – which therefore can be measured by statistical indexes and B) a generic mathematical model to quantify heterogeneity in IT landscapes. This model is based on the statistical entropy measure, which is also used in the context of anti-monopoly legislation to determine the market power of firms and serves as basis to estimate product or firm diversification. The usability of the definition and the proposed model are demonstrated in two IT organizations. We hope that the proposed conceptualization of heterogeneity and the quantification model can be used especially in the fields of IT Governance, Vendor Management, Enterprise Architecture Management, and IT Standardization; since those research streams typically address – sometimes in passing – "heterogeneity" (of decision outcomes, vendors, architectural components, standards, and so forth).

The remainder of the paper is structured as follows: The next section provides an overview of heterogeneity in IS literature. Afterwards we present the model to quantify heterogeneity in IT landscapes. In the fourth section we show by the application of the proposed approach in two real IT landscapes, how the presented model allows identifying possible targets for selective measures to reduce heterogeneity. Finally, we give a short summary of our findings, highlight the main limitations of the approach and propose avenues for further research.

Related Work: Heterogeneity in IT Landscapes

The term "heterogeneity" is often used in the context of IT. Nevertheless, in related literature only a few approaches to define and classify the possible types of heterogeneity in IT landscapes exist. Based on a literature review in the TOP 25 MIS Journals (we used AIS (2012) for orientation), we found 306 articles with reference to "heterogeneity" / "homogeneity" in IT landscapes. Hereby, we explicitly included various elements of an IT landscape (like network, application, software, etc.) in the respective search strings. Figure 1 provides an overview of occurrences of the term "heterogeneity" / "homogeneity" in combination with the search terms.

[1] An information system is a socio-technical system which consists of elements that exchange information. These elements can be classified in different element types (e.g., human and technical) (Hall and Hagen 1969). In this article we use the term "IT landscape" to refer to all technical elements of an information system.

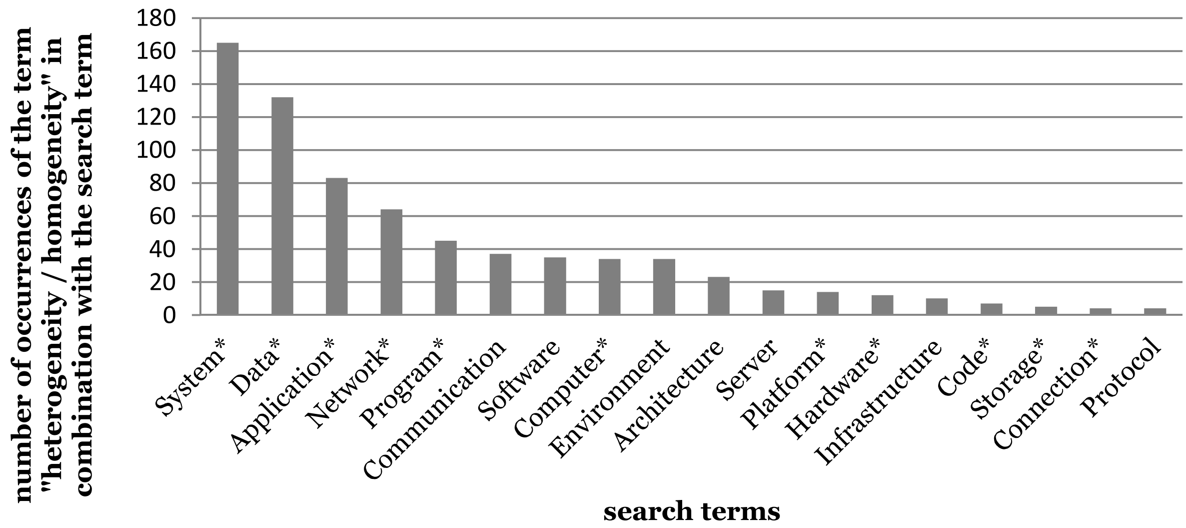

Figure 1: Overview of articles with reference to "Heterogeneity" / "homogeneity" in combination with search terms

Despite the fact that the term heterogeneity is used in various articles (cf. Figure 1), only few publications are dedicated to the topic heterogeneity itself. We identified three streams of literature regarding heterogeneity: 1) "costs and benefits of IT heterogeneity", 2) "solutions of problems related to IT heterogeneity", and 3) "classifications of IT heterogeneity". The first stream of literature evaluates *costs and benefits related to heterogeneity* (e.g., Notkin et al. 1988; Notkin et al. 1987). Since various advantages are associated with a certain level of heterogeneity, a reason for heterogeneity may also be a conscious decision of the IT architect or the CIO: If, for example, vendor and technology heterogeneity are considered, a high degree of heterogeneity reduces possible lock-ins (e.g., Cousins and Spekman 2003; Rottman and Lacity 2006). If decision makers choose to implement a best of breed system to maximize the fulfilled functional requirements, this system most likely shows a high level of vendor diversity (Light et al. 2000). Note that in some cases a best of breed IT system may also be homogeneous, however, the following simple example illustrates that this is an extremely rare case: Assume that two vendors exist, which both offer one component for each IT function that should be covered. If we assume that each of these vendors is able to supply the best component for each IT function with the same probability, e.g., $p = 0.5$, the probability that the best of breed system is homogeneous regarding the vendors is $p^{number\ of\ IT-functions}$. For 10 IT functions this probability is $0.5^{10} \approx 0.00098$.

A second part of literature proposes *solutions* to problems caused by special types of heterogeneity; for example, Bouguettaya et al. (1995) develop a hyper distributed database concept to address conflicts when heterogeneous and autonomous databases are integrated. Härder et al. (1999) present an approach to solve conflicts of structural heterogeneity in databases by introducing a mapping language. Dey (2008) proposes a method to address the entity heterogeneity problem. Park and Ram (2004) develop a framework to automatically detect and resolve semantic conflicts in heterogeneous information systems. Enterprise Architecture Management and usage of IT standards as ways to reduce heterogeneity are discussed in (e.g., Boh and Yellin 2006; Ross et al. 2006).

The third stream of literature relates to *classifications* of heterogeneity types – as there has been various research on data heterogeneity, most existing classifications are based on this perspective. For example, Chatterjee and Segev (1995) consider two types of heterogeneity in databases: structural and semantic heterogeneity – and list sources of different types of heterogeneity. However, there are few studies that present holistic approaches to classify heterogeneity characteristics within IT landscapes. Sheth and

Larson (1990) present types of heterogeneity illustrated in a hierarchical model divided in heterogeneity related to database systems, operating systems and hardware. In another work, Sheth (1999) divides the heterogeneity of information systems into *system heterogeneity* (e.g., diversity of hardware, operating system, database management system) and *information heterogeneity* (e.g., semantic or syntactical heterogeneity of data). Hasselbring (2000) distinguishes between heterogeneity on technical and conceptual level.

Considering the plurality of types of heterogeneity in IT landscapes (cf. Figure 1) and existing classifications (e.g., Chatterjee and Segev 1995 ; Hasselbring 2000; Sheth 1999) we propose the following generic definition: *Heterogeneity in IT landscapes is a statistical property and refers to the diversity of attributes of elements in the IT landscape.* Note that this definition addresses the core of "heterogeneity" and considers all types of elements in an IT landscape. Deliberately, the term "attributes" allows to address various forms of heterogeneity for each element, for example vendors, products and product variants as well as element specific, mostly technical or logical, characteristics. To the knowledge of the authors it is a novel approach to consider heterogeneity in IT landscapes as statistical property. As will be shown, this approach is advantageous since our generic definition A) can be applied to all types of heterogeneity and B) allows the usage of well-studied measures from other disciplines (esp. economics).

Generic Model to Quantify Heterogeneity in IT Landscapes

In this section, a generic model to quantify heterogeneity in an IT landscape is introduced. The model can be based on different measures of heterogeneity (see next subsection) which can be chosen depending on the application of the model.

A Measure of Heterogeneity

When considering a specific type of heterogeneity and a specific element of an IT landscape, heterogeneity can be understood as a frequency distribution. For example, the heterogeneity of vendors (attribute) for all workstations (element type) in a company can be expressed graphically as the relative or absolute frequency of vendors (histogram) as shown in Figure 2. Note that this holds true for all types of heterogeneity (e.g., "heterogeneity of hardware suppliers", "heterogeneity of software vendors", "heterogeneity of used technologies", and "semantic heterogeneity in databases").

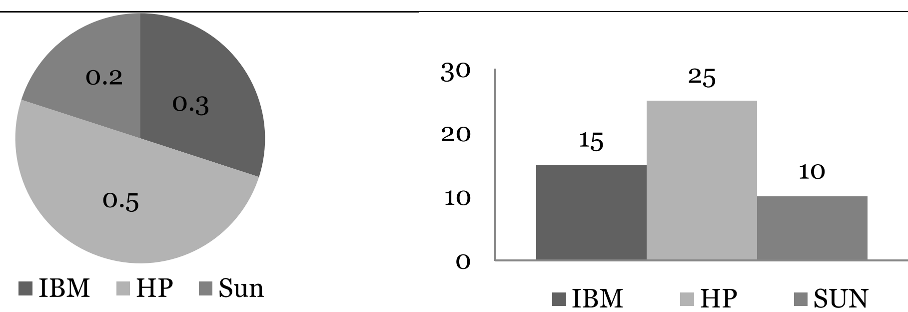

Figure 2. Heterogeneity expressed as a relative and absolute frequency distribution

Based on this consideration, we can define two requirements for a measure of heterogeneity:[2]

- The greater the amount of different values (e.g., vendors) the higher the level of heterogeneity.
- The lower the disparity between different values the higher the heterogeneity.

With regard to the first requirement, an IT landscape with workstations from five vendors would be considered , all other things being equal, as more heterogeneous than an IT landscape with two vendors. Furthermore, a distribution where 50 % of workstations are supplied by vendor A and 50 % of

[2] Similar requirements are stated in (Gollop and Monahan 1991) for a measure of firm diversification.

workstations are supplied by vendor B is more heterogeneous than a distribution where 99 % of all workstation are supplied by vendor A and just 1 % by vendor B.

A possible way to measure heterogeneity with respect to those requirements is using concentration measures.[3] Concentration measures are used in many different economic sectors, for instance to measure the market power of firms (Kwoka Jr 1985; U.S. Department of Justice and the Federal Trade Commission 2010) and to illustrate the product or business segment diversification of firms (Marfels 1975; Troutt and Acar 2005). A popular index of market power which is calculated based on market shares is the Herfindahl-Hirschman-Index (also known as Herfindahl Index, or HHI, see Hirschman (1964)):

$$HHI = \sum_{i=1}^{n} f_i^2 \quad with\ f_i = \frac{x_i}{\sum_{i=1}^{n} x_i}$$

In this formula f_i represents the relative market share of firm i of n firms where x_i denotes the absolute market share of firm i. The Herfindahl-Hirschman-Index increases, a) with a lower amount of firms, and b) with a higher disparity between these firms, i.e., a high index value indicates a high concentration. Therefore, $1 - HHI$ satisfies the two above stated requirements. Note that, a similar index is used to measure diversity in biological ecosystems (Hill 1973; Peet 1974; Simpson 1949).

We transfer these ideas from the domain of anti-trust legislation and diversity of biological ecosystems to heterogeneity in IT landscapes: The market share of firm i can be interpreted as the relative frequency of characteristic i of the considered element of the IT landscape. Thus, the Herfindahl-Hirschman-Index of the distribution presented in figure 1 equals 0.38. The lower bound $1/n$ of the index value is reached by an equal distribution of market share while the upper bound 1 is reached by a monopoly. Low relative frequency values have only a marginal impact on the Herfindahl-Hirschman-Index. If for example the two distributions $\{50\ \%, 50\ \%\}$ and $\{50\ \%, 48\ \%, 2\ \%\}$ are compared there is only a marginal impact on the Herfindahl-Hirschman-Index (0.5 compared to 0.48). A concentration ratio that avoids this property (and was mainly proposed by this reason) is the Entropy Measure (Jacquemin and Berry 1979), which is also used to measure the diversification of firms or the georgraphic concentration of industries (Garrison and Paulson 1973):

$$EM = \sum_{i=1}^{n} f_i \ln\left(\frac{1}{f_i}\right)$$

If the Entropy Measure is used to compare the two aforementioned distributions $\{50\ \%, 50\ \%\}$ and $\{50\ \%, 48\ \%, 2\ \%\}$ the resulting values differ stronger, i.e., 0.68 and 0.78. The Entropy Measure takes the minimum value 0 in a "monopoly" and reaches its maximum of $ln(n)$ for n values with an equal distribution. The Entropy Measure increases, a) the greater the amount of different values and b) the lower the disparity between these values. As this behavior exactly match the previously stated requirements, the Entropy Measure (and also the 1 - Herfindahl-Hirschman-Index, which is known as the Berry Index (Berry 1974)) can serve as a measure for heterogeneity in the proposed model (see next sub section).

The numbers-equivalent Entropy Measure is defined as $EM_A = \exp(EM)$ (Baldwin et al. 2001). EM_A is equivalent to the number of values that would lead to the same value of E for an equal distribution of EM_A values. $EM_A = 5$ means, for example, that the same value of Entropy Measure would be reached by a distribution with five values, each with a share of 20%. Note that this property can be used to facilitate the interpretation of the index values but further conclusions have to be drawn with care: The distributions $\{5\ \%, 5\ \%, 10\ \%, 80\ \%\}$ and $\{50\ \%\ ,\ 50\ \%\}$ lead for example to nearly the same value of the numbers-equivalent Entropy Measure (2.03 respective 2.0) but could be related to completely different benefits and costs. This effect could be mitigated by stating the number of firms (i.e. classes) that were considered in addition to the Entropy Measure respectively the numbers-equivalent Entropy Measure.

[3] Concentration measures can be classified in absolute and relative concentration measures. In opposition to absolute measures, relative measures (e.g., the Gini coefficient, Lorenz curve) do not consider the absolute quantity of values (i.e. don't fulfill the first requirement).

Mathematical Formulation of a Model to Quantify Heterogeneity

In this section, we present a mathematical formulation of a model to quantify heterogeneity of an IT landscape based on the Entropy Measure (see previous sub section). As a first step, the set of types of elements of the examined IT landscape (*ET)* has to be defined:

$$ET = \{et \mid et\ is\ a\ element\ type\ of\ the\ IT - Landscape\}$$

For each element type $et \in ET$, HT^{et} is an tuple containing $n^{et} \in \mathbb{N}$ types of heterogeneity that are associated with this element type:

$$HT^{et} = (ht_1^{et}, \ldots, ht_{n^{et}}^{et}) | \forall et \in ET$$

For instance, for the element type "computer", $HT^{Computer}$ could be defined as:

$$HT^{Computer} = (Architecture, Performance, Vendor, Product)$$

For each type of heterogeneity HT_i^{et} a set HV_i^{et} of characteristic values has to be defined:

$$HV_i^{et} = \{hv \mid hv\ is\ a\ charateristic\ value\ of\ ht_i^{et} \wedge hv \notin \{0,1\}\} | \forall i \leq n^{et} \forall et \in ET$$

The set $HV_3^{Computer}$, i.e., the characteristic values for heterogeneity of vendors (see figure 1) of computers could be defined as:

$$HV_3^{Computer} = \{HP, IBM, DELL\}$$

The considered IT Landscape can be divided in different organizational units $o \in O$.[4]

$$O = \{o \mid o\ is\ an\ organizational\ unit\ of\ the\ firm\}$$

The set E is defined as set of elements of the examined IT landscape. An element $e \in E$ is defined as 3-tuple:

$$e = (o, et, A) | o \in O \wedge et \in ET \wedge A = (A_1, \ldots, A_{n^{et}}) \wedge (A_i \ \in \ HV_i^{et} | \forall i \leq n^{et})$$

The first part of the tuple relates to the organizational unit of the element, i.e., "where" the element is used. Note that, this organizational unit can for example correspond to a division in the company or a domain in an architecture model. The second part of the tuple specifies the element type $et \in ET$, e.g., computer or application software. The third part refers to a tuple A which contains all characteristic values regarding the respective heterogeneity types HT^{et} of the element type. For a computer supplied by the vendor "DELL" the corresponding element e' might be:

$$e' = (Department\ A\ , Computer, (x86\ , Low\ Performance\ Class, DELL, Precision\ 1500))$$

In order to quantify the heterogeneity of elements of element type $et \in ET$ in an organizational unit $\in O$, it has to be specified what *type of heterogeneity* and how diversifications based on *other types of heterogeneity* have to be considered. The choice of the examined type of heterogeneity and potential constraints are coded in an examination tuple Ex^{et}:

$$Ex^{et} = (ex_1^{et}, \ldots ex_{n^{et}}^{et}) \left| (ex_i^{et} \in \ HV_i^{et} \cup \{0,1\} | \forall i \leq n^{et}) \wedge et \in ET \wedge \left(\sum_{i \in \{1 \ldots n^{et}\}} \left| \{ex_i^{et}\} \cap \{1\} \right| \right) = 1 \right.$$

The examined type of heterogeneity is marked by the value 1 at the respective position. Note that exactly one element of the tuple Ex^{et} takes the value 1. For all other types, ex_i^{et} indicates a non-varying characteristic value of the corresponding type of heterogeneity ht_i^{et}. In the case of no constraints, ex_i^{et} is chosen as 0. The examination tuple $Ex^{Computer} = (x86, 0,1,0)$ allows the heterogeneity of vendors

[4] Note that the calculated index values may have to be aggregated (e.g., due to hierarchical relationships between organizational units).

($ht_3^{computer} = vendor$) on the for the element type $Computer$ to be examined by exclusively allowing the architecture "$x86$". If the heterogeneity of DELL computers ($ht_4^{computer} = product$) with the architecture "$x86$" should be examined then the respective examination tuple would be $Ex^{Computer} = (x86, 0, Dell, 1)$.

Based on the examination tuple, the respective relative frequencies (note that these relative frequencies correspond to the relative market shares in the last sub-section) can be determined by the following function:

$$f(o, et, Ex^{et}, hv) = \frac{\left|\left\{(o, et, A) \in E \middle| \left(A_i = ex_i^{et} \middle| \forall i \; with \; ex_i^{et} \notin \{0,1\}\right) \wedge \left(A_i = hv \middle| \forall i \; with \; ex_i^{et} = 1\right)\right\}\right|}{\left|\left\{(o, et, A) \in E \middle| \left(A_i = ex_i^{et} \middle| \forall i \; with \; ex_i^{et} \notin \{0,1\}\right)\right\}\right|}$$

Note that elements which are determined by the examination tuple Ex^{et} are taken into account. The heterogeneity of an organizational unit $o \in O$ for an element type $et \in ET$ with respect to an examination tuple Ex^{et} can be determined by using any measure for concentration (see section last sub-section). The following formulation of the model is based on the Entropy Measure:

$$Heterogeneity(o, et, Ex^{et}) = \sum_{hv \in S(o,et,Ex^{et})} f(o, et, Ex^{et}, hv) \; ln\left(\frac{1}{f(o, et, Ex^{et}, hv)}\right)$$

The set $S(o, et, Ex^{et})$ contains all characteristic values of HV_i^{et} which should be considered, i.e., where $ex_i^{et} = 1$ and is defined as:

$$S(o, et, Ex^{et}) = \{hv | (hv \in HV_i^{et} | i \; with \; ex_i^{et} = 1) \wedge f(o, et, Ex^{et}, hv) \neq 0\}$$

The constraint $f(o, et, Ex^{et}, hv) \neq 0$ avoids the consideration of characteristic values with a relative frequency of 0. The constraint particularly applies, if an examination tuple is chosen in a way that one type of heterogeneity is constrained by a second type. For example, it would not make sense to examine the heterogeneity of vendors while constraining the value of the type "product" to a fixed characteristic value.

Based on the literature review we propose a definition of heterogeneity in an IT landscape as a statistical property – since this covers the essential concept of all in literature discussed types of heterogeneity. Therefore, heterogeneity can be measured by statistical indexes. The presented generic mathematical model allows quantifying various types of heterogeneity in IT landscapes.

Application of the Model

The applicability of the proposed definition and measure was evaluated in IT organizations of two different companies. The approach was applied to various element types of an IT landscape: 1) types and manufactures of "computer" and "office software" as well as 2) vendors of "application software".

Application of the Model 1: Government Organization

The presented model has been applied to a subset of the IT landscape (456 elements) of a government organization (> 100,000 employees and > 1,000 IT employees). The data presented in the following has been anonymized. The goal of the application was to determine the organizational units with high degree of heterogeneity regarding the used element types "computer" and "office software" – this information is especially relevant to the organization, since the analyzed computers serve as front-end of a central Enterprise Resource Planning System (ERP).

During the application we examined five organizational units, i.e., $O = \{o_1, o_2, o_3, o_4, o_5\}$ and various element types. To reduce complexity we focus on the results related to personal computers and office software, i.e., $ET = \{computer, office \; software\}$. For both element types we collected data regarding the vendor and the product. For the element type *computer* we also considered the "Computer Type", i.e., Laptop computer or desktop computer and "Performance"[5] of the respective computer. Therefore $HT^{Computer}$ is defined as:

$$HT^{Computer} = (Computer\ Type, Vendor, Product, Performance)$$

For the *office software* we additionally considered the version of the software, e.g., which version of Microsoft Office 2003 is used (Microsoft 2010):

$$HT^{Office\ Software} = (Vendor, Product, Version)$$

Information regarding the heterogeneity of the Table 1 shows the result of the examination of the element type computer.

Table 1: Heterogeneity of element type "computer"

Type of heterogeneity /Examination Tuple	o_1	o_2	o_3	o_4	o_5	o_1 - o_5
Number of elements (Computers)	83	165	20	132	56	456
Computer TypeEx^{Computer} $= (o_i, \text{Computer}, (1,0,0,0))$	0.59 (1.80; 2)	0 (1; 1)	0 (1; 1)	0.66 (1.94; 2)	0 (1; 1)	0.44 (1.55; 2)
Vendor Ex^{Computer} $= (o_i, \text{Computer}, (0,1,0,0))$	1.08 (2.94;4)	0.48 (1.61;2)	0.33 (1.38;2)	1.05 (2.87;2)	0.47 (1.6;2)	0.89 (2.44;4)
Product Ex^{Computer} $= (o_i, \text{Computer}, (0,0,1,0))$	1.38 (3.97;5)	0.66 (1.93;3)	0.64 (1.9; 3)	1.33 (3.77;5)	0.75 (2.12;3)	1.11 (3.04;5)
Performance Ex^{Computer} $= (o_i, \text{Computer}, (0,0,0,1))$	1.01 (2.74; 3)	0.92 (2.5; 3)	0.86 (2.37; 3)	1.09 (2.99; 3)	0.79 (2.21; 3)	1.09 (2.99; 3)

Each cell contains the Entropy Measure and in brackets the respective numbers-equivalent Entropy Measure as well as the quantity of characteristic values for the considered heterogeneity type. The entry "0.59 (1.80; 2)" in the first cell (i.e. the heterogeneity of computer types in organizational unit 1) represents a value of 0.59 for the Entropy Measure, which corresponds to an numbers-equivalent Entropy Measure of 1.80. Note that the in this organizational unit 2 types of computers (i.e. laptops and desktops) are used – therefore the theoretical maximum of the numbers-equivalent Entropy Measure is 2.

The analysis regarding the results for the element type "computer" (Table 1) reveals that the highest heterogeneity for all types can be found at the organizational units o_1 and o_4. Taking the numbers-equivalent Entropy Measure of o_4 for the heterogeneity of computer types as an example (1.94), it turns out that the corresponding heterogeneity is nearly equivalent to an equal distribution of two computer types (which is close to the theoretical maximum, since the characteristic values of this type of heterogeneity are the laptop and desktop). The numbers-equivalent Entropy Measure of 1 for the units o_2 o_3 , and o_5 indicates that there is only one type of computer used (only the desktop computer). Furthermore the high vendor and product entropy measures for the organizational units o_1 and o_4 show that especially in these departments multiple products of different vendors are in use. Regarding performance, all organizational units are taking a value of numbers-equivalent Entropy Measure close to 3. Taking into consideration that all machines were classified in three categories it can be concluded that the IT landscape consists of machines that are nearly uniformly distributed over the three performance classes rather than providing the same performance.

[5] In order to classify the computers in three different classes (low, middle, high performance) we used benchmark tables.

Table 2: Heterogeneity of element type "office software"						
Type of heterogeneity /Examination Tuple	o_1	o_2	o_3	o_4	o_5	o_1 - o_5
Number of elements (Office Software Instances)	83	165	20	132	56	456
Vendor $Ex^{office\ software} = (o_i, Office\ Software, (1,0,0))$	0.23 (1.26;2)	0 (1;1)	0.42 (1.53;2)	0 (1;1)	0 (1;1)	0.09 (1.09;2)
Product $Ex^{office\ software} = (o_i, Office\ Software, (0,1,0))$	0.74 (2.14;3)	0.23 (1.26;2)	0.42 (1.53;2)	0.16 (1.18;2)	0 (1;1)	0.36 (1.43;3)
Version $Ex^{office\ software} = (o_i, Office\ Software, (Microsoft, Office\ 2003,1))$	1.4 (4.04;6)	0.61 (1.85;4)	0.5 (1.65;3)	0.32 (1.37;2)	0.47 (1.6;4)	0.55 (1.74;7)

Table 2 shows measures regarding the heterogeneity of office software. It turns out that office software supplied by a single vendor is used in organizational units o_2, o_4, and o_5. Apart from this, all units except o_5 are using different products. The highest product diversification can be found in unit o_1 (2.14). For the heterogeneity of software versions, the scope was limited to Microsoft Office 2003 products. Obviously there is heterogeneity of versions at all organizational units. Again the highest heterogeneity appears at unit o_1. The numbers-equivalent Entropy Measure of 4.041 indicates a heterogeneity that is equal to a number of 4 uniformly distributed software versions within this unit.

The model and the above stated results confirmed the (already existing) plan of the CIO to implement a new lifecycle management for computer products and to utilize a leasing contractor for the hardware in order to reduce the performance heterogeneity. The results regarding the heterogeneity of the office software versions triggered a further investigation of the related costs and benefits of this type of heterogeneity. Furthermore the model facilitated the communication between the different organizational units and is used as key performance indicator to present the status quo regarding the heterogeneity of the IT landscape to the CIO.

Application of the Model 2: Passenger Transportation Company

The second application of the proposed model was conducted in cooperation with the IT organization of a large-scale client enterprise (> 40.000 employees) operating in the passenger transportation sector. The company relies on a complex IT landscape that is structured by a domain model which clusters software components in domains and subdomains. This domain model is based on an industry standard which ensures high intra-domain coherency and low inter-domain dependency of (sub-)domains.

We examined 147 software components of the IT Landscape, i.e., $ET = \{Software\ Component\}$, in five functional domains, i.e. $O = \{o_1, o_2, o_3, o_4, o_5\}$ and collected data regarding the vendor of the software components, i.e. $HT^{Software\ Component} = (Vendor)$. The goal of this study was to determine the overall vendor related heterogeneity of the five examined domains and to identify domains with exceptional high degree of heterogeneity – both aspects were especially relevant for the company's vendor management team. Note that the underlying trade-off regarding "vendor heterogeneity" of software components is also discussed in a rich literature stream. It has been shown that the "optimal" degree of multi-sourcing (i.e. degree of vendor heterogeneity) depends on various factors: A high degree of multi-sourcing increases competition among the vendors (Lacity and Willcocks 1998), reduces strategic risks (Rottman and Lacity 2006), is favorable in case of rapid change in global supplier markets (Levina and Su 2008), and decreases the degree of dependency on a distinct vendor (Cousins and Spekman 2003). On the other hand a relative small number of vendors induces high-quality customer-vendor relationships (Bakos and Brynjolfsson 1993), reduces coordination efforts, and improves resource utilization (Cousins and Spekman 2003). Note that the goal of the presented analysis was not to determine the optimal degree of vendor heterogeneity, but to monitor the company's decision regarding the multi-sourcing strategy over time.

The 147 analyzed software components are supplied by 69 distinct vendors[6] and Table 3 shows the result of the heterogeneity analysis – each cell contains the Entropy Measure and in brackets the respective numbers-equivalent Entropy Measure as well as the number of vendors in the respective domain. All in all, the company's IT landscape is fairly homogeneous (cf. last column of Table 3): This is reflected by an Entropy Measure value of 1.39 (i.e., a numbers-equivalent value of approximately 4 vendors). This overall homogeneity is rooted by the company's strategy to rely – whenever economical reasonable – on a distinct vendor or a small set of vendors – note that this is conform with the results of (e.g., Bakos and Brynjolfsson 1993; Kaiser and Buxmann 2012; Rottman and Lacity 2006).

Table 3: Heterogeneity of element type "application" (domain perspective)

Type of heterogeneity /Examination Tuple	o_1	o_2	o_3	o_4	o_5	o_1 - o_5
Number of elements (Software Component)	35	25	20	14	53	147
$\text{Ex}^{Software\ Component} = (\text{o}_\text{i}, Software\ Component, (1))$	1.52 (4.59; 25)	0.80 (2.22; 14)	0.20 (1.22; 11)	0.73 (2.08; 7)	1.18 (3.25; 25)	1.39 (4.0; 69)

The two domains with the highest heterogeneity in the IT landscape are domain 1 and domain 5. Domain 1 supports the company's multi-channel strategy. This domain contains 35 software components that are supplied by 25 vendors and the Entropy Measure takes a value of 1.52 (i.e. numbers-equivalent Entropy Measure of 4.59). Since customer flexibility and experience during sales and service processes are an important part of the company's differentiation strategy, various marketing channels are supported by distinct applications (often provided by niche vendors), which leads to a high – but justified by business requirements – heterogeneity for this domain.

Domain 5 – the second domain with respectively high heterogeneity – supports the operational backbone of the transport company, i.e., a crucial part of the company's IT Landscape. A possible explanation of the high vendor heterogeneity is the sheer complexity of the supported business processes. We further investigated this domain and calculated the heterogeneity values for the seven sub-domains (cf. Table 4).[7] Note that in Table 4, $O = \{o_{5.1,}, o_{5.2}, o_{5.3,}, o_{5.4}, o_{5.5} o_{5.6,}, o_{5.7}\}$.[8] On this lower aggregation level, the sub-domains of domain 5 are fairly homogenous (except of $o_{5.3,}$). The high vendor heterogeneity in sub-domain $o_{5.3,}$ indicates potential for optimization and consolidation and we therefore suggested a detailed investigation by the vendor management team of this domain.

Table 4: Heterogeneity of element type "application" (sub-domains of Domain 5)

Type of heterogeneity /Examination Tuple	1	2	3	4	5	6	7
Number of elements (Software Components in sub-domain of domain 5)	10	3	27	3	4	4	2
$\text{Ex}^{Software\ Component} = (\text{o}_{5.\text{i}}, Software\ Component, (1))$	0.34 (1.40; 6)	0.29 (1.34; 3)	1.42 (4.1; 17)	0.15 (1.16; 3)	0.0 (1.0; 1)	0.30 (1.34; 2)	0.13 (1.2; 2)

---

[6] Note that we weighted the values of $f(o, et, Ex^{et}, hv)$ (i.e. the "relative market share" of the vendors) with the relative magnitude of the "license volume" of the respective vendor.

[7] The sub-domains of the other four top-level domains differed not substantially from the heterogeneity of the respective top-level domains.

[8] Note that $o_{x,y}$ denotes the subdomain y of domain x.

The presented application of the model showed that the preferred strategy of the company to rely on a single (or small set of vendors) is reflected by the IT landscape. The proposed approach to quantify IT heterogeneity allows the IT-Management of the respective company to monitor the compliance to its strategy for each domain and over time. Furthermore, this case elucidated that the different abstraction levels of the IT architecture (i.e., the overall architecture, domains, and sub-domains) have to be considered – heterogeneity on a higher level could be rooted in sets of homogeneous elements on a lower abstraction level.

Discussion of Results

Theoretical implications

We propose to conceptualize heterogeneity in an IT landscape as a statistical property – which therefore can be measured by statistical indexes. This conceptualization is flexible enough to incorporate heterogeneity regarding all attributes and elements on the different levels of IT architecture. The proposed model can therefore serve as foundation to quantify heterogeneity in semantics of databases, vendors of software components, versions of applications, performance of hardware systems, etc. Thereby, the presented conceptualization can be used in a wide array of theoretical contexts. Furthermore the proposed model also offers a comprehensive way to operationalize (e.g., for empirical studies) the IT heterogeneity of enterprises.

Based on this foundation, studies that investigate companies' management on different types of IT heterogeneity in fields like IT Governance (e.g., Weill and Ross 2004; Weitzel et al. 2006), Vendor Management (e.g., Cousins and Spekman 2003; Lacity and Willcocks 1998; Rottman and Lacity 2006), usage of IT standards (e.g., Boh and Yellin 2006; Weitzel et al. 2006), and Enterprise Architecture Management (e.g., Richardson et al. 1990; Ross et al. 2006; Tamm et al. 2011) can refer to the proposed conceptualization. This conceptualization also paves the way to identify antecedents as well as consequences of IT heterogeneity. On the one hand, the presented conceptualization could be integrated in research on e.g., Enterprise Architecture Management (e.g., Tamm et al. 2011) and IT Governance (e.g., Weill and Ross 2004) as potential antecedents of IT heterogeneity. For example, it might be interesting which IT Government regime leads to which types and degree of IT heterogeneity. On the other hand, this work can help to better understand the connection between IT heterogeneity and, for example, IT value (e.g., Boh and Yellin 2006; Bradley et al. 2012; Tamm et al. 2011). Based on the proposed conceptualization of IT heterogeneity, future empirical studies could analyze the relation between heterogeneity / homogeneity of different element types in an IT landscape and various business benefits, such as flexibility and efficiency (e.g., Bradley et al. 2012; Schmidt and Buxmann 2011). In summary, the understanding of antecedents and possible outcomes of heterogeneous IT landscapes could particularly contribute to the literature on IT Governance and Enterprise Architecture Management by advising IT Governance regimes and Enterprise Architecture patterns that may lead to desirable degrees of heterogeneity.

Practical implications

As showed in the two presented applications of the model, the proposed approach allows practitioners to quantify a wide array of IT heterogeneity types – this information can be useful for the CIO (Case 1), Enterprise Architecture Management Team (Case 1 & Case 2) and the Vendor Management Team (Case 2). Based on the model it is possible to asses A) the overall heterogeneity of the IT landscape and to B) identify organizational units that show exceptional degrees of IT heterogeneity. Therefore, the proposed model can be used to monitor the realization of strategies regarding various heterogeneity types.

Note that it is possible to apply the proposed model only to selected parts rather than on a company's entire IT landscape (Case 1 & Case 2). Especially in such cases, the reasonability of the model's application is not only dependent on the size of the company (or the selected part of the IT landscape) – it also matters, how complex the considered part of the IT landscape (with respect to the investigated heterogeneity type) is. The proposed model can therefore be useful for relatively small IT landscapes (if they are sufficiently complex).

In the two applications of the model the following "best practice" for the definition of measures to manage a certain type of IT heterogeneity emerged. For each organizational unit the decision maker should consider the following two questions: *1.) Exists a business reason for the observed degree of heterogeneity in this organizational unit?* Note that the specific cost-benefit trade-off induced by the analyzed type of heterogeneity should be considered (as for example discussed in Case 2). Furthermore, some business strategies justify a certain degree of heterogeneity (cf. Domain 1 in Case 2) – if there is no business requirement for the observed degree of heterogeneity, then: *2.) Is the high degree of heterogeneity in the analyzed domain caused by a set of homogeneous sub-domains?* It could be possible, that a top-level structure shows an exceptional high degree of heterogeneity but a detailed analysis reveals that existing sub-structures itself are homogeneous (cf. Domain 2 in Case 2). If there is no justified and homogenous sub- structure, then: *the definition of measures to adjust heterogeneity is necessary* (cf. Case 1).

Conclusion, Limitations, and Further Research

The two main contributions of the article are A) the definition of heterogeneity in an IT landscape as a statistical property – which therefore can be measured by statistical indexes and B) a generic mathematical model to quantify heterogeneity in IT landscapes. We propose the following definition: *Heterogeneity in IT landscapes is a statistical property and refers to the diversity of attributes of elements in the IT landscape.* The generic model to quantify heterogeneity allows to incorporate different measures of heterogeneity (e.g., the Herfindahl-Hirschman-Index and the Entropy Measure proposed by Jacquemin and Berry (1979)) and supports the analysis of different types of heterogeneity in IT landscapes. The model can be used to quantify heterogeneity for all elements of an IT landscape, e.g., heterogeneity of semantics in databases, software vendors in application architectures, suppliers of hardware, and performance of clients. Since the heterogeneity can be compared across different organizational units the model can serve as basis for an enterprise wide key performance indicator system. The CIO could use the proposed model to monitor heterogeneity of IT landscapes in organizational units in the same way as the Department of Justice monitors market concentration in industry sectors. We proved the applicability of the proposed model in 1) a government organization (> 100,000 employees and > 1,000 IT employees) and 2) a passenger transport company (> 40,000 employees). Beside this practical contribution, the proposed conceptualization of IT heterogeneity and the approach to quantify the heterogeneity can be incorporated in various research streams that address costs and benefits related to IT heterogeneity as well in research areas that analyze the management of special types of IT heterogeneity (or IT heterogeneity in general) like, IT Governance, Enterprise Architecture Management, Vendor Management, IT Standardization, etc.

The proposed model is a first approach to quantify heterogeneity in IT landscapes. A main limitation is the assumption of deterministic data regarding the distributions of the characteristic values, since various cases exist where the data collection may be costly and time consuming. Therefore, a decision support system that facilitates the data collection and also helps to interpret and compare heterogeneity results might be valuable.

Avenues for further research are especially the investigation of the benefits and costs related to certain types of heterogeneity in IT landscapes, since the heterogeneity in "versions of office software" may imply different (and for most organizations also less relevant) types of costs and benefits than semantic heterogeneity in databases. These insights regarding the economic effects of certain types of heterogeneity in combination with the proposed measure can serve as first cornerstone during the analysis of the economically "optimal" degree of heterogeneity in IT landscapes.

Acknowledgements

First of all, the authors thank the anonymous reviewers for their valuable suggestions. Furthermore, we would like to thank Alexander Schütz for his valuable and constructive comments during the application of the proposed model in the case companies. We also thank our interview partners – whose names cannot be mentioned here without violating the anonymity of the case companies – for their openness and time spent on this work.

References

AIS. 2012. "Mis Journal Rankings." Retrieved 04.01.2012, 2012, from http://ais.affiniscape.com/displaycommon.cfm?an=1&subarticlenbr=432

Bakos, J.Y., and Brynjolfsson, E. 1993. "Information Technology, Incentives, and the Optimal Number of Suppliers," *J. Manage. Inf. Syst.* (10:2), pp 37–53.

Baldwin, J.R., Beckstead, D., and Caves, R. 2001. "Changes in the Diversification of Canadian Manufacturing Firms (1973-1997): A Move to Specialization," *Statistics Canada*.

Berry, C.H. 1974. *Corporate Growth and Diversification*. Princeton: Princeton University Press.

Boh, W.F., and Yellin, D. 2006. "Using Enterprise Architecture Standards in Managing Information Technology," *Journal of Management Information Systems* (23:3), pp 163–207.

Bouguettaya, A., Papazoglou, M., and King, R. 1995. "On Building a Hyperdistributed Database," *Information Systems* (20:7), pp 557-577.

Bradley, R.V., Pratt, R.M.E., Byrd, T.A., Outlay, C.N., and Wynn, J.D.E. 2012. "Enterprise Architecture, It Effectiveness and the Mediating Role of It Alignment in Us Hospitals," *Information Systems Journal* (22:2), pp 97-127.

Chatterjee, A., and Segev, A. 1995. "Rule Based Joins in Heterogeneous Databases," *Decision Support Systems* (13:3-4), pp 313-333.

Cousins, P.D., and Spekman, R. 2003. "Strategic Supply and the Management of Inter-and Intra-Organisational Relationships," *Journal of Purchasing and Supply Management* (9:1), pp 19-29.

Dey, D. 2008. "Entity Matching in Heterogeneous Databases: A Logistic Regression Approach," *Decision Support Systems* (44:3), pp 740-747.

Garrison, C.B., and Paulson, A.S. 1973. "An Entropy Measure of the Geographic Concentration of Economic Activity," *Economic Geography* (49:4), pp 319-324.

Gollop, F.M., and Monahan, J.L. 1991. "A Generalized Index of Diversification: Trends in U.S. Manufacturing," *The Review of Economics and Statistics* (73:2), pp 318-330.

Hall, A.D., and Hagen, R.E. 1969. "Definition of System," in: *Organizations - Systems, Control and Adaptation,* J.A. Litterer (ed.). New York: Wiley, pp. XII, 417 S.

Härder, T., Sauter, G., and Thomas, J. 1999. "The Intrinsic Problems of Structural Heterogeneity and an Approach to Their Solution," *The VLDB Journal* (8:1), pp 25-43.

Hasselbring, W. 2000. "Information System Integration," *Communications of the ACM* (43:6), pp 32-38.

Hill, M.O. 1973. "Diversity and Evenness: A Unifying Notation and Its Consequences," *Ecology* (54:2), pp 427-432.

Hirschman, A.O. 1964. "The Paternity of an Index," *The American Economic Review* (54:5), pp 761-762.

Jacquemin, A.P., and Berry, C.H. 1979. "Entropy Measure of Diversification and Corporate Growth," *The Journal of Industrial Economics* (27:4), pp 359-369.

Kaiser, J., and Buxmann, P. 2012. "Organizational Design of It Supplier Relationship Management: A Multiple Case Study of Five Client Companies," *Journal of Information Technology* (27:1), pp 57-73.

Kwoka Jr, J.E. 1985. "The Herfindahl Index in Theory and Practice," *Antitrust Bulletin*.

Lacity, M.C., and Willcocks, L.P. 1998. "An Empirical Investigation of Information Technology Sourcing Practices: Lessons from Experience," *MIS Quarterly*, pp 363-408.

Levina, N., and Su, N. 2008. "Global Multisourcing Strategy: The Emergence of a Supplier Portfolio in Services Offshoring," *Decision Sciences* (39:3), pp 541-570.

Light, B., Holland, C., Kelly, S., and Wills, K. 2000. "Best of Breed It Strategy: An Alternative to Enterprise Resource Planning Systems," *8th European Conference on Information Systems (ECIS 2000)*, Vienna, Austria.

Marfels, C. 1975. "A Bird's Eye View to Measures of Concentration," *Antitrust Bulletin*.

Microsoft. 2010. "How to Check the Version of Office 2003 Products." Retrieved 05.01.2012, 2012, from http://support.microsoft.com/kb/821549/en-us

Notkin, D., Black, A.P., Lazowska, E.D., Levy, H.M., Sanislo, J., and Zahorjan, J. 1988. "Interconnecting Heterogeneous Computer Systems," *Communications of the ACM* (31:3), pp 258-273.

Notkin, D., Hutchinson, N., Sanislo, J., and Schwartz, M. 1987. "Heterogeneous Computing Environments: Report on the Acm Sigops Workshop on Accommodating Heterogeneity," *Communications of the ACM* (30:2), pp 132-140.

Park, J., and Ram, S. 2004. "Information Systems Interoperability: What Lies Beneath?," *ACM Transactions on Information Systems* (22:4), pp 595-632.

Peet, R.K. 1974. "The Measurement of Species Diversity," *Annual Review of Ecology and Systematics* (5, pp 285-307.

Richardson, G.L., Jackson, B.M., and Dickson, G.W. 1990. "A Principles-Based Enterprise Architecture: Lessons from Texaco and Star Enterprise," *MIS Quarterly*, pp 385-403.

Ross, J.W., Weill, P., and Robertson, D. 2006. *Enterprise Architecture as Strategy: Creating a Foundation for Business Execution*. Harvard Business Press.

Rottman, J.W., and Lacity, M.C. 2006. "Proven Practices for Effectively Offshoring Proven Practices for Effectively Offshoring It Work," *MIT Sloan Management Review* (47:3), pp 56-63.

Schmidt, C., and Buxmann, P. 2011. "Outcomes and Success Factors of Enterprise It Architecture Management: Empirical Insight from the International Financial Services Industry," *European Journal of Information Systems* (20, pp 168-185.

Sheth, A.P. 1999. "Changing Focus on Interoperability in Information Systems. From System, Syntax, Structure to Semantics," *Interoperating Geographic Information Systems*, pp 5-29.

Sheth, A.P., and Larson, J.A. 1990. "Federated Database Systems for Managing Distributed, Heterogeneous, and Autonomous Databases," *ACM Computer Surveys* (22:3), pp 183-236.

Simpson, E. 1949. "Measurement of Diversity," *Nature* (163:4148), p 688.

Tamm, T., Seddon, P.B., Shanks, G., and Reynolds, P. 2011. "How Does Enterprise Architecture Add Value to Organisations?," *Communications of the Association for Information Systems* (28:1), p 10.

Troutt, M.D., and Acar, W. 2005. "A Lorenz-Pareto Measure of Pure Diversification," *European Journal of Operational Research* (167:2), pp 543-549.

U.S. Department of Justice and the Federal Trade Commission. 2010. "Horizontal Merger Guidelines." Retrieved 05.01.2012, from http://www.justice.gov/atr/public/guidelines/hmg-2010.html

Weill, P., and Ross, J. 2004. *It Governance: How Top Performers Manage It Decision Rights for Superior Results*. Harvard Business Press.

Weitzel, T., Beimborn, D., and König, W. 2006. "A Unified Economic Model of Standard Diffusion: The Impact of Standardization Cost, Network Effects, and Network Topology," *Management Information Systems Quarterly* (30:Special Issue on Standard Making), pp 489–514.